Exploration of Biological and Computational Modalities of Scientific Creativity

Thesis Submitted by

Sayantani Ghosh

Doctor of Philosophy (Engineering)

Artificial Intelligence Laboratory

Department of Electronics and Tele-Communication Engineering

Faculty Council of Engineering and Technology

Jadavpur University, Kolkata -700032, India

2025

Index No: 238/23/E

TITLE OF THE THESIS:

EXPLORATION OF BIOLOGICAL AND COMPUTATIONAL MODALITIES OF SCIENTIFIC CREATIVITY

Name, Designation and Institution of Supervisor:

Prof. Amit Konar

Professor

Department of Electronics & Tele-communication Engineering

Jadavpur University

Kolkata-700032

India

List of Publications

List of Publications Related to Thesis

JOURNALS

1. **S. Ghosh**, A. Konar, and A.K. Nagar, "Cognitive Assessment of Scientific Creative-Skill by Brain-Connectivity Analysis Using Graph Convolutional-Interval Type-2 Fuzzy Network," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 16, no. 5, pp. 1872-1886, Oct. 2024.

2. **S. Ghosh** and A. Konar, "Decoding the Scientific Creative-Ability of Subjects Using Dual Attention Induced Graph Convolutional-Capsule Network," *Applied Soft Computing*, vol. 161, p.111769, 2024.

3. **S. Ghosh** and A. Konar, "Computational Creativity by Diversity-Optimized Intelligent Search: An Automatic Approach to Artificial Synthesis of Trigonometric Identities," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 55, no. 6, pp. 4385-4395, June 2025.

4. **S. Ghosh**, A. Konar and A. K. Nagar, "Assessment of Scientific Creative-Potential by Near-Infrared Spectroscopy Using Brain-Network-Based Deep-Fuzzy Classifier," *IEEE Access*, vol. 13, pp. 108533-108550, 2025.

5. **S. Ghosh**, and A. Konar, "Near-Infrared Sensor-Based Assessment of Scientific Creativity Using Graph-Guided Attention-Driven Capsule Network," *IEEE Sensor Letters*, vol. 9, no. 8, August 2025.

6. **S. Ghosh**, and A. Konar, "Enhanced Generative Adversarial Network with Leaked Information (E-LeakGAN): A Deep Learning Framework for Synthesizing Computational Creativity in Scientific Domain," *Applied Soft Computing*, (re-submitted after major revision), 2025.

Sayantani Ghosh

CONFERENCE PAPERS

1. **S. Ghosh,** S. Pramanick, A. Bagchi and A. Konar, "Automatic Detection of Confidence Level of Examinees in Answering Multiple Choice Questions using P1000 Brain Signal," *2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, Chennai, India, 2019.

2. **S. Ghosh**, M. Laha and A. Konar, "P1000 Induced Brain Signal Analysis for Assessing Subjective Pain Sensitivity using Type-2 Fuzzy Classifier," *2020 International Conference on Communication and Signal Processing (ICCSP)*, Chennai, India, 2020.

3. **S. Ghosh**, M. Laha, A. Konar, P. Rakshit and A. K. Nagar, "Vowel Sound Imagery Decoding by a Capsule Network for the Design of an Automatic Mind-Driven Type-Writer," *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, UK, **IEEE Flagship Conference**, 2020.

4. **S. Ghosh**, L., Ghosh, A. Konar, and A.K. Nagar, "Assessment of Subjective Creativity Skill Using EEG Induced Capsule Network," In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, Canberra, ACT, Australia, 2020., pp. 3107-3114, **IEEE Flagship Conference**, 2020.

5. **S. Ghosh**, M. Laha, A. Konar and A. K. Nagar, "Decoding of Subjective Pain-Sensitivity by Brain Signal Analysis Using a General Type-2 Fuzzy Classifier," *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, Canberra, ACT, Australia, 2020., pp. 3107-3114, **IEEE Flagship Conference**, 2020.

6. **S. Ghosh**, A. Konar, and A.K. Nagar, "Decoding subjective creativity skill from visuo-spatial reasoning ability using capsule graph neural network," In *2021 International Joint Conference on Neural Networks (IJCNN)*, Shenzhen, China, pp. 1-8, **IEEE Flagship Conference**, 2021.

7. **S. Ghosh**, S. Bose, R. S. Chowdhury, A. Konar and A. K. Nagar, "Decoding the Creative Ability of Subjects from Aesthetic Quality Assessment Using Dual Convolution Induced Capsule Network," *2022 IEEE Symposium Series on Computational Intelligence (SSCI)*, Singapore, Singapore, **IEEE Flagship Conference**, 2022.

8. **S. Ghosh**, A. Konar and A. K. Nagar, "Hemodynamic Analysis for Assessment of Scientific Creative Potential of Subjects Using Capsule Graph Neural Network," *2022 IEEE Symposium Series on Computational Intelligence (SSCI)*, Singapore, Singapore, **IEEE Flagship Conference**, 2022.

9. **S. Ghosh,** A. Konar, and M. Laha, "Hemodynamic Analysis for Assessing Creative Skill of Subjects Using Convolutional Neural Network," In 2023 8th International Conference on Computers and Devices for Communication (CODEC), pp. 1-2, IEEE, 2023, **(IEEE Best Paper Award in the track Signal Processing)**.

Sayantani Ghosh

10. **S. Ghosh**, A. Konar and A. K. Nagar, "Analyzing the Creative Potential of Subjects Using EEG-Induced Capsule Graph Neural Network," *2024 International Joint Conference on Neural Networks (IJCNN)*, Yokohama, Japan, pp. 1-8, **IEEE Flagship Conference**, 2024.

11. **S. Ghosh,** A. Konar and A. K. Nagar, "Computational Creativity by Generative Adversial Network with Leaked Information," *2024 International Joint Conference on Neural Networks (IJCNN)*, Yokohama, Japan, pp. 1-8, **IEEE Flagship Conference**, 2024.

PAPERS PRESENTED

1. **S. Ghosh**, M. Laha, A. Konar, P. Rakshit and A. K. Nagar, "Vowel Sound Imagery Decoding by a Capsule Network for the Design of an Automatic Mind-Driven Type-Writer," *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, UK, **IEEE Flagship Conference**, 2020.

2. **S. Ghosh**, L., Ghosh, A. Konar, and A.K. Nagar, "Assessment of Subjective Creativity Skill Using EEG Induced Capsule Network," In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, Canberra, ACT, Australia, 2020., pp. 3107-3114, **IEEE Flagship Conference**, 2020.

3. **S. Ghosh**, M. Laha, A. Konar and A. K. Nagar, "Decoding of Subjective Pain-Sensitivity by Brain Signal Analysis Using a General Type-2 Fuzzy Classifier," *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, Canberra, ACT, Australia, 2020., pp. 3107-3114, **IEEE Flagship Conference**, 2020.

4. **S. Ghosh**, A. Konar, and A.K. Nagar, "Decoding subjective creativity skill from visuo-spatial reasoning ability using capsule graph neural network," In *2021 International Joint Conference on Neural Networks (IJCNN)*, Shenzhen, China, pp. 1-8, **IEEE Flagship Conference**, 2021.

5. **S. Ghosh**, S. Bose, R. S. Chowdhury, A. Konar and A. K. Nagar, "Decoding the Creative Ability of Subjects from Aesthetic Quality Assessment Using Dual Convolution Induced Capsule Network," *2022 IEEE Symposium Series on Computational Intelligence (SSCI)*, Singapore, Singapore, **IEEE Flagship Conference**, 2022.

6. **S. Ghosh**, A. Konar and A. K. Nagar, "Hemodynamic Analysis for Assessment of Scientific Creative Potential of Subjects Using Capsule Graph Neural Network," *2022 IEEE Symposium Series on Computational Intelligence (SSCI)*, Singapore, Singapore, **IEEE Flagship Conference**, 2022.

Sayantani Ghosh

7. **S. Ghosh,** A. Konar, and M. Laha, "Hemodynamic Analysis for Assessing Creative Skill of Subjects Using Convolutional Neural Network," In 2023 8th International Conference on Computers and Devices for Communication (CODEC), pp. 1-2, IEEE, 2023, **(IEEE Best Paper Award in the track Signal Processing)**.

8. **S. Ghosh,** A. Konar and A. K. Nagar, "Analyzing the Creative Potential of Subjects Using EEG-Induced Capsule Graph Neural Network," *2024 International Joint Conference on Neural Networks (IJCNN)*, Yokohama, Japan, pp. 1-8, **IEEE Flagship Conference**, 2024.

9. **S. Ghosh,** A. Konar and A. K. Nagar, "Computational Creativity by Generative Adversial Network with Leaked Information," *2024 International Joint Conference on Neural Networks (IJCNN)*, Yokohama, Japan, pp. 1-8, **IEEE Flagship Conference**, 2024.

OTHER PUBLICATIONS

1. R.S. Chowdhury, S. Bose, **S. Ghosh**, and A. Konar, "Attention Induced Dual Convolutional-Capsule Network (AIDC-CN): A deep learning framework for motor imagery classification," *Computers in Biology and Medicine*, vol. 183, p.109260, 2024.

2. L. Ghosh, **S. Ghosh**, A. Konar, P. Rakshit and A. K. Nagar, "Decoding of EEG Signals Using Deep Long Short-Term Memory Network in Face Recognition Task," *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, Bangalore, India, pp. 477-483, 2018.

3. M. Laha, **S. Ghosh**, A. Bagchi, S. Pramanick and A. Konar, "Decoding of Brain Signals to Detect Perceived Color-Stimuli using Convolutional Neural Network," *2019 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, Chennai, India, pp. 425-429, 2019.

4. R. Das, S. Goswami, **S. Ghosh**, M. Laha, C. Debnath and A. Konar, "Brain Signal Analysis for Mind Controlled Type-Writer Using a Deep Neural Network," *2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET)*, Chennai, India, pp. 149-153, 2020.

5. I. Paul, **S. Ghosh** and A. Konar, "Voice Command Decoding for Position Control of Jaco Robot Arm using a Type-2 Fuzzy Classifier," *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, Bangalore, India, pp. 1-6, 2020.

6. C. Debnath, M. Das, **S. Ghosh** and A. Konar, "Hemodynamic Analysis for Touch

Sayantani Ghosh

Induced Object Recognition using Convolutional Neural Network," *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*, Bangalore, India, pp. 1-6, 2020.

7. S. Bose, **S. Ghosh**, A. Konar, and A.K. Nagar, "Decoding the Confidence Level of Subjects in Answering Multiple Choice Questions Using EEG Induced Capsule Network," In *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1-8, **IEEE Flagship Conference**, 2021. *(IEEE Brain Best Paper Award Runner Up)*.

8. M. Laha, **S. Ghosh** and A. Konar, "Exploration of Depth Perception in Human Binocular Vision using EEG-Based Neuro-Fuzzy Classifier," 2023 *8th International Conference on Computers and Devices for Communication (CODEC)*, Kolkata, India, pp. 1-2, 2023.

9. S. Datta, S. Sengupta, T. Das, S. Dasgupta, **S. Ghosh** and A. Konar, "A Deep Learning Approach for Position Control of a Mobile Robot," 2023 *8th International Conference on Computers and Devices for Communication (CODEC)*, Kolkata, India, pp. 1-2, 2023.

10. A. Nandi, R. Halder, **S. Ghosh** and A. Konar, "Cluster Shape Detection Using Dual Attention Induced Convolutional-Clustering Algorithm," *2024 IEEE International Conference on Information Technology, Electronics and Intelligent Communication Systems (ICITEICS)*, Bangalore, India, pp. 1-6, 2024.

11. A. Konar, and **S. Ghosh**, "Computational Creativity by Heuristic Search and Machine Learning", *Proceedings of the 2024 Sixth Doctoral Symposium on Intelligence Enabled Research (DoSIER 2024)*, Dhupguri, Jalpaiguri, West Bengal, India, November 28-29, 2024.

12. S. Sanyal, A. Bhattacharjee, D. Konar, **S. Ghosh** and S. Chatterjee, "ASIC Design of an Interval Type-2 Fuzzy Logic Engine for Control Application," *2025 Devices for Integrated Circuit (DevIC)*, Kalyani, India, pp. 101-106, 2025.

PATENTS: Nil

Sayantani Ghosh

"Statement of Originality"

I, Sayantani Ghosh, registered on 28/03/2023 (Index No. 238/23/E) do hereby declare that this thesis entitled "**Exploration of Biological and Computational Modalities of Scientific Creativity**" contains literature survey and original research work done by the undersigned candidate as part of doctoral studies.

All information in this thesis have been obtained and presented in accordance with existing academic rules and ethical conduct. I declare that, as required by these rules and conduct, I have fully cited and referred all materials and results that are not original to this work.

I also declare that I checked this thesis as per the "Policy on Anti Plagiarism, Jadavpur University, 2019", and the level of similarity as checked by iThenticate software is 5%.

Signature of Candidate: Sayantani Ghosh Date: 16.9.25

Certified by Supervisor:
(Signature with date, seal)

A Konar 16/09/25

PROF. AMIT KONAR
CO-ORDINATOR
INTELLIGENT AUTOMATION
& ROBOTICS
E.T.C.E. DEPTT., J.U., KOL-32

Prof. Amit Konar

Professor

Department of Electronics and Tele-communication Engineering

Jadavpur University

Kolkata-700032, India

CERTIFICATE FROM THE SUPERVISORS

This is to certify that the thesis entitled "**Exploration of Biological and Computational Modalities of Scientific Creativity**" submitted by **Sayantani Ghosh,** who got her name registered on 28/03/2023 (Index no.: 238/23/E) for the award of Ph.D. (Engineering) degree of Jadavpur University, is absolutely based upon her own work under the supervision of **Prof. Amit Konar** and that neither her thesis nor any part of the thesis has been submitted for any Degree/Diploma or any other academic award anywhere before.

Akonar
16/09/25

PROF. AMIT KONAR
CO-ORDINATOR
INTELLIGENT AUTOMATION
& ROBOTICS
E.T.C.E. DEPTT., J.U., KOL-32

Prof. Amit Konar

Professor

Department of Electronics and Tele-communication Engineering,

Jadavpur University

Kolkata-700032, India.

Acknowledgement

I extend my sincere gratitude to the administration of Jadavpur University for providing a conducive academic and vibrant environment for pursuing the present doctoral research. I am also deeply indebted to Prof. *Sudhabindu Ray*, Head of the Department of Electronics and Tele-Communication Engineering (ETCE), Prof. *Manotosh Biswas*, former Head of the Department, and Prof. *Mrinal Kanti Naskar*, all from Jadavpur University, for their invaluable technical guidance and moral support during the course of this Ph.D. research. Their encouragement and unwavering support have been instrumental in the successful completion of this work, for which I remain profoundly grateful.

I owe my sincere and heartfelt gratitude to my supervisor, Prof. *Amit Konar*, who has been a friend, philosopher, and guide throughout the entire journey of this research. He has provided unswerving support at every stage of the thesis, ranging from the theoretical aspects of model design and brain-imaging experiments to the interpretation of results, as well as the writing and review of research papers derived from this work. His mentorship has taught me to explore new avenues and to approach unknown problems from multiple perspectives in search of feasible solutions. His caring attitude and constant encouragement have been a source of motivation, inspiring me to dedicate my time and energy wholeheartedly to this Ph.D. research.

The realization of this extensive laboratory work would not have been possible without the wholehearted cooperation and support of my co-researchers and teammates: Mrs. *Mousumi Laha*, Mrs. *Lidia Ghosh*, Mr. *Arnab Rakshit*, and Ms. *Baishali De* in our AI Laboratory. Their presence made my stay both enjoyable and memorable, and taught me the art of working as part of a team. I would also like to acknowledge the contributions of the postgraduate students: Mr. *Apurba Nandi*, Ms. *Priyanka Paul*, Mr. *Rupayan Halder*, Ms. *Ipsita Paul*, and Ms. *Chandrima Debnath*, and the undergraduate students: Mr. *Shirsha Bose*, Mr. *Ritesh Sur Chowdhury*, Mr. *Shraman Pramanick*, and Mr. *Anurag Bagchi*, who closely interacted with me in connection with their research in the AI Laboratory and extended their unswerving support.

I owe my deepest gratitude to my parents for their constant support, encouragement, and unconditional love throughout the course of my life and academic journey. Words such as ‘thanks’ and ‘gratitude’ are far too small to express the depth of my feelings for my father, Mr. *Tapan Kr. Ghosh*, and my mother, Mrs. *Soma Ghosh*. Their unwavering faith in me, their

sacrifices, and their active support and inspiration have been the true driving forces behind the successful completion of this thesis. It is their patience, moral strength, and continuous motivation over the long years of my research that have enabled me to overcome challenges and bring this work to fruition in its present form.

I also express my heartfelt gratitude to the Almighty for showering me with blessings, for being my ultimate guide and protector, and for granting me the strength to persevere through every difficulty. I am deeply thankful for the privilege of being surrounded by such wonderful people, whose presence, affection, and encouragement have remained a constant source of inspiration in both my personal life and academic pursuits.

Finally, I extend my sincere thanks to all my well-wishers who, in various ways, have contributed directly or indirectly to the successful completion of this work. Their encouragement, guidance, and goodwill have continually strengthened my determination throughout this journey. I remain profoundly grateful for their enduring support, which will continue to inspire my future endeavors.

Sayantani Ghosh

Sayantani Ghosh

Artificial Intelligence Lab.,
Department of Electronics and Tele-Communication
Engineering, Jadavpur University, Kolkata-32, India

Preface

Traditionally, creativity has been closely associated with the arts, encompassing diverse forms of human expression such as music, painting, poetry, performance, and even culinary practices. Scientific creativity, however, emerges as a distinct dimension of the well-accepted notion of creativity. Unlike artistic creativity, scientific creativity deals with the discovery and/or innovation of knowledge within the realm of science. It may be involve the formulation of new theories, the design of novel experiments, and similar other contributions that aids in shaping human progress. For instance, scientific creativity has been the driving force behind groundbreaking discoveries such as the laws of motion, the theory of evolution, and the structure of DNA, as well as transformative innovations including electricity, vaccines, and space exploration.

Scientific creativity can be studied from two modalities: biological and computational. The biological modality focuses on understanding the neural underpinnings of creativity by examining the interaction of different brain regions or modules that function either independently or collaboratively during various stages of the creative process such as genesis of novel ideas, their elaboration, refinement, and eventual transformation into feasible outcomes. In contrast, the computational modality deals with the artificial synthesis of creativity by a machine through the utilization of different computational techniques/algorithms. It is exemplified by the automated generation of original ideas, genesis of unconventional solutions, and the development of innovative formulations in the scientific domain, thereby simulating aspects of human creative thought.

The thesis deals with the analysis of the above two modalities that help in exploring both the neural underpinnings and the artificial genesis of scientific creativity. The first part of the thesis focuses on the biological modalities of scientific creativity, where the research challenge lies in analyzing the interconnections among different brain lobes that participate during the generation of a scientific creative outcome, and subsequently classifying individuals into different levels of creativity. The main steps involved in fulfilling the above objective include: i) acquisition of brain signals, ii) transforming the acquired signals into connectivity maps/networks, iii) analysis of the connectivity maps to understand the activation of different brain lobes during a creative outcome, and iv) classification of the formulated networks into different levels of creative potential.

As discussed, the first step involved in the exploration of biological modalities of creativity is the acquisition of brain signals from the scalp of participants. Brain signal acquisition devices can be broadly classified into two categories based on their

resolution: i) high temporal resolution and ii) high spatial resolution. Electroencephalographic (EEG) devices provide high temporal resolution by recording the brain's electrical activity through electrodes. In contrast, functional magnetic resonance imaging (fMRI) offers high spatial resolution by capturing detailed structural and functional information of brain regions, although its high cost limits practical use. As an alternative, the present work employs functional near-infrared spectroscopy (fNIRS), a spatial resolution dominant device that measures changes in oxygenated and deoxygenated hemoglobin concentrations in cortical regions, thereby enabling localized monitoring of brain activation at lower cost and with greater feasibility than fMRI. Following signal acquisition, the data are transformed into a connectivity network by computing correlations among pairs of acquisition channels, facilitating the exploration of interconnections among different brain lobes during the said cognitive task. Network-based features are then extracted to identify the active brain regions or Brodmann Areas (BAs) involved in the creative process. Finally, these features are classified using a suitable classifier to assign subjects to different levels of scientific creative potential.

The second part of the thesis addresses the computational aspect of scientific creativity, i.e., its automated synthesis by a machine or computer. While the artificial synthesis of scientific creativity encompasses a wide range of applications, such as the development of new mathematical theorems, the generation of novel formulas in physics, and the like, the present thesis exemplifies this modality through the automatic generation of new mathematical identities. Here, the main research challenge lies in designing a system capable of automatically generating new mathematical identities that can serve as chapter-end exercises, while ensuring a significant degree of diversity and/or variation compared to existing textbook problems.

Two distinct approaches have been investigated for the automatic generation of diverse mathematical identities. The first approach is based on an extended version of the best-first search (BFS) strategy, conducted over an OR-tree structure, with a provision for estimating the degree of diversification of the generated mathematical identities with respect to their predecessors. The most notable aspect of this framework lies in preserving diversity among the generated identity problems during successive tree expansions, while ensuring that the selected mathematical identities remain novel across multiple executions of the search strategy. The second approach is grounded in a modified Generative Adversarial Network (GAN) based framework. In the above framework, the generator deals with the creation of new mathematical identity problems, while the discriminator determines their originality by categorizing them as real (i.e., a novel and correct identity) or fake (i.e., similar to an

existing sample) with respect to a given dataset. Thus, the aforesaid adversarial scenario provides an opportunity for the generator to successively improve so that the mathematical identities generated are not only valid but also novel. An additional provision of this framework lies in the implicit guidance of the discriminator that aids the generator to maintain diversity among the newly generated identities, ensuring that they differ meaningfully from existing textbook problems.

The thesis is organized into seven chapters. Chapter 1 provides a comprehensive review of existing approaches to assess the biological modalities of scientific creativity, as well as methods for its computational synthesis. From Chapter 2 onwards, the thesis is divided into two parts: Part I (Chapters 2-4) presents the candidate's original contributions under the biological modalities of scientific creativity, while Part II (Chapters 5-6) focuses on the computational modalities of scientific creativity, also representing the candidate's original contributions. Specifically, Chapter 2 covers the categorization of subjects into distinct levels of scientific creative potential based on their spatial reasoning ability through brain network analysis using an fNIRS device. Chapter 3 examines the differentiation between creative and analytical mindsets through analysis of the Eureka effect using EEG acquisition system. Chapter 4 addresses the classification of subjects according to their levels of scientific creative ability based on analogical reasoning potential using fNIRS-based brain network analysis. Chapter 5 presents a depth-limited BFS algorithm designed to generate diverse mathematical identity problems suitable for inclusion as chapter-end exercises in mathematics textbooks while Chapter 6 introduces a GAN based framework in which new mathematical identities are generated using leaked information transmitted from the discriminator to the generator to enhance problem diversity. The final chapter i.e., Chapter 7 reviews the research works undertaken, highlights the conclusions, and offers final remarks on the usability of the outcomes, while also outlining potential future extensions of the proposed works. Extracts of the main contributions of the thesis presented in Chapters 2, 3, 4, 5, and 6 are summarized below.

Chapter 2 focuses on the classification of the scientific creative potential of individuals into four distinct grades based on their spatial reasoning ability. Spatial reasoning, defined as the capacity to mentally manipulate an object's attributes such as shape, orientation, pattern, and size, is a key cognitive process that strongly influences scientific creativity. For instance, an architect aiming to design a new surface development algorithm requires strong spatial reasoning skills, while the production of new formulations in fields of mathematics such as trigonometry, calculus, and arithmetic also relies heavily on advanced spatial ability. Thus, spatial reasoning can serve as an early marker of creativity in the scientific domain. In this

chapter, different levels of creative potential are identified through a popular spatial reasoning task referred to as Mental Paper Folding (MPF), using brain connectivity networks constructed from fNIRS data. The above methodology is carried out in three phases. In the first phase, brain connectivity networks are built using Pearson's correlation to capture the interrelationships among different brain modules. The second phase involves the extraction of three centrality features (degree, closeness, and betweenness) to identify the most active BAs engaged in the MPF task. In the third phase, the extracted features are classified into four levels of creative potential using a proposed Graph Convolutional-Interval Type-2 Fuzzy Network (GC-IT2FN).

The novelty of the proposed classifier lies in three aspects. First, a self-attention mechanism is introduced within the architecture to allow the graph convolutional layers to focus on the most relevant nodal features to enhance classifier performance. Second, a new activation function called Logish is employed after graph convolution to improve both accuracy and stability of the classifier. Third, a novel mapping function is designed to select the most promising region within the Footprint of Uncertainty (FOU) of the used fuzzy sets, thereby mitigating the effect of uncertainty intrinsic within brain data. Experimental evaluation shows that the proposed GC-IT2FN achieves a classification accuracy of 97.54% and also performs significantly better than its conventional counterparts. In addition, extensive ablation studies also confirm the individual contribution and effectiveness of each module included in the classifier.

Chapter 3 investigates the differentiation between creative and analytical mindsets through the analysis of brain responses recorded using EEG. The motivation for the above study stems from the growing importance of scientific creativity in industrial research and development, where innovative thinkers are in high demand, alongside the equally vital role of analytical individuals in fields such as investigation, academia, and management. Unfortunately, utilizing only behavioral measures to distinguish between the above two cognitive styles are often unreliable, which makes the analysis of neural responses a more powerful alternative. In this work, participants undertook a creativity test designed to elicit the Eureka effect, a sudden insight experience that marks the emergence of creative thought. Based on their brain activity during the task, individuals were categorized into four groups: Analytical, High Creative, Medium Creative, and Low Creative. The classification framework was implemented in two stages. First, brain connectivity maps were constructed from EEG data using Pearson's correlation to capture functional interactions between different brain regions. Next, three centrality features (degree, closeness, and betweenness) were extracted from the connectivity maps and provided as input to a novel classifier.

The proposed classifier model referred to as Dual Attention Induced Graph Convolutional-Capsule Network (DAIGC-CapsNet), integrates a Graph Convolutional Network (GCN) which learns representations of connectivity-based features, with a Capsule Network (CapsNet) which performs the final classification. The design of DAIGC-CapsNet includes two key innovations. First, a dual attention mechanism is introduced, where the first one guides the convolutional layers to emphasize the most critical nodes, while the second one ensures only that the most relevant information is transmitted from the primary capsules to the class capsules. Second, a Sparsemax-based routing algorithm is incorporated to strengthen the coupling between the adjacent capsule layers, thereby enhancing both interpretability and stability of the model. Experimental analysis demonstrates that the proposed framework achieves an average classification accuracy of 96.76%, significantly outperforming the state-of-the-art (SOTA) baselines. Statistical validation using the Friedman's test further confirms the superior performance of DAIGC-CapsNet.

Chapter 4 deals with the classification of subjects into five levels of scientific creative ability based on their proficiency in analogical reasoning. Analogical reasoning refers to the capability to identify structural or functional similarities between one or more distinct concepts to create something new and plays a vital role in influencing scientific creativity. For instance, Niels Bohr compared the structure of an atom to the solar system that laid the groundwork for quantum mechanics. Similarly, William Harvey discovered the blood circulation in the human body by drawing an analogy with water being pumped through a building. Again, René Descartes revolutionized mathematics by inventing the coordinate system by linking geometry and algebra. Thus, the above instances demonstrate how analogical reasoning contributes to the emergence of novel scientific insights.

The classification task in this chapter is carried out by analyzing the brain connectivity patterns of subjects engaged in Raven's Advanced Progressive Matrices (RAPM) test, a widely used analogical reasoning test, with data acquired through fNIRS. The proposed framework progresses through three phases. In the first phase, brain connectivity networks are constructed using Wavelet Transform Coherence (WTC). The second phase involves the extraction of three node-based features, namely strength, efficiency, and betweenness, to identify the most active BAs during the RAPM task. In the third phase, these features are classified into five grades of creative potential using a novel classifier model referred to as Enhanced Graph Convolution Induced Type-2 Fuzzy Classifier (EGC-IT2FC).

The proposed classifier introduces five important innovations. First, an enhanced graph convolution operation is developed to capture both local and global structural characteristics of the brain network, enabling more comprehensive feature learning

by the classifier. Second, a new activation function named Smish is applied after the convolution layer to improve classification accuracy. Third, instead of a conventional pooling layer, a one-dimensional spatial convolution is employed to preserve essential information from the convolved features. Fourth, a new mapping function is designed to reduce uncertainty among the spatial-convolved features. Fifth, the Takagi-Sugeno-Kang (TSK) fuzzy reasoning mechanism is incorporated to minimize computational expense. Experimental results obtained confirm that the proposed framework achieves superior performance in comparison to the traditional approaches and yields a classification accuracy of 97.78%. Furthermore, statistical validation using the Wilcoxon's signed rank test further supports the robustness and effectiveness of the proposed classifier.

Chapter 5 introduces the first approach to artificially synthesizing computational creativity, exemplified through the automatic generation of new mathematical identities. An extended best-first search (BFS) strategy is proposed, designed to assess the diversity of newly generated identities with respect to their predecessors. This strategy operates on an OR-tree structure in which nodes represent new mathematical identities or creative outputs, and edges represent the substitution rules applied to derive child nodes from parent nodes. Two algorithmic variants are developed under this framework: Relative Diversity-Guided Best-First Search (RDGBFS) and Absolute Diversity-Guided Best-First Search (ADGBFS). These variants differ in the cost functions used to evaluate the diversity of the generated node (identity). The RDGBFS maximizes diversity relative to the immediate parent, while ADGBFS maximizes the difference between a node's diversity score and a penalty term with reference to the root node.

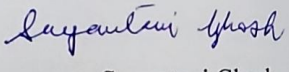
A key contribution of this work lies in preserving diversity during iterative expansions of the search tree, while also ensuring novelty across different runs of the algorithm. The chapter further explores several theoretical properties of RDGBFS and ADGBFS, clarifying their individual significance as well as their comparative strengths in handling varied scenarios. Additionally, the soundness and logical completeness of the algorithms are also established, providing evidence of the robustness of the proposed algorithms. Experimental evaluations demonstrate that both the algorithms are able to generate identities that are substantially more diverse than traditional textbook problems, while also surpassing existing SOTA techniques. Their practical utility is highlighted through successful deployment in creating chapter-end exercises across multiple domains, including propositional logic, set theory, and algebra, thereby underscoring the broader potential of the proposed framework.

Chapter 6 focuses on the exploration of computational creativity in the scientific domain by introducing a more modern framework compared to the one presented in Chapter 5, again demonstrated through the automatic generation of diverse trigonometric identity problems. The motivation for this framework arises from the limitations of earlier approaches such as search based algorithms, template driven methods, and even fine tuned Large Language Models (LLMs) which often yield problems that are too similar to standard textbook examples. This lack of diversity is largely due to the restricted set of transformation rules, the rigidity of templates, or the absence of explicit mechanisms for encouraging variety in the generated outputs. To overcome the above shortcomings, a novel framework called Enhanced Generative Adversarial Network with Leaked Information (E-LeakGAN) is proposed. This framework retains the classical GAN setup of a generator and a discriminator engaged in adversarial training. The generator attempts to create new identity problems, while the discriminator evaluates them by distinguishing between authentic samples and generated ones. Through this iterative competition, the generator progressively improves its ability to produce both correct and original outputs. Beyond simple classification, the discriminator in E-LeakGAN plays an additional role by leaking high-level features back to the generator. These features serve as guiding signals, ensuring that the generated identities remain both coherent and diverse. Additionally, the generator here also departs from traditional GANs by adopting a hierarchical architecture composed of two modules, the manager and the worker. The manager processes the leaked features to construct a goal vector that captures key trigonometric relationships and provides strategic direction for the generation process. The worker then uses this goal vector to assemble identity components in a logically consistent and mathematically meaningful manner. Thus, the integration of leaked information with a hierarchical generator design enables the framework to move beyond surface level variations and facilitates the creation of semantically rich and structurally diverse mathematical identities.

Since the leaked features act as implicit guidance for the generator to ensure diversity in the outcomes, the proposed framework incorporates an enhanced discriminator module built on a convolutional neural network (CNN) architecture. The discriminator network of the E-LeakGAN model is organized into three main layers. The first layer applies a convolution operation to the input, which consists of a manually curated dataset of standard textbook identity problems. This operation extracts both semantic and syntactic features from the data. The convolution is followed by a new activation function called Trish, which produces small but informative outputs even for negative inputs. This property allows the network to capture subtle variations and learn more nuanced and complex patterns from the

convolved features. The second layer introduces a dual attention mechanism designed to highlight the most relevant information in the convolved features. This mechanism contains two sequential modules, namely channel attention and spatial attention. The channel attention module identifies important semantic features such as function types, algebraic structures, factorization patterns, and reciprocal relations. The spatial attention module then focuses on the syntactic placement of these features, for instance, whether they appear on the left hand side or right hand side of the identity. The attention maps produced by these two modules are fused with the original convolved features, creating refined feature representations that allow the network to concentrate more effectively on critical semantic and syntactic content. The third layer applies a novel pooling method called eigenvector guided pooling (EGP). The role of this layer is to reduce the spatial dimensions of the features while retaining the most meaningful semantic and syntactic content related to trigonometric identities, such as reciprocal relationships and complementary functions. EGP achieves this by projecting the attention enhanced convolved features onto the eigenvector associated with the largest eigenvalue, thereby aligning them in the direction of maximum semantic-syntactic relevance.

The experimental evaluation undertaken reveals the superior performance of the E-LeakGAN framework compared to its traditional counterparts. Additionally, the proposed model not only generates accurate and diverse trigonometric identity problems but also demonstrates robustness and adaptability when applied to other mathematical domains such as algebra and calculus. Thus, the generalizability of the E-LeakGAN model highlights its potential as an effective tool for the automatic generation of mathematical problems to be used as chapter-end exercises and/or worksheets, thereby enhancing the overall quality of mathematics education.



Sayantani Ghosh

Artificial Intelligence Lab.,
Department of Electronics and Tele-Communication Engineering, Jadavpur University, Kolkata-32, India

Table of Contents

Chapter 1: An Overview of Scientific Creativity and Its Models

Part I: Biological Modalities of Scientific Creativity

Chapter 2: Assessment of Scientific Creative-Skill from Spatial Reasoning Ability Using Graph Convolutional-Interval Type-2 Fuzzy Network

Chapter 4: Assessment of Scientific Creative-Potential from Analogical Reasoning Ability Using Brain-Network Based Deep-Fuzzy Classifier

Part II: Computational Modalities of Scientific Creativity

Chapter 5: Computational Creativity by Diversity-Optimized Intelligent Search: An Automatic Approach to Artificial Synthesis of Trigonometric Identities

Chapter 6: Enhanced Generative Adversarial Network with Leaked Information (E-LeakGAN): A Deep Learning Framework for Synthesizing Computational Creativity in Scientific Domain

Chapter 7: Conclusions and Future Directions

List of Figures

List of Tables

1

An Overview of Scientific Creativity and Its Models

The chapter begins with a philosophical overview of creativity, highlighting the well-established perspectives of Wallas, Guilford, Rhodes, and Kaufmann. It then moves on to define various types of creativity, followed by an in-depth exploration of scientific creativity and the associated philosophical viewpoints. The chapter further examines the biological modalities of scientific creativity, including brain network modeling, Markov models, and information theory-based frameworks. It then delves into computational approaches to scientific creativity, such as random search, inductive learning, and the use of large language models to simulate creative processes. Finally, the scope of the thesis is elaborated in detail, offering readers both the motivation behind the work and a comprehensive overview of the topics addressed.

1.1 INTRODUCTION

Creativity refers to the ability to think beyond conventional patterns, envisioning novel possibilities that challenge established norms or build upon existing knowledge. In other words, it denotes the capability to generate ideas, solutions, or works that are original and unique within a specific context [1-3]. It is not merely the ability to come up with something new but also involves ensuring that the ideas generated are meaningful, practical, and relevant to the context in which they arise [4-5]. The genesis of novel ideas is often influenced by several cognitive factors. For example, spatial reasoning [6-7] facilitates the mental manipulation and visualization of concepts in ways that might not be immediately apparent, leading to the generation of new patterns or solutions. Similarly, analogical reasoning [8-9] fosters the ability to draw connections between seemingly unrelated concepts, thereby facilitating the transfer of insights across domains and enabling the emergence of innovative ideas. Additionally, deductive reasoning [10-11], the ability to draw conclusion from a set of premises, also influences the formulation of novel theories or hypotheses. Thus, creativity is a concoction of diverse cognitive skills that helps to bring forth something new and significant, ultimately contributing to innovation and progress.

The word “creativity” originates from the Latin word “creare”, which means “to bring forth” [12-13]. Initially, the notion of creativity was not linked to human innovation but was associated with divine acts and the generative forces of nature. Early interpretations of creativity were thus deeply spiritual, often regarded as expressions of the divine power. This perspective is vividly illustrated in the Hebrew Bible, which begins with the narrative of God creating the heavens and the earth out of nothing (*ex nihilo*)—a supreme act of the omnipotent authority. Since humans were believed to be created in God’s image, they were thought to possess a limited capacity to generate something unique or novel. Human beings were seen as fulfilling divine commands, such as being fruitful and multiplying. Consequently, human creativity was viewed as subordinate and constrained, operating within the framework of the divine guidance [14].

During the Renaissance, the traditional view that creativity was solely the act of the omnipotent underwent a transformative shift. This period marked the historical era where creativity started being recognized as a human endeavor [15]. Technological breakthroughs, such as the printing press, revolutionized the way ideas were dispersed, making knowledge more accessible and fostering a new path of intellectual and artistic exploration. This era also witnessed the discovery of the New World and the expansion of global trade, leading to the emergence of modern capitalism [16]. Furthermore, the Renaissance period also saw the evolution

scientific innovation and exploration [17]. Figures such as Galileo Galilei and Nicolaus Copernicus began to challenge the reigning dogma of science, clearing the way for accepting fact-based reasoning. Again, one of the great minds of the Renaissance was Leonardo da Vinci, who exemplified his excellence in a wide variety of disciplines. His masterpieces include the "Mona Lisa" and "The Last Supper", which expressed his artistic brilliance, while his research on flying machines and other military inventions contributed to the development of knowledge in both science and engineering. Therefore, this period laid down the foundation for future generations to be creative in a wide variety of realms, starting from art and science to business and technology.

Following the Renaissance, the Industrial era began where creativity became a driving force behind innovation and progress, particularly in manufacturing, engineering, and design [18]. The rapid technological advancements of the 18th and 19th centuries, such as the steam engine, mechanized textiles, and the assembly line, showcased fresh approaches to improving efficiency, productivity, and quality. Visionary inventors like James Watt and Henry Ford emerged, whose creative ideas led to transformative breakthroughs that reshaped economies and societies on a global scale [19]. After the Industrial Era, the Digital Era followed, where creativity evolved in response to the vast expansion of digital technologies, like the internet, computer and other information systems [20]. Thus, the Digital era paved the way for utilizing creativity in diverse fields such as artificial intelligence (AI), digital art, online content creation and many more. In other words, the Digital era extended creativity beyond physical inventions, to include digital experiences, software solutions, and algorithms, thus expanding the scope and accessibility of creative potential.

The rest of the thesis is divided into thirteen sections. Section 1.2 discusses the philosophical models of creativity. Section 1.3 examines the types of creativity, and Section 1.4 outlines the philosophical contributions to scientific creativity. Section 1.5 discusses the biological modalities of scientific creativity. Sections 1.6, 1.7, and 1.8 cover brain network modeling, Markov models, and information theory-based models for representing the biological modality of scientific creativity, respectively. Section 1.9 discusses the computational modality of scientific creativity. Sections 1.10, 1.11, and 1.12 address the random search approach, inductive learning, and large language models (LLMs) used to synthesize computational creativity, respectively. Finally, Section 1.13 reviews the existing literature on both the biological and computational modalities of scientific creativity, while Section 1.14 outlines the scope of the thesis.

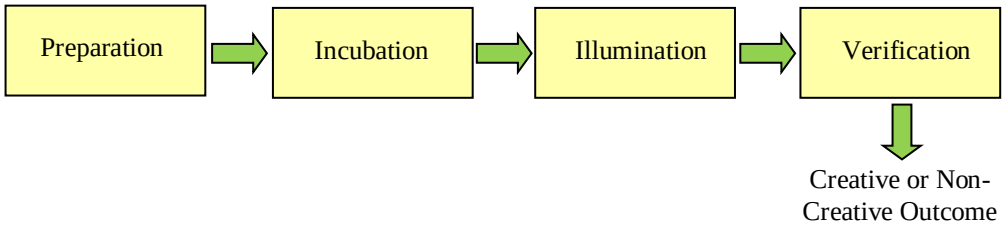


Fig. 1.1. Block diagram of the Wallas's model of creativity

1.2 PHILOSOPHICAL MODELS OF CREATVITY

Literature [21-28] highlights several philosophical models of creativity that explain the processes involved in achieving a creative outcome. This section provides a selective illustration of some widely recognized models of creativity.

1.2.1 The Wallas Model of Creativity

The Wallas Model of Creativity [24] is one of the most influential models that describes the creative process. This model involves a four-stage process that an individual goes through when engaged in a creative task. The stages of his model are illustrated in Fig. 1.1. and described as below.

(a) **Preparation**: The preparation stage is characterized by gathering information and learning about the problem at hand. This is a period where one can study or analyze the problem from different angles or perspectives. Thus, this phase acts as the first building block for the creative process as it involves knowledge construction and understanding.

(b) **Incubation**: In this stage, an individual goes astray and allows the subconscious mind to work on the problem. In other words, this stage often occurs outside of conscious awareness, where the mind continues to process the ideas in the background. Incubation may involve relaxation, mind-wandering, daydreaming, or engaging in unrelated activities that give the mind the freedom to make connections to generate new insights.

(c) **Illumination**: The illumination stage is where the creative idea or solution finally emerges, suddenly out of the blue. This stage is often characterized by the "Aha!" or the "Eureka" moment in which bits of information suddenly fall into place. In other words, illumination phase involves the sudden realization of the solution pertaining to the problem at hand.

(d) **Verification**: In this stage, an individual tests and verifies the solution of the problem found during the illumination stage. In other words, this stage makes sure that the idea created is useful and feasible in real world.

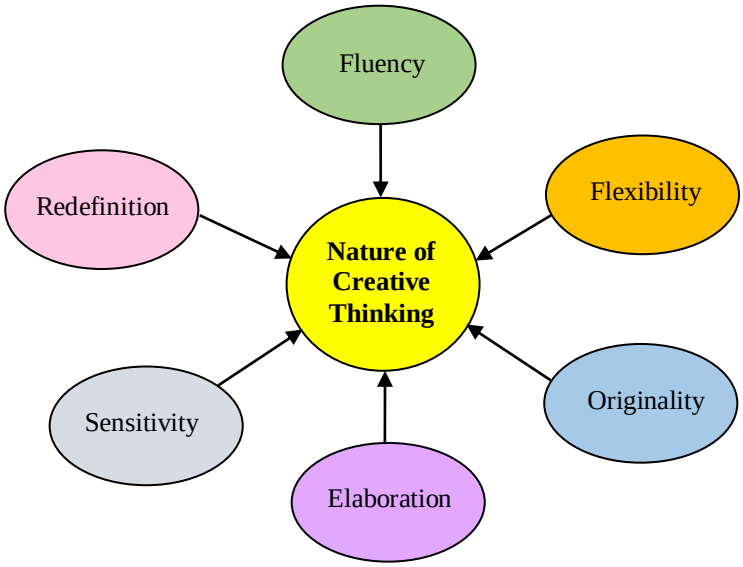


Fig. 1.2. Cognitive abilities influencing creative thinking according to Guildford's model

1.2.2 Guildford's Model of Creativity

Guildford's Model of Creativity, proposed by psychologist J.P. Guildford in the 1950s [25], remains one of the most influential frameworks for understanding creativity. His model asserts that creativity is a multifaceted and complex process, measurable and comprehensible through various cognitive components. Guildford hypothesized that creativity involves both convergent and divergent thinking, as detailed below.

(a) **Divergent Thinking**: This type of thinking involves the generation of a variety of possible solutions to open-ended problems. Guildford suggested that the emergence of numerous ideas or solutions without immediate evaluation helps to foster originality in thought process.

(b) **Convergent Thinking**: This type of thinking involves the selection of the most promising solution to a problem from a set of multiple possible options. For example, providing a single solution to a brain-teaser demonstrate the concept of convergent thinking.

Guildford's model also underscores several cognitive abilities that contribute to creative thinking, illustrated in Fig. 1.2 and elaborated as follows.

(i) **Fluency**: The capacity to generate a large number of ideas.

(ii) **Originality**: The capability to produce novel and unique ideas.

(iii) **Flexibility**: The capability to change perspectives and approach problems from different angles.

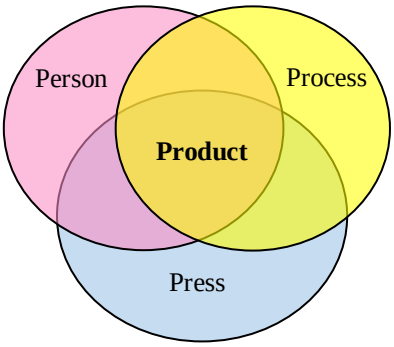


Fig. 1.3. Venn diagram depicting the four P model of creativity

(iv) **Elaboration**: The capability to expand or ornate ideas with great details.
(v) **Sensitivity**: The ability to recognize gaps in existing knowledge and identify the opportunities to fulfill such voids.
(vi) **Redefinition**: The capability to reinterpret problems in new ways that can shed light upon unique insights.

Thus, Guildford's model summarizes that creativity is not a single trait but rather a combination of different cognitive skills.

1.2.3 The Four P Model of Creativity

James Melvin Rhodes, a popular creativity researcher of mid 1900s, developed the four P model of creativity in 1961 [26-27]. In this model, he proposed that a creative product emerges by combining three key important components which includes person, process and press. Thus, product, person, process and press can be regarded as the main components of a creative model. Each of these components is elaborated below.

(a) **Person**: This constituent refers to the subject who participates in the creative process. It includes the individual's capabilities, traits, and personality aspects such as inquisitiveness, openness to new ideas, flexibility etc., that contribute towards creative thinking. Domain-specific knowledge or expertise also plays an important role in shaping a subject's creative ability.
(b) **Process**: Rhodes viewed creativity as a process which involves a series of phases. These phases generally include identification of a problem, critical thinking, idea illumination and verification. However, the concept of process is often not linear but rather involves multiple iterations and feedbacks for refining the idea generated.
(c) **Press**: It refers to the environment within which an individual tries to generate something novel. Such an environment includes the physical surrounding of the person (for example, workspace, residing area etc.) as well as the social ambience

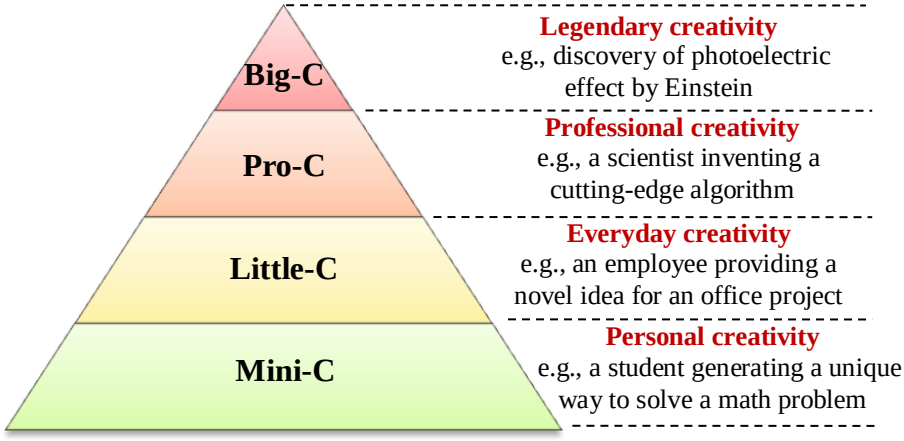


Fig. 1.4. Types of creativity according to the four C model

(such as influence of family and peers). Thus, press plays a vital role in either influencing the creative thought process by providing motivation or inhibiting it by imposing different constraints.

(d) **Product**: It refers to the final output of the creative thought process which is formed by the amalgamation of person, process and press as shown in Fig. 1.3. So, a product may be a new piece of painting, a unique mathematical formulation, a novel music composition or something entirely new.

1.2.4 The Four C Model of Creativity

The Four C Model of Creativity was proposed by James C. Kaufman and Ronald A. Beghetto in 2009 [28]. It provides a framework that defines the different types of creativity and outlines how they relate to one another in the development of a creative outcome. Before discussing the connections between the different varieties of creativity, a brief overview of the types is presented below and illustrated in Fig. 1.4.

(a) **Mini-C (Personal Creativity)**: It refers to the production of creative outcome at the individual level which is insightful and/or meaningful to the person but may not be widely recognized as creative by others. For instance, a student generating a unique way of solving a math problem or an individual creating a personal style of art are some examples of mini-c creativity.

(b) **Little-C (Everyday Creativity)**: It refers to the illustration of creativity in day to day tasks. It involves solving common problems or tasks in novel ways which is not necessarily associated intellectual expertise.

(c) **Pro-C (Professional Creativity)**: It refers to the demonstration of innovation by individuals who are experts in a particular field. for instance, a scientist conducting

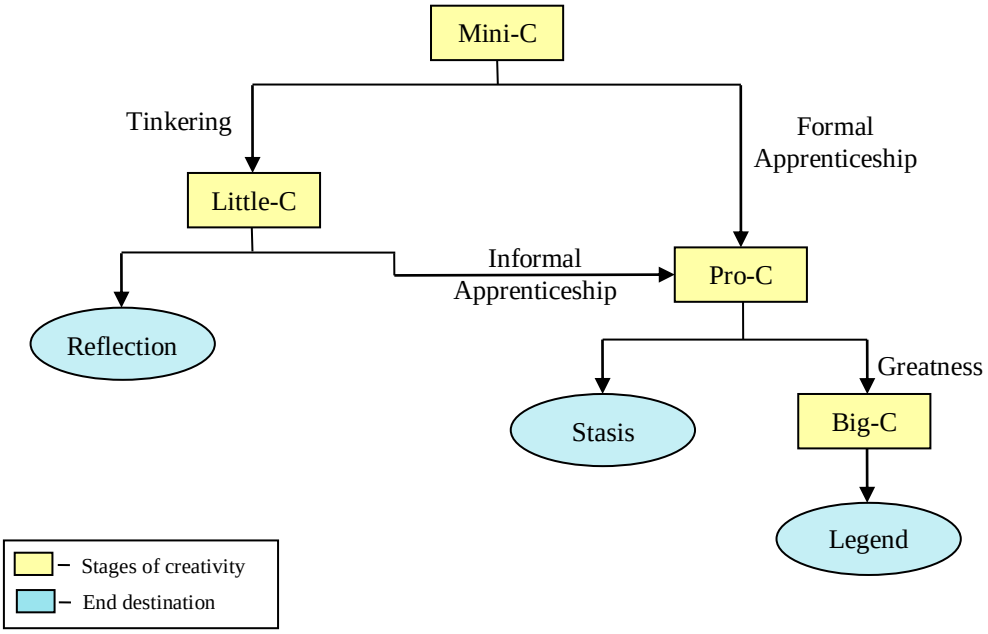


Fig. 1.5. Illustration of the four C model of creativity

cutting-edge research, a business professional developing new strategies, or an artist producing original works for a gallery are considered as little-C creativity.

(d) **Big-C (Legendary Creativity)**: It refers to the illustration of groundbreaking contributions to a particular domain. For instance, the discovery of the law of photoelectric effect by Einstein, Marie Currie's discovery of polonium and radium, and the Mona Lisa painting by Da Vinci are considered as illustrations of Big-C creativity.

Fig. 1.5 illustrates the Four C model. Here, creativity begins at the Mini-C stage, which represents the personal level of creative expression. From this point, individuals may undergo one of the two transitional stages. The first stage involves the Pro-C level, where an individual might reach by undergoing any formal apprenticeship ranging for a few months to years. Alternatively, an individual may reach the Little-C level via self-guided experimentation or "tinkering". Upon reaching the Little-C stage, two potential paths emerge. The first path leads to the stage of Pro-C where an individual might reach through an informal apprenticeship where the guidance is provided by an experienced mentor or colleague . The second path involves the stage of reflection where an individual can conclude their creative journey without pursuing it at a professional level. Again at the Pro-C level, individuals face two possible outcomes. Here, the first outcome involves a stage of stasis where some people might conclude their careers without making further notable contributions. On the contrary, the second outcome describes the illustration of exceptional creativity that can lead an individual to a level of greatness. Among

these, a selected few may, over time, be recognized by future generations as having reached the extraordinary stage of creativity i.e., Big-C level where the pinnacle of creative achievement is acquired.

1.3 TYPES OF CREATVITY

Creativity can be broadly categorized into artistic creativity and scientific creativity, with the former often referred to as non-scientific creativity [29]. The details regarding each type of creativity are discussed below.

1.3.1. Artistic Creativity

Artistic creativity [30] can be defined as pursuing new ideas in different fields of arts like visual arts, performance arts, culinary arts and others. Each artist expresses their thoughts, ideas and/or beliefs through the aforementioned domains. For instance, a painter utilizes color, shape, and/or texture to make paintings which appeal to spectators on a sensory as well as an emotional basis. Similarly, a musician uses sound, rhythm, and melody to call forth different types of emotions in its audience. Again, a dancer/actor imparts life into stories utilizing gestures, dialogue, and other forms of action. Thus, artistic creativity involves the process of bringing forth new concepts into tangible forms through mediums like the strokes of a painter, the lyrics of a composer, and the like, that leads towards enriching the cultural and emotional fabric of the society.

1.3.2. Scientific Creativity

Scientific creativity [31] can be defined as the concoction of imagination, logical reasoning, and empirical evidence, that enables an individual to push the boundaries of what is known or already existing and unravel novel possibilities across various scientific domains. Unlike artistic creativity which focuses on demonstrating novelty across different artistic realms, scientific creativity accelerates the generation of new knowledge that fuels innovation and helps to shape the future of humanity.

Creativity in scientific realm is explorable in numerous forms, such as the development of new theories, the invention of novel technologies, the discovery of unknown phenomena, and the creation of innovative methodologies or algorithms. For instance, in medicine, scientific creativity can lead to the discovery of novel drugs that can cure autoimmune disorders such as Hashimoto Thyroiditis, Grave's disease, Lupus and the like [32]. Again in physics, it can give rise to theories that

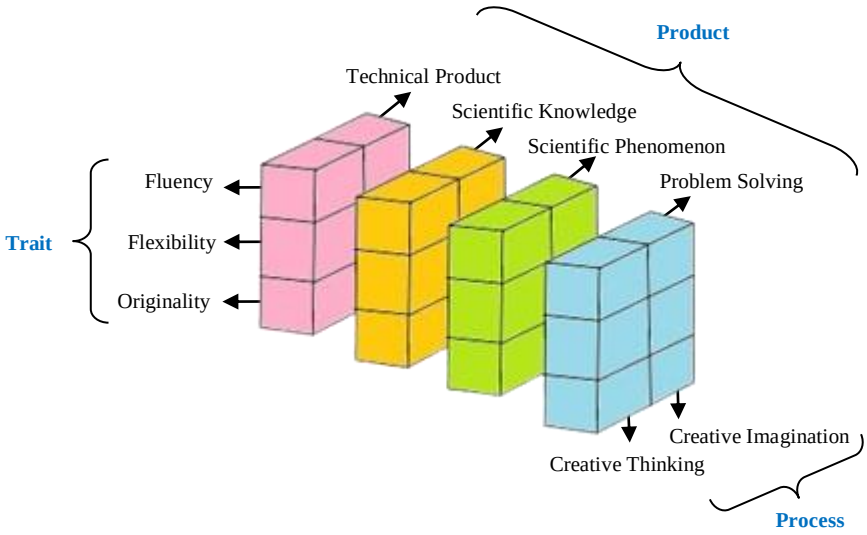


Fig. 1.6. The 3D depiction of scientific structure creativity model (SSCM)

redefine the understanding of the universe's fundamental principles [33]. Similarly in engineering, scientific creativity fosters the development of new products that address different challenges and improve the quality of life [34].

In essence, scientific creativity acts as a keystone of progress in all fields of science that leads to industrial, technological and/or societal progress. Since the thesis is based on scientific creativity, the present chapter will explore this concept in detail, without elaborating further on non-scientific creativity.

1.4 PHILOSOPHICAL CONTRIBUTIONS TO SCIENTIFIC CREATVITY

A few philosophical models of scientific creativity can be found in the existing literature [35-36] that attempt to explain the cognitive processes and underlying principles involved in scientific discovery and innovation. The following sub-section discusses them in detail.

1.4.1 The Scientific Structure Creativity Model (SSCM)

The Scientific Structure Creativity Model (SSCM) [35] is a three-dimensional (3D) framework that provides a multidimensional perspective on understanding the concept of scientific creativity. In other words, the present model emphasizes the relation among various components that lead to an innovative thought process, which include:

(a) **Trait**: This focuses on the individual characteristics that contribute to creativity which involves the following components.

(i) Fluency-The capability of producing an ample amount of ideas.
(ii) Flexibility-The capability of adapting and shifting perspectives while solving problems.
(iii) Originality-The capability to produce novel and unique solutions.
(b) **Process**: This relates to the cognitive mechanisms involved in producing a creative outcome. It involves the following two components.
(i) Creative Thinking- Logical approaches to generate something new.
(ii) Creative Imagination- The ability to envision possibilities beyond immediate reality.
(c) **Product**: This involves the final outcome of a creative process. It included the following four components.
(i) Technical Products- Tools created to address specific needs.
(ii) Scientific Knowledge- Contributions that expand the present knowledge of science.
(iii) Scientific Phenomena- Insights into natural or scientific processes.
(iv) Problem Solving- Development of solutions to novel and complex challenges.

The SSCM is represented as a three-dimensional structure consisting of 24 cells [35-36] (depicted by Fig. 1.6), created by the intersection of the above discussed three core elements. Each cell corresponds to a distinct combination of the trait, process and product components of scientific creativity, some of which are discussed as follows.
(a) **Fluency × Creative Thinking × Technical Product**: Describes brainstorming multiple ways to improve an existing system.
(b) **Fluency × Creative Imagination × Technical Product:** Represents producing multiple designs for tools or systems based on imaginative thought.
(c) **Fluency × Creative Thinking × Scientific Knowledge:** Describes generating a variety of ideas using scientific knowledge through logical reasoning.
(d) **Flexibility × Creative Thinking × Problem Solving:** Focuses on adapting diverse thought approaches to solve scientific problems.
(e) **Originality × Creative Imagination × Technical Products:** Represents producing novel and unique designs for tools or systems based on imaginative thought.

1.4.2 The Cognitive Model of Scientific Creativity (CMSC)

The cognitive model of scientific creativity (CMSC) developed by Jongwon Park [37-38] is an influential philosophical framework that seeks to explain the cognitive processes underlying scientific creativity. According to this model, a creative

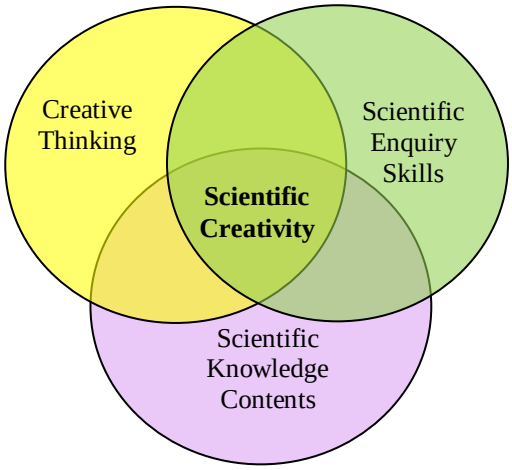


Fig. 1.7. Venn diagram depicting the cognitive model of scientific creativity

outcome in the scientific domain arises from the amalgamation of the following three cognitive factors (see Fig. 1.7), which are elaborated below.

(a) **Creative Thinking:** Creative thinking is a fundamental element of scientific creativity, enabling the generation and/or refinement of original and valuable ideas in the scientific domain. It is a product of the interaction of three cognitive processes which include divergent thinking, convergent thinking, and associational thinking. Divergent thinking, the capability of producing abundant applicable solutions or ideas for a given problem, traces its roots from Guilford's Model of Creativity, as discussed in Section 1.2.2. On the contrary, convergent thinking focuses on narrowing down the possible solutions of the given problem to identify the most effective one, which also traces its roots in the Guilford's model. Building on the above-established foundations, Park incorporated associational thinking as the third component of creative thinking. Associational thinking involves the formation of connections between seemingly unrelated concepts that aids in the recombination of knowledge from different domains. A notable example of associational thinking involves the discovery of DNA's double-helix structure, where cross-disciplinary knowledge from molecular biology, X-ray diffraction data, and Chargaff's rules [39] came together to provide revolutionary insights.

(b) **Scientific Inquiry Skills**: The steps of scientific inquiry form a structured process through which one investigates and answers questions about the unknown, yet to be discovered or invented. This systematic framework guides the collection and interpretation of evidence, ultimately advancing scientific research and knowledge. As shown in Fig. 1.8, scientific inquiry begins with observation—identifying a phenomenon, pattern, or unusual occurrence that warrants investigation. After observation, the formulation of a question follows, which guides

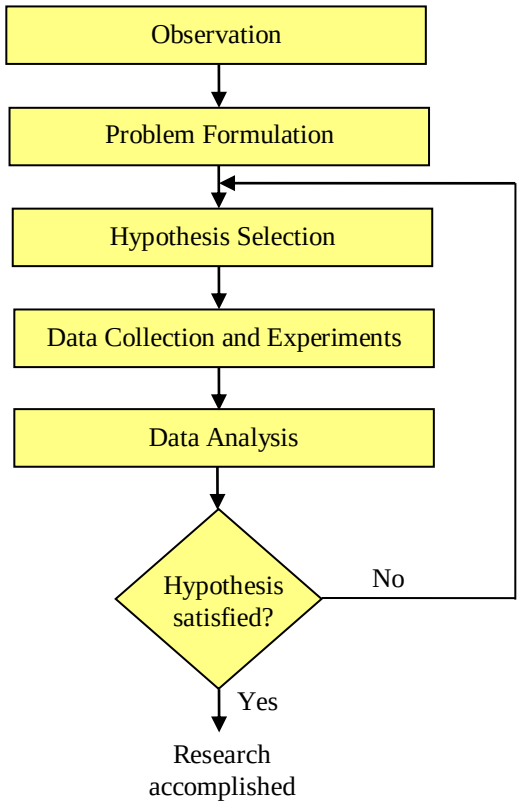


Fig. 1.8. Flow chart illustrating the steps involved during a scientific inquiry

the direction of the research. Next comes the proposal of a hypothesis—a tentative, testable explanation. The experimentation stage involves designing tests, selecting variables, and determining methods of measurement. Once data is collected, it needs to be analyzed to see if it supports or contradicts the hypothesis. Finally, one can draw conclusions based on the analysis, deciding whether the hypothesis is supported or refuted, and suggesting further research or adjustments. If the hypothesis is disproved, the process restarts with a new hypothesis, continuing until new insights emerge. In the worst case, if the entire observation is found to be incorrect, the process restarts from the beginning.

(c) **Scientific Knowledge**: Scientific knowledge plays a foundational role in fostering creativity, providing a rich reservoir of information, principles, and insights that spark new ideas and innovations. It enables individuals to build upon existing theories, identify gaps in knowledge, and explore areas that remain unexplained.

1.4.3 Chance-Configuration Theory of Scientific Creativity

The Chance-Configuration Theory of Scientific Creativity, proposed by Dean Keith Simonton [40], offers a framework for understanding how creative breakthroughs

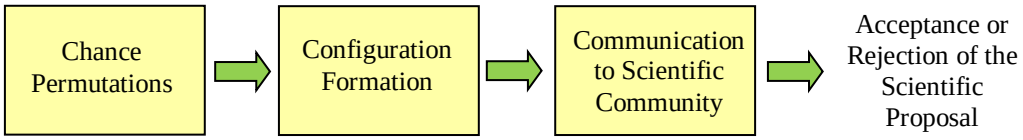


Fig. 1.9. Block diagram illustrating the process involved in chance configuration theory

occur in science. According to this theory, such breakthroughs emerge through a three-step process, which is depicted in Fig. 1.9 and can be outlined as follows.

(a) **Chance Permutations:** The theory suggests that creativity arises from operations performed on fundamental mental elements, such as facts, ideas, or concepts. These elements are combined in various ways through a process of "chance permutations." While this process involves randomness, it does not signify pure chaos. Instead, it represents a scenario where numerous potential combinations are explored, each carrying a low but non-zero probability of yielding success. This approach mirrors the generation of diverse ideas or solutions, which can then be evaluated for their practicality and effectiveness.

(b) **Configuration Formation:** Among the permutations, some combinations stand out as stable and coherent, referred to as "configurations." These stable configurations are more likely to capture conscious attention and undergo further refinement.

(c) **Social Communication and Acceptance:** Once a promising configuration is identified, the next challenge lies in effectively communicating it to the scientific community. The acceptance of this configuration depends on factors such as its clarity, relevance, and the innovator's ability to present the idea in a manner that aligns with societal norms and demonstrates its potential benefits to society.

In addition to outlining the three-step process responsible for scientific breakthroughs, Simonton proposed a mathematical formulation [40-41] to describe an individual's creative productivity within the framework of the chance-configuration theory. Before presenting this formulation, two key terms need to be defined. The first is ideation rate, which refers to the rate at which chance permutations occur, arising from operations performed on fundamental mental elements such as facts, ideas, or concepts. The second is elaboration rate, which denotes the rate at which the generated ideations are further articulated and refined. Based on these two definitions, the mathematical formulation of creative productivity is described by equation (1.1).

$$r(t) = k(e^{-pt} - e^{-qt}) \tag{1.1}$$

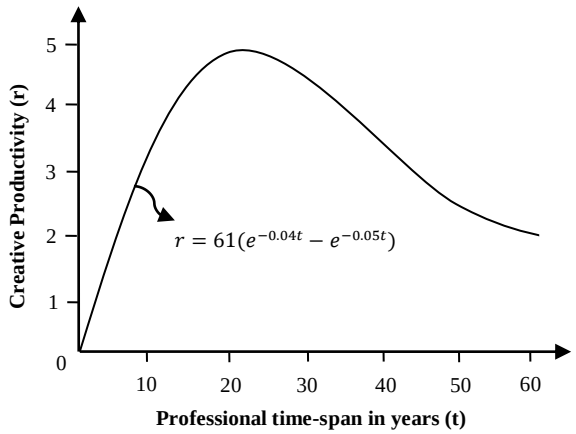


Fig. 1.10. Creative productivity as a function of professional time-span as illustrated by Simonton in [40-41]

where, $r(t)$ denotes the creative productivity at time *t*, *p* represents the ideation rate, *q* signifies the elaboration rate, k refers to the maximum number of contributions an individual is able to produce and *t* indicates the professional time-span of the individual.

It is important to note that Simonton's mathematical formulation was derived from an extensive analysis of datasets [42-44] encompassing individuals from various fields, including the arts, sciences, and notable achievers such as elites and laureates. This formulation resulted in the creation of a graph (see Fig. 1.9) illustrating the trajectory of creative productivity over time and age. The graph indicates that creative output initially rises steeply, reaches a peak following a decelerating curve, and then gradually tapers off, asymptotically approaching zero. Notably, the highest levels of creative productivity are observed during a professional span of 20–30 years, typically corresponding to the age range of 40-50 years.

1.4.4 Conceptual Blending Theory in Mathematical Invention

Conceptual Blending Theory, first introduced by Gilles Fauconnier and Mark Turner [45-46], provides a framework for understanding how new ideas emerge by combining elements from distinct conceptual domains. This theory suggests that knowledge from two unrelated domains, referred to as domain 1 and domain 2, can be blended to create novel insights, solutions, or constructs that were not previously apparent. For example, in the field of technology, the blending of domain 1: communication and domain 2: computing led to the creation of the internet—a transformative medium for global connectivity. Similarly, the integration of domain

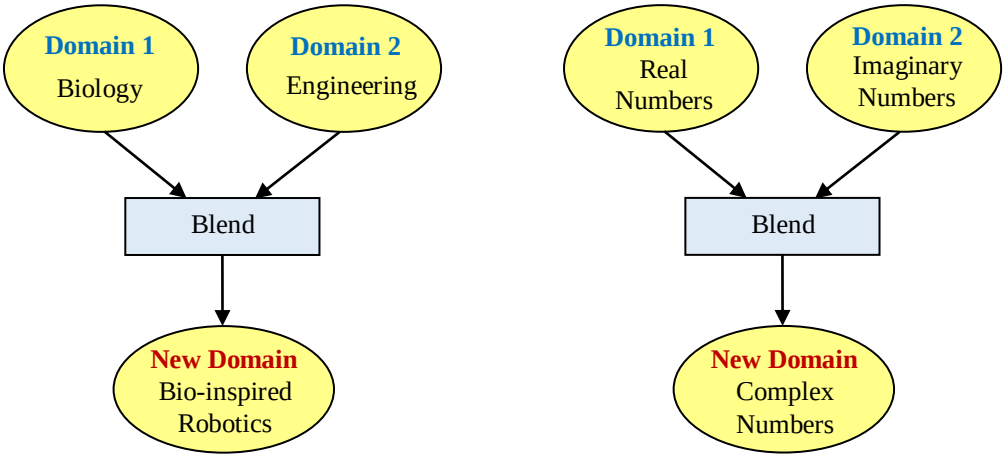


Fig. 1.11. Illustration of conceptual blending theory for creation of new domains (a) creation of bio-inspired robotics domain (b) creation of complex numbers domain in mathematics

1: biology and domain 2: engineering has given rise to bio-inspired robotics (see Fig. 1.11.(a)), where machines mimic biological processes to solve engineering challenges.

According to the Blending Theory, it plays a vital role in driving innovation within mathematics, a discipline that underpins all realms of science [45-50]. The creation of new mathematical constructs frequently involves identifying patterns, drawing analogies, and merging seemingly unrelated domains to solve abstract problems or extend existing frameworks. Numerous examples in mathematics demonstrate the application of this theory, such as the development of calculus through the fusion of algebra and infinitesimals, which enabled the formalization of rates of change and accumulation. Another example is the evolution of fractions, achieved by blending the concepts of numbers and division to represent parts of a whole effectively. The invention of complex numbers similarly exemplifies conceptual blending, integrating real and imaginary numbers (see Fig. 1.11(b)).

1.5 BIOLOGICAL MODALITY OF SCIENTIFIC CREATIVITY

The human brain comprises several modules/regions, each having one or more biological functions. For handling any problem, these modules in the brain operate either independently or collaboratively, depending on the intricacy and/or characteristics of the task. In the context of scientific creativity, different brain regions dynamically engage in various phases of generating a creative outcome, from the initial genesis of novel ideas to their refinement and practical application. This dynamic interaction is what defines the biological modality of scientific creativity, i.e., the structured sequence in which different brain regions contribute to the genesis

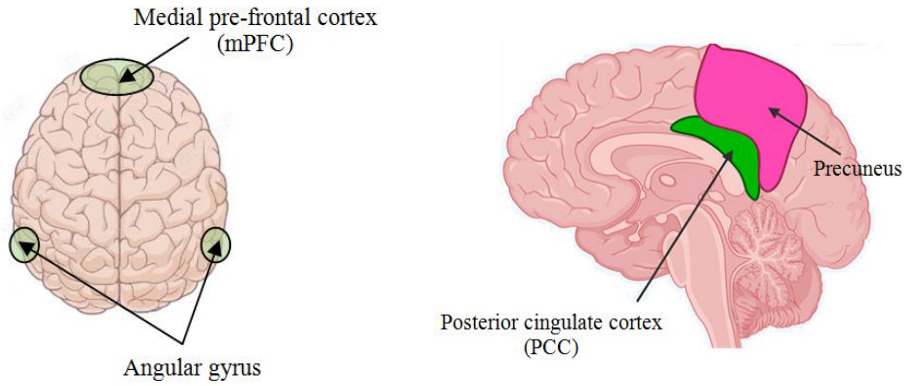


Fig. 1.12. Illustration of the brain lobes involved in Default Mode Module. (a) 3D view of mPFC and angular gyrus and (b) saggital view of the PCC and precuneus

of something new. In this context, three important brain modules have been reported in literature [51-52] which is discussed in details below.

1.5.1. Default Mode Module

The default mode module (DMM) plays an essential role in scientific creativity [53-55] and consists of the medial prefrontal cortex (mPFC), posterior cingulate cortex (PCC), precuneus and the angular gyrus (see Fig. 1.12). These brain regions collectively contribute to various cognitive processes responsible for scientific creativity, which include such as divergent thinking, intuitive insight, logical and spatial reasoning. The mPFC is responsible for divergent thinking that involves the generation of multiple ideas to a given open-ended problem. Divergent thinking is quite important for a scientific creative solution since it enables scientists to create new hypotheses by considering various possible options. The precuneus, located in the parietal lobe, contributes towards logical and spatial reasoning. The aforementioned reasoning techniques help in scientific creativity by aiding researchers to manipulate abstract concepts and visualize theoretical models. The angular gyrus is a key region associated with intuitive thinking, which has been responsible for many scientific breakthroughs. Intuitive thinking often enables the brain to recognize hidden patterns and novel connections that may not be immediately apparent through logical analysis alone. The PCC is present in the posterior part of the cingulate cortex as shown in Fig. 1.12.(b). PCC is considered as the central region of the DMM as it integrates the information from the different brain regions including the ones stated above. The above information integration is crucial for producing the final creative outcome, as it enables the synthesis of different cognitive functions into feasible innovative solutions. In other words, the

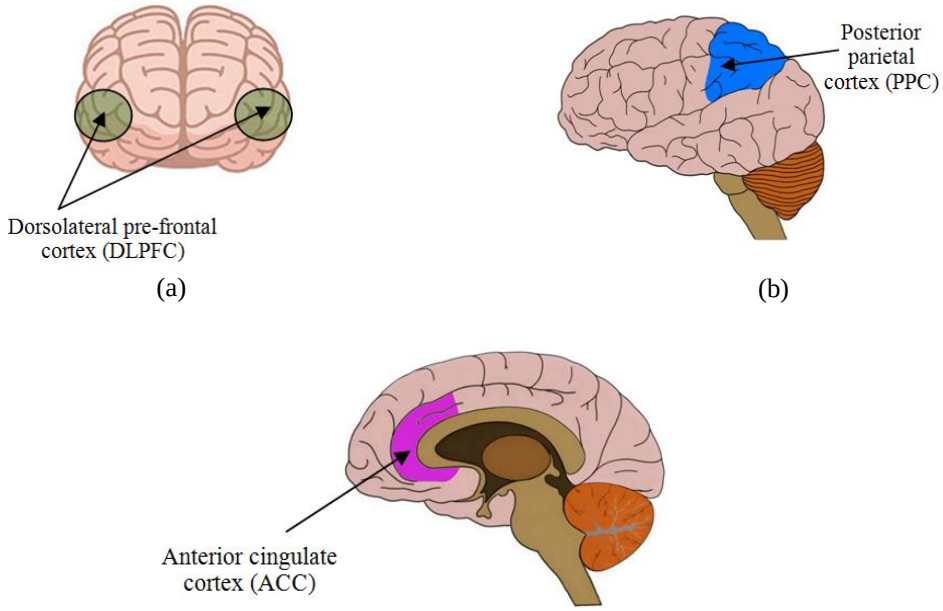


Fig. 1.13. Illustration of the brain lobes involved in Executive Control Module. (a) 3D front view of DLPFC, (b) 3D left view of PPC and (c) saggital view of ACC

PCC's connections with other brain regions enables the interaction between memory, abstract thought, and self-generated ideas, making it a critical component of the scientific creative thought process.

1.5.2 Executive Control Module

The executive control module (ECM) [56-57] functions antagonistically to the DMM [58] by refining and evaluating the outcomes of the cognitive processes controlled by the DMM. While the DMM facilitates the generation and exploration of novel ideas, the ECM ensures the generated ideas are logically feasible, and can be translated into actionable hypotheses. The ECM consists of three brain lobes which include dorsolateral prefrontal cortex (DLPFC), anterior cingulate cortex (ACC), and posterior parietal cortex (PPC) and is illustrated in Fig. 1.13. The DLPFC helps in organizing and/or structuring the creative insights generated by the DMM into well-defined and testable hypotheses. It also supports abstract thinking and long-term goal planning, ensuring that creative ideas are practically feasible and can be empirically validated. The ACC plays an important role in error detection, hypothesis revision, and cognitive flexibility. It enables scientists to evaluate the validity of creative ideas by identifying inconsistencies or conflicts in reasoning. The PPC ensures that the spatial reasoning task is applied in a structured and goal-directed manner thereby

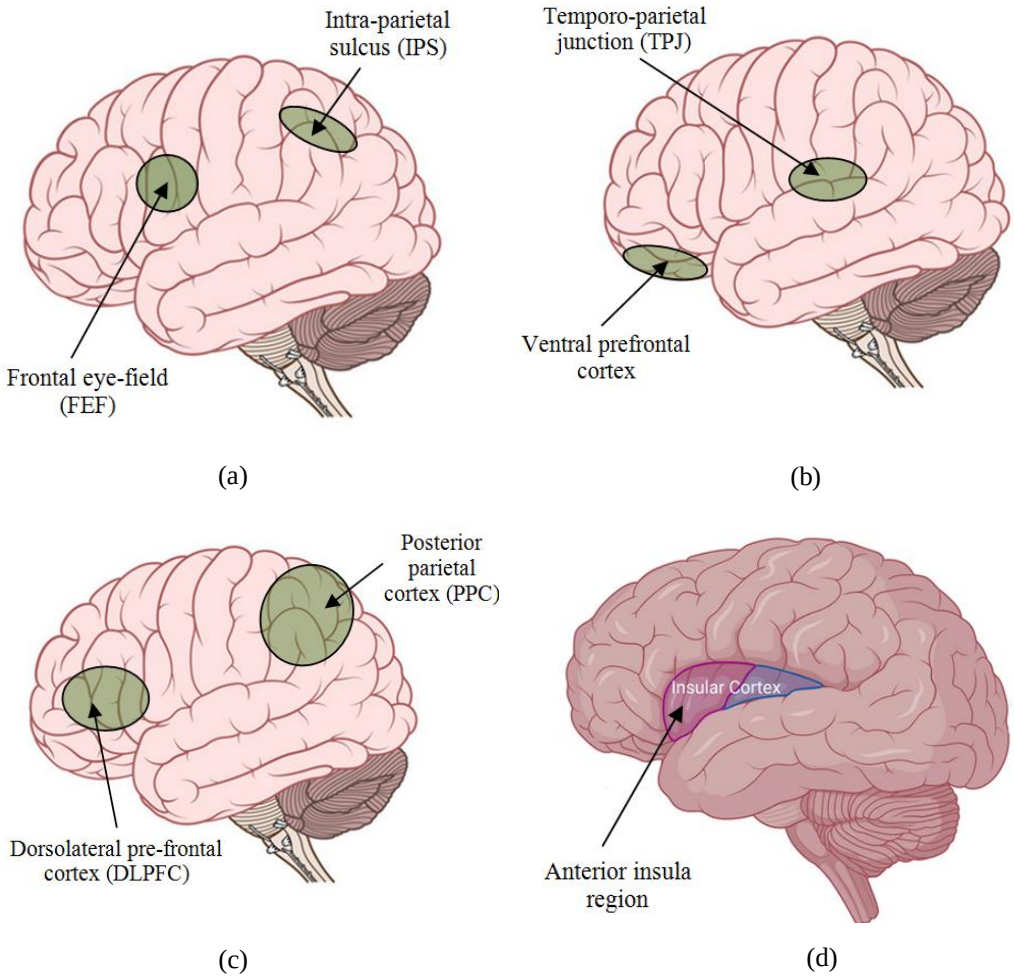


Fig. 1.14. Illustration of the brain lobes involved in the Attention Modules. (a) 3D left view of DASM, (b) 3D left view of VASM, (c) 3D left view of FPSM and (d) anterior insular region of COSM

enabling cognitive control during a problem-solving process. It helps integrate information, direct attention to relevant details, and facilitate precise calculations, all of which contribute to structured scientific inquiry. Thus, the ECM and DMM interact complementarily, with DMM facilitating free-thinking and novel idea generation while ECM providing the necessary cognitive control and analytical rigor that transforms spontaneous creative thoughts into structured and scientifically sound information.

1.5.3 Attention Module

The attention module (AM) also plays a significant role in contributing towards a scientific creative outcome [59-60]. Its main function involves regulating focus, selectively filtering information, and dynamically shifting between different cognitive states. In other words, it helps subjects to manage cognitive resources

efficiently, ensuring that relevant information is processed while distractions are minimized. The AM consist of four sub-modules [61] which are illustrated in Fig. 1.14 and described as follows.

(a) **Dorsal Attention Sub-Module (DASM)**: It involves the intra-parietal sulcus (IPS) and frontal eye fields (FEF) (see Fig. 1.14 (a)). Its primary function is to deeply engage an individual towards the novel solution of the scientific creative problem.

(b) **Ventral Attention Sub-Module (VASM)**: It includes the temporo-parietal Junction (TPJ) and ventral prefrontal cortex of the brain (depicted in Fig. 1.14 (b)). It helps to detect unexpected, novel, or important information. In other words, it enables individuals to notice anomalies in data, leading to unexpected discoveries.

(c) **Fronto-Parietal Sub-Module (FPSM)**: It comprises the DLPFC and PPC brain regions (see Fig. 1.14 (c)) and helps to switch between different cognitive states such as from spontaneous insight (as in DMM) to structured thinking (as in ECM).

(d) **Cingulo-Opercular Sub-Module (COSM)**: It helps to maintain long-term focus and detects errors in reasoning and comprises the ACC (as shown in Fig. 1.13 (c)) and anterior insula regions of the brain (as shown in Fig. 1.14 (d)). Thus, it prevents distractions and ensures persistence in long-term scientific research.

1.6 BRAIN NETWORK MODELING TO EVALUATE THE BIOLOGICAL MODALITIES OF SCIENTIFIC CREATVITY

Brain network analysis [62-63] provides an interesting framework to model the biological modalities of scientific creativity [78-80]. Such an analysis examines how different brain modules interact with one another in response to a given cognitive task. In other words, the aforementioned analysis aims at exploring the functional connectivity of different brain lobes that helps to uncover the underlying mechanisms such as problem-solving, attention, divergent thinking, and the like, pertaining to a specific cognitive activity.

For the present scenario, the thesis tries to explore the connectivity patterns among different brain modules when a scientific creative task is undertaken by an individual. The aforementioned analysis involves modeling the interaction among different brain lobes as a graph, where the brain lobes are represented as nodes while their connections are denoted by edges or links. Such a graph-based representation aids to provide insights into the neural basis of scientific creativity.

The brain network analysis is undertaken in two distinct phases. The first phase involves the formulation of the connectivity matrix, popularly known as the adjacency matrix [62], from the brain signals acquired from the scalp of subjects

involved in a scientific creativity based task. The acquisition of the brain signals may be done using any popular brain activity recording techniques such as Electroencephalograph (EEG) [64], Magnetic Resonance Imaging (MRI) [65], Magnetoencephalography (MEG) [66], Positron Emission Tomography (PET) [67], functional Near Infrared Spectroscopy (fNIRS) [68] and the like. The second phase involves the quantification of the formulated brain networks via adjacency matrix to understand different properties of the acquired network. The details regarding the aforementioned two phases of brain network analysis is discussed below.

1.6.1. Construction of Brain Network by Adjacency Matrix Computation

For the computation of the adjacency matrix, the first phase involves the computation of connectivity measures from the acquired brain signals such as correlation, coherence and the like. Some of the popular connectivity measures are elaborated below.

(a) **Pearson's Correlation Coefficient (PCC)**

PCC is a widely used statistical measure in brain network analysis utilized to quantify the linear relationship between the activities of two brain signals over time [69]. In other words, it assesses how strongly the signals from two lobes fluctuate together. The correlation coefficient ranges between -1 and 1 where the coefficient value of 1 denotes positive correlation among the brain signals, coefficient value of -1 signifies negative correlation among the brain signals, and coefficient value of 0 indicates no linear relationship between the brain signals. The PCC among a pair of brain signals can be mathematically represent by (1.2).

$$\rho_{XY} = \frac{1}{h-1}\sum_{i=1}^{h}\left(\frac{x_i-\bar{x}}{\sigma_X}\right)\left(\frac{y_i-\bar{y}}{\sigma_Y}\right) \tag{1.2}$$

where, ρ_{XY} represents the correlation coefficients values among a given pair of brain signals corresponding to channels *X* and *Y*, *h* represents the number of data points pertaining to a given channel, $\bar{x}, \bar{y}, \sigma_X$ and σ_Y represent the mean and standard deviation values of the data points of the channels *X* and *Y* respectively.

(b) **Phase Lag Index (PLI)**

PLI measures how consistently the phase of one brain signal lags behind another over time [70]. It is widely used in EEG/MEG studies to assess brain network interactions while minimizing the influence of volume conduction [71]. A PLI value of 0 indicates no connectivity among a pair of brain lobes, PLI value of 1 signifies strong connectivity among a pair of brain lobes while a PLI value of 0.5 indicates

weak connectivity or random fluctuations. The PLI value among a pair of brain signals is mathematically represented by (1.3).

$$PLI = |\langle sign(\sin(\Delta\phi(t_m)))\rangle| \tag{1.3}$$

where, $\Delta\phi(t_m)$ denotes the phase difference between two signals at time point *m*, *sign* depicts the signum function, $\langle\cdot\rangle$ signifies the mean value acquired over all time points and $|\cdot|$ indicates the absolute value. $\Delta\phi(t_m)$ is calculated using (1.4).

$$\Delta\phi(t_m) = \phi_x(t_m) - \phi_y(t_m) \tag{1.4}$$

where, $\phi_x(t_m)$ and $\phi_y(t_m)$ represent the instantaneous phases for signals *x* and *y* respectively evaluated using Hilbert Transform [70].

(c) **Wavelet Transform Coherence (WTC)**

WTC is a useful technique utilized to evaluate the time-frequency relationship between two brain signals [72]. Unlike traditional coherence techniques based on Fourier analysis, that consider the evaluating signals to be stationary, WTC is well-suited for non-stationary signals like EEG, fNIRS, etc. In essence, it provides insight into how two signals are correlated over time and across different frequency bands, making it a valuable tool for studying brain connectivity. WTC values ranges between 0 and 1, where 1 indicates strong coherence among a pair of brain signals. The WTC between a pair of brain signals is evaluated using (1.5).

$$W_n(s) = \frac{\left|\left\langle s^{-1} W_n^{XY}(s)\right\rangle\right|^2}{\left|\left\langle s^{-1} W_n^{X}(s)\right\rangle\right|^2 \left|\left\langle s^{-1} W_n^{Y}(s)\right\rangle\right|^2} \tag{1.5}$$

where, *s* indicates the wavelet scale, *n* depicts the time instance, $W_n^X(s)$ and $W_n^Y(s)$ represents the continuous wavelet transform of X_n and Y_n signals respectively, and $\langle\cdot\rangle$ signifies the smoothing operation [72-73].

For *N* number of channels, the connectivity measures evaluated using PCC, PLI, WTC, or any other technique yield an $N\times N$ matrix, where each value signifies the strength of the connection among every pair of channels. The second and final phase in the formulation of the adjacency matrix involves the transformation of the acquired $N\times N$ matrix into either a weighted or binary graph using thresholding. The adjacency matrix of the weighted graph is evaluated by (1.6).

$$A_{ij}^{w} = \begin{cases} \beta_{ij}, & if\ \beta_{ij} \geq \alpha \\ 0, & otherwise \end{cases} \tag{1.6}$$

where, A_{ij}^{w} represents the weighted adjacency matrix values of the brain network, β_{ij} signifies the connectivity values acquired from PCC, PLI, WTC or any other technique while α denotes a predefined threshold. The adjacency matrix of the binary graph is computed by (1.7).

$$A_{ij} = \begin{cases} 1, & if\ \beta_{ij} \geq \alpha \\ 0, & otherwise \end{cases} \tag{1.7}$$

where, A_{ij} denotes the adjacency matrix values for the binary network. The pre-defined threshold can be selected based on a given experiment. However, this creates brain networks each having different number of links where some have dense connections while others having sparse connections [62]. This disparity leads to non-uniform analysis among a given set of networks. To address such an issue, the threshold value is often adaptively adjusted to ensure that all networks have a fixed number of links [74].

It may be noted that the process of thresholding for determining the adjacency matrix of a brain network is a crucial step. This process ensures that only strong, meaningful connections are considered while weak connections arising due to noise/artifacts are ignored [62]. Moreover, if thresholding is not utilized, a fully connected adjacency matrix is created which increases the computational cost of analyzing complex brain network properties such as modularity, small-worldness, centrality and the like [76-77].

1.6.2 Quantitative Evaluation of the Constructed Brain Network

The acquired brain networks using the adjacency matrix computation as discussed above need to be further analyzed to identity the different connectivity patterns among the subjects engaged in a scientific creative task. Such an analysis is possible by utilizing some quantitative measures that helps to characterize different network properties [62], [75]. Thus, by comparing various network properties across different subjects, researchers can uncover the brain connectivity patterns associated with scientific creativity. The network properties are usually categorized into two classes: local and global. The two categories of network properties are discussed below.

1.6.2.1 Local Brain Network Properties

Local brain network properties sheds a light upon the connectivity and functional organization of different brain modules/lobes with respect to the overall brain network, thereby providing finer details regarding the functionality of specific brain regions pertaining to creativity. In other words, local properties provide deeper insights into the intricate brain architecture by focusing on how different brain regions contribute to information flow, processing, and the like. The detailed list of local properties can be found in [62], [75]. Some of the popular local network properties are presented below instead of the entire cohort due to space limitations.

(a) **Degree**

The degree of a node within a brain graph/network represents the total number of direct links/edges that a particular brain region or node has with other regions. It is a fundamental measure in graph theory and provides insight into the relevance of a node within the network. The degree D_i of a node i in a binary network is represented as

$$D_i = \sum_{j \in N} A_{ij} \tag{1.8}$$

On the contrary, the weighted degree D_i^w of a node i in a weighted network is denoted by (1.9).

$$D_i^w = \sum_{j \in N} A_{ij}^w \tag{1.9}$$

(b) **Closeness centrality**

Closeness centrality analyzes how easily a node can reach the rest of the nodes within a network. It is defined as the reciprocal of the sum of the shortest path distances from a node to all other nodes in the graph. The closeness centrality of a node in a binary network is denotes as

$$C_i = \frac{N-1}{\sum_j d_{ij}} \tag{1.10}$$

where, C_i represents the closeness centrality of node i and d_{ij} indicates the minimal path distance amongst nodes i and j. The closeness centrality of a node in a weighted graph follows the same formulation in (1.10) except d_{ij} involves the minimal weighted distance amongst nodes i and j.

(c) **Betweenness centrality**

Betweenness centrality quantifies how often a node lies on the smallest paths amongst other nodes within a network. It helps identify nodes that act as bridges or hubs in a network, facilitating information flow. The betweenness centrality B_i of a node i in a binary graph is evaluated by (1.11).

$$B_i = \frac{2}{(N-1)(N-2)} \sum_{j \neq h \neq i} \frac{\sigma_{hj}(i)}{\sigma_{hj}} \tag{1.11}$$

where, σ_{hj} and $\sigma_{hj}(i)$ signifies the number of smallest paths amongst nodes h and j and the number of shortest paths amongst nodes h and j passing through node i respectively. The weighted version of betweenness centrality follows the same formulation in (1.11) expect the shortest paths consider the weights in the network.

(d) **Clustering coefficient**

The clustering coefficient of a node measures the extent to which its neighbors are also connected to each other. It quantifies local connectivity and helps identify highly clustered regions in a network. Its mathematical formulation for a binary graph is depicted by (1.12).

$$L_i = \frac{2F_i}{D_i(D_i - 1)} \tag{1.12}$$

where, L_i represents the clustering coefficient of node *i* and F_i indicates the number of edges between the neighbors of node *i* that forms triangles. The weighted version of (1.12) is denoted by (1.13).

$$L_i^w = \frac{1}{D_i(D_i - 1)} \sum_{j,k} (w_{ij} w_{ik} w_{jk})^{1/3} \tag{1.13}$$

where, L_i^w represents the weighted clustering coefficient of a node *i*, and w_{ij}, w_{ik}, w_{jk} denote the edge weights between nodes *i* and *j*, *i* and *k* and *j* and *k* respectively.

(e) **Local efficiency**

Local efficiency measures how well a node's neighbors communicate with each other when the node itself is removed. In other words, local efficiency helps identify robust functional clusters that can maintain communication even if a node is damaged. Its mathematical formulation is depicted by (1.14).

$$G_i = \frac{1}{D_i(D_i - 1)} \sum_{j,k \in S(i)} \frac{1}{g_{jk}} \tag{1.14}$$

where, G_i represents the local efficiency of node *i*, $S(i)$ denotes the set of neighbors of node *i*, g_{jk} indicates the smallest path length amongst nodes *j* and *k* after removing node *i*. the weighted version of (1.14) follows the same formulation except p_{jk} is replaced by the weighted smallest path length between nodes *j* and *k* after removing node *i*.

1.6.2.2 Global Brain Network Properties

Global brain network properties explore how the brain operates as a cohesive or integrated system, capturing large-scale patterns of connectivity and information flow across different brain lobes. Unlike local properties, which focus on individual brain modules, global properties provide a system-wide perspective on the functional organization of brain modules pertaining to any cognitive task such as scientific creativity. These properties help to identify the brain's overall efficiency, resilience, and communication capacity, which are essential to understand various cognitive processes involved during a scientific creative outcome. The complete list of global

properties can be found in [62]. Some of the global properties of brain network are detailed below.

(a) **Global efficiency**

Global efficiency of a brain network describes how effectively information is transmitted across the entire brain. In other words, it assesses the ease with which different brain lobes communicate among themselves thereby making it a crucial indicator of cognitive function, processing speed, and network resilience. Higher global efficiency of a network implies that brain lobes can exchange information more quickly, while lower global efficiency is often associated with poor cognitive activity. It is mathematically defines as the inverse of the mean of the minimal path length amongst all pairs of nodes within the network and is expressed by (1.15).

$$G = \frac{1}{N(N-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \tag{1.15}$$

where, G represents the global efficiency of the network. The above formulation can be applied to weighted graphs by introducing the weighted shortest path length in place of the shortest path length.

(b) **Assortativity**

Assortativity assesses whether highly connected regions tend to link with other highly connected regions. The above assessment is performed by computing the degree assortativity coefficient for a pair of nodes *u* and *v* and is mathematically expressed by (1.16).

$$R = \frac{E^{-1} \sum_i u_i v_i - \left[E^{-1} \sum_i \frac{1}{2}(u_i + v_i) \right]^2}{E^{-1} \sum_i \frac{1}{2}(u_i^2 + v_i^2) - \left[E^{-1} \sum_i \frac{1}{2}(u_i + v_i) \right]^2} \tag{1.16}$$

where, R represents the degree assortativity coefficient, E represents the total number of edges in the network, u_i and v_i denote the degree of nodes that are linked by the i^{th} edge. The assortativity values ranges between 1 to -1 where a value of 1 signifies that higher degree nodes are exclusively connected to other high degree nodes and vice-versa while a value of -1 indicates disassortativity i.e., higher degree nodes preferentially connect with low degree nodes and vice-versa.

(c) **Modularity**

Modularity is a key global property of a brain network that describes the extent to which the network is organized into distinct modules or communities i.e., groups of

nodes that are more densely connected to each other than to the rest of the network. A highly modular brain network reveals specialized processing within different modules while maintaining efficient global integration among them. It can be mathematically depicted by (1.17).

$$Q = \frac{1}{2E}\sum_{ij}(A_{ij} - e_{ij})\delta(c_i, c_j) \tag{1.17}$$

where, Q denotes the modularity index, e_{ij} indicates the number of edges between nodes i and j, c_i and c_j represents the modules pertaining to nodes i and j respectively and $\delta(c_i, c_j)$ signifies the Kronecker delta function which has a values 1 if nodes i and j pertain to the same module, otherwise has a value 0. The formula in (1.17) can be extended to a weighted graph by replacing the binary edges values with weighted edges.

(d) **Small World Coefficient**

The small-world property describes a network structure that balances high local clustering with short characteristic path lengths. This structure allows efficient information processing by supporting both segregated (local) and integrated (global) communication. In brain networks, the small-world architecture is crucial for efficient communication with minimal wiring costs, flexible cognitive processing through local specialization and global integration, and robustness against cognitive disruptions. The small world coefficient can be mathematically represented as

$$\lambda = \frac{L_{avg} / L_{random}}{H_{avg} / H_{random}} \tag{1.18}$$

where, L_{avg} represents the average clustering coefficient of the network, H_{avg} denotes the average shortest characteristic path length among two nodes, L_{random} and H_{random} signifies the clustering coefficient and the path length of an equivalent random network. It may be notes that the small world property uses a random network as a reference because it helps determine whether a given network exhibits a balance between local clustering (like regular lattices) and global integration (like random graphs) [62]. A network is considered small world if the value of $\lambda > 1$. The average clustering coefficient of the network and average shortest characteristic path length among two nodes are denoted by (1.19) and (1.20) respectively.

$$L_{avg} = \frac{1}{N}\sum_{i}\frac{2F_i}{D_i(D_i - 1)} \tag{1.19}$$

$$H_{avg} = \frac{1}{N(N-1)}\sum_{i \neq j} d_{ij} \tag{1.20}$$

1.7 INFORMATION THEORY BASED MODELING TO ANALYZE THE BIOLOGICAL MODALITIES OF SCIENTIFIC CREATIVITY

One of the key components of scientific creativity is divergent thinking which refers to the ability to explore multiple possible solutions to a given problem. Information theory provides a quantitative framework for analyzing the biological modalities of divergent thinking. In other words, information theory based approach helps to distinguish an individual capable of divergent thinking from the one lacking such a cognitive mechanism by evaluating their brain signals.

The most important metric utilized for the aforementioned analysis in information theory is entropy. The entropy of brain signals measures the variability and complexity of neural signals, reflecting how dynamically information is processed during a divergent thinking task [81-82]. High entropy of brain signals is associated with a more flexible, exploratory cognitive state, allowing for the generation of diverse and/or unconventional ideas, which is central to divergent thinking. Conversely, lower entropy of brain signals indicates a more structured and rigid neural state thereby highlighting a more predictable brain activity.

The present section discusses some widely used entropy based approaches to examine the changes in brain signals during divergent thinking.

(a) **Approximate Entropy**

Approximate Entropy (ApEn) has been widely applied in neuroscience to analyze the brain dynamics under different cognitive states and conditions [83]. In other words, it is used to measure the complexity and randomness in time-series data (e.g., EEG signals) by quantifying predictability i.e., how similar patterns repeat in the brain signal. In the context of divergent thinking, higher ApEn values are associated with a more dynamic and flexible neural state, allowing for spontaneous and diverse thought processes. Conversely, lower ApEn values suggest more structured and predictable brain activity, often observed in individuals engaged in routine problem-solving tasks. ApEn can be mathematically expressed as

$$ApEn\,(f,b,h) = \psi^{f}(b) - \psi^{f+1}(b) \tag{1.21}$$

where, f denotes the length of data segment used for comparison, b indicates the similarity threshold, typically set to a percentage of the standard deviation of the brain data and h represents the total number of data points. The measure of regularity for sequence of length f is represented by $\psi^{f}(b)$ and is mathematically denoted as

$$\psi^{f}(b) = \frac{1}{h-f+1}\sum_{i=1}^{h-f+1}\ln(h_i) \tag{1.22}$$

where, h_i quantifies the proportion of data points that remain close to *b* when comparing all possible segments of length *f* in the time series.

(b) **Sample Entropy**

Sample Entropy (SampEn) [84] is a modified version of Approximate Entropy (ApEn) that is used to quantify the complexity and irregularity of time-series data, such as brain signals. It measures how unpredictable a time series is by assessing the likelihood that similar patterns remain similar when an additional data point is introduced. Unlike ApEn, SampEn eliminates self-matching, making it more robust and consistent across different datasets. The SampEn is mathematically evaluated by (1.23).

$$SampEn(f,b,h) = -\ln\left(\frac{U}{V}\right) \tag{1.23}$$

where, *U* represents the number of similarity patterns of length f +1 while *V* indicates the number of similarity patterns of length *f*.

(c) **Permutation Entropy**

Permutation Entropy (PermEn) [85] is a computationally efficient and noise-resistant method for measuring the randomness or regularity of a time series. Unlike other entropy measures that utilize amplitude values of signals, PermEn relies on the ordinal patterns (i.e., ranked order of values). For a given time series $X = \{x_1, x_2, ..., x_h\}$, the PermEn of order *o* is evaluated as

$$PermEn(o) = \sum p(\pi) \log p(\pi) \tag{1.24}$$

where, π denotes the possible ordinal patterns (permutations) of length *o*, and $p(\pi)$ represents the probability of occurrence of each permutation π. Again here, higher PermEn values are associated with divergent thinking while lower values indicate more rigid thought process.

(d) **Spectral Entropy**

Spectral Entropy (SpecEn) [86] is another information-theoretic measure that quantifies the randomness or regularity of a signal based on its power spectral density (PSD)**.** It is derived from Shannon entropy but applied in the frequency domain, making it useful for analyzing signals physiological time series data such as EEG, fNIRS and the like. For a given time series $X = \{x_1, x_2, ..., x_h\}$, SpecEn is computed as

$$SpecEn = -\sum_{i=1}^{K} P(k_i) \log P(k_i) \tag{1.25}$$

where, K denotes the number of frequency components, k_i represents the i^{th} frequency and $P(k_i)$ indicates the normalized PSD at frequency k_i which is depicted by (1.26).

$$P(k_i) = \frac{S(k_i)}{\sum_{j=1}^{K} S(k_j)} \tag{1.26}$$

where, $S(k_i)$ represents the PSD at frequency k_i. A higher value of SpecEn indicates that the signal contains a broad range of frequencies with approximately equal power, indicating randomness and complexity, which is the characteristic linked to divergent thinking.

(e) **Multiscale Entropy**

Multiscale Entropy (ME) is an advanced information theoretic measure that evaluates the complexity of a time series across multiple temporal scales [87-92]. Unlike traditional entropy measures discussed above that performs signal analysis at a single scale, ME takes into account randomness of signal when it is coarse-grained over different time resolutions. This makes ME particularly useful for analyzing brain signals, (such as EEG) where meaningful patterns may emerge at different scales. For a given time series $X = \{x_1, x_2, ..., x_h\}$, ME is computed in two stages. In the first stage, a coarse-grained time series $Z^{(\tau)}$ is constructed using (1.27).

$$Z_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad j = 1,2,....., \frac{h}{\tau} \tag{1.27}$$

where, τ represents a scaling factor that controls how much the signal is compressed. In the second stage, for each coarse-grained time series $Z^{(\tau)}$ the SampEn is calculated using (1.23). Higher ME values across different scales indicates the time series maintains complexity across multiple time resolutions which implies divergent thinking while lower values represent constrained cognitive states.

1.8 MARKOV MODEL APPROACH TO EXPLORE THE BIOLOGICAL MODALITIES OF SCIENTIFIC CREATVITY

Markov models [93-97] provide a powerful framework to capture the temporal dynamics of brain activity during the production of a novel outcome [98-101]. Such models can quantify the flexibility and variability of neural state switching while generating a novel idea by modeling transitions between discrete functional brain states that is extracted from time-varying connectivity patterns. Creative individuals are hypothesized to exhibit richer, less stereotyped transition structures, reflected in

denser transition matrices and higher entropy. The mathematical steps involved for implementing a first order Markov model from brain signals are detailed below.

(a) **Formation of Brain Connectivity Matrix**: The first step involved in formulating a Markov model is evaluation of a brain connectivity matrix or then adjacency matrix. Such a matrix can be constructed using any of the methods discussed in Section 1.6.1. Thus, for N number of channels of any brain signal acquisition device (such as EEG, fNIRS etc.) produces an $N \times N$ connectivity or adjacency matrix A for the first trial or experimental instance as shown in (1.28).

$$A^1 = \begin{bmatrix} a_{11}^1 & \cdots & a_{1N}^1 \\ \vdots & \ddots & \vdots \\ a_{N1}^1 & \cdots & a_{NN}^1 \end{bmatrix} \tag{1.28}$$

(b) **Construction of a Column Vector**: All the elements with the adjacency matrix is stacked into a column vector excluding the self-connections. Thus, the above computation yields $M^1 \in \Re^{N \times (N-1)}$ as shown in (1.29).

$$M = \begin{bmatrix} a_{12}^1 \\ a_{13}^1 \\ a_{14}^1 \\ \vdots \\ a_{N(N-1)}^1 \end{bmatrix} \tag{1.29}$$

The steps in (a)-(b) is repeated for all T number of trails or experimental instances. This produces T number of column vectors. Next, each of the column vectors is stacked along the rows of a vector $M' \in \Re^{S \times N(N-1)}$ as shown in (1.30).

$$M' = \begin{bmatrix} a_{12}^1 & a_{13}^1 & \cdots & a_{N(N-1)}^1 \\ a_{12}^2 & a_{13}^2 & \cdots & a_{N(N-1)}^2 \\ \vdots & \vdots & \vdots & \vdots \\ a_{12}^S & a_{13}^S & \cdots & a_{N(N-1)}^S \end{bmatrix} \tag{1.30}$$

(c) **Evaluation of clusters**: k-means clustering [102-103] is applied upon M' to create k-number of clusters. After this, each row vector of M' is assigned to its respective cluster. The above computation thus yields a state sequence after transposition operation and is denoted by (1.31).

$$s = [s_1, s_2, \ldots\ldots, s_T] \tag{1.31}$$

(d) **Estimation of State Transition Probabilities**: The state transition probability matrix $O \in \Re^{k \times k}$ whose every element is evaluated using (1.32).

$$o_{ij} = P(s_{t+1} = j \mid s_t = i) \tag{1.32}$$

Creative individuals tend to have denser (less sparse) Markov matrices, because their brains transition more flexibly between different connectivity states as compared to

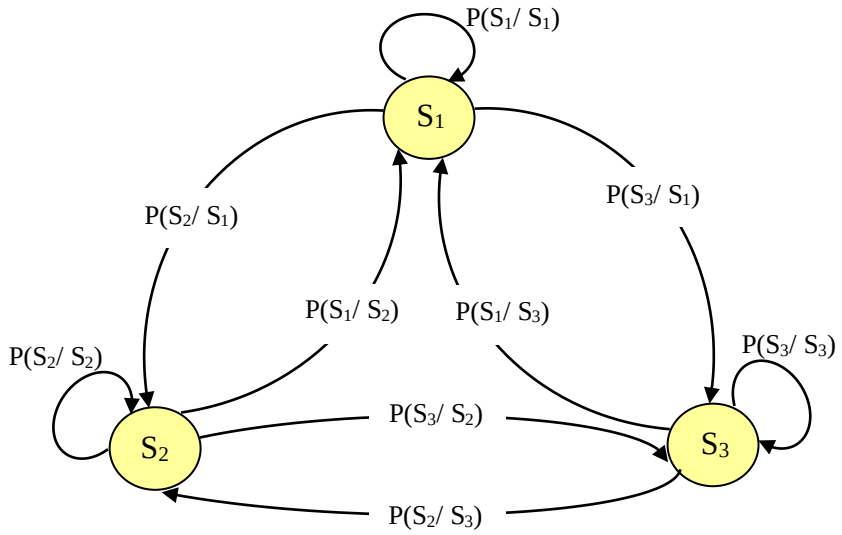


Fig. 1.15. Illustration of first order Markov process to capture the temporal dynamics of brain activity during the production of a novel outcome

non-creative ones. The first order Markov process for three state transitions is depicted in Fig. 1.15. As an additional step, after the computation in (1.32), the entropy of each state can be evaluated using the methods described in Section 1.9.

1.9 COMPUTATIONAL MODALITY OF SCIENTIFIC CREATIVITY

Computational modality of scientific creativity refers to the utilization of various computational methods to generate original ideas, unconventional solutions, or innovative formulations in scientific realm. In other words, it involves simulating aspects of human creativity through different computational techniques to synthesize something new. These computational techniques may include evolutionary algorithms [104], neural networks [105], symbolic reasoning [106], and generative models [107], among others [108-109]. Such methods are capable of autonomously generating hypotheses [110], designing experiments [111], uncovering latent patterns within complex datasets [112], and even proposing entirely new theoretical frameworks [113]. Thus by leveraging these tools, computational modality of creativity enhances the scope and efficiency of scientific inquiry, often extending beyond the limitations of traditional human cognition.

Within this modality, two key aspects are central which include diversity and novelty. While novelty of a creative outcome is assessed by the inherent originality of the creation with respect to its predecessors [114], diversity refers to the intrinsic differences among the creative competitors within the same cohort [115]. In other words, novelty captures how much a solution/idea departs from what has been previously seen or known, whereas diversity characterizes the spread and variation

among current ideas, models, or approaches being explored simultaneously. Thus, in the computational context, the dual emphasis on diversity and novelty enables the design of algorithms and models that simulate or support aspects of human creativity. In essence, by encoding the principles of variation and originality, computational systems can meaningfully contribute to the creative process i.e., either autonomously or in collaboration with human experts.

The possible approaches to synthesize computational modality of scientific creativity are discussed in details in the following sections.

1.10 EXPLORING COMPUTATIONAL MODALITY OF SCIENTIFIC CREATIVITY BY RANDOM SEARCH

Random search [116-119] plays a vital role in exploring the computational modality of scientific creativity, enabling machines to discover novel solutions through unbiased exploration. Unlike deterministic algorithms that follow predefined paths, random search allows machines to investigate the solution space in pursuit of interesting outcomes, without relying on prior knowledge or assumptions. The aforesaid approach embraces uncertainty and variability that leads to the generation of diverse hypotheses, unconventional ideas, or surprising insights. In other words, random search reflects the cognitive flexibility observed in humans, who are capable of producing novel ideas by randomly exploring the solution space from multiple perspectives to arrive at something new.

It is evident from history that random search has been the prime motivation behind many groundbreaking scientific discoveries and innovations. For example, discovery of penicillin happened by chance [120]. Alexander Fleming while working with Staphylococcus bacteria observed the accidental growth of mold in a petri dish had inhibited the bacteria growth. Instead of considering the above scenario as a failed experiment, Fleming's more in-depth analysis of the findings resulted in the invention of the first antibiotic that revolutionized modern medicine. Again, Wilhelm Röntgen while experimenting with cathode rays noticed a fluorescent light from a screen in the room. This aberration eventually resulted in the discovery of X-rays, which became a cornerstone of diagnostic imaging [121]. Similarly, the discovery of the principles of microwave oven also happened by accident when Percy Spencer noticed that the radar equipment melted a chocolate bar in his pocket. Instead of ignoring this odd result, Spencer investigated further which lead to the invention of the most useful kitchen technology [122]. The aforesaid instances exemplify how random exploration can reveal new possibilities that structured goal-driven research might overlook.

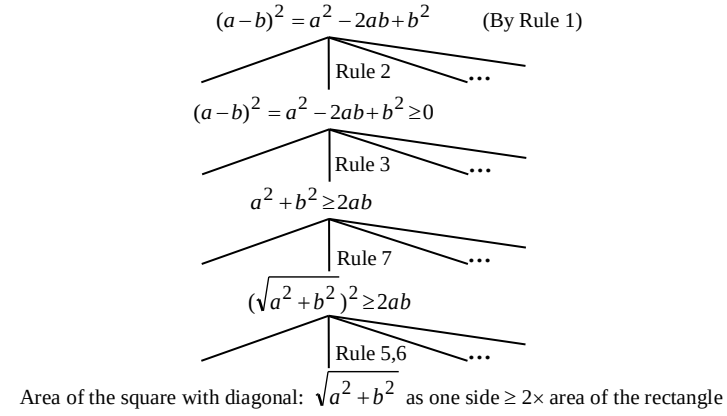


Fig. 1.16. Illustration of Random search process

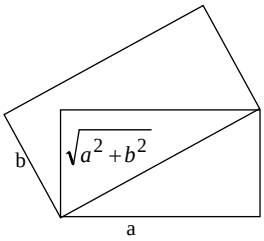


Fig. 1.17. Parameters of the square and the rectangle

To demonstrate the concept of random search, let the following set of rules be considered which may be applied in a randomly ordered sequence to a mathematical identity to derive new/meaningful results.

Rule 1: $(a-b)^2 = a^2 - 2ab + b^2$, $for\, a, b \in$ Real space.

Rule 2: $x^2 \geq 0$, $\forall x \in$ Real numbers.

Rule 3: If x=y then x ± expression = y ± expression, for any expression of real variables or constants.

Rule 4: If a and b are sides of a rectangle then area of the rectangle $= 2ab$.

Rule 5: If a and b are sides of a rectangle then diagonal of the rectangle $= \sqrt{a^2+b^2}$.

Rule 6: If a is the side of a square then area of the square a^2.

Rule 7: $(\sqrt{a})^2 = a$ for real a.

For instance, if one begins with the identity (Rule 1) and randomly apply rules to expand it gradually until interesting results are found as shown in Fig. 1.16. One may stop at a result that the area of the square with one side equal to the diagonal of the rectangle with sides a and b ≥ 2× area of the rectangle. The aforementioned result (illustrated in Fig. 1.17) is a new property which is not well-known in mathematics. So, this result is indeed a creative outcome in scientific domain.

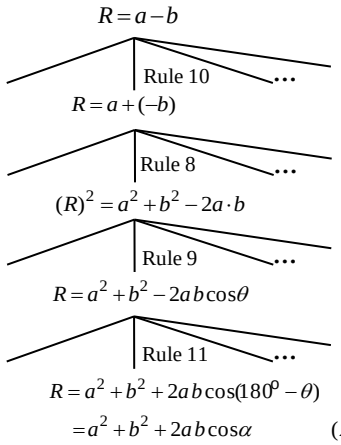


Fig. 1.18. The search tree to derive a property of obtuse-angled triangle

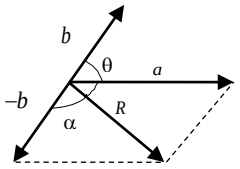


Fig. 1.19. The parameters of $R = a - b$

Now another example is considered to demonstrate the scope of random search in deriving an existing property of obtuse-angled triangle. If *a* and *b* are two vectors then Rule 1 can be restructured as

Rule 8: $(a-b)^2 = (a-b)\cdot(a-b)$
$= a^2 + b^2 - 2ab.$

Let some more additional rules are considered as follows.

Rule 9: $a \cdot b = |a||b|\cos\theta$, where θ is the angle between vectors *a* and *b*.

Rule 10: $a - b = a + (-b)$
= vector sum R of a and $(-b)$, where $-b$ is obtained by drawing b oppositely.

Rule 11: $\cos(180^0 - \theta) = -\cos\theta$

The search tree is expanded with a goal to derive interesting (new) properties. Let the root node be $R = a - b$ as depicted by Fig. 1.18 and its parameters is illustrated in Fig. 1.19. It is evident from the figure that the end result is $R^2 = a^2 + b^2 + 2ab\cos\alpha.$ The acquired result is an important (well-known) property of obtuse-angled triangle with obtuse angle=α. Thus, the above example illustrates that many important new derivations in mathematics may have emerged as a result of a random search process. In other words, mathematical discoveries, such as novel identities or proofs, are often not purely the result of structured/pre-guided reasoning but can stem from

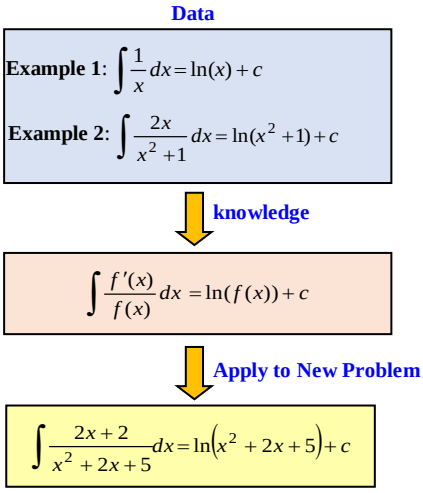


Fig. 1.20. Illustration of inductive learning by machine in Calculus domain

exploratory, trial and error based approaches that eventually lead to elegant formulations.

However, the process of random search in pursuit of a scientific creative outcome often involves navigating an extremely large and/or complex solution space. This can make the discovery process inefficient or computationally prohibitive. To address this challenge, heuristic algorithms can be employed to intelligently prune the search space so that the search explores more promising regions which in turn significantly reduces the number of unproductive paths. Such heuristics can be based on domain knowledge, adaptive learning etc., ultimately enhancing the efficiency and effectiveness of the creative search process.

1.11 DESIGNING COMPUTATIONAL MODALITY OF SCIENTIFIC CREATIVITY BY INDUCTIVE LEARNING

An alternative approach to model the computational modality of scientific creativity involves solving new problems by generalizing from a given set of specific instances. The aforementioned approach is referred to as inductive learning [123-124] that focuses on identifying patterns, similarities, or recurring structures in problem-solving strategies within or across various scientific domains to solve an unknown problem. In other words, inductive learning enables machines to extract abstract principles that can be transferred and applied to novel and/or previously unseen problems.

Thus, inductive learning enables a machine to move beyond rote memorization or rule-following and instead form generalized concepts from limited data. Furthermore, inductive learning mimics a key aspect of human's scientific creative potential which involves the capacity to reframe known solutions in new ways and

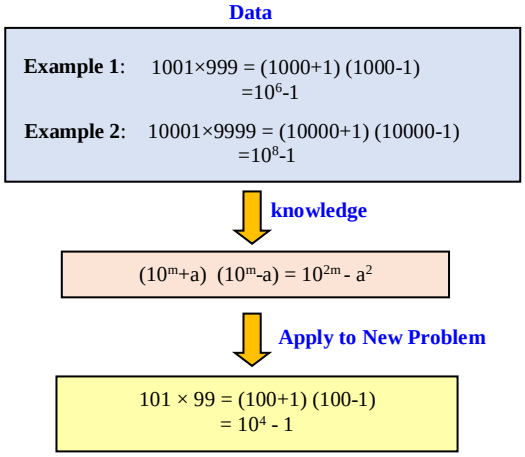


Fig. 1.21. Illustration of inductive learning by machine in Algebraic domain

repurpose existing methods to tackle unfamiliar questions. The aforesaid technique of knowledge abstraction and its transfer to new domains/problems is central to generating original hypotheses, constructing innovative models, and discovering previously unrecognized connections [125].

Suppose a machine learns some examples of solving integrals as shown in Fig. 1.20. In Example 1, during integration, the denominator is $f(x) = x$ and its derivative $f'(x) = 1$ is present in the numerator. This integration results in $\ln(x) + c$, where x is $f(x)$. Similarly, in Example 2, the denominator is $f(x) = x^2 + 1$ and its derivative $f'(x) = 2x$ is present in the numerator. This integration yields $\ln(x^2 + 1) + c$, where $x^2 + 1$ is $f(x)$. The similarity between these examples lies in the rule that the integration of $f'(x)/f(x)$ produces the natural logarithm of $f(x)$ plus a constant. By recognizing this commonality, the machine learns the underlying pattern and applies this knowledge to solve new problems as illustrated in Fig. 1.20. This ability to generalize and apply learned principles demonstrates a basic level of creativity through inductive learning.

Another example of developing the computational modality of scientific creativity through inductive learning in algebraic domain is illustrated in Fig. 1.21. Here, the machine learns from Example 1 that the multiplication of 1001 and 999 can be expressed as the product of two quantities which are (1000+1) and (1000-1). This product results in the simplified form 10^6-1. Again from Example 2, the machine learns that the multiplication of 10001 and 9999 can be expressed as the product of two quantities which are (10000+1) and (10000-1), and this product can be further simplified to 10^8-1. From these examples, the machine finds the underlying similarity between solving the above types of problems by acquiring the knowledge that the product of (10^m + a) and (10^m - a) results in the square of the base (i.e., 10^m) minus the square of the difference from the base. Now, utilizing the

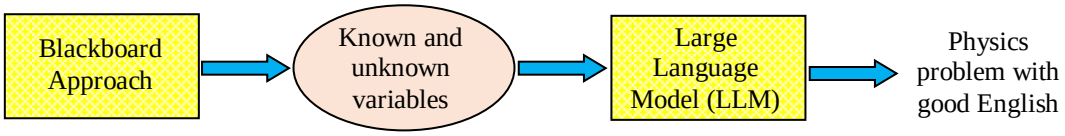


Fig. 1.22. Block diagram to illustrate physics problem generation using blackboard technique of AI and LLM

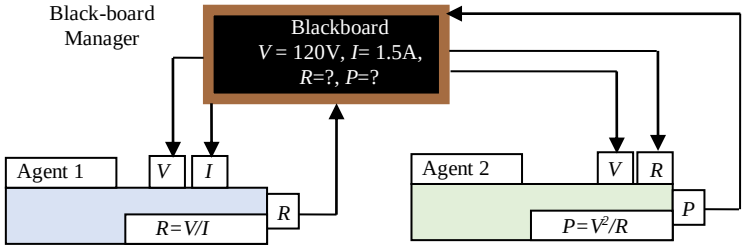


Figure 1.23. Illustration of Blackboard approach for generating a physics problem from Electricity domain

above acquired knowledge if the machine is able to solve a new problem, then the machine is said to be creative.

1.12 SYNTHESIZING THE COMPUTATIONAL MODALITY OF SCIENTIFIC CREATIVITY BY LARGE LANGUAGE MODELS

With the advent of AI, there has been a sharp rise in the design of different language models (LLMs) [126-128] such as ChatGPT [129], LLaMA [130], Gemini [131], and others [132-133]. These models are capable of performing intelligently across a wide range of tasks which includes natural language understanding, content generation, translation, question-answering, and the like. However, despite their intelligence, these models still fall short when it comes to generating novel content, particularly in the scientific domain [134]. This limitation arises primarily because current models are trained to predict and replicate patterns based on vast amounts of existing data, rather than to invent radically new paradigms. As a result, they excel at recombining known concepts but struggle with conceptual breakthroughs that lie beyond the bounds of their training data [135]. Thus, to explore the computational modality of scientific creativity using LLMs for the autonomous generation of diverse scientific problems, an additional module is required to handle the numerical aspects, such as computing the known and unknowns variables, while the language-related components can be managed by the LLMs. The numerical component can be efficiently addressed using the blackboard approach of AI [136]. The basic block diagram of the above approach is depicted in Fig. 1.22.

Central to the blackboard approach is the blackboard manager. The blackboard itself acts as a shared workspace that displays initial variables, making this

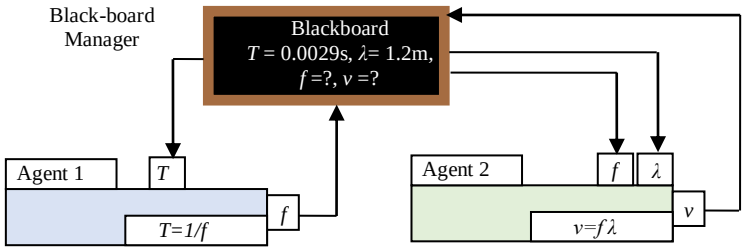


Figure 1.24. Illustration of Blackboard approach for generating a physics problem from wave mechanics domain

information accessible to various computational agents. For example (see Fig. 1.23), suppose the blackboard manager displays the values of voltage *V* and current *I*. These values are picked up by Agent 1, which uses them to compute the value of resistance *R*. Once calculated, Agent 1 returns the value of *R* to the blackboard manager, which then updates the blackboard to display *V*, *I*, and *R*. Subsequently, Agent 2 accesses the values of *V* and *R* from the blackboard and uses them to compute the power *P*. Once the value of *P* has been calculated, an LLM is utilized to generate a physics problem which can be stated as follows.

Problem 1: "An electric heater is rated for use with *V*=120 volt household power source. When it is operating, it draws a current *I*=1.5 A. Calculate the resistance *R* of the heater and also the power *P* dissipated by the heater."

The above process can again be utilized to generate a physics problem from a different domain such as wave mechanics. As shown in Fig. 1.24, let The blackboard displays the initial variables for a wave which include time-period *T* and wavelength λ. These values are picked up by Agent 1, which uses them to compute the frequency *f*. Once calculated, Agent 1 returns the value of *f* to the blackboard manager, which then updates the blackboard to display *T*, λ, and *f*. Subsequently, Agent 2 accesses the values of λ and *f* from the blackboard and uses them to compute the velocity *v*. Once the value of *v* has been calculated, an LLM is utilized to generate a physics problem which can be stated as follows.

Problem 2: "A sound wave is travelling through air with a wavelength λ =1.2m. The period of the sound wave is *T*=0.0029s. Calculate the frequency *f* of the sound wave and also determine the velocity *v* of the sound wave in air."

Algorithm 1.1 outlines the blackboard approach utilized for generating problems in physics. The process begins by initializing different agents, each responsible for computing a specific mathematical equation. An iteration counter *r* is also initialized. The agents operate in parallel, each checking whether all but one variable in their assigned equation is known. If so, they compute the unknown variable and submit it to the blackboard manager. The manager then updates the shared blackboard with

Algorithm 1.1: Problem generation in Physics using Blackboard Approach

1. Initialize Agents $1, 2, \ldots . N, N \in \mathbb{R}$, for computing the mathematical equations $1, 2, \ldots . N, N \in \mathbb{R}$. Set iteration $r = 0$.
2. **For** each Agent do in parallel
   i) If all variables in the equation are known except one, compute the unknown variable using Agents
   ii) Submit the value of the newly computed variable to the blackboard manager.
   iii) $r \leftarrow r + 1$
   End-For
3. The blackboard manager updates the newly computed variables on the blackboard.
4. i) If $r \leq$ maximum number of variables and $\alpha \leq$ pre-defined runtime, loop through Step 2.
   ii) If $r >$ run-time, stop.
5. If condition 1 holds and r = maximum number of variables, then the blackboard manager passes on the known variables and unknown variables to an LLM to construct a question.

the newly computed values. This process is repeated as long as the number of iterations r remains within a predefined limit and the total runtime α is below a user-defined threshold. Once the maximum number of variables is reached and the runtime constraint is satisfied, the blackboard manager forwards both the known and unknown variables to an LLM which formulates a physics question based on the data.

1.13 REVIEW OF BIOLOGICAL AND COMPUTATIONAL MODALITIES OF SCIENTIFIC CREATIVITY

This section first reviews the existing literature on the biological modalities of scientific creativity, followed by a review of related work on the computational modalities of scientific creativity.

1.13.1 Review on Biological Modalities of Scientific Creativity

The review on the biological modalities of scientific creativity can be divided into three main parts: i) assessing human creativity in science from behavioral responses, ii) evaluating scientific creative potential through convergent problem-solving tasks, and iii) analyzing the creative potential based on divergent thinking ability. It is important to note that, although review on the behavioral responses of subjects during scientific creative task does not directly fall under the domain of biological modalities, the behavioral markers such as solution fluency, cognitive flexibility, originality, etc., can provide valuable insights into the underlying cognitive processes of scientific creativity and act as indirect indicators of creative potential.

Thus, such measures can guide future investigations that utilize biological modalities to further explore the neural underpinnings of scientific creativity.

1.13.1.1 Review on Assessing Human creativity in Science from Behavioral Responses

This sub-section discusses some interesting works that assess the mathematical creativity of subjects based on subjects based on their behavioral responses. Such studies can support future investigations into the neural phenomena underlying scientific creativity by utilizing brain signal acquisition techniques to better understand the behavioral responses produced during mathematical creative tasks.

Leikin et al. [137] assessed the mathematical creativity of three classes of students (gifted, proficient and regular) from 10^{th}-11^{th} grade using Multiple Solution Tasks (MSTs). In these tasks, students were encouraged to solve a given mathematical problem in as many different ways as possible. This scheme aimed to measure the originality of each generated problem-solving approach using the 2-4-6 scoring system (where, 2 points are awarded for a conventional solution, 4 points for a moderately novel solution, and 6 points for a novel solution). Beyond novelty, their approach also addressed fluency (i.e., the number of different solutions a student could provide) and flexibility (i.e., the variety in the methods used, such utilizing an algebraic strategy to solve a geometry problem).

Levav-Waynberg et al. [138] extended the work in [137] by proposing a new mathematical formulation, instead of the 2-4-6 scoring scheme, that could quantitatively evaluate fluency, flexibility and originality, and also integrate them to create a final creativity score for each subject. The formulation for flexibility and is computed (1.33).

$$F = \sum_{i=1}^{n} F_i \tag{1.33}$$

where, n denotes the number of feasible solutions evaluated by a student for a given problem (also, known as fluency), and U_i denotes the flexibility score in range [0.1, 10] of the i^{th} solution. The originality of each generated solution is evaluated using (1.34).

$$O = \sum_{i=1}^{n} O_i \tag{1.34}$$

where, n denotes fluency and O_i denotes the originality score in range [0.1, 10] of the i^{th} solution depending on the level of novelty of the generated solution. The final creativity score C of a subject is computed using (1.35).

$$C = \sum_{i=1}^{n} F_i \times O_i \tag{1.35}$$

It is important to note that the work in [138] had been performed for MSTs only from geometric domain. Thus, the authors in [139] utilized the formulations from [138] to assess the mathematical creative potential of subjects in the algebraic domain. Similarly, the authors in [140] utilized the formulations from [138] to assess the mathematical creative potential of subjects across a wide range of mathematical domains. Additionally, Lev et al. [141] employed the formulation in [138] to understand the connection between mathematical giftedness and mathematical creativity.

Ugur Sak et al. [142] developed a mathematical formulation for assessing scientific creativity by calculating the fluency and flexibility of responses in the domains of biology, interdisciplinary science, chemistry, physics, and ecology, based on tasks involving hypothesis generation, hypothesis testing, and evidence evaluation. The mathematical formulation utilized here is referred to as the Creativity Quotient (CQ) which was derived from [143] and is expressed by (1.36).

$$CQ = \log_2\{(1+n_1)(1+n_2).....(1+n_c)\} \tag{1.36}$$

where, n denotes the number of correct responses pertaining to a given category while c indicates the total number of categories. On the contrary, Pasztor et al. [144] developed a mathematical formulation for assessing the creative ability of subjects in science using (1.37).

$$O = \left[1 - \frac{L+b}{2n}\right]^{14} \tag{1.37}$$

where, O denotes the originality score of an answer, L indicates the number of responses pertaining to a single domain, b represents the number of responses in a sub-category and n indicates the total number of responses. The index of 14 is utilized in (1.37) to keep the originality score within the range [0, 1].

1.13.1.2 Review on Assessing the Scientific Creative Potential of Subjects through Convergent Problem-Solving Tasks

Convergent problem-solving tests comprise problems that are designed to evaluate a subject's ability to identify a single correct solution based on logical reasoning within a pre-defined set of constraints [145]. Although these tests emphasize arriving at one solution, the strategies employed may vary, involving either an analytical approach or a creative insight-driven process [146-148]. Analytical problem solving involves a deliberate and systematic search through the problem space. On the contrary, insight based problem solving is characterized by the spontaneous and

often unexpected emergence of a solution, typically accompanied by a pronounced "Eureka" or "Aha!" experience. This sudden cognitive breakthrough is considered a critical mechanism underlying the operation of a creative thought process. This sub-section reviews the literature that attempts to distinguish between creative and analytical individuals while solving convergent problems through brain signal analysis.

Dandan et al. [149] employed functional magnetic resonance imaging (fMRI) to investigate the brain activity of participants as they solved technical problems in science and engineering. Their analysis revealed that the left dorsolateral prefrontal gyrus and left angular gyrus were actively engaged when participants experienced the Eureka effect during problem-solving. In a similar vein, Hao et al. [150] conducted an fMRI-based study, which inferred the active involvement of the middle temporal gyrus and the middle occipital gyrus while participants tackled scientific problems using an insight-driven approach.

A significant body of research has investigated brain activity during insight-driven problem solving, particularly in the context of Chinese riddles [151-157]. For instance, Qiu et al. [151] identified the activation of the quadrate lobule of Foville, left inferior and medial frontal gyri, inferior occipital gyrus, and the little brain (i.e., cerebellum) during insight-based solutions. Similarly, Tian et al. [152] confirmed the involvement of the left middle frontal gyrus, left superior temporal gyrus, right cerebellum, bilateral claustrum, and left post-central gyrus for the aforesaid task. Again, Zhao et al. [153] found that the initial stages of solving Chinese riddles through insight were associated with activations in the left superior temporal pole, temporal gyrus, anterior cingulate cortex, medial frontal gyrus, right inferior frontal gyrus, para-hippocampal gyrus, bilateral gyri, and middle temporal gyrus. On the contrary, the later stages of insight based solutions involved activations in the olfactory bulb, medial frontal gyrus, anterior cingulate cortex, medial frontal gyrus, inferior parietal gyrus, right putamen, amygdala, bilateral medial temporal gyrus, hippocampus, and angular gyrus. Additionally, Luo et al. [154] conducted an event-related potential (ERP) study and observed negative ERP deflections, particularly in the N300-500 and N1100-1300 ranges, during insight-based problem-solving. Shen et al. [155] noted that an early P200 in the fronto-central regions was associated with the pre-conscious realization of a mental impasse during the Eureka effect, while a subsequent P300 reflected the sustained focus during such a mental block. Furthermore, Xing et al. [156] reported that insight strategies were linked to negative N300-500 ERP deflections across most brain regions, while positive P600-1100 ERP deflections were found in the frontal, fronto-central, and central regions. Finally, Zhao et al. [157] demonstrated that insight-based approaches to solving both

Semantic Riddles (SR) and Homophonic Riddles (HR) elicited a N350-500 ERP response across the central brain regions.

1.13.1.3 Review on Assessing the Scientific Creative Potential of Subjects from Divergent Thinking Ability

The Alternate Uses Task (AUT) [21], [158] is a widely used measure to assess divergent thinking, which refers to the ability to generate multiple and/or varied solutions to a given problem. Although the AUT may/may not involve scientific or mathematical content, it effectively captures divergent thinking which is considered as a critical component in scientific creativity. This sub-section discusses existing works focused on identifying the creative ability of subjects through brain signal analysis during their performance on AUT.

Fink et al. [159] performed an EEG and fMRI-based study to analyze the brain activation patterns of subjects while solving AUTs. Their EEG signal analysis revealed that the generation of novel ideas involved increased alpha band power over the frontal and parietal regions. Moreover, the fMRI analysis indicated strong activation of the left frontal regions and parietotemporal brain areas. In a similar vein, Vartanian et al. [160] revealed the activation of prefrontal cortex while solving AUTs through fMRI signal analysis. Again, Roberts et al. [161] inferred the active participation of temporal pole, hippocampus, rostrolateral prefrontal cortex, dorsolateral prefrontal cortex, insula, and anterior cingulate cortex while performing an fMRI study on subjects solving AUTs.

There has been a significant amount of research on detecting the neural underpinnings of the AUT using EEG [162-167]. For instance, Fink et al. [162] performed an EEG study on AUTs and inferred that the original idea generation was associated with higher upper band alpha activity (10–12 Hz) in the prefrontal region of the brain. In a similar vein, Schwab et al. [163] revealed that original idea generation was characterized by an increase in alpha power at the commence of idea generation followed by a declination and finally a re-rise of alpha power preceding the response over parietal and temporal regions of the right hemisphere. On the contrary, Agnoli et al. [164] revealed alpha synchronization over the frontal, central, and temporal regions during the production of unique thoughts in the first phase of divergent thinking, while the later phases of divergent thinking were characterized by alpha synchronization over the centro-parietal regions. Again, Kroger et al. [165] reported the elicitation of N400 signal when subjects produced novel uses of a given object in AUTs. Additionally, Rataj et al. [166] reported the activation of upper alpha band from parieto-occipital regions and lower alpha band from left anterior parts of the brain during AUTs. They also reported the elicitation of N400 signal

during the production of original ideas. Finally, Hartog et al. [167] conducted an AUT-based EEG study on engineering students and revealed the elicitation of the N400 signal when participants were able to provide novel uses for the portrayed objects.

1.13.2 Review on Computational Modalities of Scientific Creativity

The review of computational approaches to scientific creativity can be divided into three main categories: i) automatic generation of mathematical word problems, ii) automated scientific knowledge discovery, and iii) automatic generation of new theorems and conjectures. The details of each of the above categories are provided below.

1.13.2.1 Review on Automatic Math Word Problem Generation

Automatic word problem generation that ensures diversity with respect to existing problems aims to produce varied and novel problems by altering context, structure, and difficulty using generative models. This capability is vital to computational creativity, as it enables machines to simulate human-like innovation and adaptability in educational content creation. This sub-section discusses the existing literature on the aforesaid domain.

Polozov et al. [168] proposed an automatic mathematical word problem generation scheme that can generate problems depending on teacher-defined mathematical constraints and student-defined narrative preferences. Teacher-defined constraints specify the mathematical structure of the problem, such as the required operations (e.g., addition, multiplication), and/or specific equation patterns (e.g., x=?+3×?). On the other hand, student-defined narrative preferences determine the context of the story such as the desired literary setting (e.g., fantasy, science fiction), characters (e.g., their names, genders), and relationships among characters (e.g., friends, adversaries). The proposed system first encodes the tutor and student defined constraints and uses an Answer-Set Programming (ASP) [169] model to synthesize both a mathematical expression and a logical story graph that jointly satisfy all constraints. It ensures narrative coherence through a system of discourse tropes i.e., logical templates enforcing plausible storylines. The logical graph is then realized into a math word problem using a natural language model. The basic block diagram of the above model is depicted by Fig. 1.25.

The work in [168] relied on rigid templates and/or manually crafted data, which in turn limited the variability among the generated word problems. This limitation

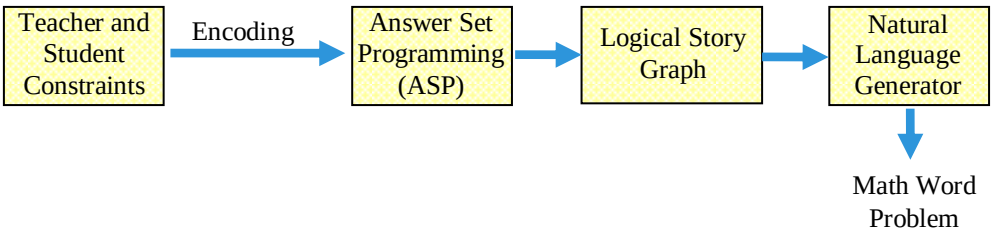


Figure 1.25. Block diagram of math word problem generation scheme proposed by Polozov et al.

was mitigated by Zhou et al. [170], an automatic math word problem generator based on a neural network was designed referred to as MAGNET (Math Word Problem Generation NETwork). The prime objective of the proposed work involved the generation of math word problem W as a sequence of words $W=\{w_1,w_2,....,w_k\}$ from a given topic-word list $T=\{t_1,t_2,.....,t_n\}$ and equation template $E=\{e_1,e_2,.....,e_m\}$. The aforesaid objective can be expressed mathematically in the form of a probability distribution using (1.38).

$$P(W\mid T,E)=\prod_{i=1}^{k}P(w_k\mid w_1,w_2,....,w_{k-1},T,E) \tag{1.38}$$

The mechanism of the MAGNET model is briefly illustrated in Fig. 1.26. The model takes two inputs: an equation template and a set of topic words. The equation template is encoded using a bidirectional Gated Recurrent Unit (GRU) [171] to capture its semantic representation, while the topic words are encoded through an embedding lookup table. The encoded representations of both inputs are then fused to integrate mathematical structure with contextual information. This fused output is subsequently transformed into a math word problem using a unidirectional GRU. Additionally, the fusion module incorporates an attention mechanism to enhance the diversity of the generated problems. The proposed model was trained on the Dolphin18K dataset [172], which consists of a wide range of problems and thus helps to further improve the diversity and quality of the generated output.

However, MAGNET suffers from four key limitations. First, MAGNET's encoding mechanism using a GRU limits its ability to capture the hierarchical structure and long-range dependencies of expressions (e.g., nested operations or operator precedence). Second, its dual attention mechanism treats topic words and expressions independently, missing the inter-dependencies between topic context and mathematical structure effectively. Third, the entity-enforced loss does not guarantee logical consistency or solvability. Fourth, the model lacks a verification mechanism (e.g., a solver) to ensure the generated problem yields the correct expression or answer.

Wu et al. [173] proposed MWPGen, a neural model for generating math word problems from topic words and math expressions, addressing key limitations of

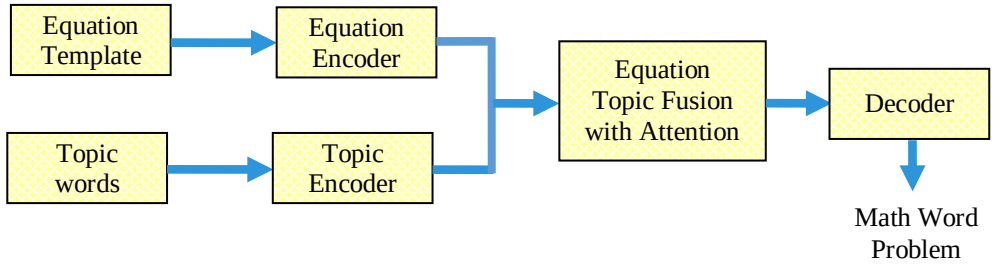


Figure 1.26. Block diagram of math word problem generation scheme proposed by Zhou et al.

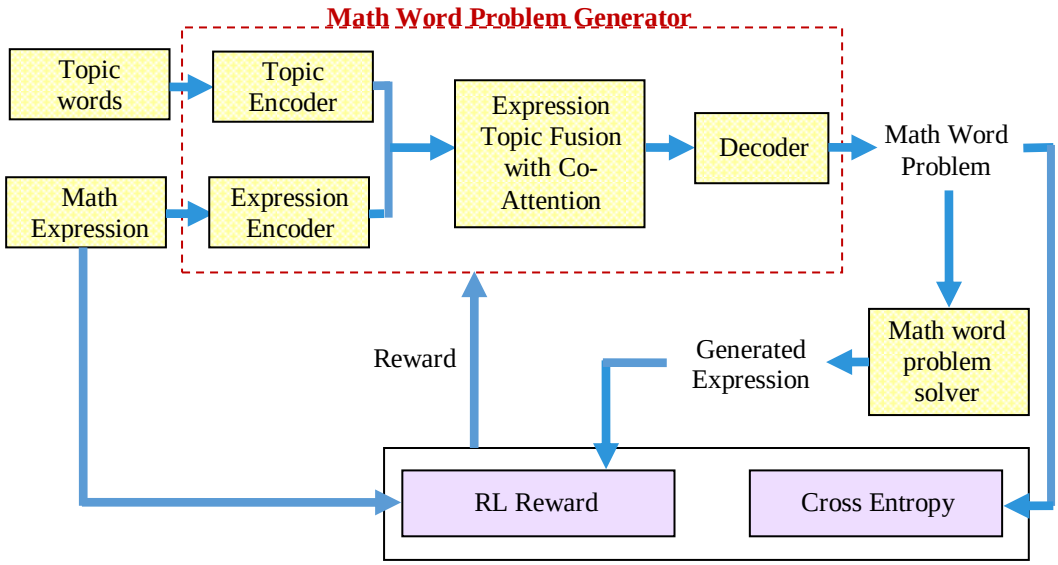


Figure 1.27. Block diagram of math word problem generation scheme proposed by Wu et al.

earlier models like MAGNET. Their approach introduces a topic-expression co-attention mechanism to jointly model the relevance between topic words and expressions, ensuring the generated problems are both topically and mathematically coherent. To better capture the structural semantics of math expressions, they represent expressions as pre-order traversals of expression trees and incorporate adjacent node embeddings (parent, left child, right child) into the encoding process. Additionally, they apply reinforcement learning, using a math problem solver to assess the solvability of generated problems and provide reward signals for fine-tuning. The model is initially trained using cross-entropy loss, which helps it learn to generate grammatically and semantically appropriate problem statements based on reference data. However, since cross-entropy does not directly optimize for problem correctness, reinforcement learning is employed in a second phase to guide the model toward generating problems that lead to correct and solvable equations. Experiments on the Math23K dataset [174] show that MWPGen produces more fluent, complete, and solvable problems than prior methods. A brief overview of the approach in [173] is illustrated in Fig. 1.27.

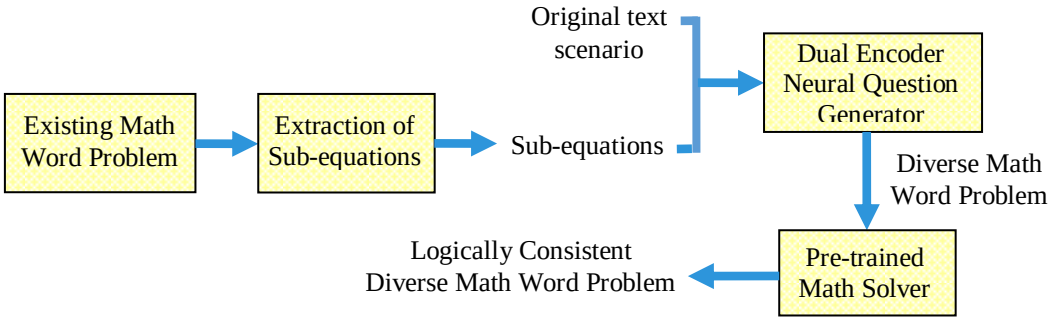


Figure 1.28. Block diagram of math word problem generation scheme proposed in [175].

A major drawback of all previous approaches is the lack of diversity in the generated problems. To address this limitation, several recent methods have been proposed to introduce greater diversity in problem generation. For instance, Zhou et al. [175] proposed a novel framework called DQGF (Diverse Questions Generation Framework) to enhance diversity in math word problems by generating multiple semantically consistent questions for a single scenario. In the proposed pipeline, diverse sub-equations are first extracted from a given math problem using heuristic rules that isolate partial computations or unit-based operations. These sub-equations are then paired with the original scenario text and fed into a dual-encoder neural question generator, which models number-level interactions to produce varied yet relevant questions. To ensure quality, a filtering module based on a pre-trained math solver is used to retain only logical and solvable question-equation pairs. This process mimics analogy learning, as the model generates alternative questions by drawing parallels between different sub-equations derived from the same scenario—effectively teaching the solver to solve by analogy. The resulting dataset, DiverseMath23K, significantly extends Math23K with question-level variety. The proposed architecture is outlined in Fig. 1.28.

In a similar vein, Chen et al. [176] proposed ControlMath, a controllable data generation framework to enhance the diversity and generalization of math word problem solvers. The method uses an iterative pipeline driven by large language models (LLMs), beginning with an equation generator that creates varied mathematical expressions using rule-based controls. These equations are then transformed into coherent problems by a Problem-Crafter Agent, and filtered by a Reverse-Agent that discards low-quality or redundant examples, resulting in ControlMathQA, a large-scale dataset of approximately 190,000 diverse, high-quality problems. Likewise, Zhang et al. [177] introduced DivKD, a diversity-enhanced knowledge distillation framework for generating multiple valid solution equations. DivKD uses an adaptive distillation strategy in which the teacher selectively transfers high-quality equation variants to a student model. The student

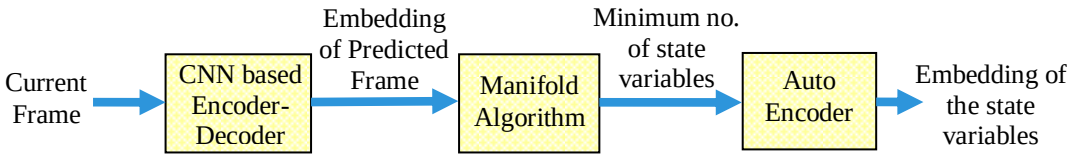


Figure 1.29. Block diagram of automated state variable estimation framework

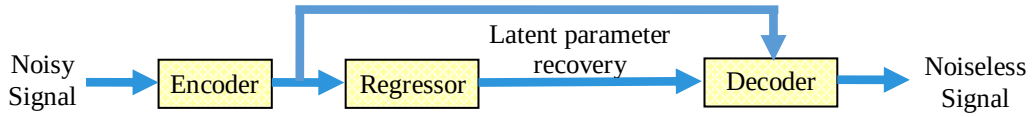


Figure 1.30. Block diagram for automated latent parameter recovery and denoising of noisy oscillating time-series.

incorporates a conditional variational autoencoder (CVAE) to model the distribution of plausible equations, enabling diverse generation at inference.

1.13.2.2 Review on Automated Scientific Knowledge Discovery

Automated scientific knowledge discovery refers to the use of different computational methods to identify patterns, formulate hypotheses, and/or uncover laws from data without any human intervention. This paradigm is central to computational creativity, as it mimics human-like discovery processes that enables machines to automate not just data analysis but also the generation of novel ideas, scientific theories, problem-solving strategies and the like. This sub-section discusses the works undertaken in the aforementioned domain.

Chen et al. [178] proposes a novel framework for the automated discovery of fundamental state variables directly from high-dimensional observational data, such as videos of physical systems. Unlike traditional approaches that assume prior knowledge of relevant variables (e.g., position, velocity), this method uses a two-stage process. First, it trains a fully convolutional encoder–decoder network to predict future video frames, and then applies manifold learning (i.e., Levina–Bickel estimator) on the latent embeddings to estimate the intrinsic dimension (ID) which includes the minimum number of variables required to fully describe the system's dynamics. Second, it constructs a compact latent space of dimension equal to the estimated ID, yielding neural state variables that encapsulate the system's evolution. Experiments undertaken revealed that the acquired variables correlate strongly with known physical quantities when available, such as angles, velocities, or energies and thus are sufficient for accurate long-term prediction. Thus the proposed approach enables unsupervised discovery of interpretable and dynamically meaningful variables, laying the foundation for automated scientific reasoning in complex

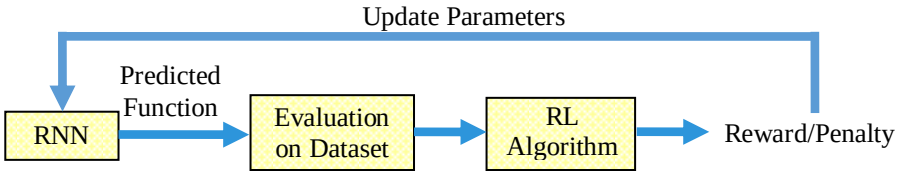


Figure 1.31. Block diagram of automated function discovery from dataset

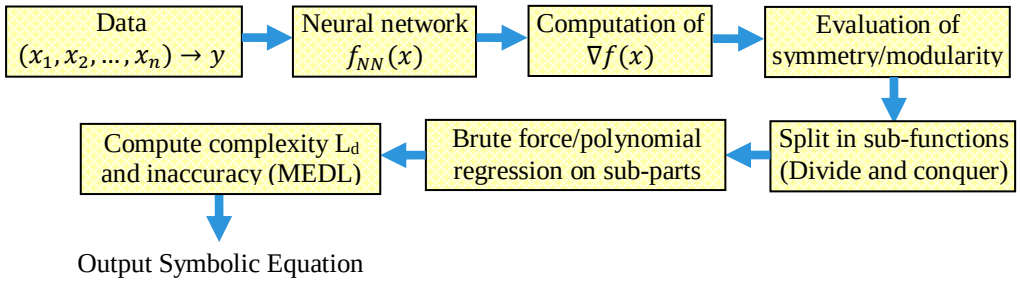


Figure 1.32. Block diagram of automated symbolic equation discovery from data

physical environments. The block diagram illustrating the above process is depicted by Fig. 1.29.

Garcon et al. [179] propose a deep neural network (DNN) architecture designed to recover unknown physical parameters and denoise noisy oscillating time-series signals, such as decaying monochromatic, amplitude-modulated (AM), and frequency-modulated (FM) sine waves, without requiring prior knowledge of the signal type. The model consists of an encoder that extracts compressed latent features from the noisy input, a regressor that predicts explicit physical parameters like frequency and phase, and a decoder that reconstructs a clean, denoised signal using both the encoder's representation and the regressor's output. By training on synthetic data generated from known signal types with randomized latent parameters, the model learns to generalize across various signal forms. The block diagram illustrating the aforementioned framework is shown in Fig. 1.30.

Peterson et al. [180] introduce a novel framework called Deep Symbolic Regression (DSR), in which a recurrent neural network (RNN) learns to generate interpretable mathematical expressions that fit a given dataset. Unlike conventional regression models, the RNN in DSR does not receive the dataset as direct input. Instead, it samples expressions token by token in a pre-order traversal of expression trees, initially in a random manner. Each generated expression is then evaluated on the dataset, and a scalar reward based on how well the expression fits the data is computed. This reward is not used as input to the RNN but serves as a learning signal to update the RNN's parameters through a reinforcement learning method known as the risk seeking policy gradient. Over time, the RNN gradually transitions from random guessing to generating high reward expressions, effectively learning a

distribution over symbolic equations that explain the observed data. The block diagram illustrating the entire process is depicted by Fig. 1.31.

Udrescu et al. [181] introduce a symbolic regression framework, AI Feynman 2.0, which recovers closed-form mathematical expressions from data by combining neural function approximation, modularity detection, and Pareto-based model selection. Given a dataset of input–output pairs $\{(x_i, y_i)\}$, the method trains a feedforward neural network $f_{NN}(x)$ to approximate the unknown function $f(x)$. This neural approximation enables the computation of gradients $\nabla f(x)$, which are then analyzed to identify structural patterns such as compositionality, generalized symmetry, and additivity. For example, if $f(x) = g(h(x))$, then the gradients ∇f and ∇h should be directionally aligned, i.e., $\widehat{\nabla f} = \pm\widehat{\nabla h}$. Such patterns allow the system to recursively decompose the original problem into simpler sub-problems. Each sub-function is then approximated using polynomial fitting or brute-force symbolic search. Candidate expressions are scored based on their inaccuracy, computed as the mean error description length (MEDL), and their complexity, measured via description length L_d. Only expressions on the Pareto frontier, representing optimal trade-offs between simplicity and accuracy, are retained. This hybrid approach enables the discovery of exact or near-exact symbolic forms, such as $\frac{1}{2}mv^2$ or $\sin(x_1 + x_2)$ from numerical data, even in the presence of noise. The block diagram describing the aforesaid process is depicted in Fig. 1.32.

1.13.2.3 Review on Automated Generation of New Theorems and Conjectures

The automatic generation of new theorems and conjectures in mathematics by intelligent systems represent a core aspect of computational creativity that enables machines to contribute original theorems or insightful conjectures in domains such as number theory, geometry, algebra, etc. This sub-section discusses the works undertaken in the aforementioned domain.

Buchberger et al. [182] introduce Theorema, a unified software system designed to support the complete process of mathematical theory development, including the invention of definitions, theorems, and algorithms, as well as their formal verification and execution. Implemented within the Mathematica [183] environment, Theorema includes a wide range of internal reasoning tools, covering predicate logic, equational reasoning, differential equations, geometry, and program verification, and also connects to external provers through a dedicated interface. A key contribution of the work is the Lazy Thinking method, which enables the synthesis of correct algorithms by analyzing failed proof attempts and generating precise specifications for unknown components. This method is illustrated through the automated

derivation of a Gröbner basis algorithm, a task that involves selecting appropriate algorithm schemata and generating supporting sub-algorithms.

Chen et al. [184] propose an automated method for discovering and proving congruence theorems involving partial sums of combinatorial sequences that can be expressed as the constant term of Laurent polynomial expressions. Let $P(x)$ and $Q(x)$ be Laurent polynomials with integer coefficients. The sequence $\{a_k\}$ is defined by (1.39).

$$a_k = CT\left[P(x)^k Q(x)\right] \tag{1.39}$$

where, CT denotes the constant term (i.e., the coefficient of x^0) in the expansion of the Laurent polynomial. For a given positive integer r and a prime number p they they study the congruence of the partial sum using (1.40).

$$S_{r,p} = \sum_{k=0}^{rp-1} a_k \bmod p \tag{1.40}$$

The main theorem (Theorem 2.1 in the paper) shows that this sum modulo p can be expressed as a linear combination of terms from a C-finite sequence $\{S_k\}$ using (1.41).

$$S_k = \text{coeffcient of } x^k \text{ in } \frac{Q(x)}{P(x)-1} \tag{1.41}$$

Let $-m$ be the lowest degree of $P(x)$, and let c_j be the coefficient of x^{-rm+j} in the expansion of $P(x)^r - 1$. Then, for any prime number $p > n$, where $-n$ is the lowest degree of $\frac{Q(x)}{P(x)-1}$, then,

$$\sum_{k=0}^{rp-1} a_k \equiv \sum_{j=0}^{rm} c_j \cdot S_{(rm-j)p} \bmod p \tag{1.42}$$

This result allows symbolic computation of such congruences using computer algebra systems such as Maple [185]. Furthermore, they prove that in many cases the values of these partial sums modulo varying primes lie in a finite set. Using these results, the authors are able to recover and generalize known congruences and discover new ones. Thus their methods implemented in a Maple package, help in the automated verification and generation of congruence identities for a broad class of combinatorial sequences.

Raayoni et al. [186] propose a novel automated framework called the Ramanujan Machine for generating conjectures on fundamental mathematical constants such as π, e, the Catalan constant, and values of the Riemann zeta function. The focus is on discovering regular formulas (RFs), especially in the form of continued fractions, using only numerical data, without prior symbolic knowledge or logical derivations. They target identities of the general form using (1.43).

$$\frac{\gamma(c)}{\delta(c)} = PCF(\alpha, \beta) \tag{1.43}$$

where, c is a fundamental constant, and $\alpha(x), \beta(x), \delta(x), \gamma(x) \in Z[x]$ are integer polynomials. The right-hand side of (1.43) is a polynomial continued fraction (PCF) defined by (1.44).

$$PCF(\alpha, \beta) = \cfrac{\alpha(0)}{\beta(0) + \cfrac{\alpha(1)}{\beta(1) + \cfrac{\alpha(2)}{\beta(2) + ...}}} \tag{1.44}$$

The first algorithm in [186] referred to as Meet-In-The-Middle for Regular Formulas (MITM-RF), enumerates expressions on both sides of Equation (1.43). It pre-computes low-precision values of the left-hand side and stores them in a hash table. It then evaluates candidate PCFs from Equation (1.44) for various polynomial pairs α, β and matches them with stored values. Promising matches are validated at high numerical precision (e.g., 2,000 digits) to reduce false positives. The second algorithm in [186] referred to as Descent&Repel, frames the search for continued fraction representations of fundamental mathematical constants as an optimization problem. It defines the loss function as

$$\ell = \left| \frac{\gamma(c)}{\delta(c)} - PCF(\alpha, \beta) \right| \tag{1.45}$$

and aims to minimize it to zero under the constraint that all polynomial coefficients are integers. The algorithm performs gradient descent to approach local minima where $\ell = 0,$ combined with a repulsion step between multiple candidate solutions to explore a broader parameter space. A final step forces convergence to integer-valued solutions. Using the above methods, the authors generated dozens of continued fraction identities, including new representations for π and e. An exemplar problem generated by their word is shown by (1.46).

$$\frac{4}{3\pi - 8} = 3 - \cfrac{1.1}{6 - \cfrac{2.3}{9 - \cfrac{3.5}{12 - .....}}} \tag{1.46}$$

The generated conjectures, though obtained without symbolic manipulation, have been proven post hoc by the community, demonstrating the potential of numerical pattern recognition in advancing mathematical discovery. The utilization of the Ramanujan Machine in [186] thus shifts the conventional paradigm: from proof-driven derivation to conjecture-driven exploration.

1.14 SCOPE OF THE THESIS

The aim of the present research can be divided into two main parts. The first part explores the biological underpinnings of scientific creativity to identify the varying levels of creative potential among subjects engaged in a creative task. This involves

acquiring functional Near-Infrared Spectroscopy (fNIRS) and Electroencephalographic (EEG) signals from the scalp of participants during the aforesaid task. These signals are then transformed into brain connectivity networks using correlation or coherence-based analyses. Subsequently, feature extraction is carried out to determine the active Brodmann Areas (BAs) involved in the present cognitive process. The extracted features are further analyzed and classified to distinguish between different levels of creative potential across individuals. The second part focuses on the computational modeling and artificial synthesis of scientific creativity by machines or computers. In this context, the goal is to enable a computer or an algorithm to develop new heuristics for searching novel scientific or mathematical solutions. It may also include the design of different learning algorithms that enables machines to abstract patterns and structures from previous data in order to generate original mathematical concepts, such as formulating new types of problems, identities, or proofs. Thus, the primary objective of the second part of the thesis involves the development of algorithms that enables machines to mimic human-like creative reasoning.

The thesis under consideration will comprise seven chapters, with Chapter 1 providing a comprehensive review of existing works on both the biological and computational modalities of scientific creativity. From Chapter 2 onwards, the thesis is divided into two main parts. Part I consists of Chapters 2, 3, and 4, which present the candidate's original contributions under the biological modalities of scientific creativity, while Part II comprises Chapters 5 and 6, focusing on the computational modalities of scientific creativity. Chapter 2 deals with the classification of different levels of creative potential in individuals within the scientific domain, based on their spatial reasoning ability, using brain connectivity patterns derived from fNIRS data. Chapter 3 focuses on distinguishing between creative and analytical cognition by analyzing the neural arousal associated with the Eureka effect during a convergent reasoning task, using EEG signals. Chapter 4 explores the classification of varying levels of creative potential based on analogical reasoning ability, again leveraging brain connectivity analysis from fNIRS recordings. Chapter 5 presents the development of a depth-limited Best-First Search (BFS) algorithm designed to generate diverse mathematical identity problems suitable for inclusion as chapter-end exercises in mathematics textbooks. Chapter 6 involves the development of a Generative Adversarial Network (GAN) framework, in which new mathematical problems are generated using leaked information transmitted from the discriminator to the generator to enhance problem diversity and structure. The concluding chapter of the thesis will outline the scope for future work as an extension of the proposed

methodologies. A summary of the main contributions presented in Chapters 2, 3, 4, 5, and 6 is provided below.

Summary of Chapter 2: Spatial reasoning ability, defined as the capacity to mentally manipulate an object's various attributes such as shape, orientation, pattern, and size, plays a crucial role as a cognitive process influencing scientific creativity. This chapter aims to identify different levels of creative potential in individuals performing a spatial reasoning task referred to as the Mental Paper Folding (MPF), using brain connectivity networks derived from functional Near-Infrared Spectroscopy (fNIRS) data. The research is structured into three main stages. In the first stage, brain connectivity networks are constructed using Pearson's correlation method to capture inter-relationships among different brain lobes. The second stage involves extracting centrality features from these networks, including degree, closeness, and betweenness, to determine the most active Brodmann Areas (BAs) engaged in the afore-said spatial reasoning task. In the third stage, the extracted centrality features are fed into a proposed Graph Convolutional-Interval Type-2 Fuzzy Network (GC-IT2FN) for the classification of creative potential into four distinct levels. The novelty of the proposed framework is threefold. First, a self-attention mechanism is introduced within the network architecture to enable the graph convolutional layers to focus on the most relevant nodes contributing to the classification task. Second, a new activation function, Logish, is employed after the graph convolution operation to improve the accuracy and stability of the classifier. Third, the classifier utilizes the most promising region within the Footprint of Uncertainty (FOU) of the fuzzy sets used in the Interval Type-2 Fuzzy Network, which helps in reducing the impact of uncertainty inherent in brain signal data. Experimental analyses conducted demonstrate the effectiveness of the proposed approach in classifying levels of scientific creative potential based on spatial reasoning performance in comparison to its traditional counterparts.

Summary of Chapter 3: There is a growing demand for scientific creative individuals in various departments of industrial research and development sectors. At the same time, analytical thinkers are essential in investigation departments, academic institutions, and management domains. However, classifying individuals as either creative or analytical based solely on behavioral responses is a challenging task. This chapter presents an approach to classify individuals into four categories (Analytical, High Creative, Medium Creative, and Low Creative) based on their brain responses recorded during participation in a creativity test grounded in a convergent thinking task. The proposed classification process is carried out in two main phases. In the first phase, a brain connectivity map is generated from the electroencephalogram (EEG) data using Pearson's correlation technique. In the

second phase, three centrality features (degree, closeness, and betweenness) are extracted from the connectivity map and passed into a classifier model to assign the appropriate category label. The classifier designed for this task synergistically combines a Graph Convolutional Network, which abstracts the brain connectivity-based centrality features, with a Capsule Network, which performs the classification, resulting in a Dual Attention Induced Graph Convolutional-Capsule Network (DAIGC-CapsNet). The main contribution of the proposed classifier involves the design of a dual attention modules and a Sparsemax function based routing algorithm. The dual attention mechanism includes a Mish Induced Attention Module (MI-AM), which helps the graph convolutional layers focus on the most significant node features, and a Fused Attention Module (F-AM), which ensures that the most relevant predictions are transmitted from the primary capsule layer to the class capsule layer. Furthermore, the coupling between the primary and class capsule layers is enhanced by a Sparsemax-based routing algorithm, which improves the interpretability and of the classifier network. Experimental results validate the proposed method's effectiveness, demonstrating its superiority over conventional approaches in accurately categorizing individuals based on their cognitive traits.

Summary of Chapter 4: Analogical reasoning ability is defined as the capacity to draw comparisons between distinct concepts based on their structural or functional similarities, and it plays a crucial role in fostering scientific creativity. This chapter presents a novel approach to evaluate the scientific creative potential of individuals by analyzing their brain connectivity patterns through fNIRS data while they engage in Raven's Advanced Progressive Matrices (RAPM), a well-known analogical reasoning test. The proposed method involves three key stages. In the first stage, brain connectivity networks are constructed using Wavelet Transform Coherence (WTC). The second stage focuses on the abstraction and analysis of three node-based network features which include strength, efficiency and betweennness. In the third stage, the extracted features are classified into five levels of creative potential using a novel model referred to as Enhanced Graph Convolution Induced Type-2 Fuzzy Classifier (EGCIFC). The classifier introduces five major innovations. First, an enhanced graph convolution operation is designed to capture both local and global structural information from the input graph, allowing the model to learn diverse patterns. Second, a new activation function called Smish is incorporated into the graph convolution layer to improve performance. Third, a one-dimensional spatial convolution layer replaces the traditional pooling layer to retain the most relevant information. Fourth, a novel mapping function is introduced to handle uncertainty among the spatially convolved vectors in the type-2 fuzzy layer. Fifth, the Takagi-Sugeno-Kang (TSK) fuzzy reasoning mechanism is used to reduce computational

cost. Experimental analysis confirms that the proposed framework outperforms conventional methods by a large margin in assessing scientific creative ability.

Summary of Chapter 5: The chapter presents an innovative approach to synthesizing computational creativity through a process inspired by deductive reasoning, with an added mechanism for evaluating the diversity of generated instances in comparison to their predecessors. This two-step process, comprising expansion and testing, is implemented using best-first search on an OR-tree structure, where the nodes represent trial solutions or new creations, and the edges indicate the substitution formula applied to a parent node to generate a child node. Two alternative versions of best-first search are explored, differing in the cost functions applied at each node to ultimately identify the optimal node within a user-specified depth as the solution to the creativity problem. The first algorithm focuses on maximizing the diversity cost of a node with respect to its immediate parent, while the second aims to maximize the difference between the diversity and a penalty cost of a node with respect to the root node. A key contribution of this research lies in maintaining diversity throughout the iterative expansions of the search tree and ensuring the novelty of the optimal solution across different runs of the program. The chapter also provides a comparative analysis of the performance of the two algorithms in terms of their applicability and effectiveness. Additional properties of the proposed algorithms, such as soundness and logical completeness, are discussed to establish their robustness. Experimental results show that the algorithms generate more diverse mathematical identities compared to conventional methods. Furthermore, the algorithms have been successfully applied to the development of chapter-end exercises in propositional logic, set theory, and algebra, demonstrating their potential for generalization across different mathematical domains.

Summary of Chapter 6: Computational creativity refers to the development of artificial systems capable of producing artifacts that exhibit significant novelty or originality, comparable to those created by human professionals. While various efforts have been made to apply computational creativity in artistic and non-scientific fields, its implementation in the scientific domain remains largely unexplored. To address this gap, the current study introduces a novel method for incorporating computational creativity within the scientific context, demonstrated through the automatic generation of new trigonometric identity problems intended for use as preliminary-level chapter-end exercises in mathematics textbooks. This objective is achieved by designing a new neural network model called Enhanced Generative Adversarial Network with Leaked Information, or E-LeakGAN. The distinctiveness of E-LeakGAN lies in its discriminator architecture, which is built on

a convolutional neural network that ensures the generation of original and high-quality trigonometric identity problems. The model works by abstracting key features from the input data and leaking or transferring these features to the generator module, thereby guiding it to produce high-quality problems based on this information. The convolutional module introduces three core innovations. First, it employs a new activation function named Trish to enhance the expressive capability of the discriminator. Second, a unique dual attention mechanism is incorporated, which utilizes channel attention to determine which semantic information is important and spatial attention to locate where this critical information is positioned syntactically. Third, an eigenvector guided pooling technique is developed to direct the attention-induced convolved features toward the most semantically and syntactically significant content. Experimental evaluations using three categories of metrics (BiLingual Evaluation Understudy, Mean Normalized Cosine Similarity, and Mean Normalized Diversity) highlight the superior performance of E-LeakGAN in comparison to conventional methods. The model also demonstrates efficacy in generating identity problems across other mathematical domains such as algebra and calculus, making it a generalizable tool for developing chapter-end exercises.

Part I

Biological Modalities of Scientific Creativity

2

Assessment of Scientific Creative-Skill from Spatial Reasoning Ability Using Graph Convolutional-Interval Type-2 Fuzzy Network

Scientific creativity refers to natural/automated genesis of innovations in science, propelling scientific, technological, industrial and/or societal progress. Mental paper folding (MPF) requires spatial reasoning, which is an important attribute to determine creative-potential of people. The chapter proposes a novel approach to determine creative potential of people from their brain-connectivity network (BCN) during their participation in MPF tasks using functional Near-Infrared Spectroscopy (fNIRS). The work involves three phases. The first phase includes construction of BCN using Pearson's correlation method. The centrality features of the nodes in the network are assessed in the second phase, and transferred to a proposed Graph Convolutional-Interval Type-2 Fuzzy Network (GC-IT2FN) in the third phase to classify the creative potential of individuals in four grades. The novelty of the work includes i) a novel self-attention mechanism in the network to guide graph convolution layers to focus on the most relevant

nodes, ii) selection of a new activation function, Logish, after graph convolution to enhance classifier-accuracy, and iii) utilizing the promising region in the Footprint of Uncertainty (FOU) of the used fuzzy sets of IT2FN-based classifier to reduce the effect of uncertainty in brain data on classifier-performance. Experiments conducted demonstrate the efficacy of the proposed framework in contrast to traditional approaches.

2.1 INTRODUCTION

Historically, creativity [1-2] has been intertwined with the domain of art, spanning the origination and realization of a myriad of artistic expressions, from music and painting to poetry, as well as encompassing a broad spectrum of fine arts, culinary pursuits, and performance arts. In contrast, scientific creativity emerges as a distinct facet within the broader conceptual framework of creativity, highlighting its unique focus on unveiling innovation within the scientific domain. Within this context, various cognitive skills play a pivotal role in shaping creative outcomes, such as divergent thinking [3], inductive learning [4], convergent analysis [5], analogical reasoning [6], and the like. Beyond the afore-mentioned cognitive skills, the ability to mentally reshape an object's different attributes (such as shape, orientation, pattern, size etc.) emerges as a pivotal element in the process of creative ideation. This cognitive process, commonly denoted as spatial reasoning [7], has demonstrated its essential role in influencing creative thinking across a spectrum of scientific domains such as mathematics [8-10], engineering [11-12], chemistry [13], medicine [14] etc. The current chapter seeks to examine the creative potential of individuals in scientific domain by assessing their spatial reasoning ability through the utilization of functional Near Infrared Spectroscopy (fNIRS).

The spatial reasoning ability of an individual can be evaluated through various cognitive tasks [15], including mental rotation of 2D or 3D figures and mental paper folding (MPF) (which involves visualizing the resulting pattern after a series of folding operations on a piece of paper). The current study focuses on the MPF task to evaluate the spatial reasoning proficiency of subjects. This task involves intricate mental rotations and a sequence of complex visual transformations, thereby providing a comprehensive measure of spatial reasoning [16].

Current literature [17-20] on the MPF task explores the connection between MPF activity and mathematical problem-solving skills. Recently, Dahm *et al*. [21] attempted to investigate the role of action imagery during MPF task. Goumopoulos *et al*., in a very recent study [22], focused on detecting the cognitive load in MPF problem solvers using Electrocardiogram (ECG) signals. In neuroscience studies related to MPF tasks, research predominantly centers on exploring the active regions

of the brain during said cognitive activity using functional Magnetic Resonance Imaging (fMRI) [23]-[26], Electroencephalogram (EEG) [27]-[28], or Positron Emission Tomography (PET) scans [16]. There exists hardly any literature on the inter-relationship between creative potential of a subject and his/her brain-connectivity during solving a MPF task. This chapter attempts to determine the brain-connectivity involved in a spatial reasoning problem, such as MPF, to assess the creative potential of the subject. In a very recent study, Ghosh *et al*. [29] focused on assessing individuals' creative potential by analyzing EEG signals associated with paper folding skills. Nevertheless, the high cost of brain scanning devices (such as fMRI, MRI, PET, etc.) and the limited spatial resolution of EEG due to volume conductivity [30] preclude their utilization for the current application. In contrast, the fNIRS device provides moderately high spatial resolution at a lower cost and entails minimal computational overhead. Hence, the selection of such a device aligns with the practical considerations for the present application.

The primary aim of the present study is to evaluate the varying levels of creative ability in individuals based on their proficiency in paper folding, employing a classifier built on fNIRS data. This objective is accomplished through a two-stage process. In the initial stage, fNIRS signals collected from subjects engaged in a MPF task are subjected to pre-processing and transformation into a brain connectivity network (BCN). This network is constructed using the Pearson's Correlation (PC) technique [31]. The selection of the PC technique is motivated by its simplicity, interpretability [32], and demonstrated efficacy across diverse domains [31], [33]-[34]. Following afore-said transformation, the obtained brain networks undergo feature abstraction using three centrality measures (degree, closeness, and betweenness) [35]-[36] to identify the Brodmann Areas (BAs) of the brain that play a central role in coordinating the overall brain network during the MPF based cognitive task. The second stage of the study involves classifying the abstracted features into four distinct degrees of creative potential: High Creative (HCR), Medium Creative (MCR), Low Creative (LCR), and Non-Creative (NCR).

The key contribution of the present research lies in the formulation of a suitable classifier model capable of handling the afore-mentioned classification task. Since the current problem deals with brain connectivity-based features, the classifier is meticulously designed on the principles of a Graph Convolution Network (GCN) [37]-[40], which converts node representations into graph embeddings through the first-order approximation of spectral graph convolutions [37]-[38]. The GCN model comprises two main components where the first part involves an automatic feature learning process using graph convolution and pooling operations, while the second part classifies the pooled vectors using a fully connected (FC) network. However, the

fNIRS response acquired from a given source is prone to variations within and across sessions due to undesirable parallel thoughts and various technical and/or physiological artifacts [41]-[42], thereby introducing uncertainty among the pooled feature vectors. To address this issue, the present work introduces a Type-2 Fuzzy Network (T2FN) [43] in conjunction with the GCN. A T2FN adeptly manages the uncertainty among feature vectors by handling intra and inter-session variations [43]-[44]. Consequently, the combined approach of GCN and T2FN aims to enhance classifier performance by mitigating the limitations associated with uncertainties introduced during sessional variations.

In this chapter, a novel classifier model, referred to as Graph Convolutional-Interval Type-2 Fuzzy Network (GC-IT2FN) has been designed to categorize individuals with varying spatial problem-solving skills into 4 distinct classes (HCR, MCR, LCR, and NCR) by leveraging the synergies of GCN and T2FN. The novelties of the GC-IT2FN model involves i) introduction of a self-attention mechanism in the GCN that guides the subsequent graph convolution layers to focus on the most significant nodes within the BCN, ii) employment of Logish [45] activation function (due to its inherent negative activation response to signals of small negative amplitude) after the graph convolution operation to enhance the classifier performance, and iii) introduction of a new policy to utilize the most promising region in the Footprint of Uncertainty (FOU) [46]-[47] of the used fuzzy sets of IT2FN to minimize the effect of uncertainty on classifier performance. The policy stated above is accomplished by incorporating a new mapping function to uplift the Lower Membership Function (LMF) of individual type-2 fuzzy sets, thereby utilizing the relatively promising upper region of the FOUs to improve classification accuracy. The efficacy of the proposed GC-IT2FN has been demonstrated through rigorous performance analysis, establishing its superiority over existing state-of-the-art (SOTA) techniques.

The subsequent sections of the chapter encompass the following content. Section 2.2 details the principles and methodologies adopted to address the current classification problem. Section 2.3 elaborates the architecture of the proposed classifier and the technique employed to rank individuals based on their creative ability. Section 2.4 presents the experiments and results of the current cognitive experiment. Section 2.5 showcases the performance of the proposed classifier in comparison to state-of-the-art techniques. Finally, Section 2.6 offers the inferences drawn from the present work.

2.2 PROPOSED FRAMEWORK

This section provides a concise overview of the experimental paradigm, with Fig. 2.1 illustrating the schematic diagram of the entire system. The experiment initiates by

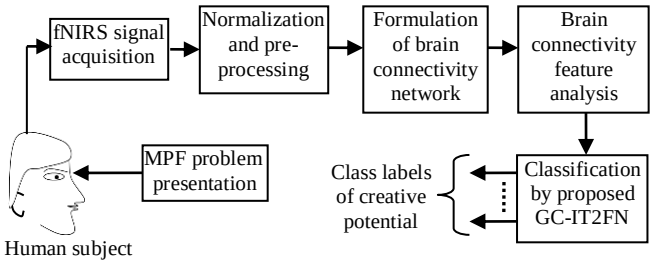


Fig. 2.1. Schematic overview of the proposed framework

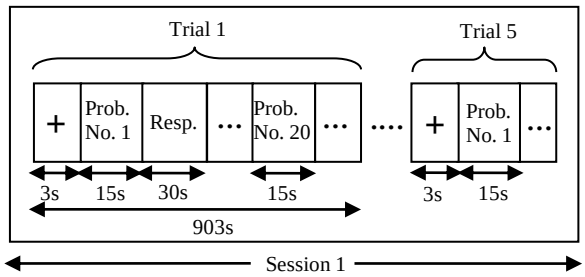


Fig. 2.2. Visual stimuli utilized for a single session of the MPF task

capturing fNIRS signals from the participants' scalp during their engagement in the MPF task. The visual stimuli structure for the current application pertaining to a single session is depicted by Fig. 2.2. A single session comprises 5 trials where each trial begins by 3-second presentation of a fixation cross. Subsequently, a 15-second time-window follows, presenting the mental problem that participants are required to solve. For the present application, the MPF task involves participants to view diagrams depicting the folding of a paper and the punching of a hole (of any shape) within it. The challenge then prompts participants to mentally envision the spatial arrangement of the holes after unfolding the paper [48]-[49]. An example of such a problem is illustrated in Fig. 2.3. After this phase, the visual stimuli consist of a 30-second time-window during which participants communicate their answers to the experimenter by drawing their attempted solution to the MPF task on a sheet of paper. The afore-mentioned process is repeated for 20 different problem sets within a single trial. Notably, a 30-second time gap is preserved between consecutive trials to alleviate the potential residual effect of the preceding stimulus. The entire experiment is composed of 5 sessions, each encompassing 5 trials. Across 10 experimental days, a total of 5000 experimental instances per subject are generated for the present scenario, calculated as follows: 20 problems/trial × 5 trials/session × 5 sessions/day × 10 days.

The fNIRS signals acquired from the acquisition phase undergo normalization, pre-processing, and transformation into a brain connectivity network for an in-depth analysis of inter-lobe interactions during the ongoing cognitive task. Subsequently, brain connectivity based centrality features are derived and employed as input to a

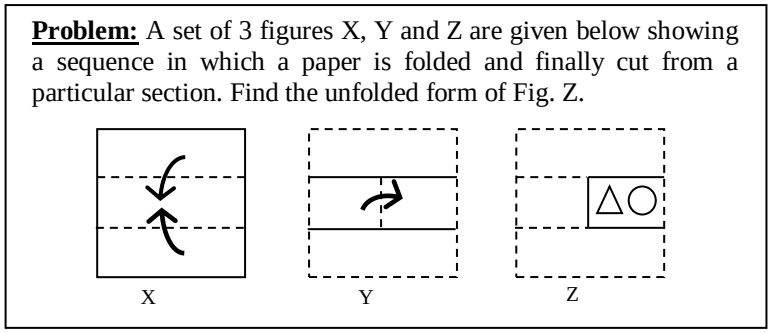


Fig. 2.3. Exemplar problem utilized for the MPF task

novel GC-IT2FS classifier, designed to categorize subjects' creative abilities into 4 distinct classes.

The 4 categories of class labels (HCR, MCR, LCR, NCR) were determined based on a scoring policy framed by 10 spatial reasoning experts. Here, each paper folding task involves 3 steps of folding and a run-through punching of the paper, where the results of unfolding the 4 segments of the paper is independent of the order of folding the paper. After a detailed discussion among the experts, a consensus about the following scoring policy was decided. For each segment of a paper-folding task, 25 marks are assigned. Again, within each segment, 5 points are reserved for correctly identifying number and type (e.g., circular/triangular) of cuts, 8 points for correctly identifying the order of type of cuts (e.g., circle followed by triangle in left-to-right order), and 12 marks for correctly identifying the orientation of the shaped cuts (e.g., inverted/upright triangles). The sub-score for each segment of a paper is evaluated first by adding the scores for above 3 items and the final score out of 100 is the sum of the sub-totals for four segments of a paper. A few samples of solutions provided by the subjects and their corresponding scores have been included in Section A.3 of the Appendix.

The detailed procedures involved in the present classification task are elucidated below.

2.2.1 Normalization of Raw fNIRS Data

The raw fNIRS signals acquired from the scalp of subjects consists of two types of blood concentrations: oxy-hemoglobin blood concentration (in mmol/L) and deoxy-hemoglobin blood concentration (in mmol/L). Let the oxy-hemoglobin blood concentration and deoxy-hemoglobin blood concentration for the λ^{th} channel corresponding to a given montage during a time-interval τ be denoted as $CHbO(\tau)$ and $CHbR(\tau)$ respectively. Since, $CHbR(\tau) < CHbO(\tau)$ for all τ from a brain region corresponding to the given montage [44], the normalization of fNIRS signals is performed by first computing (2.1) and (2.2).

$$Max(CHbO) = Max_{\tau}(CHbO(\tau) : \tau_s \le \tau \le \tau_e, \forall \lambda) \tag{2.1}$$

$$Min(CHbR) = Min_{\tau}(CHbR(\tau) : \tau_s \le \tau \le \tau_e, \forall \lambda) \tag{2.2}$$

where, τ_s and τ_e represent the starting and ending time of an experimental trial respectively corresponding to a given stimuli for a particular subject. Next, in order to evaluate the oxygen consumption during a time instant τ, the difference between the values of oxy-hemoglobin and deoxy-hemoglobin blood concentration is considered. The afore-said evaluation is represented by (2.3).

$$d(\tau) = CHbO(\tau) - CHbR(\tau) \tag{2.3}$$

The normalized value of the difference signal in (2.3) within the range of [0,1] during a time instant τ is computed using (2.4).

$$\hat{d}(\tau) = \frac{CHbO(\tau) - CHbR(\tau)}{Max(CHbO) - Min(CHbR)} \tag{2.4}$$

The data acquired from a fNIRS device, like any measurement method, it is susceptible to various forms of artifacts. The prevalent types of artifacts that deserve special consideration are: i) physiological artifacts and ii) technical artifacts [41]-[42]. Physiological artifacts are related to biological processes within the human body and can be categorized into 3 main classes [41], [50]-[52]: a) Motion Artifacts: they are caused by head movements, muscle contractions, or other bodily motions, b) Systemic Physiological Artifacts: they arise due to the fluctuations in blood pressure, heart rate, and respiration and c) Skin Blood Flow Artifacts: they occur due to changes in blood flow within the skin, such as those related to skin temperature and emotional responses like blushing. Technical artifacts in fNIRS signals are related to different aspects of instrumentation, data acquisition, and the environment in which the measurements are taken. In other words, they arise due to fluctuations in light intensity, interference from other nearby electrical devices, optode placement instability etc. [42].

To eliminate the afore-mentioned artifacts from the raw fNIRS signal, a rigorous signal processing approach is employed. Initially, the normalized difference signal $\hat{d}(\tau)$ for $\lambda = 1$ to H channels undergoes filtering using an Elliptical band-pass filter of order 4, with a pass band range set at 0.1-2 Hz. The selection of the Elliptical band-pass filter is made due to its sharp roll-off characteristics around the cut-off frequency, ensuring precise isolation of the desired frequency components while effectively attenuating noise and artifacts [53]. Subsequent to the filtering stage, Independent Component Analysis (ICA) [54] is applied to the filtered signals. This technique helps to restore the independent hemodynamic components corresponding to each channel.

2.2.2 Pre-Processing Stage

The data acquired from a fNIRS device, like any measurement method, it is susceptible to various forms of artifacts. The prevalent types of artifacts that deserve special consideration are: i) physiological artifacts and ii) technical artifacts [41]-[42]. Physiological artifacts are related to biological processes within the human body and can be categorized into 3 main classes [41], [50]-[52]: a) Motion Artifacts: they are caused by head movements, muscle contractions, or other bodily motions, b) Systemic Physiological Artifacts: they arise due to the fluctuations in blood pressure, heart rate, and respiration and c) Skin Blood Flow Artifacts: they occur due to changes in blood flow within the skin, such as those related to skin temperature and emotional responses like blushing. Technical artifacts in fNIRS signals are related to different aspects of instrumentation, data acquisition, and the environment in which the measurements are taken. In other words, they arise due to fluctuations in light intensity, interference from other nearby electrical devices, optode placement instability etc. [42].

To eliminate the afore-mentioned artifacts from the raw fNIRS signal, a rigorous signal processing approach is employed. Initially, the normalized difference signal for 1 to H channels undergoes filtering using an Elliptical band-pass filter of order 4, with a pass band range set at 0.1-2 Hz. The selection of the Elliptical band-pass filter is made due to its sharp roll-off characteristics around the cut-off frequency, ensuring precise isolation of the desired frequency components while effectively attenuating noise and artifacts [53]. Subsequent to the filtering stage, Independent Component Analysis (ICA) [54] is applied to the filtered signals. This technique helps to restore the independent hemodynamic components corresponding to each channel.

2.2.3 Formulation of Brain Connectivity Network (BCN)

In this sub-section, the pre-processed fNIRS signals are transformed into a connectivity graph to explore the interaction between the different BAs involved in the current cognitive task. The afore-said exploration begins by computing the Pearson's correlation coefficient for every pair of fNIRS channel using (2.5).

$$\rho_{XY} = \frac{1}{h-1}\sum_{i=1}^{h}\left(\frac{x_i-\bar{x}}{\sigma_X}\right)\left(\frac{y_i-\bar{y}}{\sigma_Y}\right) \tag{2.5}$$

In (5), ρ_{XY} represents the correlation coefficients values for a pair of fNIRS channels *X* and *Y*, *h* represents the number of data points pertaining to a given fNIRS channel, $\bar{x}, \bar{y}, \sigma_X$ and σ_Y represent the mean and standard deviation values of the data points of fNIRS channels *X* and *Y* respectively.

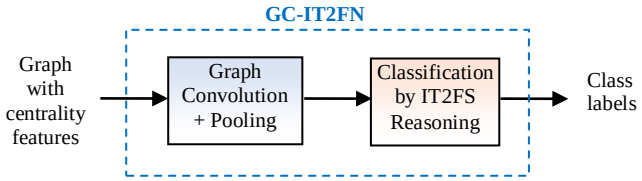


Fig. 2.4. Architectural overview of the proposed GC-IT2FN model

The above computation yields a $K \times K$ matrix representing the correlation between a pair of fNIRS channels. In the next step, the correlation coefficients are normalized using max-min normalization technique [55]. The normalized coefficients are utilized to construct an undirected (binary) graph or BCN by adopting the following strategy.

$$\alpha_{ij} = \begin{cases} 1, & if \ \hat{\rho}_{ij} \geq \zeta \\ 0, & otherwise \end{cases} \tag{2.6}$$

where α_{ij} represents the adjacency matrix values of the graph, $\hat{\rho}_{ij}$ denotes the normalized Pearson's correlation coefficients and ζ indicates a pre-defined threshold. The pre-defined threshold has been carefully chosen to ensure that all the connectivity networks/graphs share the same mean degree, facilitating their straight-forward comparability [56].

2.2.4. Feature Extraction of Brain Connectivity Network

The brain connectivity networks acquired using PC technique needs to be quantified to explore the hub nodes controlling the entire brain network. The afore-mentioned quantification is performed by abstracting three features: degree centrality (DC), closeness centrality (CC) and betweenness centrality (BC) [35]-[36].

The abstracted features are ultimately classified into four distinct classes using the GC-IT2FN classifier, the intricate architecture of which is thoroughly explained in the following section.

2.3 PROPOSED GRAPH CONVOLUTIONAL-INTERVAL TYPE-2 FUZZY NETWORK (GC-IT2FN)

The architectural overview of the proposed GC-IT2FN model is illustrated in Fig. 2.4 which comprises two main modules: i) a graph convolutional block and ii) IT2FS classifier. The functionality of each layer is elucidated below.

2.3.1 Attention Induced First Graph Convolution Layer with Logish Activation Function

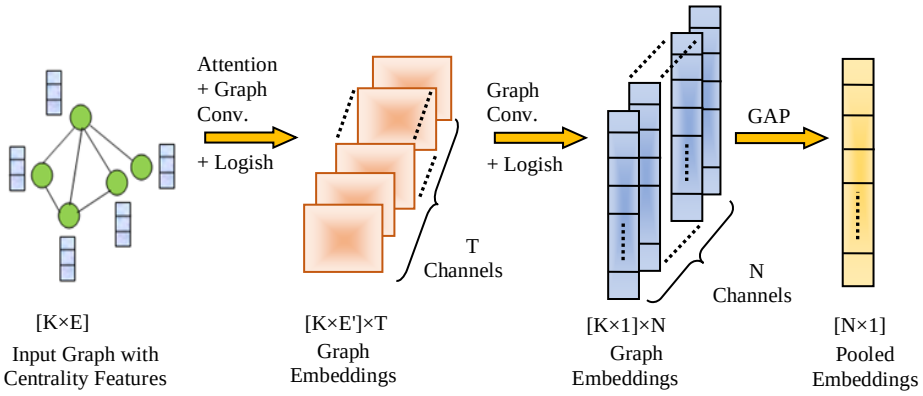


Fig. 2.5. Formation of global average pooled embeddings

The initial layer of the proposed model involves two primary operations: an attention-induced graph convolution operation and the application of the Logish activation function. The detailed aspects of the aforementioned operations are discussed below and illustrated in Fig. 2.5.

2.3.1.1 Attention Induced Graph Convolution

The traditional graph convolution operation [37] abstracts the necessary features from the input graph $P \in \Re^{K \times E}$ and transforms it into low dimensional vectors of size $K \times E'$ (where, $E' < E$) or graph embeddings using (2.7).

$$Z_j^{l+1} = \sigma(\sum_i \hat{A} Z_i^l W_{ij}^l) \tag{2.7}$$

where, $Z^l \in \Re^{K \times E}$ denotes the input graph in the l^{th} layer and $Z^0 = P$. $\hat{A} = \tilde{D}^{-1/2} \tilde{A} \tilde{D}^{-1/2}$ represents the normalized graph Laplacian [37] that consists of the structural information inherent in the graph. $\tilde{A} = A + I$ indicates the adjacency matrix comprising self-loops while $I \in \Re^{K \times K}$ signifies the identity matrix. $\tilde{D} = D + I$, where D represents the diagonal degree matrix and $\tilde{D}_{i,i} = \sum_j \tilde{A}_{ij}$. W_{ij}^l denotes the input to hidden layer trainable weight matrix. $\sigma(\cdot)$ signifies an activation function that adds non-linearity within the given neural network.

However, the present work introduces a self-attention mechanism before graph convolution operation that helps the subsequent graph convolution layers to focus on the most relevant nodes. Such a self-attention mechanism computes the attention coefficients between adjacent nodes *i* and *j* with feature vectors p_i and p_j using (2.8).

$$\phi_{ij} = \frac{1}{\exp(\| p_i - p_j \|_1)} \tag{2.8}$$

where, ϕ_{ij} denotes the attention coefficients between adjacent nodes *i* and *j* while $\| \cdot \|_1$ represents the L1 norm operation. In (2.8), the L1 norm operation assesses the similarity or proximity of feature vectors for two nodes. Consequently, a lower L1

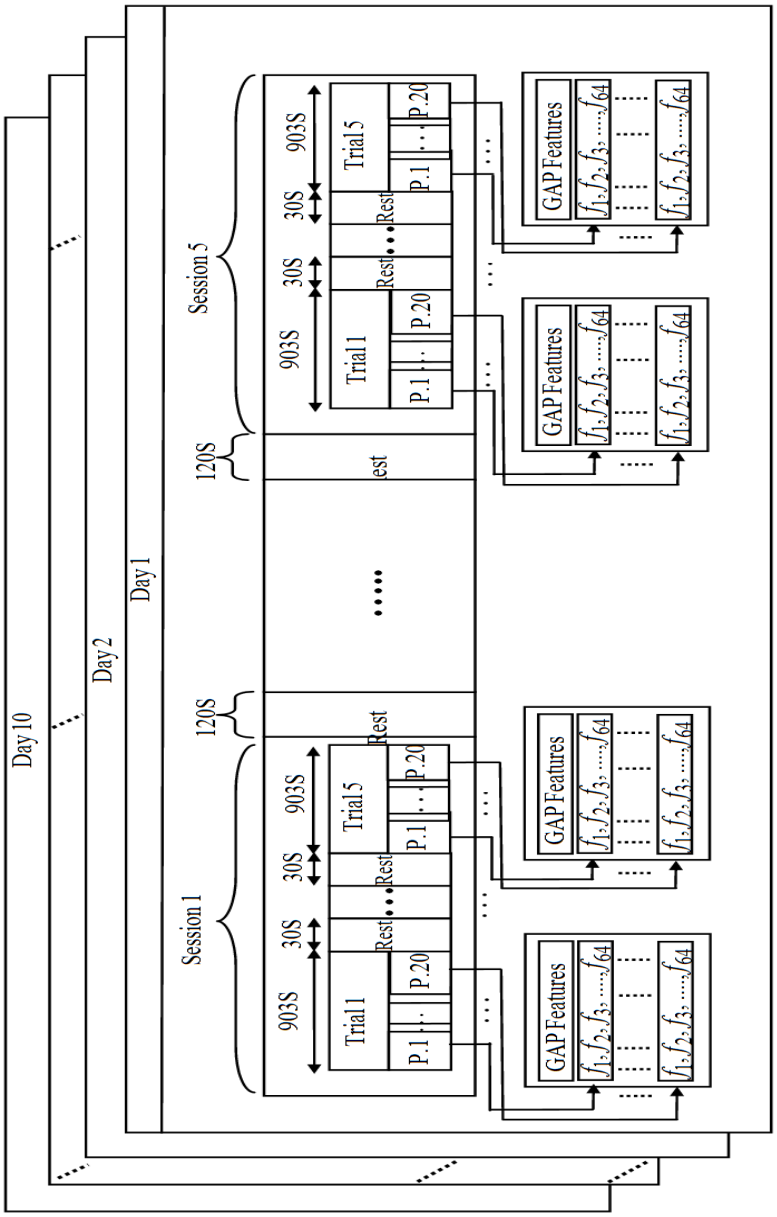


Fig. 2.6. The GAP features acquired from 5 sessions, each containing 5 trials (every trial containing 20 MPF problems), repeated over 10 experimental days.

norm value suggests greater similarity between nodes, while a higher value indicates dissimilarity. Nevertheless, to assign greater attention weights to similar nodes, the reciprocal of the L1 norm operation is considered. Furthermore, in (2.8), the result of the L1 norm undergoes an exponential operation prior to the afore-said reciprocal operation. This process ensures that the attention coefficient for the self-node ϕ_{ii} attains the maximum attention value i.e., 1, while attention coefficients among other nodes ϕ_{ij} are confined within the range (0, 1). For the current scenario, if a node m is not a neighbor of node i, then (2.8) is not applicable and thus $\phi_{im} = 0$ is considered.

The attention coefficient values for node *i* with respect to all the nodes within the graph forms a coefficient vector $\phi_i = [\phi_{i1}, \phi_{i2}, ..., \phi_{iK}]^T$. The coefficient vectors for all the nodes within the graph collectively constitute a $K \times K$ matrix $\Phi = [\phi_1, \phi_2, ..., \phi_K]$ which is symmetric i.e., $\phi_{ij} = \phi_{ji}$.

In order to make the graph convolution layers focus on the most important nodes, the matrix Φ is utilized to modify the graph Laplacian using (2.9).

$$\hat{M} = \hat{A} \otimes \Phi \tag{2.9}$$

where, $\hat{M}$ represents the modified graph Laplacian and $\otimes$ denotes the Hadamard product. After this, the modified graph Laplacian is employed in the graph convolution operation as shown in (2.10).

$$Z_j^{l+1} = \sigma(\sum_i \hat{M} Z_i^l W_{ij}^l) \tag{2.10}$$

Thus, the modified Laplacian matrix enables the convolution layers to concentrate on specific aspects of the graph structure. This, in turn, serves as a mechanism to enhance the encoding of the underlying relationships and patterns within the graph.

2.3.1.2. Utilization of Logish Activation Function

Although various conventional activation functions could have been employed to introduce non-linearity to the classifier, the authors have specifically chosen to use the Logish [45] activation function for the current scenario. This particular activation function is preferred for several reasons. Its non-linear nature ensures that the neural network does not reduce to a mere linear operation, enhancing its capacity to capture complex patterns in the data. Another distinctive feature of Logish is its unbounded nature above, effectively addressing the vanishing gradient problem and promoting a faster convergence rate for the classifier. Furthermore, the activation function is bounded below, contributing to a robust regularization effect that aids in preventing over-fitting. Additionally, the non-monotonic behavior of the Logish function (i.e., imparting small negative magnitude for small negative input value) plays a crucial role in stabilizing the training of negative values, thereby increasing the overall stability of the network. This unique set of properties associated with the Logish activation function serves as a strong motivation for its utilization in the current application, enhancing the expressive capabilities of the neural network while addressing common challenges in training and convergence. The mathematical expression of Logish is indicated by (2.11).

$$\sigma(x) = Logish(x) = x \cdot \ln(1 + sigmoid(x)) \tag{2.11}$$

2.3.2. Second Graph Convolution Layer with Logish Activation Function

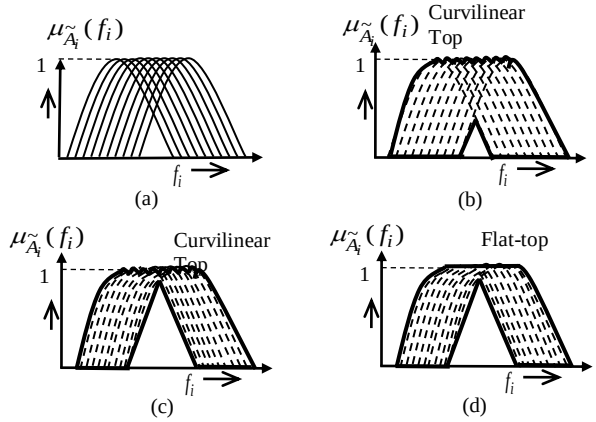


Fig.2.7. Formulation of IT2FS with modified FOU (a) Type-1 MFs for 10 days (b) Curvilinear-top based IT2FS produced by union of Type-1 MFs (c) Curvilinear-top based IT2FS with modified FOU (d) IT2FS with flat-top approximation

To capture higher-level features, another graph convolution operation (as shown in Fig. 2.5) is applied to the convolved output from the initial layer using equations (2.7) and (2.11). This strategic choice enables the network to extract and learn more complex features, enhancing its capability to understand intricate patterns within the data [57].

2.3.3. Global Average Pooling Layer

Global average pooling (GAP) is a pooling operation that reduces the spatial dimensions of a feature map to a single value per feature channel by taking the average of all values in that channel [57]. Let $V \in \Re^{K\times 1\times N}$ be the feature map that has been computed by the second graph convolution operation where $K\times 1$ denotes the size of each feature map while N represents the number of feature channels. Let the i^{th} feature vector of V having size $K\times 1$ be denoted as $V_i = [v_1, v_2, ..., v_K]^T$. The GAP operation for V_i is represented by (2.12).

$$\beta(V_i) = (v_1 + v_2 + ... + v_K)/K \tag{2.12}$$

where, $\beta(\cdot)$ represents the GAP function.

2.3.4. Interval Type-2 Fuzzy layer

The GAP operation produces a feature vector of size $N\times 1$ (where $N = 64$) for every MPF problem as shown in Fig. 2.6. The management of uncertainty introduced within these GAP vectors due to sessional variations is accomplished through an Interval Type-2 Fuzzy Network (IT2FN), the details of which are elaborated below.

2.3.4.1. Antecedent Construction of IT2FS

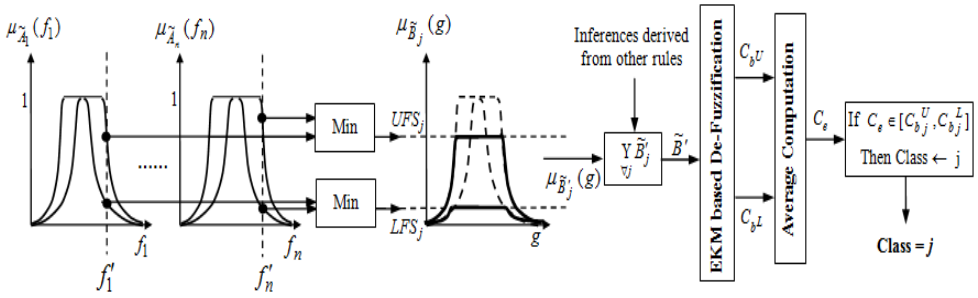


Fig. 2.8. The formulation of IT2FN

Let, $f_1, f_2, ..., f_N$ be N number of features acquired after the GAP operation. Let $\tilde{A}_i$ for i=1 to N represent an IT2FS defined as $[\overline{\mu}_{\tilde{A}_i}, \underline{\mu}_{\tilde{A}_i}]$ for a given linguistic variable f_i as the upper and lower membership function of $\tilde{A}_i$. For the present scenario, $\tilde{A}_i$ has been constructed considering both inter-session (5 sessions in an experimental day) as well as intra-session (5 trials within a session) variations in f_i (representing one average pooled feature). For the afore-said formulation, the initial step involves constructing one Type-1 Gaussian MF by determining the mean (m) and variance (σ^2) among all the f_i values obtained in an experimental day as illustrated in Section A.1. of the Appendix. The constructed Gaussian MF has the center of the base positioned at m and the two extremities situated at $m \pm 3 \times \sigma$ [58]. Since the experiment has been conducted for 10 days, 10 such Type-1 Gaussian MFs are obtained, and these are utilized for the construction of IT2FS, as depicted in Fig. 2.7(a).

An interval type-2 membership function (IT2MF) $\tilde{A}_i$ is developed by computing the union of 10 Type-1 MFs as depicted in Fig. 2.7(b). The UMF and LMF of $\tilde{A}_i$ is constructed by employing the mathematical operations denoted by (2.13-2.14).

$$UMF = \overline{\mu}_{\tilde{A}_i}(f_i) = \max(\mu_{\tilde{A}_1}(f_1), \mu_{\tilde{A}_2}(f_2), ..., \mu_{\tilde{A}_N}(f_N)) \tag{2.13}$$

$$LMF = \underline{\mu}_{\tilde{A}_i}(f_i) = \min(\mu_{\tilde{A}_1}(f_1), \mu_{\tilde{A}_2}(f_2), ..., \mu_{\tilde{A}_N}(f_N)) \tag{2.14}$$

However, the promising contribution of embedded T1-fuzzy sets to the FOU lies in the local neighborhood region of the UMF inside the FOU. To utilize this promising region of the FOU, a mapping function is introduced to level up the LMF towards the UMF, thereby reducing the area under the modified FOU as depicted in Fig. 2.7(c). The mapping function for constructing the refined LMF is denoted by (2.15).

$$LMF_{\text{Re}} = \underline{\mu}_{\tilde{A}_i'}(f_i) = \left(\overline{\mu}_{\tilde{A}_i}(f_i)^{\exp(\underline{\mu}_{\tilde{A}_i}(f_i))} \right)^{\eta} \tag{2.15}$$

where, $\eta \geq 1$ indicates a hyper-parameter that is fine-tuned during the training process. The term $\overline{\mu}_{\tilde{A}_i}(f_i)^{\exp(\underline{\mu}_{\tilde{A}_i}(f_i))}$ in equation (15) reduces the area under the FOU, with η acting as a reduction ratio. In other words, the function in (2.15) aims to decrease the area under FOU while the hyper-parameter η governs the extent of this reduction. Thus, by adjusting η, the classifier can control the effective management of variations within and across sessions to achieve improved performance. The UMF of the modified FOU remains same as (2.13) i.e., $\overline{\mu}_{\tilde{A}_i'}(f_i) = \overline{\mu}_{\tilde{A}_i}(f_i)$. It can be verified from (15) that $\overline{\mu}_{\tilde{A}_i'}(f_i) \leq \overline{\mu}_{\tilde{A}_i}(f_i)$ for all f_i.

However, during the computation of the UMF for IT2FS, a curvilinear top is generated by taking the maximum of the 10 Type-1 MFs, as illustrated in Fig. 2.7(b) and (c). To ensure the convexity of the constructed Type-2 fuzzy sets, a flat-top approximation is applied to the obtained IT2FS. This process involves connecting the peaks of the individual Type-1 MFs with a straight line characterized by a zero slope [4], [43]. The resulting flat-top approximated IT2FS is depicted in Fig. 2.7(d).

2.3.4.2. Classifier Rule

The classifcation rule R_j is presented below.

If f_1 *is* $\tilde{A}_1'$, *and* f_2 *is* $\tilde{A}_2'$ *and...and* f_N *is* $\tilde{A}_N'$, *Then g is* $\tilde{B}_j$ *lying in* $[C_{b_j^U}, C_{b_j^L}]$.Here, $\tilde{B}_j$ denotes the consequent MF of the j^{th} interval type-2 fuzzy class having class centroid within upper and lower class boundary $C_{b_j^U}$ and $C_{b_j^L}$ respectively. The consequent MF $\tilde{B}_j$ is formulated based on the intra-session and inter-sesssion variations of the scores assigned to each written response as discussed in Section 2.2. The subsequent steps in the IT2FS construction for the consequent part follow the same process as outlined for the antecedent part in the preceding sub-subsection 2.3.4.1.

2.3.4.3. Design of IT2FS Classifier

The architecture of the IT2FS based classifier is demonstrated in Fig.2.8. Let the measurements points be $f_1 = f_1', f_2 = f_2', ..., f_N = f_N'$. The Upper Firing Strength (UFS) and Lower Firing Strength (LFS) of rule j is evaluated by (2.16) and (2.17) respectively.

$$UFS_j = \min[\overline{\mu}_{\tilde{A}_1'}(f_1'), \overline{\mu}_{\tilde{A}_2'}(f_2'), ..., \overline{\mu}_{\tilde{A}_n'}(f_N')] \tag{2.16}$$

$$LFS_j = \min[\underline{\mu}_{\tilde{A}_1'}(f_1'), \underline{\mu}_{\tilde{A}_2'}(f_2'), ..., \underline{\mu}_{\tilde{A}_n'}(f_N')] \tag{2.17}$$

Now, the IT2 inference $[\overline{\mu}_{\tilde{B}_j'}(g), \underline{\mu}_{\tilde{B}_j'}(g)]$ is acquired using the computation denoted by (2.18) and (2.19).

$$\overline{\mu}_{\tilde{B}_j'}(g) = \min(UFS_j, \overline{\mu}_{\tilde{B}_j}(g)) \tag{2.18}$$

Fig. 2.9. Subject participating in the MPF based cognitive task

$$\underline{\mu}_{\tilde{B}'_j}(g) = \min(LFS_j, \underline{\mu}_{\tilde{B}_j}(g)) \tag{2.19}$$

However, in the event that multiple rules associated with either identical or different class labels in the consequent are fired, the final inference $\tilde{B}'$ is assessed by computing the union of IT2 inferences using (2.20).

$$\tilde{B}' = [\bigcup_{\forall j} \tilde{B}'_j] = [\bigcup_{\forall j} \overline{\tilde{B}}'_j, \bigcup_{\forall j} \underline{\tilde{B}}'_j] \tag{2.20}$$

where, $\cup$ denotes the union operator.

Finally, the type-reduction and defuzzification of $\tilde{B}'$ is executed using the Enhanced Karnik-Mendel (EKM) [59] algorithm to obtain C_{b^U} and C_{b^L}. Subsequently, the centroid C_e is determined by calculating the mean of C_{b^U} and C_{b^L}.

It is crucial to highlight that due to the presence of overlap among the data points of 2 or more classes, the acquired cluster centroids must lie in disjoint contiguous intervals for accurate classification of the input features $[f'_1, f'_2, ..., f'_N]$ into one of the S distinct classes (here, S = 4). Let, the interval for the j^{th} class centroid be $[C_{b_j^U}, C_{b_j^L}]$. Similarly, let the class centroids for neighboring classes j-1 and j+1 be denoted by $[C_{b_{j-1}^U}, C_{b_{j-1}^L}]$ and $[C_{b_{j+1}^U}, C_{b_{j+1}^L}]$ respectively. To ensure contiguous class boundaries, the following criteria are maintained for S-class classification.

i) $C_{b_{j-1}^L} < C_{b_{j-1}^U}$,

ii) $C_{b_{j-1}^U} = C_{b_j^L} < C_{b_j^U}$,

iii) $C_{b_j^U} = C_{b_{j+1}^L} < C_{b_{j+1}^U}$

The optimal class boundaries are acquired using random search algorithm [60] and their values are denoted in Section 2.4. Once, the contiguous class boundaries are maintained, the class inferred is j if the centroid value C_e acquired from EKM algorithm lies within the range $[C_{b_j^U}, C_{b_j^L}]$ (vide Appendix A.2).

2.3.4.4. Rank Assignment to Individuals with Respect to Different Levels of Creative Potential

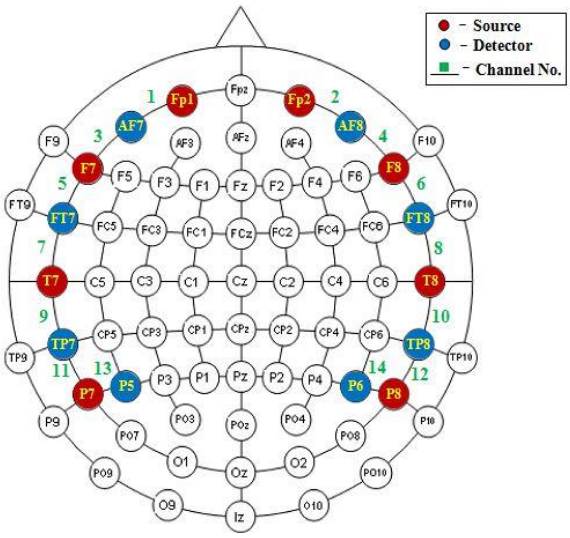


Fig. 2.10. Montage utilized for the MPF task

Participants in the current cognitive task are categorized based on the number of experimental instances classified as High Creative (HCR). In other words, if a subject has *r* experimental instances identified as HCR, their ranking is determined using the following Creativity Potential Measure (CPM).

$$CPM = \frac{r}{R} \times 100\,\% \tag{2.21}$$

where, *R* denotes the total number of experimental instances for a specific participant. Once the CPM scores for all subjects are obtained, they are sorted in descending order. Then, each subject is assigned a rank based on its CPM score, with the highest score receiving the top rank.

2.4. EXPERIMENTS AND RESULTS OF BRAIN CONNECTIVITY BASED COGNITIVE ASSESSMENT OF CREATVITY

2.4.1. fNIRS Data Acquisition

The experiment was conducted at the Artificial Intelligence Laboratory of Jadavpur University in Kolkata, India. To capture the hemodynamic response of subjects during the current cognitive experiment, a whole-brain fNIRS device (NIRScout TM) manufactured by NIRx Medical Technologies, LLC, was employed. This device utilizes 8 infrared (IR) sources and 8 detectors, operating at a sampling rate of 7.81 Hz. The placement of the source-detector pairs, or optodes, on the subjects' scalps adhered to the 10-10 international placement system as shown in Fig. 2.9. The architecture of the montage for optode placement is portrayed in Fig. 2.10 and encompasses the pre-frontal, temporal, and parietal lobes. The rationale for opting for the afore-said montage architecture is derived from the acknowledgment that a

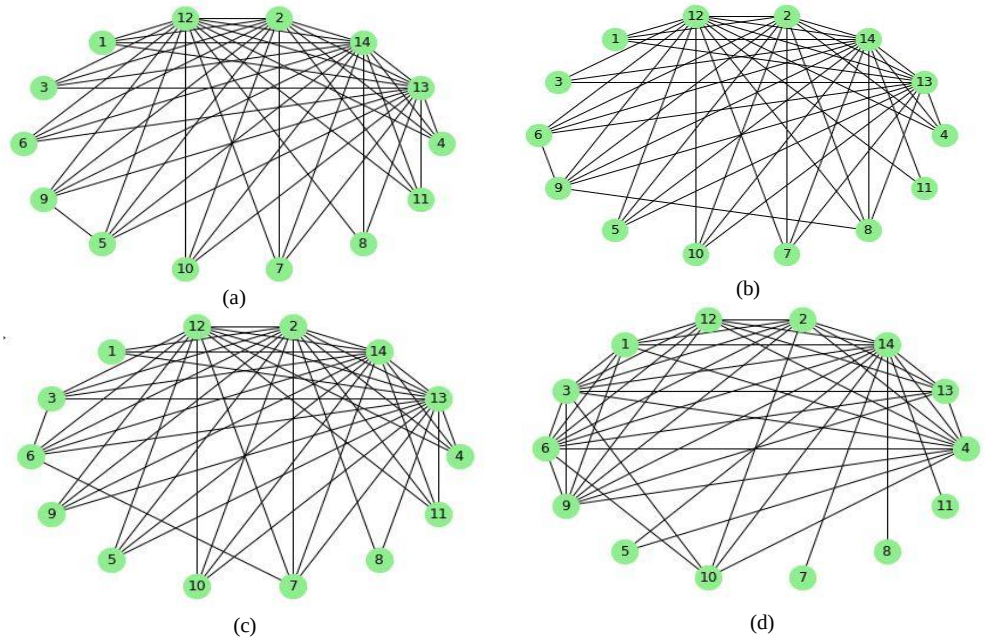


Fig. 2.11. Brain connectivity network of participants performing the MPF task: (a) Brain network of subject ID: 02 who could correctly solve the presented problem, (b) Brain network of subject ID: 23 who could moderately solve the presented problem, (c) Brain network of subject ID: 07 who could hardly solve the presented problem and (d) Brain network of subject ID: 20 who could not solve the presented problem

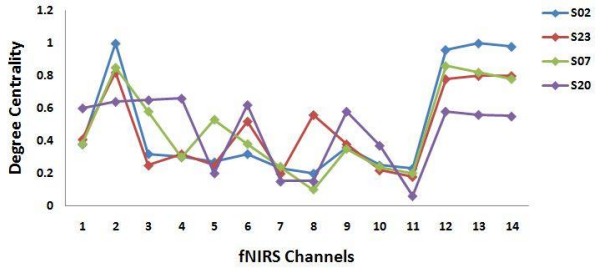
spatial reasoning task engages the working memory and executive functions of the brain [61]-[62]. A total of 64 channels were generated by the optodes, and 14 channels were chosen based on the nearest-neighbor source-detector combination.

2.4.2. Participants

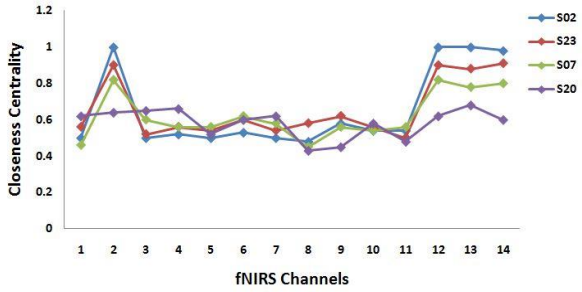
The said cognitive task involved the participation of 32 (engineering student) volunteers, consisting of 15 males and 17 females, all falling within the age range of 18 to 33 years. It is worth noting that all participants possessed normal or corrected-to-normal vision and had no previous history of neuropsychiatric or motor disorders. In accordance with the ethical guidelines and safety protocols related to the experiment, adherence to the Helsinki Declaration of 1970, revised in 2004, was maintained [63].

2.4.3. Cognitive Aspect of Brain Connectivity Features

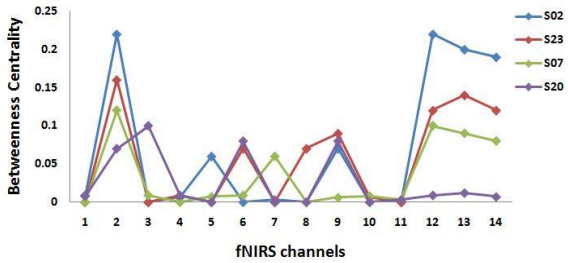
The brain connectivity graphs for the current cognitive task have been obtained by the utilization of PC technique, to capture the interactions among different brain lobes. The connectivity patterns associated with each class label were extracted using the Networkx package [64] of Python 3 and are depicted in Fig. 2.11. To gauge the topological significance of specific lobes, a quantitative assessment has been



(a)



(b)



(c)

Fig. 2.12. Centrality based feature comparison for 4 degrees of creative ability (a) degree centrality analysis (b) closeness centrality analysis (c) betweenness centrality analysis

conducted using centrality measures including degree, closeness, and betweenness. These measures facilitate the recognition of topologically central or hub lobes within the acquired brain graphs. The examination of brain topographic networks using individual centrality features is outlined below.

(a) DC Analysis: The Fig. 2.12 (a) illustrates the DC values of four subjects pertaining to HCR, MCR, LCR, and NCR classes for a single trial. Notably, for subject ID: 02 in the HCR class, the DC values are elevated over the right frontal (channel 2), right temporal (channel 12), and bilateral parietal (channels 13 and 14) regions. Additionally, a discernible trend is observed where the DC values decrease from HCR to LCR class for these specified regions. In contrast, for subjects in the NCR class, exemplified by subject ID: 20, there is no significant activation observed in the mentioned channels of the frontal, parietal, and temporal lobes. Nevertheless, moderately high DC values are evident in other nodes, such as channels 6 and 9 for MCR and channels 3 and 5 for LCR. This analysis suggests that relying solely on the DC measure provides only a partial understanding of the functional importance of

nodes. Consequently, incorporating additional measures is essential to comprehensively quantify the distinct networks.

(b) CC Analysis: In Fig. 2.12 (b), the CC values for the same subjects utilized in the DC analysis are presented. Notably, for a subject in the HCR class, elevated CC values are evident over the right frontal (channel 2), right temporal (channel 12), and bilateral parietal (channels 13 and 14) regions. Furthermore, CC values for subjects in the MCR and LCR classes also show heightened values in these specified channels, with a discernible decrease in values from HCR to LCR classes. Conversely, subjects in the NCR class exhibit no significant activation in the aforementioned channels, as indicated by lower CC values. The CC analysis supports the conclusion that the regions with elevated values can be considered topologically central. This centrality implies that these regions can effectively interact with other nodes within the network through a small number of links [35].

(c) BC Analysis: The results of the BC analysis, as illustrated in Fig. 2.12 (c), underscore the same regions for each class as identified through CC-based evaluation. The manifestation of high BC values in these identified lobes suggests their effective functional integration among various brain regions [35]-[36].

Hence, the centrality-based feature analysis mentioned above supports the proposition that, for spatial reasoning, the actively engaged brain regions include the right anterior prefrontal cortex (BA 10 corresponding to channel 2), right posterior middle temporal gyrus (BA 21 corresponding to channel 12), and bilateral regions of the posterior supramarginal gyrus (BA 40 corresponding to channels 13 and 14). These findings align with existing literature [23], [29], [61]-[62]. Additionally, a noticeable trend emerges, indicating a diminishing activation in the aforementioned Brodmann Areas (BAs) as the level of creative ability decreases from high to low levels.

2.4.4. Optimization of Classifier Parameters

Fine-tuning various parameters of the GC-IT2FN classifier model is essential for ensuring both its robustness and classification accuracy. This process is achieved by employing the random search algorithm, known for efficiently exploring the search space with minimal computational cost [60]. The model's performance has been evaluated by a 10-fold cross-validation across various sets of parameter configurations, where the dataset for each subject is partitioned into 10 disjoint sets or folds. For each combination of parameters, the model is trained on 9 folds and tested on the remaining 1 fold. This process is repeated in 10 iterations, rotating the test fold. Hence, both the training and testing sets for each subject comprise samples

Table 2.1
Comparative study of classifier performance with respect to traditional methods

Classifiers with optimal parameter settings	CA (%)	F1 (%)	Run time complexity (ms)
DCNN [65]	76.32	75.26	192.36
Chebnet [66]	78.02	78.32	156.20
GCN [37]	80.83	81.47	87.46
DGCN [67]	83.28	83.86	120.34
CayleyNet [68]	85.12	85.05	158.87
GAT [69]	88.70	89.27	96.12
AGCN [70]	91.95	92.50	95.46
Proposed GC-IT2FN	**97.54**	**97.72**	**99.06**

from the same individual, although they are distinct in each iteration. Upon attaining the highest accuracy level with all candidate settings, the best configuration is subsequently applied to the test set. The optimal parameter values of GC-IT2FN classifier utilized for the current experiment include: $E' = 2,\ T = 128,\ N = 64,\ \eta = 1.8,$ $C_{b_{NCR}^L} = 9.27,\ C_{b_{NCR}^U} = C_{b_{LCR}^L} = 23.70,\ C_{b_{LCR}^U} = C_{b_{MCR}^L} = 43.02,\ C_{b_{MCR}^U} = C_{b_{HCR}^L} = 72.56,$ and $C_{b_{HCR}^U} = 97.25.$

2.5. PERFORMANCE ANALYSIS OF THE PROPOSED CLASSIFIER MODEL

2.5.1. Relative Performance Analysis of the Proposed GC-IT2FN Classifier with Respect to Traditional Algorithms

The performance of the GC-IT2FN classifier is analyzed with the traditional algorithms using i) classification accuracy (CA) and ii) F1-score whose mathematical equations are illustrated by (2.22-2.23).

$$CA = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{2.22}$$

$$F1 = \frac{2 \times precison \times recall}{precision + recall} \tag{2.23}$$

where, T_P, T_N, F_P, F_N represents the number of true positives, true negatives, false positives and false negatives respectively, $precison = \frac{T_P}{T_P + F_P}$ and $recall = \frac{T_P}{T_P + F_N}$.

Table 2.2
Comparative study of classifier performance with respect to state-of-the-art-methods

Classifiers with optimal parameter settings	CA (%)	F1 (%)	Run time complexity (ms)
GCNN-LSTM [71]	87.78	88.11	302.86
ASGCNN [72]	90.23	89.70	372.45
GAT + BiLSTM [73]	91.34	91.63	421.23
CapsualGNN [74]	92.97	92.74	134.70
CNN-AE + IT2FR-GWO [75]	94.56	94.83	125.60
AEGCN [76]	93.25	93.61	248.27
NCGNN [77]	95.12	95.36	143.18
Proposed GC-IT2FN	**97.54**	**97.72**	**99.06**

Table 2.1 clearly demonstrates the remarkable precision with which the proposed algorithm classifies the desired class labels. Notably, the results achieved by GC-IT2FN far surpass those of the traditional techniques. Moreover, Table 2.1 highlights the run-time complexity of the proposed classifier, which stands at an impressively low 99.06 ms, making it notably faster than the majority of competitive techniques.

2.5.2. Relative Performance Analysis of the Proposed GC-IT2FN Classifier with respect to State-of-the-Art (SOTA) Algorithms

The performance evaluation of the GC-IT2FN classifier has been conducted by comparing it with recent SOTA techniques, which include various hybrid models utilized for classifying brain connectivity features. The outcomes of this comparative analysis are presented in Table 2.2. The results clearly demonstrate that the proposed classifier surpasses the performance of the SOTA techniques by a substantial margin. Additionally, the run time complexity of the proposed technique is notably lower in comparison to all the hybrid methods.

2.5.3. Ablation Study of the Proposed GC-IT2FN Classifier

A comprehensive ablation study has been conducted on the proposed GC-IT2FN classifier to assess the influence of its individual modules on overall performance. In this analytical process, the architecture of the proposed model underwent systematic modifications through the removal of specific modules, and the corresponding

Table 2.3
Ablation study of the proposed classifier

Variation in Classifier Modules	CA (%)	F1 (%)
GCN	89.47	90.03
IT2FS	88.78	88.55
Attention + GCN	92.50	92.83
GCN + IT2FS	94.36	94.64
Proposed GC-IT2FN	**97.54**	**97.72**

Table 2.4
Comparative study of Different formulations of brain Connectivity networks with respect to CA (%)

Formulations of brain connectivity network	Class Labels			
	HCR	MCR	LCR	NCR
MSC	89.52	90.04	89.73	90.11
PLV	91.36	92.46	91.60	92.29
PLI	92.08	91.22	91.54	92.16
KNN	94.20	93.75	93.85	94.32
MI	95.67	95.43	95.91	95.76
PC	**97.87**	**97.32**	**96.92**	**98.03**

results are detailed in Table 2.3. It is apparent from this table that both the CA and F1-score values decline when either the GCN or IT2FS classifier is utilized in isolation for the classification task. Conversely, the inclusion of the attention mechanism to the GCN model leads to a significant improvement in both CA and F1-score. Furthermore, employing both GCN and IT2FS results in noticeable enhancement in CA and F1-score. Notably, the highest CA and F1-score values are achieved when all three modules of GC-IT2FN (i.e., attention, GCN, and IT2FN) are utilized together. Thus, the findings from this table underscore that excluding any single unit results in a noteworthy decrease in the overall performance of the proposed model.

2.5.4. Effect of Different Formulations of Brain Connectivity Network on Classifier Performance

Table 2.5
Effect of different activation functions on the proposed model with respect to CA (%)

Activation Functions	Class Labels			
	HCR	**MCR**	**LCR**	**NCR**
Sigmoid [83]	79.60	79.48	80.03	79.76
Tanh [84]	81.05	80.87	80.96	81.12
ReLU [85]	82.90	83.36	82.82	83.20
Leaky ReLU [86]	84.55	85.14	85.06	84.61
ELU [87]	86.34	86.17	85.78	85.97
SELU [88]	87.16	87.38	87.42	87.52
Swish [89]	89.11	88.97	89.20	88.83
ELiSH [90]	92.07	91.88	92.15	91.72
Mish [91]	93.63	93.47	94.03	93.68
Logish	**97.87**	**97.32**	**96.92**	**98.03**

Table 2.6
Effect of reducing the area under the FOU on classifier performance

Class Labels	Varying Classifier Architecture			
	GCN + IT2FS with original FOU		**GCN + IT2FS with modified FOU**	
	CA (%)	**F1 (%)**	**CA (%)**	**F1 (%)**
HCR	92.83	93.25	**97.87**	**98.12**
MCR	93.14	92.90	**97.32**	**97.08**
LCR	92.56	92.77	**96.92**	**97.25**
NCR	93.12	93.23	**98.03**	**98.43**

The performance of the proposed classifier is assessed in comparison to brain connectivity network construction based on the PC method. This evaluation includes several standard methods for constructing brain connectivity networks, such as Magnitude Squared Coherence (MSC) [78], Phase Locking Value (PLV) [79], Phase Lag Index (PLI) [80], K-Nearest Neighbor (KNN) [81], and Mutual Information (MI) [82]. The outcomes of this comparative analysis are detailed in Table 2.4, revealing a noteworthy improvement in the proposed classifier's performance when

Table 2.7
Comparative study of different Connectivity Features with respect to F1-score (%)

Centrality Features	Class Labels			
	HCR	MCR	LCR	NCR
DC	93.62	92.87	93.30	93.08
CC	94.03	93.40	94.21	94.15
BC	95.27	95.74	95.45	95.92
DC+CC	94.88	94.51	94.72	94.60
DC+BC	96.20	96.02	96.11	96.33
CC+BC	96.75	96.81	96.25	97.42
DC+BC+CC	**98.12**	**97.08**	**97.25**	**98.43**

employing PC-based brain connectivity network construction in contrast to the competitor techniques.

2.5.5. Influence of Different Activation Functions on Classifier Performance

Table 2.5 presents the outcomes of CA and F1-score variations in GC-IT2FN concerning different activation functions. The table clearly indicates that the proposed classifier achieves optimal results when the Logish activation function is employed.

2.5.6. Effect of Reducing the Area under the FOU on Classifier Performance

The effect of reducing the FOU using the modified LMF (as discussed in Section 2.2) is compared against the traditional FOU formulation produced by considering the union of 10 Type-1 MFs obtained from 10 experimental days. The results of this comparison are tabulated in Table 2.6. It is evident from the table that the proposed approach of FOU reduction significantly improves both the CA and F1-score compared to the classical approach.

2.5.7. Influence of Different Brain Connectivity Features on Classifier Performance

The performance of the proposed classifier is evaluated with respect to the variation in three centrality based features: DC, CC and BC. It is observed from Table 2.7 that though CC+BC, DC+BC and BC are able to provide a fairly high F1-score value,

Table 2.8
Rank of 10 subjects based on classification results

Subject ID	CPM (%)	Rank
S27	**83.50**	**1**
S02	78.45	2
S11	76.52	3
S05	69.27	4
S08	69.11	5
S30	65.50	6
S23	59.47	7
S18	59.38	8
S09	55.29	9
S03	51.05	10

Table 2.9
Statistical Validation of the proposed GC-IT2FN Using Friedman's Test

Classifiers Algorithms	R_a	χ_F^2
DCNN	8	**27.20 (Null Hypothesis is Rejected)**
Chebnet	6.75	
GCN	6	
DGCN	3.87	
CayleyNet	5.25	
GAT	3	
AGCN	2.12	
Proposed GC-IT2FN	**1**	

DC+CC+BC yields the highest F1-score in comparison to the other feature combinations.

2.5.8. Ranking of Subjects on the Basis of Scientific Creative Potential

Each subject's rank is determined by initially calculating their CPM score, denoted in (21), and arranging them in descending order. Following this arrangement, ranks are assigned to each subject based on their CPM score, with the highest CPM score

receiving the top rank. In Table 2.8, the ranks of the 10 subjects are presented instead of the entire cohort due to space constraints.

2.5.9. Statistical Validation of the Proposed GC-IT2FN Using Friedman's Test

To statistically validate the proposed algorithm, the Friedman's 2-way non-parametric statistical test [92] has been utilized for the present application. This test assigns the lowest rank to the best-performing algorithm. The Friedman statistic, denoted by χ_F^2 is computed for (*k-1)* degrees of freedom using the following equation.

$$\chi_F^2 = \frac{12N}{k(k+1)}\left[\sum_{a=1}^{k} R_a^2 - \frac{k(k+1)^2}{4}\right] \tag{2.24}$$

where, N represents the number of datasets (i.e., the number of creative classes), k indicates the number of classifier algorithms and R_a portrays the mean rank of the p^{th} algorithm pertaining to the q^{th} dataset. The results of the Friedman's test with respect to all classification metrics for $N = 4$ and $k = 8$ has been portrayed in Table 2.9. It is apparent from this table that the Friedman statistic $\chi_F^2 = 27.20 > \chi_{7,0.95}^2 = 14.07,$ which depicts the chi-square value at 95% confidence level with 7 degrees of freedom. The outcomes of the Friedman test reveal the rejection of the null hypothesis, which asserts that all the algorithms exhibit equivalent performance. Consequently, the comparison of the algorithms must be performed by taking into account their respective ranks.

2.6 CONCLUSION

Unraveling the creative capacities of individuals in the scientific domain poses a significant challenge within the realm of cognitive neuroscience. The present study makes a noteworthy contribution to this domain by utilizing fNIRS data, with a focus on evaluating the levels of creative aptitude through brain connectivity analysis. Such an analysis, centered on centrality features (i.e., degree, closeness, and betweenness), reveals the active engagement of the right anterior prefrontal cortex (BA 10), right posterior middle temporal gyrus (BA 21), and bilateral regions of the posterior supramarginal gyrus (BA 40) during the performance of a paper folding task. Subsequently, the centrality based features are classified using a novel GC-IT2FN classifier, demonstrating precise identification of desired class labels related to creative ability and superior performance compared to its competitors.

A captivating application of the proposed scheme involves identifying creative individuals for potential placement in various research-oriented departments in industrial sectors based on their different ranks of creative ability. For example, an

individual exhibiting high creativity could be appointed to roles such as Advanced Data Analyst or Innovation Manager, where the demand for groundbreaking ideas and inventive problem-solving is paramount. A person with medium creative abilities might find a fitting role as a Research Assistant, contributing to the execution of research projects with a balance of structured methodologies and adaptable problem-solving. Meanwhile, someone with lower creative inclinations could be well-suited for a position like Quality Control Analyst, where attention to established protocols and meticulous adherence to quality standards is essential. Furthermore, the current study also nurtures the possibility of drawing biological inspiration for the creation of computational models of scientific creativity. The future scope of the proposed work may involve assessing the creative potential of individuals by exploring other cognitive phenomena that influence creative ideation in the scientific domain, such as inductive learning, analogical reasoning, and the like.

2.7 APPENDIX

A.1. Construction of the Type-1 MFs from Global Average Pooled Vectors

For each MPF problem, the current experiment has been conducted over a period of 10 days. Each day consists of 5 sessions, with each session comprising 5 trials. A single trial yields 64 global average pooled features. Table A.I provides the distribution of features across different trials and sessions for each day of the experiment.

The Type-1 Gaussian membership function (MF) is formulated by initially calculating the mean of the features from 5 trials/session, as illustrated in Fig. A.1. (a). This process results in 5 sets of mean vectors, each of size 64×1. Subsequently, the mean and variance across all the feature values of these 5 sets of mean vectors are computed to create a Type-1 Gaussian MF, depicted in Fig. A.1. (b). Since the experiment spans 10 days, 10 such Type-1 Gaussian MFs are generated, which are then utilized in designing the Interval Type-2 Fuzzy Set (IT2FS).

TABLE A.I
Trials and sessions conducted in an experimental day for a single MPF

Experimental Day	Session (SS)	Trial	Features
Day 1	SS1	1	$f_1, f_2, \ldots\ldots f_{64}$
		2	$f_1, f_2, \ldots\ldots f_{64}$
		3	$f_1, f_2, \ldots\ldots f_{64}$
		4	$f_1, f_2, \ldots\ldots f_{64}$
		5	$f_1, f_2, \ldots\ldots f_{64}$
	SS2	1	$f_1, f_2, \ldots\ldots f_{64}$
		2	$f_1, f_2, \ldots\ldots f_{64}$
		3	$f_1, f_2, \ldots\ldots f_{64}$
		4	$f_1, f_2, \ldots\ldots f_{64}$
		5	$f_1, f_2, \ldots\ldots f_{64}$
	……	……	……
	SS5	1	$f_1, f_2, \ldots\ldots f_{64}$
		2	$f_1, f_2, \ldots\ldots f_{64}$
		3	$f_1, f_2, \ldots\ldots f_{64}$
		4	$f_1, f_2, \ldots\ldots f_{64}$
		5	$f_1, f_2, \ldots\ldots f_{64}$
Day 2	SS1	1	$f_1, f_2, \ldots\ldots f_{64}$
		2	$f_1, f_2, \ldots\ldots f_{64}$
		3	$f_1, f_2, \ldots\ldots f_{64}$
		4	$f_1, f_2, \ldots\ldots f_{64}$
		5	$f_1, f_2, \ldots\ldots f_{64}$
	SS2	1	$f_1, f_2, \ldots\ldots f_{64}$
		2	$f_1, f_2, \ldots\ldots f_{64}$
		3	$f_1, f_2, \ldots\ldots f_{64}$
		4	$f_1, f_2, \ldots\ldots f_{64}$
		5	$f_1, f_2, \ldots\ldots f_{64}$
	……	……	……
	SS5	1	$f_1, f_2, \ldots\ldots f_{64}$
		2	$f_1, f_2, \ldots\ldots f_{64}$
		3	$f_1, f_2, \ldots\ldots f_{64}$
		4	$f_1, f_2, \ldots\ldots f_{64}$
		5	$f_1, f_2, \ldots\ldots f_{64}$
…….	……..	……	……
Day 10	SS1	1	$f_1, f_2, \ldots\ldots f_{64}$
		2	$f_1, f_2, \ldots\ldots f_{64}$
		3	$f_1, f_2, \ldots\ldots f_{64}$
		4	$f_1, f_2, \ldots\ldots f_{64}$
		5	$f_1, f_2, \ldots\ldots f_{64}$
	SS2	1	$f_1, f_2, \ldots\ldots f_{64}$
		2	$f_1, f_2, \ldots\ldots f_{64}$
		3	$f_1, f_2, \ldots\ldots f_{64}$
		4	$f_1, f_2, \ldots\ldots f_{64}$
		5	$f_1, f_2, \ldots\ldots f_{64}$
	……	……	……
	SS5	1	$f_1, f_2, \ldots\ldots f_{64}$
		2	$f_1, f_2, \ldots\ldots f_{64}$
		3	$f_1, f_2, \ldots\ldots f_{64}$
		4	$f_1, f_2, \ldots\ldots f_{64}$
		5	$f_1, f_2, \ldots\ldots f_{64}$

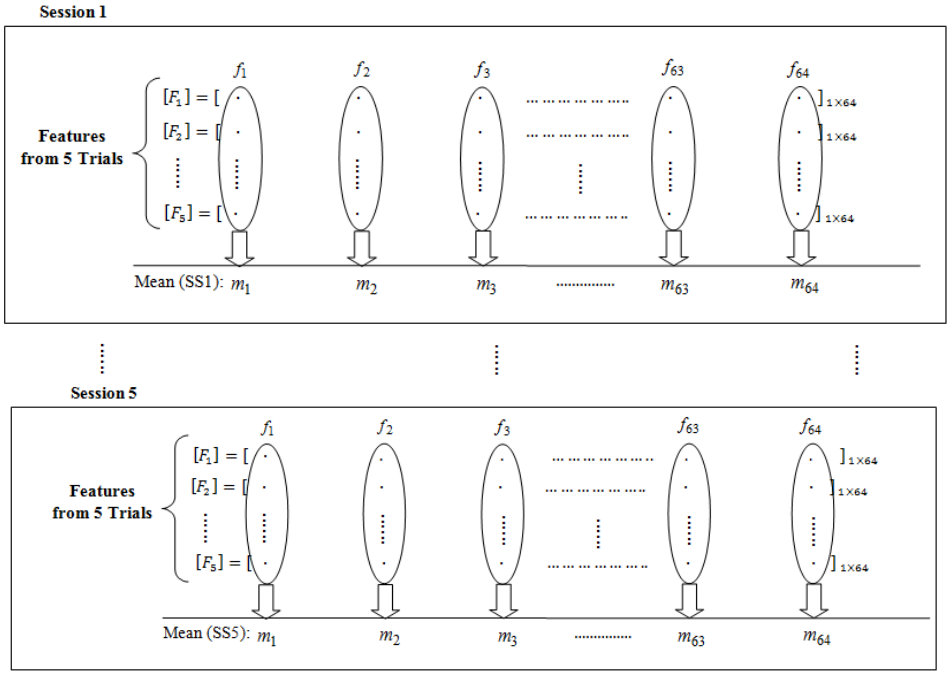


Fig. A.1. (a) Computation of mean among features across

A.2. Evaluation of Defuzzified Output from Multiple Fired Rules

For the present work, there are 4 classes of creative potential which include NCR, LCR, MCR and HCR. The consequent IT2MF for 4 creative potential labels with their initial class centroid positions are shown in Fig. A.2(a). Computation of centroid from multiple fired rules is illustrated in Fig. A.2 (b). It is evident from this figure that the class centroid here lies in $[C_{b^U_{HCR}}, C_{b^L_{HCR}}]$, indicating class =HCR.

A.3. Evaluation of Class Labels based on Expert Score

The section includes a selection of solutions by participants for Problem 1, along with their corresponding scores. This inclusion aims to facilitate an understanding of the distribution of scores among different solutions of varying quality levels. The afore-said samples are depicted in Fig. A.3.

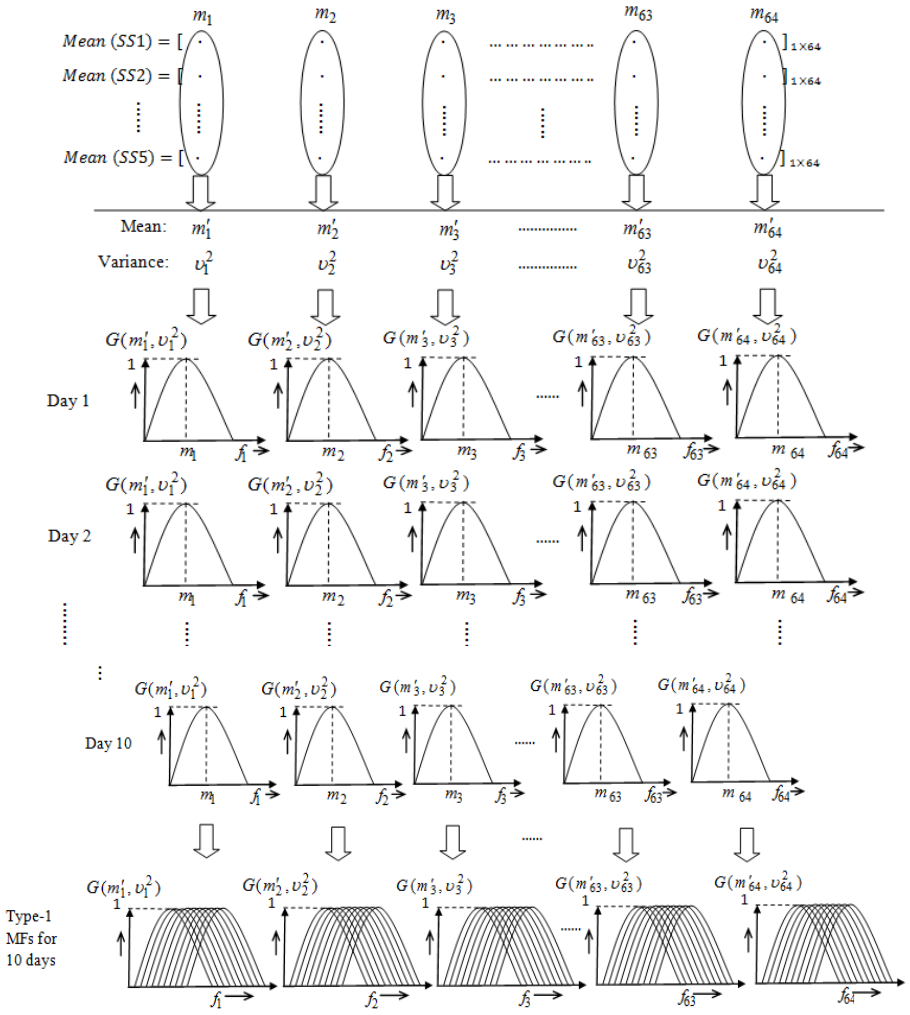


Fig. A.1. (b) Formulation of 10 Type-1 MFs

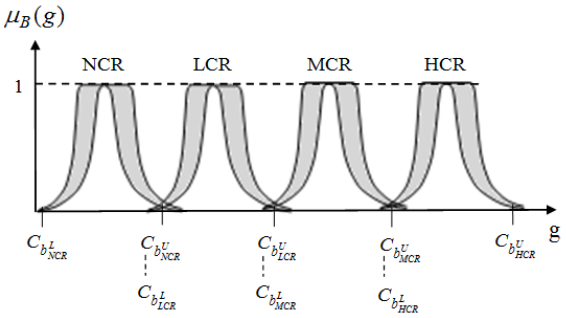


Fig. A.2. (a) Consequent IT2 MF for 4 classes (NCR, LCR, MCR, HCR) with initial class centroid positions

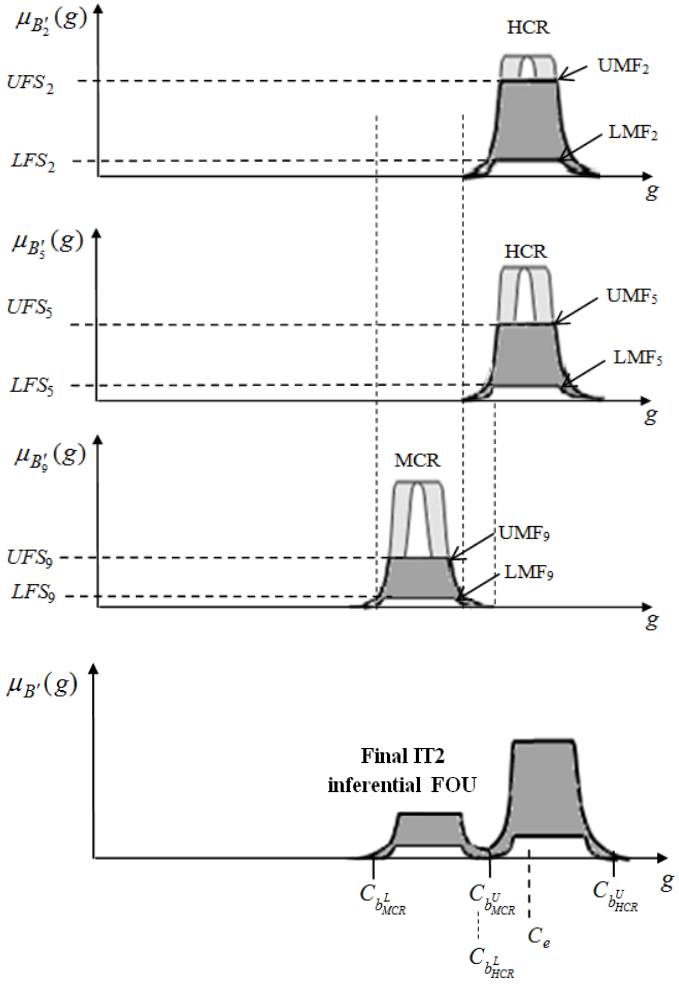
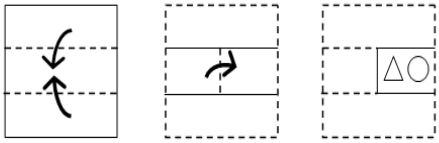


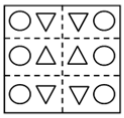
Fig. A.2. (b) Defuzzified output C_e falls in class HCR after firing of rules 2, 5 and 9

Problem 1: A set of 3 figures X, Y and Z are given below showing a sequence in which a paper is folded and finally cut from a particular section. Find the unfolded form of Fig. Z.



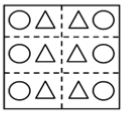
Class Label HCR with Score Range (75,100]

Performance of Subject ID: 02



Scores				
Stage-1	**Stage-2**	**Stage-3**	**Stage-4**	**Total**
5+8+12=25	5+8+12=25	5+8+12=25	5+8+12=25	100

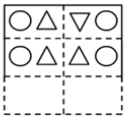
Performance of Subject ID: 11



Scores				
Stage-1	**Stage-2**	**Stage-3**	**Stage-4**	**Total**
5+8+12=25	5+8+12=25	5+8+0=13	5+8+0=13	76

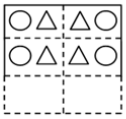
Class Label: MCR with Score Range (50,75]

Performance of Subject ID: 18



Scores				
Stage-1	**Stage-2**	**Stage-3**	**Stage-4**	**Total**
5+8+12=25	5+8+12=25	5+8+6=19	0	69

Performance of Subject ID: 23



Scores				
Stage-1	**Stage-2**	**Stage-3**	**Stage-4**	**Total**
5+8+12=25	5+8+12=25	5+8+0=13	0	63

Class Label: LCR with Score Range (25,50]

Performance of Subject ID: 07

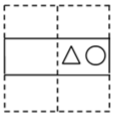
Scores				
Stage-1	**Stage-2**	**Stage-3**	**Stage-4**	**Total**
5+8+12=25	5+8+0=13	0	0	38

Performance of Subject ID: 32

Scores				
Stage-1	**Stage-2**	**Stage-3**	**Stage-4**	**Total**
5+8+12=25	5+8+12=25	0	0	50

Class Label: NCR with Score Range (0,25]

Performance of Subject ID: 20



Scores				
Stage-1	**Stage-2**	**Stage-3**	**Stage-4**	**Total**
5+8+12=25	0	0	0	25

Performance of Subject ID: 14

Scores				
Stage-1	**Stage-2**	**Stage-3**	**Stage-4**	**Total**
5+0+12=17	0	0	0	17

Fig. A.3. Determination of Class Labels with respect to expert score for problem 1

3

EEG-Based Analysis of the Eureka Effect for Discriminating Creative and Analytical Cognition Using Dual Attention Induced Graph Convolutional-Capsule Network

With the advent of technology, there has been an increase in demand of creative individuals in various departments of Research and Development (R&D) sectors. On the contrary, analytical minded people are more demanded in academia, management and investigation departments. Unfortunately, the categorization of creative and analytical individuals from their behavioral responses is a challenging task. This chapter makes a humble attempt to perform the aforesaid categorization task by classifying subjects into 4 class labels: Analytical, High Creative, Medium Creative and Low Creative from the brain responses during their participation in a convergent thinking based task. The above classification problem is undertaken in two important phases. First, a brain connectivity network is constructed from the electroencephalographic response acquired from the scalp of subjects during convergent thinking based cognitive activity using Pearson's correlation technique. Second, three network-centrality features (degree, closeness

and betweenness) are abstracted from the connectivity networks and finally classified using a novel classifier referred to as Dual Attention Induced Graph Convolutional-Capsule network (DAIGC-CapsNet). The proposed classifier leverages one Graph Convolution Network (GCN) that learns the brain connectivity features and one Capsule Network (CapsNet) that performs the classification of the desired labels. The main contribution of the DAIGC-CapsNet lies in the design of: i) dual attention mechanism that aids in steering the graph convolution layers towards the most prominent centrality features as well as assures the transference of the significant predictions to the higher capsule layers, and ii) Sparsemax based dynamic routing mechanism that boosts the coupling strength between the lower and higher capsule layers. The proposed DAIGC-CapsNet has shown favourable results in classifying the required class labels in comparison to its conventional counterparts. Additionally, statistical validation undertaken utilizing the Friedman's test also affirms the superior performance of the DAIGC-CapsNet.

3.1 INTRODUCTION

The demand for creative individuals is increasing, particularly in scientific research, the innovation departments of software industries, and other Research and Development (R&D) sectors [1-2]. These individuals are valued for their ability to showcase innovative ideas across diverse scientific disciplines, thereby propelling technological progress and expanding the boundaries of human knowledge. Conversely, investigation departments [3-4], academia [5-6], and management sectors [7] seek analytical thinkers who can interpret complex data, solve problems methodically, and make data-driven decisions. Recent research [8-11] on differentiating creative individuals from the analytical ones focuses on the behavioral response of subjects undertaking a convergent problem-solving test. A convergent problem-solving test involves an evaluative framework in which a participant finds a single correct solution to a specific problem using logic and reasoning from a given set of information or constraints, rather than exploring multiple potential solutions [9]. However, such tests consist of problems whose solution can be explored using either an analytical approach or a creative insight based strategy. The former technique demands searching the problem space in a conscious and deliberate manner [12]. The latter cognitive style involves the unanticipated emanation of solution in one's mind accompanied by a sense of the "Eureka" or "Aha!" moment, coupled with a corresponding neural reward mechanism [13]. This cognitive

phenomenon sparked by Eureka moment, is regarded as the pivotal element that governs an individual's creative thought process [10-14]. Detecting creative individuals among other examines undertaking such tests is a challenging task as a correct response to the presented problem cannot identify the underlying neural mechanism. The present study seeks to address the afore-mentioned issue by utilizing electroencephalogram (EEG) based analysis.

The current research aims at differentiating creative individuals along with their degrees of creative potential (high, medium and low) from the analytical ones during a convergent problem solving task. To attain the afore-said objective, the present study has been developed within the framework of pattern recognition, encompassing three primary phases namely, pre-processing, feature extraction and classification. The first phase involves the elimination of undesirable noise and physiological artifacts [15-17] from the raw EEG signals. The second phase encompasses the abstraction of features from the pre-processed EEG signals to understand the most discriminating characteristics of the signal. This step holds immense significance, as the extracted features will serve as a guiding foundation for the classifier algorithm, aiding in the categorization of the intended class labels. The prime motivation of the second phase is fulfilled in two steps. The first step involves the construction of a brain connectivity network [18] from the pre-processed signals using Pearson's correlation coefficient (PCC) [19] technique. This approach aims to delve into the functional interplay existing among various brain regions, thereby unraveling their intricate relationships [20]. The second step revolves around the extraction of three connectivity-based centrality features (degree, closeness and betweenness) to identify the Brodmann Areas (BAs) of the brain acting as hubs that control the overall functionality of the network [20-21]. The final phase of the current study involves classifying the abstracted brain connectivity features to distinguish creative individuals, along with their varying levels of creative potential from those with analytical traits.

The main contribution of this research lies in the development of an appropriate classifier model capable of addressing the aforementioned classification challenge. As the research primarily revolves around brain connectivity features, the classifier has been specifically crafted using the principles of a Graph Neural Network (GNN) [22]. The remarkable accomplishments of Graph Neural Networks (GNNs) in learning node representations have ignited a surge of interest in deep learning methodologies aimed at leveraging node embeddings derived from GNNs [23-25]. However, within this process, the learned representation of each node is often treated as a collection of individual scalar features, rather than as a unified vector. For example, Zhang et al. [26] employed element-wise max-pooling on node

embeddings when crafting graph embeddings, while Verma et al. [27] generated graph embeddings by computing the element-wise covariance of all nodes. These methods signify that the authors captured the node features in scalar form when crafting graph embeddings, a practice that might not adequately retain the node-based feature properties of a graph. In order to tackle the afore-mentioned issues, the current study utilizes a Graph Convolution Network (GCN) [28-29] for the formulation of graph embeddings where the learned representation of each node is treated as a vector via spectral graph convolution [28]. Moreover, GCNs have recently surged in popularity for accurately classifying brain connectivity features. For instance, Wang et al. [82] developed an innovative model called AMCNN-DGCN (Attention-based Multi-scale Convolutional Neural Network-Dynamical Graph Convolutional Network) to enhance driving fatigue detection. This model integrates attention-based multi-scale temporal convolutions for abstracting significant patterns from raw EEG signals. Subsequently, it employs DGCNs to learn spatial filters, with the adjacency matrix being dynamically determined based on the data, enabling efficient capture of essential channel relationships. This approach, in contrast with conventional methods, offered a more adaptive and effective solution for exploiting EEG data in fatigue detection. A recent study conducted by Gao et al. [83] introduced a novel EEG-GCN model that integrates spatio-temporal attention mechanisms along with self-adaptive graph convolutional networks for EEG-based emotion recognition tasks. Their performance evaluation demonstrated the significant superiority of the proposed approach over traditional methods. Jia et al. [84] introduced a novel GCN architecture aimed at enhancing seizure prediction from EEG data. This architecture integrated graph convolution layers for abstracting node features, global mean pooling layers to consolidate the abstracted node features, and dense layers for classification. Overall, their approach demonstrated improved prediction accuracy and reduced network size in comparison to traditional approaches. Qin et al. [85] utilized a graph convolution network to discern early major depressive disorder (MDD) from fMRI data. Their model incorporated three graph convolution layers, followed by a global average pooling layer and a linear layer with Softmax activation function. The experimental findings underscored the efficacy of their classifier, outperforming baseline models. Thus, the promising results of GCN in classifying brain connectivity features, owing to its intrinsic ability to treat the learned representation of each node as a vector, have motivated the present authors to utilize it for the current classification problem.

In order to create high-quality graph embeddings, it is imperative to not only discern distinct structures surrounding each node but also to preserve intricate attributes like position, direction, and connections. GCNs utilized for classification

problems often struggle to meet the above requirement due to the involvement of the pooling operation after the convolution layer that leads to the loss of fine-grained information regarding the feature attributes. Therefore, to preserve distinct feature attributes, the present work employs a Capsule Network, often denoted as CapsNet [30], in conjunction with the GCN. The rationale behind the employment of CapsNet stems from its nature of being a vector-based neural network [30-31], that provides a more efficient method for retaining both node and graph property information. Due to the afore-said attractive feature, numerous CapsNet variants have widely been utilized to improve brain signal classification. For example, Wang et al. [86] introduced a spatio-temporal attention-induced capsule network tailored for the P300-based brain-computer interface (BCI) speller system relying on P300 event-related potential. This network precisely discerns a subject's mental response in the BCI speller by leveraging the distinctive neural signatures associated with P300 responses elicited during user selections from the interface, outperforming traditional methods. Chen et al. [87] devised a Self-Attentive Channel-Connectivity Capsule Network (SACC-CapsNet) for EEG-based driving fatigue detection, demonstrating a significant enhancement in fatigue detection compared to conventional approaches. Li et al. [88] developed a 1D-CapsuleNet leveraging large kernel dilation convolution and capsule network to enhance fNIRS-based mental arithmetic and motor imagery classification. Yang et al. [89] proposed a self-attention-based CapsNet model for estimating driver vigilance from single-channel EEG data, showcasing superior classification performance over baseline models. Du et al. [90] employed a multi-layer 3-dimensional convolution technique alongside a capsule network for enhanced motor-imagery classification from EEG data. Therefore, the fusion of GCN and CapsNet to create high-quality graph embeddings offers a promising solution to address the challenges associated with attribute retention.

In this chapter, a novel classifier model referred to as Dual Attention Induced Graph Convolutional-Capsule Network (DAIGC-CapsNet) has been designed to categorize individuals with varying problem-solving skills as high creative, medium creative, low creative, or analytical by leveraging the synergies of GCN and CapsNet. The first layer of the DAIGC-CapsNet consists of a novel Mish [32] Induced Attention Module (MI-AM) that helps to steer the graph convolution layers, directing their focus towards the most pertinent nodes, thereby improving the decision-making process of the classifier. After this, graph convolution operation is employed upon the attention induced node features to form graph embeddings. The graph embeddings are transformed into primary capsules that capture higher-level representation of features. Next, a novel attention mechanism termed as Fused Attention Module (F-AM) is introduced that highlights the most important aspects of

the primary capsule which helps to transfer the most relevant information to the class capsules from the primary capsule layers. Such an attention mechanism consists of two parallel attention layers that evaluate the attention maps along the channel and spatial dimensions of the primary capsule. This configuration enables each layer to discern "what" and "where" to prioritize within the channel and spatial axes of the primary capsules. Then, the channel and spatial attention maps are combined with the primary capsule to create a high quality refinement of the features. Finally, the fused attention induced primary capsules sends its predictions to the class capsule for identifying a specific class label using a novel Sparsemax [33] based dynamic routing algorithm. This routing technique is computationally efficient and also helps to increase the coupling strength between the primary and class capsules.

Experimental analysis undertaken infers that the proposed classifier model is able to categorize the desired class labels accurately and also outperforms the state-of-the-art techniques. Besides, statistical validation performed also confirms the superlative results portrayed by the proposed model. Additionally, a scoring scheme has also been developed to rank each individual categorized as creative depending on their levels of creative ability.

The main contributions of the present work are summarized as follows.

1. Utilization of brain connectivity analysis, incorporating three centrality features (degree, closeness, and betweenness), to identify the prominent BAs, acting as hubs, associated with each class label, namely high creative, medium creative, low creative, and analytical.
2. Development of a novel classifier model based on GCN and CapsNet referred to as Dual Attention Induced Graph Convolutional-Capsule Network (DAIGC-CapsNet) for distinguishing three classes of creative ability (high, medium and low) from the analytical ones. The novelty of the DAIGC-CapsNet model is listed below.
    a) Introduction of a new Mish Induced Attention Module (MI-AM) that aids in guiding the graph convolution layers to concentrate on the most important nodes and ignore the irrelevant ones. This strategic focus plays a pivotal role in elevating the classifier's performance.
    b) Employment of a novel Fused Attention Module (F-AM) on the primary capsules to ensure that the most significant predictions are transferred to the higher or class capsule layers. This module first highlights the significant features across channel and spatial dimensions of the primary capsules and subsequently combines the resultant channel and spatial attention-driven features with the original primary capsules. Such a fusion results in a more refined

version of the primary capsules that enhances the information propagation to the higher layers of the network by acquiring the knowledge of when to accentuate or disregard a specific information.

c) Introduction of a Sparsemax based routing algorithm which is computationally efficient and helps to enhance the coupling strength among the primary and higher/class capsule layers.

3. Demonstration of the efficacy of the proposed DAIGC-CapsNet classifier on three data sets through various ablation studies and comparative analysis with the state-of-the-art algorithms.
4. Design of a scoring scheme to rank the different categories of creative individuals on the basis of their creative ability levels.

The remainder of the manuscript is organized as follows. In Section 3.2, the existing research pertaining to convergent problem-solving tasks is presented. Section 3.3 outlines the principles and methodologies adopted for conducting the classification task. The design of the proposed classifier is elaborated in Section 3.4, while Section 3.5 portrays the scoring scheme designed to rank individuals with creative abilities. Section 3.6 encompasses the conducted experiments for the current cognitive task, along with the results of the analytical study to demonstrate the effectiveness of the proposed model. The statistical validation of the proposed classifier can be found in Section 3.7. The advantages and limitations intrinsic to the proposed strategy are discussed in Section 3.8, whereas Section 3.9 provides a compilation of the inferences drawn from the present study.

3.2 RELATED WORKS

Several research studies have been conducted to analyze the brain activations during insight based convergent problem solving. Dandan et al. [34] captured functional magnetic resonance imaging (fMRI) data from subjects who were involved in solving different technical problems from science and engineering. Such an analysis inferred the active participation of left dorsolateral prefrontal gyrus (BA 9) and the left angular gyrus (BA 39) during insight based problem solving strategy. Another fMRI based study [35] confirmed the active participation of middle temporal gyrus and the middle occipital gyrus while participants solved problems from bioscience using an insight based approach. Aziz-zadeh et al. [36] reported the involvement of bilateral insula, Broca's area, right prefrontal cortex and anterior cingulate cortex when subjects solved anagrams utilizing an insight induced strategy. Jung-Beeman et al. [37] performed two independent experiments using EEG and fMRI to detect the neural mechanism involved in insight problem solving method. The EEG analysis in [37] revealed a sudden burst of high-frequency (gamma-band)

neural activity from the right anterior temporal area 0.3s prior to insight based solutions. Moreover, high alpha power was also observed from the right parieto-occipital region when a subject is trying to solve the problem using an insight induced strategy. Additionally, for the insight approach, the fMRI analysis in [37] reported increased activity in right anterior superior temporal gyrus and right parieto-occipital regions. Ghani et al. [38] inferred the involvement of high alpha power from right temporal lobe when participants employed an insight based strategy to solve word association problems. Yu et al. [39] reported that, for insight based problem solvers, high alpha band power was observed during both the preparation and solving phases of a problem. Minami et al. [40] revealed that beta-band desynchronization in the parietal–posterior regions reflects a problem being solved by insight based method. Lu et al. [41] inferred that the Eureka effect was associated with enhanced coherence of oscillations in the alpha and theta frequency ranges across widely dispersed cortical regions, with a predominant emphasis on the right hemisphere. Additionally, the authors also observed that the afore-said increase in coherence was associated with decrease in beta band power in parietal and central lobes and also a decrease of alpha power in frontal and occipital regions. Furthermore, a decline in fractal dimensionality was also observed predominantly in the right hemisphere. Bartolome et al. [42] reported that significant changes in alpha and theta frequencies from the right temporal lobe, and alpha, theta, and gamma frequencies from the right parietal lobe, were associated with the insight effect. Bieth et al. [43] inferred that the Eureka effect was linked with high alpha and gamma power from left temporal lobe and high theta power from frontal lobe. Oh et al. [13] reported that a neural reward processing was associated with the Eureka effect that emerged from the orbitofrontal cortex. Ohkuma et al. [44] performed a functional near infrared spectroscopy (fNIRS) based analysis which confirmed that the "Aha!" moment was associated with activation of right dorsolateral prefrontal cortex (DLPFC) and right superior temporal gyrus (STG).

Numerous research studies have explored the analysis of brain activity while individuals employ an insight-based strategy to solve Chinese logogriphs (riddles) [45-53]. Qiu et al. [45] inferred that insight based solutions involved the activations from precuneus, left inferior/middle frontal gyrus, inferior occipital gyrus, cerebellum while Tian et al. [46] confirmed the involvement of left medial frontal gyrus, the left superior temporal gyrus, the right cerebellum, the bilateral claustrum and the left post central gyrus for the said cognitive task. Zhao et al. [47] reported that the early period of insight based solutions for Chinese riddles was associated with activations over left superior temporal pole, inferior temporal gyrus, anterior cingulate cortex, middle frontal gyrus, right inferior frontal gyrus, para-hippocampal

gyrus, bilateral gyrus, and middle temporal gyrus. On the contrary, the later period of insight based solutions depicted activations over left olfactory, middle frontal gyrus, anterior cingulate cortex, medial frontal gyrus, inferior parietal gyrus, right putamen, amygdala, bilateral middle temporal gyrus, hippocampus, and angular gyrus. Luo et al. [48] performed an event related potential (ERP) based study that confirmed negative ERP deflections i.e., N300-500 and N1100-1300 during insight-related problem solving. Shen et al. [49] revealed an early fronto-central P200 was associated with the pre-conscious awareness of mental impasse during elicitation of insight and a P300 was associated with fixed attention when such a mental impasse was formed. Xing et al. [50] observed that insight based strategies involved a negative N300-500 ERP for most of the brain regions and a positive P600–1100 ERP in frontal, fronto-central and central brain regions. Zhao et al. [51] confirmed that the insight based strategy for solving semantic (SR) and homophonic (HR) punny riddles elicited a N350-500 over the central brain lobe. Another study [52] on solving Chinese anagrams revealed that insight based solutions elicited a more positive ERP deflection i.e., P400–600 and P640-780 than analytical solutions from the fusiform gyrus and right superior temporal gyrus respectively. Moreover, Zhao et al. [53] confirmed that insight based problem solving of Chinese character generation task elicited positive ERP deflections i.e., P500-700 and P900-1300 from the centro-parietal brain region after the onset of the stimuli.

The existing works primarily examined the brain activation patterns (such as ERPs or frequency band changes in case of EEG, hemodynamic response in case of fNIRS/fMRI) during the phase of arousal of the Eureka effect. However, these works did not study brain-connectivity and its interpretations associated with the Eureka effect. Brain connectivity analysis, which provides insights into the composite behavior of interactions among multiple brain lobes in response to a cognitive task, is expected to yield more information than brain activations alone [18, 21, 54]. Thus, the present work aims to uncover the correlation between brain connectivity and the Eureka effect, as well as to classify the degree of creativity based on the intensity of the neural reward mechanism acquired from brain connectivity during the arousal of the Eureka effect.

3.3 PROBLEM FORMULATION AND APPROACH

This section offers a condensed overview of the entire experimental paradigm. Fig. 3.1 illustrates the schematic diagram of the entire system. The experiment

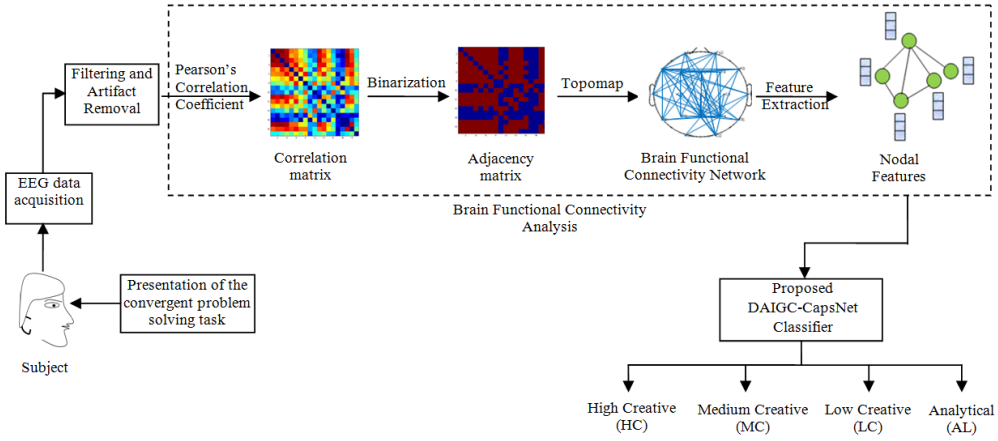
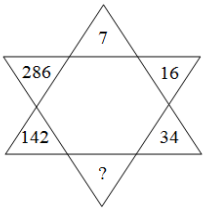


Fig. 3.1. Overview of the proposed paradigm

commences by acquiring EEG signals through $Ag/AgCl_2$ electrodes placed on the scalp of participants engaged in the convergent problem solving task. The visual stimuli employed for the current application consist of a fixation cross for 3 seconds, which assists the subjects in directing their concentration towards the experiment. This period is followed by a 15 second time window containing the mental problem to be solved. Though there exists several categories of convergent problem solving tasks, the test problems chosen for the present application encompass diverse mathematical puzzles such as match stick problems, number puzzles, calendar puzzles and the like [10-11]. This choice aligns with the current research's objective of operating within the realm of scientific creativity. Table 3.1 illustrates some of the exemplar problems used for the current cognitive task. After the problem solving time window, a 10 seconds time frame is presented where the subject verbally responds to the experimenter regarding the solution and also confirms whether the solution appeared mentally with/without the experience of Eureka/insight effect. Moreover, during this phase, for subjects who experienced insight, also rates the effect of the neural reward mechanism into three levels: high, medium and low. Next, the brain responses corresponding to correct oral answers have been considered for further evaluation, while the remaining responses have been discarded.

The raw EEG signals acquired from the scalp of subjects are pre-processed to abolish the effects of artifacts and other noise components. Next, the pre-processed signals are transformed into a graphical structure referred to as a brain functional connectivity network utilizing Pearson's correlation coefficient (PCC) technique due to its simplicity and interpretability [19, 55-56]. This technique first computes the correlation among every EEG channel pair and then binarizes the acquired correlation values to create a brain connectivity network. In the next step, three

Table 3.1
Exemplar convergent problems used for the cognitive experiment.

Category	Test problem
Match stick problem	Fix the following equation by moving only one match stick. 
Number Puzzle	Find the missing number. 
Calendar puzzle	What was the day on 1st January, 1901?

centrality based graphical features (degree centrality, closeness centrality, betweenness centrality) are extracted from the brain networks to encode the structural-functional relationships amongst different brain lobes. The acquired graph features are finally transferred to the proposed DAIGC-CapsNet module to categorize four distinct cognitive classes: Analytical (AL), High Creative (HC), Medium Creative (MC) and Low Creative (LC). This classifier produces four probabilities, each representing the likelihood of an experimental instance of a subject belonging to one of the four classes. The class with the highest probability for a given experimental instance of a subject is regarded as the class for that instance. The details of the principles and methodologies embraced to accomplish the afore-said task are discussed below.

3.3.1 Construction of Brain Functional Connectivity Network

The artifact free EEG signals are used to construct a network comprising nodes and edges to explore the relationship among different EEG channels. Such a construction begins with the computation of Pearson's correlation coefficient [19] that aims to analyze the degree of association among a pair of EEG channels each consisting of *s* data samples using the following equation.

$$\rho_{XY} = \frac{1}{s-1}\sum_{i=1}^{N}\left(\frac{x_i-\bar{x}}{\sigma_X}\right)\left(\frac{y_i-\bar{y}}{\sigma_Y}\right) \tag{3.1}$$

where, ρ_{XY} denotes the correlation coefficient values between channels *X* and *Y*, *N* signifies the total number of nodes within the network, $\bar{x}$ and $\bar{y}$ denote the mean

values while σ_X and σ_Y indicate the standard deviation values of the respective data points of channels *X* and *Y* respectively.

The correlation coefficient values for N number of electrodes produces a $N \times N$ matrix. In the next step, the computed correlation values are normalized within a range of [0,1] and then utilized to create an undirected and unweighted network whose adjacency matrix α_{ij} is expressed as

$$\alpha_{ij} = \begin{cases} 1, & if \ \tilde{\rho}_{ij} \geq \xi \\ 0, & otherwise \end{cases} \tag{3.2}$$

where, $\tilde{\rho}_{ij}$ represents the normalized correlation coefficient value among the i^{th} and j^{th} element of the correlation matrix and ξ denotes a predetermined threshold value. Generally, selecting a single threshold value for the construction of connectivity networks creates structures with different network densities which in turn lead to imbalanced comparative study. Thus, to have a fair comparison among different brain networks, the threshold value is chosen such that all the networks comprise a fixed network density [57].

3.3.2 Feature Extraction of Brain Connectivity Network

The feature extraction procedure pertaining to the domain of brain connectivity involves encoding of latent patterns that can highlight the importance of a brain region. The present work utilizes three centrality based features [20-21] namely degree centrality (DC), closeness centrality (CC) and betweeness centrality (BC) to understand a better interpretation of the acquired structure of the brain networks.

(a) Degree Centrality

The degree centrality (DC) of a node *i* corresponding to an undirected and unweighted network refers to the total number of links incident to *i* and is expressed as

$$D_i^c = \frac{1}{N-1} \sum_{j \in N_a} \alpha_{ij} \tag{3.3}$$

Where, α_{ij} represents the adjacency matrix and N_a denotes the set of all nodes within the network.

(b) Closeness Centrality

The closeness centrality of a node *i* is defined as the inverse of the sum of the shortest path length from *i* to all other nodes within the network and is represented as

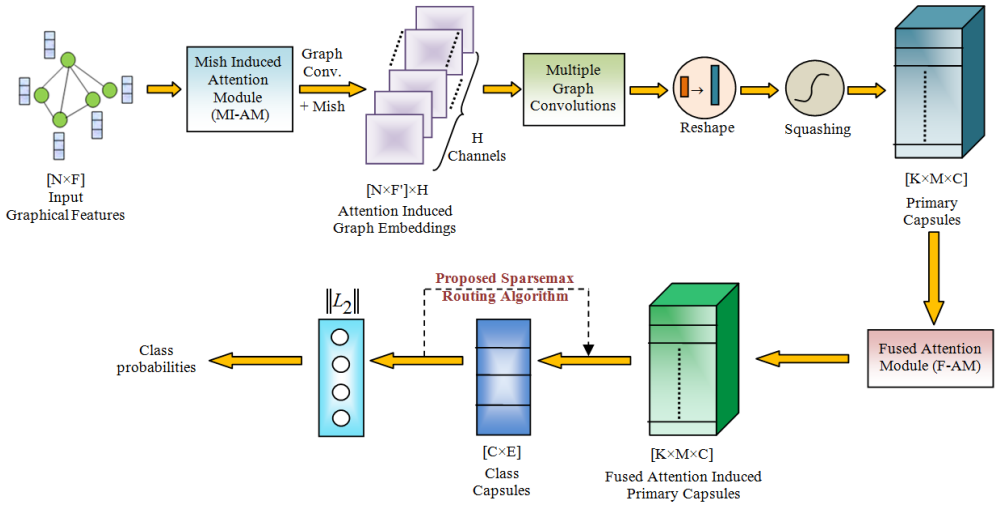


Fig. 3.2. Architectural overview of the proposed Dual Attention Induced Graph Convolutional-Capsule Network (DAIGC-CapsNet)

$$C_i^c = \frac{N-1}{\sum_{j \subset N_a, j \neq i} l_{ij}} \tag{3.4}$$

where, l_{ij} denotes the length of the shortest path from i to j.

(c) Betweeness Centrality

The betweenness centrality (BC) of a node i indicate the portion of all shortest paths within the network that traverse through that specific node. This metric can be evaluated using (3.5).

$$B_i^c = \frac{2}{(N-1)(N-2)} \sum_{j \neq h \neq i} \frac{t_{hj}(i)}{t_{hj}} \tag{3.5}$$

where, t_{hj} and $t_{hj}(i)$ represents number of shortest paths between h and j and the number of shortest paths between h and j passing through i respectively.

3.4 CLASSIFIER DESIGN

The architecture of the proposed classifier is illustrated in Fig. 3.2. The functionalities involved in each stage of the classifier module are described below.

3.4.1 Mish Induced Attention Module (MI-AM)

The first layer of the proposed network consists of a Mish induced attention module (MI-AM), depicted by Fig. 3.3, that guides the graph convolution layers to focus on the most significant features of the brain network in order to enhance the learning capacity of the classifier. In other words, the functionality of the present attention

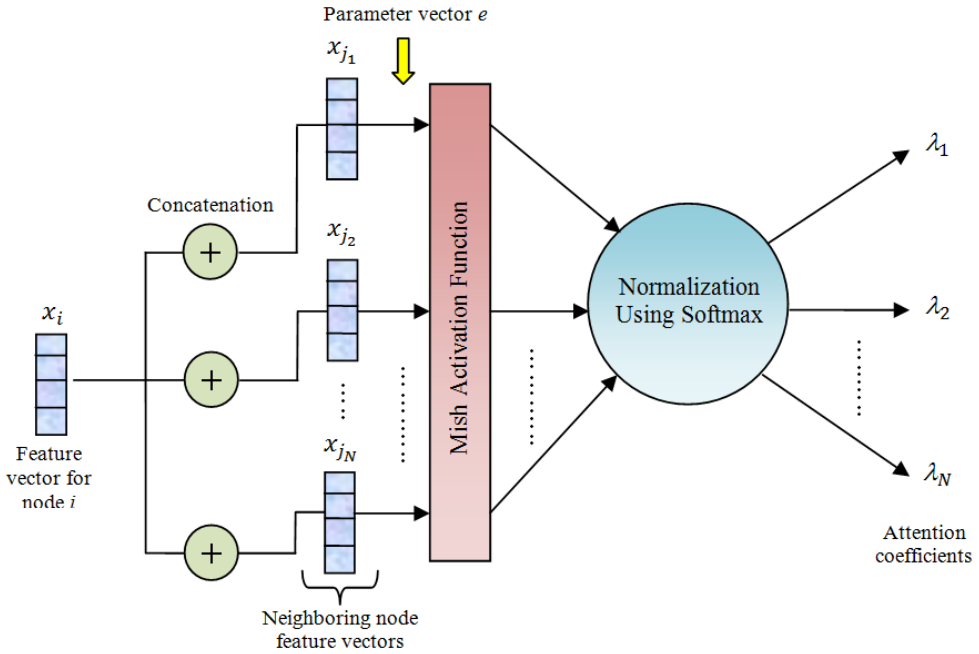


Fig. 3.3. Diagrammatic representation of the Mish Induced Attention

module can be divided into two steps. The first step involves the evaluation of attention coefficients of different node based features depending on their relevancy and importance. The second step consists of combining the evaluated attention coefficient with the graph Laplacian [58] to form an attention induced graph Laplacian. The acquired attention induced graph Laplacian is utilized in the second layer of the network for improving learning ability of the classifier.

The MI-AM utilized for the present application is inspired from [59] with one major modification. In [59], the first step of the attention mechanism begins by computing the attention coefficients a_{ij} between two adjacent nodes i and j using (3.6).

$$a_{ij} = e^T (x_i \| x_j) \tag{3.6}$$

where $e \in \Re^{2F}$, represents a $1 \times 2F$ parameter vector consisting of weights to mark the importance of different node based features, x_i and x_j indicates the features vectors of nodes i and j respectively while || signifies the concatenation operation. In the next step of [59], the computed attention coefficients are passed through a LeakyReLU activation function and then normalized using the Softmax function. The LeakyReLU activation function and the afore-said computation are illustrated by (3.7) and (3.8) respectively.

$$\sigma_{Leaky\,\mathrm{Re}\,LU}(x) = \begin{cases} \alpha x & if \;\; x \le 0 \\ x & if \;\; x > 0 \end{cases} \tag{3.7}$$

$$\lambda_{ij} = Soft\max(a_{ij}) = \frac{\exp(\sigma_{Leaky\,ReLU}(a_{ij}))}{\sum_{j \in N_i} \exp(\sigma_{Leaky\,ReLU}(a_{ij}))} \tag{3.8}$$

where, N_i denotes all the neighbors of node *i* and $\sum_{j \in N_i} \lambda_{ij} = 1$. However, the utilization of the attention coefficients using LeakyReLU activation function has two main drawbacks. First, the predictions for negative input values are inconsistent as α in (3.7) is a constant rather than a learnable parameter [60-61]. Second, the gradient evaluation by LeakyReLU for negative values is a small value due to which over-fitting issues arise during training [62]. In order to combat the above drawbacks, the present work utilizes the Mish [32] activation function to compute the attention coefficients. Mish activation function is unbounded above and bounded below which resolves the over-fitting issues during training by strong regularization effect. Furthermore, Mish possesses the advantageous non-monotonic characteristic that aids in retaining the negative input as negative outputs, thereby enhancing the flow of gradients. Also, the smoothness property of Mish function plays a pivotal role in improving the predictions as it facilitates effective optimization and promotes generalization. Moreover, Mish has shown significant improvements on a variety of deep learning datasets as compared to LeakyReLU activation function [32, 63]. Thus, the above characteristics of Mish have motivated the authors to utilize it for the present scenario. The Mish activation function and its utilization in evaluating the attention coefficients are depicted by (3.9) and (3.10) respectively.

$$\sigma_{Mish}(x) = x \cdot \tanh(\ln(1 + e^{x})) \tag{3.9}$$

$$\lambda_{ij} = Soft\max(a_{ij}) = \frac{\exp(\sigma_{Mish}(a_{ij}))}{\sum_{j \in N_i} \exp(\sigma_{Mish}(a_{ij}))} \tag{3.10}$$

The attention coefficients from all the nodes within a brain connectivity network form an attention matrix $\lambda = [\lambda_1, \lambda_2, ..., \lambda_N]^T$ which is real and symmetric [59]. The computed attention coefficient modifies the graph Laplacian using (3.11) which will be utilized for the convolution operation in the second layer of the classifier model.

$$\hat{A} = \tilde{D}^{-1/2}\,\tilde{A}\,\tilde{D}^{-1/2} * \lambda \tag{3.11}$$

where, $\tilde{A} = A + I$, A denotes the adjacency matrix and I represents the identity matrix, $\tilde{D} = \sum_j \tilde{A}_{ij}$ denotes the diagonal degree matrix, $*$ indicates the Hadamard product.

3.4.2 Graph Embedding using Graph Convolution

The attention induced graph Laplacian obtained from the previous step is utilized in the graph convolution process to form low dimensional graph embeddings of the input graph with centrality features using (3.12).

$$Z_j^{l+1} = \sigma\left(\sum_i \hat{A} Z_i^l W_{ij}^l\right) \tag{3.12}$$

where, $Z^l \in \Re^{N \times F}$ represents the input node based features at the l^{th} layer, $\hat{A}$ indicates the attention induced graph Laplacian, $W_{ij}^l \in \Re^{F \times F'}$ signifies the filters weights among the i^{th} channel in l^{th} layer and the j^{th} channel in $(l+1)^{th}$ layer, $\sigma(\cdot)$ denotes the Mish activation function and $Z^{(0)} = X$ (where $X \in \Re^{N \times F}$ is the input feature vector at the initial layer).

3.4.3 Primary Capsules

The low dimensional graph embeddings are transformed into primary capsules utilizing 3 main steps. The first step involves the employment of Q number of graph convolution operations on the $N \times F'$ low dimensional graph embedding. The second step involves the application of reshape operation on the graph convolved output. The third step involves the employment of squashing operation [30] on the reshaped output to form primary capsules $P \in \Re^{K \times M \times C}$ where K denotes the number of primary capsules, M indicates the dimension of each capsule and C represents the total number of classes.

3.4.4 Fused Attention Module (F-AM) for Primary Capsules

The motivation of the present Fused Attention Module (F-AM) lies in highlighting the important information within the primary capsules so that the most significant predictions are transferred to the higher/class capsules in the subsequent layer. The steps involved in transforming the primary capsules into attention induced capsules are discussed below and illustrated in Fig. 3.4.

First, the primary capsules are reshaped into $R \times 1$ dimension where R denotes the product of K, M and C i.e., $R = K \times M \times C.$ The reshaped output is transferred to 2 parallel sub-modules that independently aids in focusing on the salient aspects of the data while suppressing the irrelevant ones. The first sub-module referred to as the channel attention while the second sub-module is referred to as the spatial attention. The function of each sub-module is as follows.

(a) **Channel Attention**: The main function of this sub-module is to generate an attention map by leveraging the inter-channel relationships within the features. Since, each channel of a feature map signifies a feature detector [30, 64], the afore-said approach revolves around identifying "what" holds significant within the given

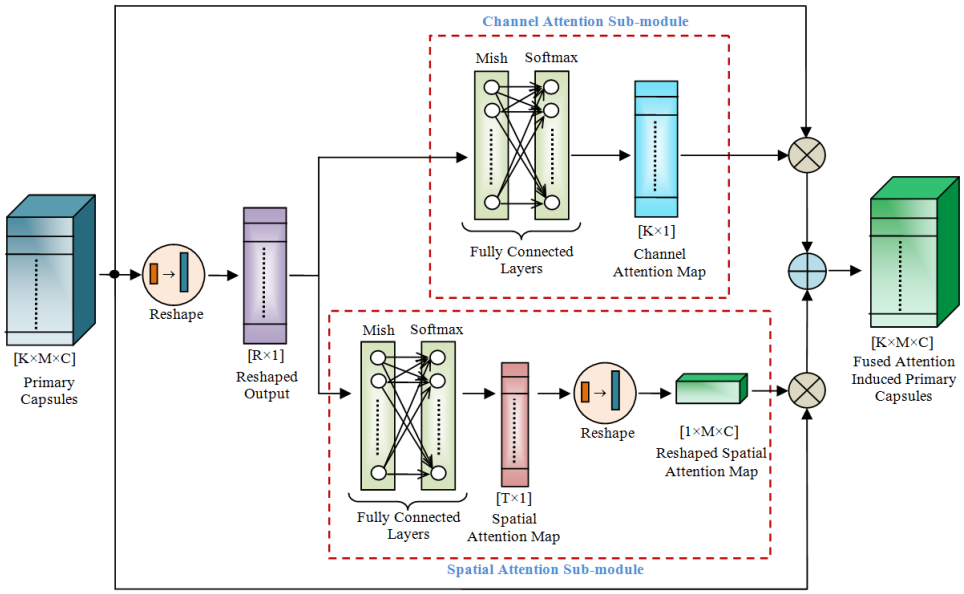


Fig. 3.4. Architecture of the proposed Fused Attention Module (F-AM) for primary capsules

input data. The channel attention sub-module first transfers the reshaped output to two-fully connected layers guided by Mish and Softmax activation functions respectively. Next, the output of the fully-connected layers denoted by $J \in \Re^{K\times 1}$ is element wise multiplied along the dimensions of K with the primary capsules which is mathematically expressed by (3.13).

$$A_c = P \otimes J \tag{3.13}$$

where, $A_c \in \Re^{K\times M\times C}$ represents the channel attention induced primary capsules and $\otimes$ signifies the element wise multiplication operator.

(b) **Spatial Attention**: A spatial attention map is created by harnessing the inter-spatial relationships within features. Unlike channel attention, which emphasizes on "what" constitutes meaningful information, spatial attention centers on identifying the informative regions or "where" certain relevant parts are located. The spatial attention sub-module first transfers the reshaped output to two-fully connected layers guided by Mish and Softmax activation functions respectively. After this, the output of the fully connected layers are reshaped into $O \in \Re^{1\times M\times C}$ and multiplied element wise along the dimensions of $M \times C$ with the primary capsules to form spatial attention induced capsules. This operation is mathematically expressed by (3.14).

$$A_s = P \otimes O \tag{3.14}$$

where, $A_s \in \Re^{K\times M\times C}$ denotes the spatial attention induced primary capsules.

(c) **Fused Attention**: The channel and spatial attention sub-modules are summed together using (15) to form a fused attention induced primary capsules. This refined version of primary capsule surpasses the capabilities of the original capsules by

Algorithm 3.1: Sparsemax Routing Mechanism

1: Input to Routing Mechanism ($\hat{u}_{j|i}, r, l$)
2: for all primary capsules *i* in layer *l* and class capsules *j* in layer $(l+1)$: $c_{ij} \leftarrow 1$
3: **for** *r* iterations **do**
4: for all class capsules *j* in layer $(l+1)$: $s_j \leftarrow \sum_i c_{ij}\hat{u}_{j|i}$
5: for all class capsules *j* in layer $(l+1)$: $v_j \leftarrow Squash\ (s_j)$
6: for all primary capsules *i* in layer *l* and class capsules *j* in layer $(l+1)$: $b_{ij} \leftarrow b_{ij} + \hat{u}_{j|i} \cdot v_j$
7: for all primary capsules *j* in layer *l*: $c_i \leftarrow Sparse\max(b_{ij})$
return v_j

concentrating on pertinent features across both spatial and channel dimensions thereby enhancing the model's capacity to prioritize relevant information within the input data.

$$A_f = \delta A_c + (1-\delta) A_s \tag{3.15}$$

where, $A_f \in \Re^{K \times M \times C}$ denotes the fused attention induced primary capsules, δ represents a hyper-parameter that is tuned during the training phase and $+$ indicates the addition operation.

3.4.5. Class Capsules

The final layer of the classifier model consists of the class capsules. The fused attention induced primary capsules sends its predictions to the class capsules using the Softmax based dynamic routing algorithm [30] for categorizing a specific class label. However, the traditional Softmax based dynamic routing algorithm encounters several shortcomings [64]. First, the Softmax-based approach restricts the variability of routing coefficients, resulting in prediction probabilities that remain uniform even after multiple iterations. Second, the Softmax mechanism fails to provide any sparsity among the prediction probabilities, (i.e., the resulting probabilities after application of Softmax always has full support) causing irrelevant information to be transmitted from primary capsules to secondary/class capsules. Third, Softmax function makes the routing algorithm to be quite computationally expensive. The afore-mentioned issues are effectively addressed by substituting the Softmax function with Sparsemax [33] in the routing algorithm. The Sparsemax function for computing the coupling coefficient is depicted by (3.16).

$$c_{ij} = Sparse\max(c_{ij}) = \underset{p \in \Delta^{G-1}}{\arg\min} \parallel p - b_{ij} \parallel \tag{3.16}$$

where, c_{ij} denotes the coupling coefficients, b_{ij} indicates the log prior probabilities that a primary capsule i will transfer information to a secondary capsule j, and Δ^{G-1} represents the $G-1$ dimensional simplex and p signifies the posterior prediction probability of labels.

The utilization of Sparsemax introduces sparsity among the prediction probabilities leading to the transference of only relevant predictions from the primary to the secondary capsule layers. Moreover, the sparse characteristic of the afore-said function breaks the restrictive range of prediction probabilities that was present due to Softmax. Furthermore, adopting Sparsemax in the routing algorithm also contributes to improved computational efficiency. Sparsemax is notably lightweight and straightforward to implement [33, 65], which translates to faster processing times during the routing process. The modified dynamic routing algorithm based on Sparsemax is denoted by **Algorithm 3.1** where all the notations signify the same meanings as indicated in [30].

Finally, the output of the class capsule is utilized to identify a distinct class label by evaluating its L2 or Euclidean norm.

3.4.6. Margin Loss Function

The prime objective of the margin loss function [30] involves the evaluation of prediction efficacy of the network. It is defined as

$$\Im_k = T_k \max(0, d^+ - \| v_k \|)^2 + \gamma(1-T_k) \max(0, \| v_k \| - d^-)^2 \tag{3.17}$$

where, T_k is a highlighter variable that aids in denoting the absence/presence of an active class label, d^+ and d^- represents the upper and lower boundary parameters respectively while the regularization parameter is indicated by γ.

3.5 RANKING THE CREATIVE ABILITY OF SUBJECTS

The chapter deals with two important aspects: i) classification of an experimental instance of a given subject into one of the four classes: Analytical (AL), High Creative (HC), Medium Creative (MC) and Low Creative (LC), and ii) ranking of a subject whose all experimental instances have been classified as creative (including High, Medium and Low Creative). Now, for the present experiment, there are a total of $M = 2000$ experimental instances available to test the creative potential of each subject. Let m be the number of times the subject is declared as High Creative. Then, a metric called Creative Ability Measure (CAM) is evaluated for the selected subject using (3.18).

$$CAM = \frac{m}{M} \times 100\% \tag{3.18}$$

Fig. 3.5. Experimental setup

Now, if there are S subjects who are experimentally judged as creative, those subjects are sorted in descending order based on the CAM score. The resulting list of subjects is thus arranged in descending order of their ranks of creative ability. For example, the subject occupying the i^{th} position in the list has rank i.

The formula (18) used for rank computation is used later in the Experiments and Results section to evaluate the rank of the subjects identified as creative.

3.6 Experiments And Results

The present section illustrates the detailed description of the experimental paradigm utilized for the present cognitive task. Moreover, it also portrays the results of different experimental analysis undertaken to prove the efficacy of the proposed classifier model.

3.6.1 EEG Data Acquisition

A 21-channel EEG system manufactured by Nihon-Kohden, having a sampling frequency of 500 Hz, has been used to capture the electrical activity of subjects volunteering for the present cognitive experiment. The experiment has been performed with temporal (T_3, T_4, T_5 and T_6), frontal (F_3, F_4, F_7, F_8, F_z), motor cortex (C_3, C_4, C_z), parietal (P_3, P_4, P_z), occipital (O_1, O_2) and prefrontal (Fp_1 and Fp_2) electrodes. The electrode placement strategy followed the 10-20 convention [66-67]. The earlobe electrodes (A_1 and A_2) were used as reference channels while Fp_z was utilized as ground electrode. The experimental setup is depicted by Fig. 3.5.

3.6.2 Participants

The experiment has been conducted on 3 datasets, namely, Artificial Intelligence Laboratory Database (AILDB) [68], North-Bengal Database (NBDB) [69], and Western India Database (WIDB) [70]. Each dataset consists of 32 volunteers and is

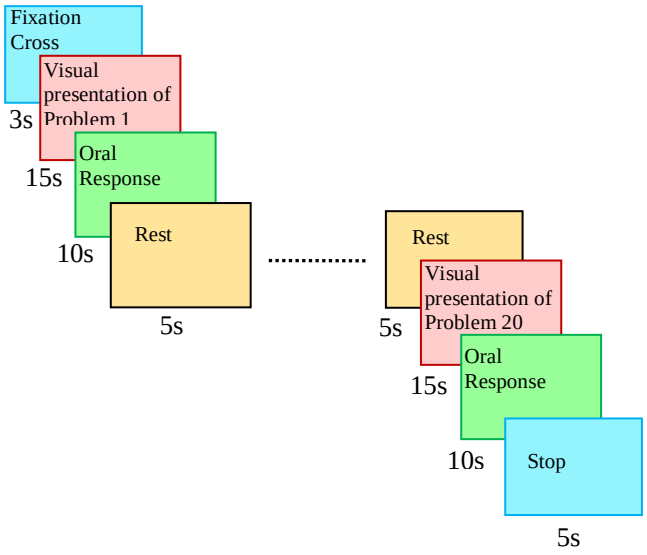


Fig. 3.6. Structure of stimulus presentation for a single trial of the cognitive experiment

representative of different regional groups of people residing in India. Regrettably, no other dataset was found online that had been prepared by other researchers for testing the scientific creative potential of subjects. The volunteers of AILDB (consisting of 18 males and 14 females), NBDB (consisting of 12 males and 20 females), and WIDB (consisting of 17 males and 15 females) belonged to the age groups of 20-36 years, 18-38 years, and 22-40 years, respectively. Normal or corrected-to-normal vision was possessed by all the volunteers, and they had no previous record of neuropsychiatric and/or motor disorders. The ethical norms and other safety precautions related to the experiment have been upheld in accordance with the Helsinki Declaration of 1970, revised in 2004 [71].

3.6.3 Stimuli preparation and presentation

The visual stimuli layout for a single trial is portrayed in Fig. 3.6. As mentioned previously, the mental problem solving phase comprises 15 seconds duration. Each trial consisted of 20 problem sets. Thus, for each data base consisting of 32 subjects, 32 × 20 problems/trial × 20 experimental days × 5 trial/ day = 64,000 experimental instances have been generated.

3.6.4 Pre-processing and Eradication of Artifacts

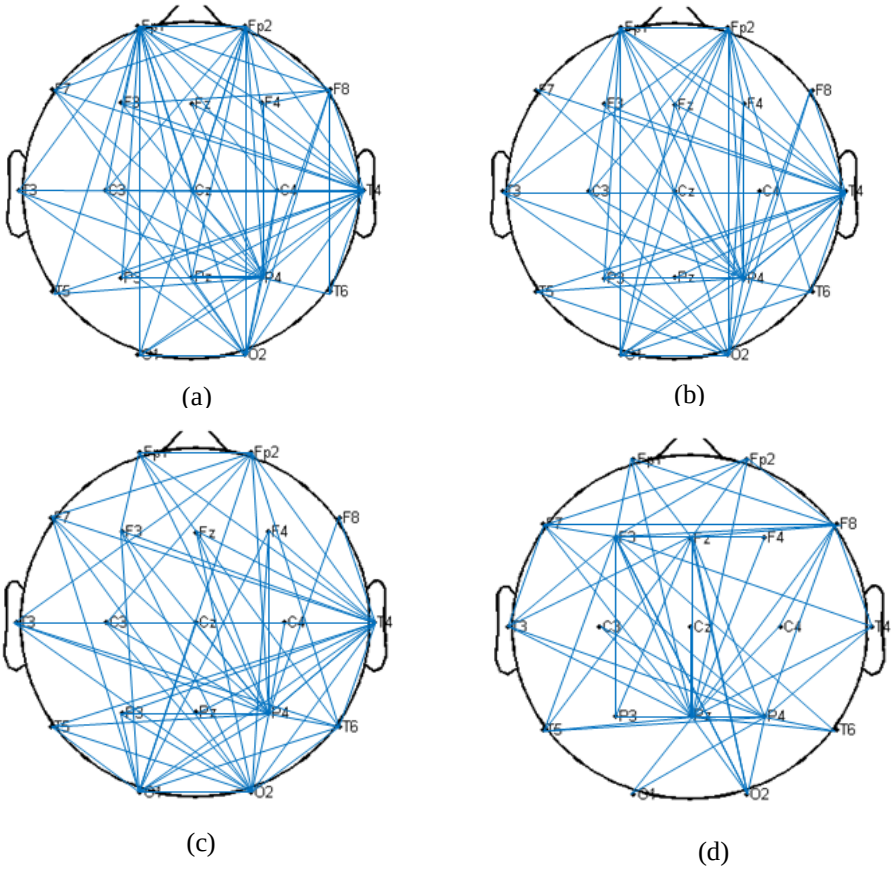
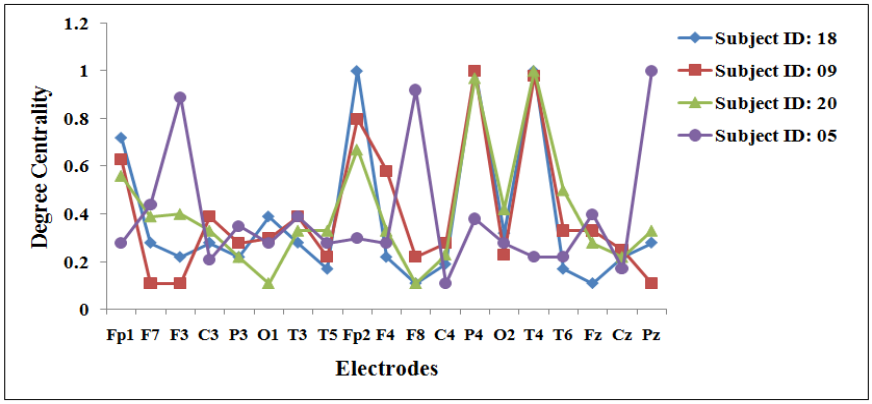


Fig. 3.7. Topology of brain connectivity networks while solving convergent problems (a) network of subject ID: 18 who experienced insight with an enhanced neural reward effect (b) network of subject ID: 09 who experienced insight with a moderate neural reward effect (c) Network of subject ID: 20 who experienced insight without any neural reward effect (d) Network of subject ID: 05 who did not experience an insight effect

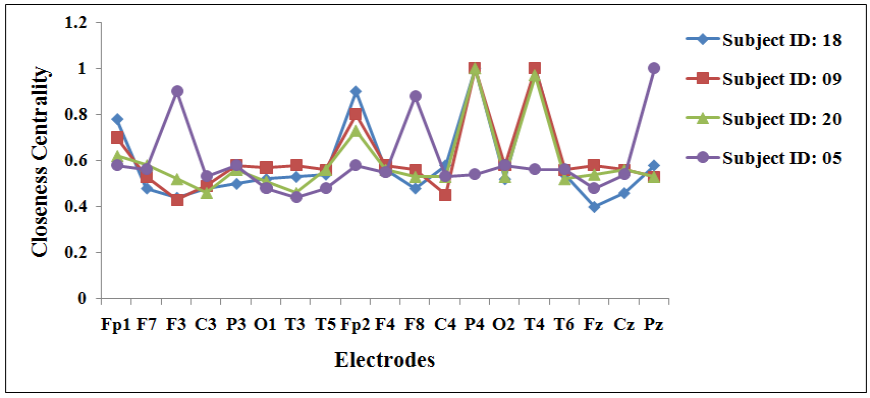
The prime motive of this experiment concerns the elimination of numerous physiological artifacts from the acquired EEG signals which include ocular movements, breathing motion, cardiac pulse and the like [15-17]. The said objective is fulfilled by first applying an Elliptical Band pass filter of order 8 with a pass band range of 0.5-40 Hz. Next, to abolish noise components from the filtered signal, Independent Component Analysis (ICA) [72-73] has been employed that restores 19 independent components pertaining to 19 EEG channels.

3.6.5 Experiment 1 (Identification of hub regions of the brain using centrality features)

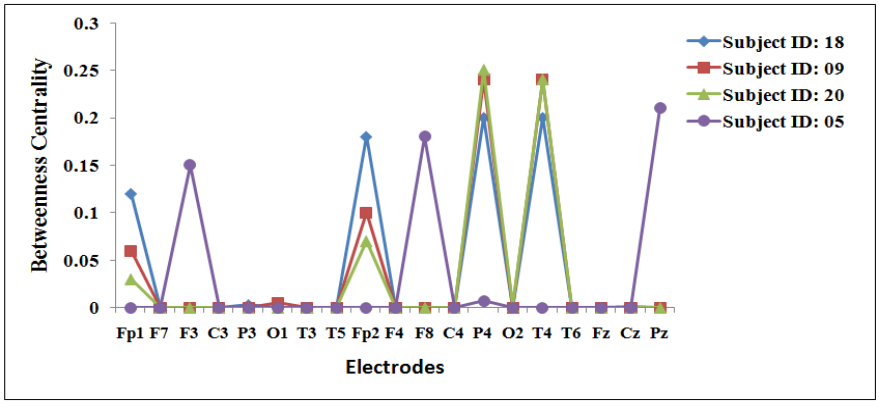
This experiment attempts to investigate the active brain areas participating in the present cognitive activity by identifying the hub lobes. Such an investigation commences with the construction of the brain connectivity network using PCC technique. Fig.3.7 depicts the topographic representation of the brain networks of four subjects belonging to HC, MC, LC and AL classes obtained using the Matlab



(a)



(b)



(c)

Fig. 3.8. Centrality based features analysis of brain connectivity networks for four subjects pertaining to class HC,MC, LC and AL (a) Degree Centrality analysis (b) Closeness Centrality analysis (c) Betweenness Centrality analysis

based Brain Connectivity Toolbox (BCT) [20]. Next, three distinct types of node-based centrality features (DC, BC and CC) are extracted to effectively quantify the intricate topographic structures of the brain obtained from PCC analysis. The analysis of the brain topographic networks using each centrality feature is discussed below.

(a) **DC Analysis**: The DC values of 4 subjects belonging to AL, HC, MC, LC classes for a single trial is depicted by Fig. 3.8(a). It is observable from the figure

that the DC values for subject ID: 05 corresponding to AL class are significantly high over frontal (F_3, F_8) and parietal (P_z) regions. On the other hand, the DC values for subjects belonging to the creative classes (HC, MC, LC) are quite high over parietal (P_4), temporal (T_4) and pre-frontal (Fp_1 and Fp_2) areas. Moreover, the DC values for the prefrontal area decreases from HC to LC class. Additionally, some other nodes (F_4 for MC and T_6 for LC) also depict moderately high DC values. The aforementioned discussion infers that the DC measure alone provides only a partial account of the functional importance of nodes and thus other measures are required to effectively quantify the different networks.

(b) **CC Analysis**: Fig. 3.8 (b) shows CC values for the same subjects utilized for the DC analysis. For AL class, high CC values are observed over frontal (F_3, F_8) and parietal (P_z) regions. On the contrary, the CC values for the creative classes depict notably elevated values across the parietal (P_4) and temporal (T_4) regions, along with heightened values in the prefrontal (Fp_1 and Fp_2) area which reduces from high to low class. CC analysis confirms that the aforementioned nodes (for each class) having elevated values can be regarded as topologically central owing to their capability to engage with other nodes within the network through a minimal number of links [21].

(c) **BC Analysis***:* The results of BC analysis as shown in Fig. 3.8(c) highlights the same regions for each class as detected by CC based evaluation. The depiction of high BC values by the aforesaid lobes suggests their ability to efficaciously integrate functionally across diverse brain regions [20-21].

Thus, the aforementioned node based feature analysis support the suggestion that for analytical thinking, the active brain regions involved are the left dorsolateral prefrontal cortex (F_3 corresponding to BA 09), right medial frontal area (F_8 corresponding to BA 46), and medial superior parietal lobule (P_z corresponding to BA 07) . The above findings are consistent with literature [11]. On the other hand, for creative thinkers, the right supramarginal gyrus (P_4 corresponding to BA 40) and right medial temporal gyrus (T_4 corresponding to BA 21) are most active due to the occurrence of insight effect and are coherent with the existing reports [11, 12, 13]. Moreover, for each degree of creative class, the bilateral orbitofrontal cortex (Fp_1, Fp_2 corresponding to BA 10) also depicts active participation which indicates the involvement of neural reward mechanism [13]. Nonetheless, a discernible trend emerges indicating a decrease in orbitofrontal activation as the degree of creative ability changes from high to low levels. This intriguing observation implies a potential alteration in the neural reward mechanism across varying degrees of creative individuals, with the highest value observed in the HC subject and the lowest in LC subject.

Table 3.2
Parameter settings of the proposed classifier.

Parameter	Optimal value
H	256
F'	3
Q	32
K	1216
M	8
E	16
r	3
δ	0.45
γ	0.5
d^+	0.9
d^-	0.1

In summary, the experimental analysis conducted above confirms a pairing of two cognitive aspects that define the three degrees of creative ability. The HC class is characterized by the presence of the insight effect alongside an enhanced neural reward mechanism, while the MC class exhibits the insight effect with a moderate neural reward mechanism. Conversely, the LC class is characterized by the occurrence of the insight effect with either negligible or no neural reward mechanism. In contrast, since the AL class is distinguished by different hub regions of the brain compared to the creative classes, it signifies the absence of both the insight effect and the neural reward mechanism.

3.6.6 Experiment 2 (Optimal experimental settings)

The parameters of the proposed network involve the dimension of primary capsules, number of channels in each layer and such like. Achieving optimal parameter selection is vital to ensure the model's robustness and highest classification accuracy. To achieve this, the parameters are fine-tuned using the random search algorithm [74]. To evaluate the model's performance, a ten-fold cross-validation test is conducted (where the training and testing data are divided in a ratio of 9:1) for various sets of parameter configurations. Once the highest accuracy level is accomplished with all the candidate settings, the best one is utilized for the test set.

The number of epochs, learning rate and optimizer utilized for the present classifier are 50, 0.001 and ADAM [75] respectively. The rest of the optimized parameters are outlined in Table 3.2.

3.6.7 Experiment 3 (Relative performance analysis of the proposed DAIGC-CapsNet classifier)

Table 3.3
Comparative evaluation of the proposed classifier with other standard Classifiers for AILDB.

Classifier Algorithm with optimal parameter settings	Class Labels								
	HC			MC			LC		
	CA (%)	F1-Score (%)	κ	CA (%)	F1-Score (%)	κ	CA (%)	F1-Score (%)	κ
GGNN [76]	72.02	72.85	0.62	73.28	74.06	0.60	70.86	71.55	0.58
GCN [28]	78.75	76.02	0.69	75.94	75.52	0.64	74.28	75.04	0.61
CayleyNet [77]	76.69	78.23	0.67	79.03	78.96	0.67	77.64	78.01	0.65
GAT [78]	82.26	81.96	0.71	82.58	82.74	0.73	83.20	83.63	0.72
AGCN [59]	85.40	86.51	0.74	86.40	86.57	0.75	86.51	86.75	0.74
DGCN [79]	87.58	88.72	0.79	86.29	87.06	0.80	87.07	87.21	0.78
GCAPS-CNN [27]	88.32	88.84	0.84	88.52	88.68	0.85	89.33	88.73	0.83
CapsGNN [80]	91.53	91.62	0.88	90.76	90.92	0.87	91.47	91.77	0.88
DAIGC-CapsNet	**97.20**	**97.76**	**0.91**	**96.85**	**96.71**	**0.90**	**97.03**	**97.15**	**0.91**

The relative performance of the proposed classifier has been evaluated on the basis of three classification metrics: Classification Accuracy (CA), F1-score and Kappa (κ) coefficient which are defined by equations (3.19-3.21).

$$CA = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.19}$$

$$F1-score = \frac{2TP}{2TP + FP + FN} \tag{3.20}$$

$$Kappa\ (\kappa) = \frac{P_o - P_e}{1 - P_e} \tag{3.21}$$

where, $P_o = \frac{TP + TN}{N}$, $P_e = \frac{TP + FN}{N} \cdot \frac{TP + FP}{N} + \left(1 - \frac{TP + FN}{N}\right) \cdot \left(1 - \frac{TP + FP}{N}\right)$, $N = (TP + TN + FP + FN)$

and *TP*, *TN*, *FP*, *FN* indicates the number of true positives, true negatives, false positives and false negatives respectively. Table 3.3 depicts the comparative analysis of the proposed classifier with other state-of-the-art (SOTA) deep learning models for HC, MC, LC classes.

The results presented in Table 3.3 highlight the GGNN classifier's poor performance, primarily due to its limitation in treating the learned representation of each node as scalar values rather than vectors. Conversely, GCN and CayleyNet demonstrate superior performance compared to GGNN, thanks to their inherent capability to regard node representations as vectors through spectral-based graph convolution. Additionally, other SOTA classifiers, such as GAT, AGCN, and DGCN, which are enhanced versions of GCN, exhibit moderately improved results. However, none of these techniques achieve high values of CA, F1-score, and kappa coefficient, as they fail to preserve various properties of node-based features effectively. GCAPS-CNN integrates the concept of capsules with graph convolution but fails to meet expectations due to its practice of compressing capsules back to scalar features between layers, thus undermining the core purpose of capsules. CapsGNN emerges as the second-highest performing classifier by effectively

Table 3.4
Ablation study of the proposed DAIGC-CapsNet model with respect to CA (%).

Classifier Architecture	Datasets							
	AILDB				NBDB			
	HC	MC	LC	AL	HC	MC	LC	AL
CapsNet	69.32	70.10	68.74	70.12	70.23	69.30	69.21	71.02
F-AM + CapsNet	72.36	71.45	73.02	72.20	72.33	72.84	72.11	73.07
GCN	78.75	75.94	74.28	75.03	76.41	76.52	74.05	75.86
MI-AM+GCN	83.32	84.20	84.52	83.63	83.54	83.75	83.90	82.43
GCN + F-AM + CapsNet	92.56	93.14	93.30	93.12	91.81	92.38	91.96	93.05
MI-AM+ GCN +CapsNet	94.60	94.32	94.92	95.02	94.86	94.50	94.22	94.65
DAIGC-CapsNet	**97.20**	**96.85**	**97.03**	**96.72**	**96.28**	**97.06**	**96.66**	**97.23**

treating each node's learned representation as a vector through graph convolution and preserving vector properties via capsule networks. However, its inability to incorporate an attention module before graph convolution prevents it from claiming the highest SOTA position. Although it features an attention module after the primary capsule layer that highlights the relevant information to be transferred to class capsule layers, this module highlights only information along channel dimensions. The proposed module addresses CapsGNN's limitations through a dual attention mechanism, thereby achieving the best classification results.

3.6.8 Experiment 4 (Ablation study of the proposed DAIGC-CapsNet model)

An ablation study has been conducted to understand the significance of each module within the proposed classifier. To achieve this aim, six different variations of the proposed model have been tested by iteratively removing specific modules in varying proportions to assess their impact on the overall performance. The results of the ablation study are depicted in Table 3.4 and explained in detail below.

(a) CapsNet: This module considers only the CapsNet model with Mish activation function and Sparsemax routing algorithm for classifying the desired class labels. It is evident from Table 3.4 that the classifier yields poor classification results compared to the other variants for all three datasets. Such a result indicates the inability of the CapsNet model to learn and decode brain-connectivity/graph-based features.

(b) F-AM+ CapsNet: This module combines the fused attention block with CapsNet. It exhibits an improvement over the results of the CapsNet module, as the fused attention block focuses on significant predictions to be transferred to the class capsules. Despite outperforming CapsNet, the results indicate that it still struggles to efficiently classify the desired labels due to its inability to decode graph-based features.

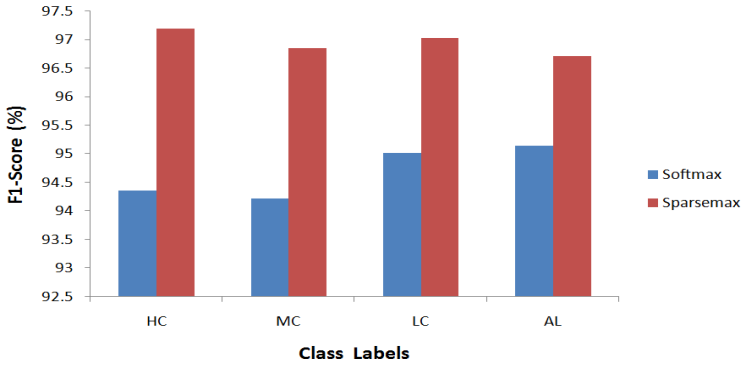


Fig. 3.9. Comparison of Sparsemax routing mechanism with the traditional Softmax routing mechanism for AILDB with respect to F1-Score.

(c) GCN: This module employs only the GCN architecture with Mish activation and fully connected layer for the classification task. As evident from Table 3.4, the current module demonstrates a significant enhancement in CA value compared to previous counterparts (namely, CapsNet and F-AM+ CapsNet), attributed to its inherent ability to proficiently learn the intricate structures within the input graph for classifying the desired labels.

(d) MI-AM+GCN: This module integrates the Mish-induced attention block with GCN. Table 3.4 demonstrates that the current module improves classifier performance compared to GCN by a notably substantial margin. This outcome underscores the importance of the attention module, which adeptly concentrates on the most influential nodes, thereby enhancing classifier performance.

(e) GCN+F-AM+CapsNet: This module integrates GCN and the fused attention block with CapsNet. The combined trio demonstrates a noteworthy enhancement in CA compared to previous counterparts in (a), (b), (c), and (d). These results suggest that such improvement stems from the effective utilization of the three modules. Specifically, the GCN module is employed for abstracting the graphical features, the fused attention highlights the significant predictions to be transferred to the class capsules, while the CapsNet module manages the classification task. However, it still fails to achieve the best result since the Mish-induced attention module is missing, which helps the graph convolution layers focus on the most relevant nodes for enhancing the performance of the classifier.

(f) MI-AM+GCN+CapsNet: This module combines the Mish-induced attention block with GCN and also includes CapsNet without the fused attention. This combination significantly enhances CA compared to the previous ablation variants. Such improvement occurs due to the effective synergy among the three modules. The Mish-induced attention block with GCN transforms the traditional graph convolution operation into an attention-sensitive graph convolution operation,

Table 3.5
Comparative study of different centrality features for AILDB.

Features	Classes					
	HC		MC		LC	
	CA (%)	F1-Score (%)	CA (%)	F1-Score (%)	CA (%)	F1-Score (%)
DC	92.37	92.50	91.86	92.04	92.11	92.08
CC	93.56	93.21	93.27	93.35	93.16	93.33
BC	95.23	95.54	96.06	95.87	95.20	95.27
DC+CC	94.85	94.73	95.29	95.03	95.58	95.41
DC+BC	96.32	96.27	96.54	96.66	96.05	96.13
CC+BC	96.84	96.71	96.33	96.45	96.16	96.07
DC+CC+BC	**97.20**	**97.76**	**96.85**	**96.71**	**97.03**	**97.15**

enabling the classifier to focus on the most significant nodes within the graph while CapsNet handles the classification of the graphical features. However, it fails to provide the highest CA value since the fused attention module in CapsNet is missing, which highlights the relevant predictions to be transferred to the class capsule layer.

(g) DAIGC-CapsNet: This is the proposed module that combines the characteristics of the previous ablation variants in (e) and (f), resulting in the highest CA values across all datasets, as illustrated in Table 3.4. These results indicate that the combination of MI-AM, GCN, F-AM, and CapsNet modules in appropriate synergy highlights the contribution of each module towards enhancing classifier performance.

3.6.9 Experiment 5 (Comparative analysis of Sparsemax routing with traditional Softmax routing technique)

Fig. 3.9 displays the comparison between the Sparsemax and Softmax routing techniques in terms of F1-score. The results indicate that the proposed Sparsemax routing mechanism outperforms the traditional Softmax routing technique, leading to significant improvements in F1-score for various classes. It can be observed from this figure that the proposed routing mechanism shows an improvement of 2.84%, 2.63%, 2.01% and 1.58% for HC, MC, LC and AL classes respectively in comparison to the traditional routing technique. These findings suggest that the F1-score values are low when using the Softmax-based routing, as Softmax restricts the variability of the routing coefficients, resulting in prediction probabilities that remain uniform after several iterations. Conversely, the utilization of the Sparsemax-based routing technique allows significant variation among the prediction probabilities by introducing sparsity among the prediction values, i.e., higher values for significant prediction probabilities and zero for the irrelevant ones. This helps to enhance the coupling strength among the primary and class capsules, which in turn enhances the classifier performance.

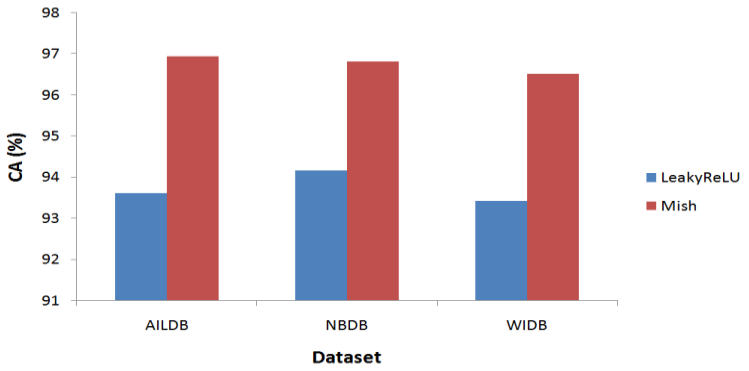


Fig. 3.10. Comparison of Mish Induced Attention Module (MI-AM) with the traditional LeakyReLU Induced Module with respect to overall CA.

3.6.10 Experiment 6 (Comparative analysis of DAIGC-CapsNet with varying centrality features)

The performance evaluation of the proposed classifier takes into account variations in three centrality-based features: degree centrality (DC), closeness centrality (CC), and betweenness centrality (BC). Analyzing the results in Table 3.5 reveals some interesting findings. The classifier's performance notably suffers when utilizing DC-based features alone. This outcome arises because DC-based features primarily emphasize the density of connections among lobes, rather than clearly delineating the topologically central lobes. In contrast, CC-based features demonstrate some improvement over the previous result by distinguishing the topologically central lobes from the rest in the graph. Therefore, combining DC and CC-based features further enhances classifier performance due to the introduction of diversity in features. A diverse feature set helps the classifier produce more reliable predictions as it learns from a richer representation of the input data rather than relying solely on an individual feature. Conversely, employing BC-based feature alone yields slightly better results than DC+CC, suggesting that they are slightly more discriminative in identifying the central lobes in the graph. Consequently, combinations such as CC+BC and DC+BC feature pairs demonstrate reasonably high accuracy values. However, these combinations still fall short of achieving the highest classifier performance because the limited diversity among feature values hampers the classifier's ability to effectively decode graphical patterns corresponding to each class. Thus, the feature combination DC+CC+BC yields the highest performance results, as the classifier benefits from learning different characteristics of the same input graph from a highly diverse feature set.

3.6.11 Experiment 7 (Comparative analysis of Mish induced attention with traditional LeakyReLU attention module)

Table 3.6
Ranking of subjects identified as creative from AILDB.

Subject ID	CAM (%)	Rank
S18	**89.23**	**1**
S23	82.19	2
S11	78.36	3
S09	71.22	4
S01	70.72	5
S19	70.30	6
S32	67.12	7
S16	63.86	8
S07	60.85	9
S25	55.34	10

Fig. 3.10 illustrates the performance comparison between the Mish and LeakyReLU based attention modules. The results indicate that applying the Mish activation function in the attention module leads to significant improvements in Classification Accuracy (CA) for three datasets: AILDB, NBDB, and WIDB. Specifically, when using the Mish activation function, AILDB dataset shows an improvement of 3.33% in CA compared to using LeakyReLU. Moreover, NBDB and WIDB datasets demonstrates an enhancement of 2.45% and 3.12% in CA over the LeakyReLU counterpart. The improved results portrayed by the Mish-induced attention module stem from the smooth and non-monotonic properties of the Mish function. The non-monotonic feature of Mish ensures that negative attention coefficients are assigned small negative magnitudes, aiding in improved expressivity and gradient flow. Furthermore, the smoothness of Mish contributes to enhancing the generalization capacity of the classifier. Conversely, LeakyReLU, lacking non-monotonic or smoothness properties, performs poorly compared to the Mish function.

3.6.12 Experiment 8 (Ranking of subjects according to their creative ability measure)

To rank individuals identified as creative from the total pool of creative subjects, the process begins by computing their CAM score, indicated by (18). Subsequently, all CAM scores are arranged in descending order. Following this, ranks are assigned

Table 3.7
Statistical Validation of the proposed classifier using Friedman's Test.

Classifier Algorithms	R_a	χ_F^2
GGNN	8.89	**22.88 (Null Hypothesis is Rejected)**
GCN	7.44	
Cayley Net	7.56	
GAT	6	
AGCN	4.89	
DGCN	3.44	
GCAPS-CNN	3.67	
CapsGNN	2.11	
Proposed DAIGC-CapsNet	**1**	

based on each subject's CAM score, with the lowest rank value awarded to the individual with the highest CAM score. In Table 3.6, due to space constraints, the ranks of 10 creative subjects from AILDB are displayed, demonstrating the effectiveness of this approach in systematically recognizing and distinguishing creative potential among individuals. Thus, the results presented highlight the structured nature of this methodology in evaluating and differentiating creative abilities.

3.7 STATISTICAL VALIDATION OF THE PROPOSED CLASSIFIER USING FRIEDMAN'S TEST

The statistical validation of the proposed algorithm has been conducted using the Friedman's 2-way non-parametric statistical test [81]. This test provides the lowest rank to the best-performing algorithm. The Friedman static score χ_F^2 for $(k-1)$ degrees of freedom is evaluated using (3.22).

$$\chi_F^2 = \frac{12N}{k(k+1)}\left[\sum_{a=1}^{k} R_a^2 - \frac{k(k+1)^2}{4}\right] \tag{3.22}$$

where, N indicates the number of datasets, k denotes the number of classifier algorithms and R_a represents the average rank of the x^{th} algorithm corresponding to the y^{th} dataset. Table 3.7 depicts the results of the Friedman's test with respect to all classification metrics for N = 3 and k = 9. It is observed from Table 3.7 that the Friedman static score $\chi_F^2 = 22.88 > \chi_{8,0.95}^2 = 15.51$, which represents the chi-square value at

95% confidence level with 8 degrees of freedom. The results of the Friedman test indicate that the null hypothesis, which states that all the algorithms are equivalent in terms of performance, is rejected. Therefore, it is essential to further analyze and compare the algorithms based on their respective ranks.

3.8 DISCUSSION

The prime motivation of the proposed work is to distinguish the creative potential of individuals from analytical thinkers. The afore-said approach is quite appropriate for such a task with respect to the following features discussed below.

(a) **Robustness analysis:** The present work involves the classification of creative individuals into three grades: high, medium and low using a DAIGC-CapsNet. The main challenge of a classification task lies in tackling the over-fitting and/or under-fitting issues. If such issues persist in a classification task, then it is directly reflected upon the classifier's performance. However, the present classifier has been tested on a variety of subjects and is able to provide an average classification accuracy of 96.20% which proves its robustness in combating the above-stated challenge.

(b) **Convenience:** The proposed work has been carried out with minimalistic hardware setup which includes a 21-channel EEG device and a processor. Moreover, the experiment is independent of the participant's age and gender. Thus, the above-mentioned characteristics make the system quite convenient in analyzing the creative skills of an individual.

(c) **Benefits:** The prime motivation of the present work involves the identification of creative individuals from the analytical ones for possible recruitment to different innovation-intensive sectors and research projects. Furthermore, the biological inspiration derived from the proposed scheme holds significant potential for developing computational models of scientific creativity. This inspiration can be utilized to enable the autonomous creation of novel problems in physics, the mathematical derivation of new theorems, and the like.

(d) **Efficiency:** The results of performance analysis depicted in Section 3.6 portray the efficacy of the proposed model in handling the present multi-class classification problem.

(e) **Computational time:** The process of accurately categorizing an unknown class is accomplished within an efficient time frame of 1.36 seconds. This performance is achieved utilizing Matlab 2020b software on a high-performance computing platform equipped with an Intel Core i7-9750H Processor @2.6 GHz and 16GB RAM, complemented by a powerful Nvidia GeForce RTX 2080 GPU with 8GB memory. Such a proficient setup ensures a seamless experience in analyzing

problem-solving strategies, enabling researchers and practitioners to gain valuable insights with ease.

(f) **Flexibility:** The proposed approach has been applied on a large number of participants who have different problem solving strategies.

(g) **Reliability:** The afore-mentioned methodology is deemed reliable for real-world implementation, especially considering that it has already been tested with participants using an EEG device from a reputable company. The approach has already been successfully implemented and proven to be robust and efficient. Therefore, reliability is not a major concern in this context.

Though the proposed model has several advantageous characteristics, it still suffers from a few limitations. These limitations are outlined as follows.

(a) The visual stimuli are presented for 15 seconds. It is presumed that brain activation is confined within 15 seconds interval from the onset of stimuli presentation.

(b) The joint excitations of the right supramarginal gyrus (located in the Parietal lobe) and the right medial temporal gyrus (located in the Temporal lobe) are considered due to the Eureka effect only.

The detection of scientific creative-ability of subjects thus is contingent on the above assumptions. The first assumption has been experimentally validated through subjective reports of experiencing the Eureka effect within 15 seconds. The second assumption was corroborated by observing responses in the supramarginal gyrus and medial temporal gyrus when subjects experienced the Eureka effect within this time-frame. Thus, the present experimental framework satisfied both the assumptions. However, the possibility of a subjective feeling of the Eureka effect after 15 seconds cannot be entirely ruled out.

An alternative possibility of false triggering might appear, in case the subject incidentally shows an evidence of Eureka effect by joint excitations in the lobes mentioned in Assumption (b), but the response is due to parallel thoughts or muscular artifacts. In the present experiment, the authors encountered one such subject (for problem no. 2 out of 20 problems) who claimed not to have experienced the Eureka effect but showed evidence of joint activations in the said lobes of Assumption (b). As the subject denied experiencing the Eureka effect, the authors disregarded his brain response for that problem.

3.9 CONCLUSION

Distinguishing creative individuals from analytical thinkers is an important problem in neurocognitive science. The present study makes a significant contribution to this area by focusing on identifying creative individuals and assessing their levels of

creative aptitude compared to analytical counterparts during a dual-strategy based convergent problem-solving task. Brain connectivity based analysis undertaken using centrality features reveals the active participation of the left dorsolateral prefrontal cortex (BA 09), right medial frontal area (BA 46), and medial superior parietal lobule (BA 07) for a subject utilizing the analytical strategy for problem solving. On the contrary, subjects experiencing insight with an enhanced neural reward mechanism during problem solving depicted the active involvement of the right supramarginal gyrus (BA 40), right medial temporal gyrus (BA 21) and bilateral orbitofrontal cortex (BA 10). The aforesaid BAs exhibited similar activation for subjects experiencing insight with moderate or no neural reward effect but the activation strength progressively diminished in the bilateral orbitofrontal cortex. The acquired centrality based connectivity features are classified using a novel DAIGC-CapsNet classifier to discriminate creative individuals along with their degrees of creative ability from the analytical ones. The novelty of the present classifier includes the addition of a Mish Induced Attention Module (MI-AM) to focus on the most important node features, the inclusion of a Fused Attention Module (F-AM) that ensures the transmission of pivotal predictions from the primary to the class capsule layers and the incorporation of a Sparsemax based routing algorithm to enhance the linking strength between the primary and class capsule layers. Experimental analysis including various ablation studies showcases the promising results of the proposed classifier with respect to the traditional classifier techniques. Furthermore, the results of statistical validation also depict the exceptional performance exhibited by proposed approach.

4

Assessment of Scientific Creative-Potential from Analogical Reasoning Ability Using Brain-Network Based Deep-Fuzzy Classifier

Scientific creativity involves the inherent or computational generation of innovations in science that drives both industrial and societal advancements. Solving the Raven's Advanced Progressive Matrices (RAPM) test requires analogical reasoning, a critical cognitive factor for evaluating creative potential of individuals. The current chapter proposes a novel approach to detect the scientific creative skill of individuals by investigating their brain network patterns acquired using functional Near-Infrared Spectroscopy (fNIRS)) while participating in the RAPM test. The present proposal consists of three phases. First, the brain network construction is undertaken utilizing the Wavelet Transform Coherence (WTC) technique. Second, three node based network features are abstracted from the brain networks. Third, the abstracted features are classified by a novel Enhanced Graph Convolution Induced Type-2 Fuzzy Classifier (EGCIFC) model into five

degrees of creative potential. The novelty of the classifier model includes: i) development of an improved graph convolution operation that aids in capturing the local as well as global structural information from the input brain network, ii) employment of a new activation function, Smish, to boost the performance of the classifier, iii) utilization of one-dimensional spatial convolution layer as a replacement for the pooling layer to retain the most significant information within the convolved embeddings, iv) design of a novel mapping function to mitigate the effect of uncertainty among the spatial convolved vectors in the type-2 fuzzy layer and v) utilization of Takagi-Sugeno-Kang (TSK) based fuzzy reasoning to minimize computational overhead. Experimental analysis implemented highlights the efficacy of the proposed model compared to the conventional methods.

4.1 INTRODUCTION

Scientific creativity [1]-[2] refers to the ability to transcend conventional thinking and generate novel ideas that foster progress in science, technology, industry, and/or society. It acts as a driving force behind scientific breakthroughs that helps to expand the frontiers of human insight. There exists several cognitive factors that are responsible for shaping creative outcomes in scientific realm which include spatial reasoning [3], inductive learning [4], convergent thinking [5] and many others [6]-[7]. In addition to the aforementioned cognitive factors, the ability to draw comparisons between two or more distinct concepts based on their structural or functional similarities, plays a crucial role in facilitating scientific creativity [8]. This ability, referred to as analogical reasoning, has shown to be responsible for several major discoveries and innovation in science. For instance, Niels Bohr drew an analogy between the structure of an atom and the solar system that laid the foundation for quantum mechanics. Similarly in biology, Louis Pasteur explained the working principle of antibodies through an analogy of a lock and key mechanism that helped in the development of vaccines. Likewise, René Descartes revolutionized mathematics by introducing the concept of coordinates by drawing similarities between geometry and algebra. Thus, analogical reasoning, as illustrated by the above examples, strongly influences the genesis of novel outcomes in scientific domain. The present work aims to assess the scientific creative skill of individuals from their analogical reasoning ability using functional Near Infrared Spectroscopy (fNIRS).

The current research employs the Raven's Advanced Progressive Matrices (RAPM) [9] test to examine the scientific creative ability of individuals. Though RAPM test is a popular choice for examining the fluid intelligence of subjects [10],

the motivation for using it for the present application derives from its basis in analogical reasoning [11]-[12]. In the aforesaid test, individuals are required to identify patterns, relationships, and logical structures among abstract visual elements for solving a missing piece of a figure. This process thus mirrors analogical reasoning where recognition of similarities and their correspondences between different elements is crucial to find solutions.

Existing literature [13]-[14] on the RAPM test focuses on examining the cognitive strategies required to solve problems associated with such tests. In a different vein, Friedman et al. [15] utilized the Electroencephalography (EEG) technique for evaluating the cognitive load experienced by subjects during the Raven's test. Alternatively, a recent study by Xu et al. [16] employed an EEG based classification algorithm to distinguish between confused and non-confused mental states of subjects while solving the Raven's test. Other research [17]-[19] on the aforesaid test focuses on exploring the active brain regions involved during the problem solving process utilizing functional Magnetic Resonance Imaging (fMRI). In a similar context, Amin et al. [20] and Jawed et al. [21] independently employed EEG signal analysis to examine the active brain regions and the dominant frequency bands involved in the present cognitive task. On the contrary, Ociepka et al. [22] focused solely on examining the influence of different frequency bands of EEG signals on the Raven's problem solving task. Additionally, Luo et al. [23] and Amin et al. [24] independently performed EEG based experiments to investigate the participation of different Event related Potentials (ERPs) while solving the Raven's test.

Unfortunately, studies exploring the relationship between a subject's creative potential and their brain connectivity during the RAPM task remain largely unexplored. This study aims to fill this gap by examining the fNIRS based brain connectivity associated with analogical reasoning in RAPM tasks to assess the subject's creative potential. In a very recent study, Ghosh et al. [25] examined the creative ability of subjects during the Raven's test by utilizing EEG signal analysis. However, in the present context, brain scanning methods such as fMRI are impractical due to their high cost, while the limited spatial resolution of EEG caused by volume conduction [26] further restricts its use. On the other hand, fNIRS provides a cost-effective solution with moderately high spatial resolution and low computational requirements. Therefore, its selection is an appropriate choice for the present application.

The objective of the current research is to evaluate individual differences in creative ability through the RAPM task. The above evaluation is carried out in three phases. In the first phase, the fNIRS signals recorded from the scalp of participants

engaged in the RAPM task are pre-processed and converted into a brain connectivity network using the Wavelet Transform Coherence (WTC) [27] method. In the second phase, three node based features are extracted from the acquired brain networks which include node strength (NS), node efficiency (NE) and node betweenness (NB) [28]. The rationale behind the feature extraction process is to examine the Brodmann Areas (BAs) of the brain that serve as primary controllers of the entire network during the processing of the RAPM task. In the third phase, the extracted features pertaining to each node is utilized to classify the creative ability of participants into five distinct levels: Extremely Creative Thinker (ECT), Superbly Creative Thinker (SCT), Fairly Creative Thinker (FCT), Mildly Creative Thinker (MCT) and Conventional Thinker (CVT).

The prime contribution of the present work involves the development of a classifier model that is specifically designed to handle the above classification problem. As the present wok deals with brain connectivity networks, the classifier model is constructed on the framework of a Graph Convolution Network (GCN) [29]-[30], a popular deep learning approach for classifying graph based data [31]-[32]. A classical GCN model consists of two main modules. The first module abstracts the features form the input graph via graph convolution and pooling while the second module classifies the abstracted features using fully connected layers [30]. However, fNIRS signals collected from a given source are susceptible to fluctuations both within and across sessions due to parallel thoughts, extrinsic and/or intrinsic artifacts [33]-[34] that introduces uncertainty within the abstracted features [35]-[37]. To combat the aforesaid issue, the present classifier architecture incorporates the concept of Interval Type-2 Fuzzy (IT2F) sets within the GCN model.

In this chapter, a novel classifier model, referred to as Enhanced Graph Convolution Induced Type-2 Fuzzy Classifier (EGCIFC) has been designed to classify the fNIRS signals of individuals into five distinct levels of creative potential (i.e., ECT, SCT, FCT, MCT and CVT). The first layer of the proposed EGCIFC model performs a novel enhanced graph convolution upon the input brain network to convert it to low-dimensional embeddings that aids in capturing both the local and global structural information from the graph. This operation helps the classifier model to learn diverse patterns from the same input, thereby improving its performance. The enhanced graph convolution is followed by the utilization of a Smish [38] activation function that helps to introduce non-linearity within the model. The rationale of choosing the Smish function lies in its ability to effectively impart negative activation to small negative amplitude values which in turns improves classification accuracy. The second layer of the EGCIFC model again performs

enhanced graph convolution operation along with Smish function upon the acquired embeddings to capture higher-level information. The third layer of the proposed model applies a one-dimensional (1D) spatial convolution operation on the output of the second layer instead of the traditional pooling operation. This approach addresses two major drawbacks of pooling. First, pooling fails to capture the most significant information from embeddings because it does not utilize kernel-based operations with trainable weights. Second, common pooling techniques either lose finer details by retaining only the most prominent embeddings (as in max pooling) or over-smooth the embeddings by blending all information (as in mean pooling). The fourth and final layer of the proposed model comprises an IT2F classifier that categorizes the spatial convolved vectors into the desired class labels.

The uniqueness within the IT2F layer of the proposed model consists of two important aspects. First, a new approach has been developed to diminish the influence of uncertainty on classifier performance. This approach involves the design of a suitable mapping function that uplifts the lower membership function (LMF) to harness the most significant region of the Foot print of uncertainty (FOU) in the type-2 fuzzy set. Second, the structure of fuzzy rules for classification is governed by the Takagi-Sugeno-Kang (TSK) [39] technique rather than the traditional Mamdani [40] approach. In the Mamdani approach, the consequent consists of a fuzzy quantified proposition for the output variable. In contrast, the TSK technique uses a linear function of the antecedent variables as the consequent. This allows the output variable to be directly obtained in defuzzified form, eliminating the need for additional defuzzification and type-reduction [41], thereby reducing computational expense.

Performance analysis undertaken against the state-of-the-art (SOTA) algorithms demonstrates the proposed model's effectiveness in classifying the different degrees of creative potential. Comprehensive comparative studies, including ablation based analysis, further highlight the proficiency of the proposed approach. Additionally, statistical validation also confirms the model's efficacy in performing the classification task. Furthermore, a scoring scheme has also been designed to rank each experimental participant in accordance to their degrees of creative potential based on the classification results.

The main contributions of the current research are summarized as follows.

i) Employment of brain connectivity based analysis with three node based features (strength, efficiency and betweenness) to investigate the BAs actively engaged for different classes of individual (Extremely Creative Thinker (ECT), Superbly Creative Thinker (SCT), Fairly Creative Thinker (FCT), Mildly Creative Thinker (MCT) and Conventional Thinker (CVT)) while solving the RAPM task.

ii) Formulation of a novel classifier model referred to as Enhanced Graph Convolution Induced Type-2 Fuzzy Classifier (EGCIFC) for categorizing the desired class labels of creative potential. The novelties within the proposed model are listed below.

a) Design of an enhanced graph convolution operation capable of capturing both local and global structural information from the input graph that enables the classifier to learn diverse patterns from the same graph.

b) Utilization of the Smish activation function to improve classification accuracy.

c) Introduction of a 1D spatial convolution layer after the enhanced graph convolution, instead of pooling, to retain the most significant information within the convolved embeddings.

d) Design of a novel mapping function in the Interval Type-2 fuzzy layer to minimize the uncertainty among the spatial convolved vectors.

e) Employment of TSK based fuzzy reasoning for handling the present classification task in order to reduce computational cost.

iii) Demonstration of the efficacy of the proposed EGCIFC model through rigorous comparative analyses against SOTA techniques.

iv) Development of an evaluation scheme to rank individuals based on their degrees of creative potential.

The rest of the chapter is arranged as follows. Section 4.2 discusses the steps required to perform the present classification task. Section 4.3 delves into the formulation of the proposed classifier model. Section 4.4 outlines the experimental paradigm and discusses the results of the cognitive experiment. Section 4.5 portrays the comparative studies conducted to demonstrate the efficacy of the proposed model, while Section 4.5 provides the conclusions drawn from the chapter.

4.2 PROBLEM FORMULATION AND APPROACH

This section presents a brief summary of the cognitive experiment undertaken for the present research. Fig. 4.1 represents the block diagram of the proposed experimental paradigm. The experiment begins by acquiring the brain signal of participants using the fNIRS device who have volunteered for the RAPM task. The RAPM task utilized for the current problem consists of 4 difficulty levels: extremely hard, hard, medium and easy [10], [42]. An exemplar problem used for the present experiment for medium difficulty level is portrayed in Fig. 4.2. Each difficulty level consists of 5 problems. A single session of the experiment pertaining to each difficulty level consists of 5 trials. The structure of the visual stimuli used for the current cognitive activity for a single session is depicted by Fig. 4.3. The stimulus begins with a fixation cross of 3 seconds followed by a time-frame of 15 seconds where the RAPM

problem needs to be solved mentally by the participant. After this, a time-frame of 10 seconds is provided where the participant verbally informs the experimenter about the answer to the RAPM task. The aforesaid process is repeated for four more number of problems. Thus, for 10 experimental days and 5 sessions in a day, 5 problems/trial × 5 trials/session × 5 sessions/day × 10 days produces 1250 experimental instances corresponding to each participant for a given difficulty level.

After the brain signal acquisition phase is over, the signals are normalized, pre-processed and converted into a brain connectivity network using the WTC technique. The nodes in the aforesaid network represent the BAs corresponding to a given montage while the edges represent the connections between the nodes. Subsequently, three node based features (strength, efficiency and betweenness) are abstracted from the formulated brain networks and analyzed to identify the BAs acting as prime controllers for the present cognitive task. The abstracted nodal features from each experimental instance are utilized by the proposed EGCIFC model to categorize five levels of creative ability of subjects: i) ECT (able to solve very high-level problems), ii) SCT (able to solve high-level problems), iii) FCT (able to solve medium-level problems), iv) MCT (able to solve easy-level problems), and v) CVT (unable to solve problems of any difficulty level). Finally, the classification results are used to develop a ranking scheme that ranks each participant relative to others. The details of the aforementioned phases are discussed below.

4.2.1 Evaluation of Normalized Cerebral Oxygen Exchange

The first step in analyzing the acquired raw fNIRS signals is to assess the net oxygen consumption in the brain's cortical region, quantified by the metric known as cerebral oxygen exchange (COE) [43]-[44]. The steps required for computing the aforementioned metric are elaborated below.

The fNIRS data collected from the scalp of participants captures two key blood concentration measurements: oxy-hemoglobin (O_2Hb) and deoxy-hemoglobin (HHb), both expressed in mmol/L. Let $\Delta O_2Hb(t)$ and $\Delta HHb(t)$ represent the changes in concentrations of O_2Hb and HHb respectively, corresponding to a channel of the given montage at time t during an experimental instance. Since, $\Delta O_2Hb(t) > \Delta HHb(t), \forall t$ [35], [37], COE at any time t is evaluated as,

$$C_o = \Delta O_2Hb(t) - \Delta HHb(t) \tag{4.1}$$

The normalized value of COE within the range [0,1] is computed by (4.2).

$$\hat{C}_0 = \frac{\Delta O_2Hb(t) - \Delta HHb(t)}{^{Max}\Delta O_2Hb(t) - ^{Min}\Delta HHb(t)} \tag{4.2}$$

where, $^{Max}\Delta O_2Hb(t)$ indicates the maximum value of $\Delta O_2Hb(t)$ at time t while $^{Min}\Delta HHb(t)$ signifies the minimum value of $\Delta HHb(t)$ at time t.

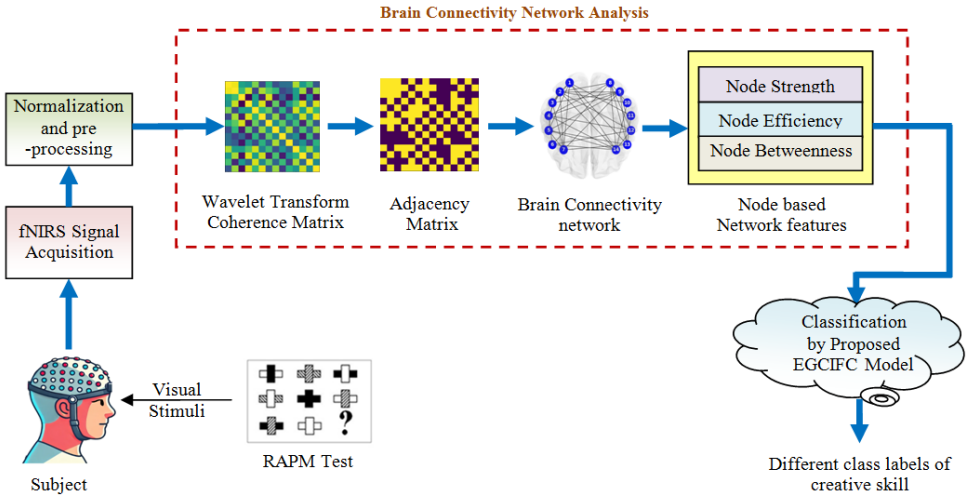


Fig. 4.1. Block diagram of the proposed experimental framework illustrating the main modules utilized for classifying different degrees of creative potential of subjects

4.2.2 Elimination of Different Artifacts from Raw fNIRS Signals

After normalizing the net oxygen consumption induced fNIRS signals, the next step involves eliminating various artifacts that may have been introduced into the raw signals during the acquisition stage. The artifacts affecting the fNIRS signals can be broadly classified into two categories [33]-[34]: i) extrinsic artifacts and ii) intrinsic artifacts. Extrinsic artifacts originate from external factors unrelated to the subject's physiological state. Extrinsic artifacts can be of two types: a) Technical: these involve instrumental noise, which may result from variations in the equipment's sensitivity, calibration issues, or inconsistencies in optode placement on the scalp, leading to differences in signal detection across trials or participants, and b) Environmental: these refer to external conditions that can impact fNIRS signal readings, such as ambient light interference, temperature fluctuations, and other similar factors. Intrinsic artifacts in fNIRS signals are primarily physiological noises and fluctuations that disrupt the accurate measurement of cerebral hemodynamics. These include cardiac pulsation, respiration, blood pressure variations, Mayer waves (low-frequency blood pressure oscillations linked to vasomotor activity that impact blood oxygenation), and motion artifacts (disturbances in fNIRS signals caused by physical movement).

A comprehensive signal processing approach is implemented to eliminate the aforementioned artifacts from the fNIRS signal. Initially, the normalized values of COE for all channels are filtered using a Chebyshev band-pass filter of order 6, with a passband range set at 0.01 to 3.5 Hz. The selection of the Chebyshev band-pass

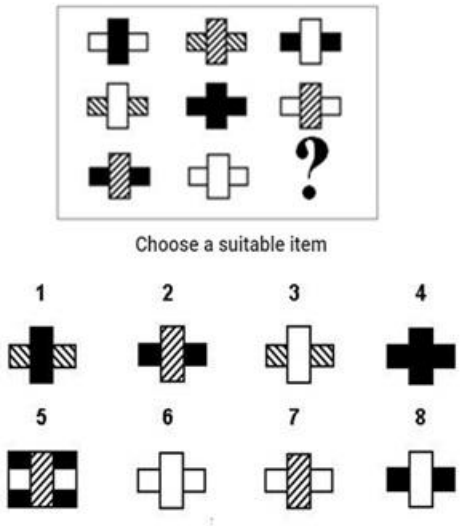


Fig. 4.2. An exemplar RAPM problem of medium difficulty level

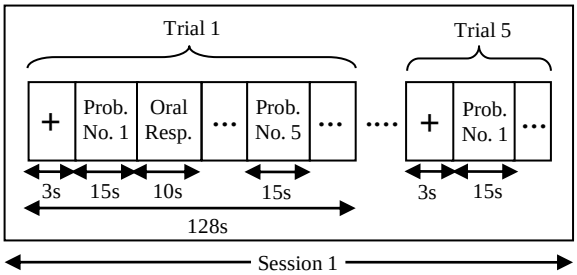


Fig. 4.3. Structure of visual stimuli used for the RAPM task for a given difficulty level pertaining to a single session

filter is made due to its ability to achieve a sharper roll-off, enhancing the separation of frequency components while maintaining a manageable level of ripple within the pass-band [45]. Independent Component Analysis (ICA) [46] is employed on the filtered signals. This technique effectively isolates the independent hemodynamic components associated with each channel, facilitating improved signal clarity and reliability for subsequent analysis.

4.2.3 Formation of Brain Connectivity Networks

The filtered fNIRS signals are processed using wavelet transform coherence (WTC) to explore their behavior across time and frequency domains. This approach quantifies the cross-correlation between two time-series signals over varying time and frequency scales [27], [47]. The WTC between two fNIRS channels, X and Y, at the time instance n (denoted as X_n and Y_n) is mathematically defined as follows.

$$V_n(s) = \frac{\left|\left\langle s^{-1} V_n^{XY}(s)\right\rangle\right|^2}{\left|\left\langle s^{-1} V_n^{X}(s)\right\rangle\right|^2 \left|\left\langle s^{-1} V_n^{Y}(s)\right\rangle\right|^2} \tag{4.3}$$

where, s represents the wavelet scale, n denotes the time instance, $V_n^X(s)$ and $V_n^Y(s)$ indicates the continuous wavelet transform of X_n and Y_n respectively, $|\cdot|$ denotes the absolute value and $\langle\cdot\rangle$ represents the smoothing operation [47]. For the present application, the Daubechies wavelet [48] is utilized as the mother wavelet for evaluating the WTC among the pair of FNIRS channels.

Using the operation in (3), a $P \times P$ matrix is generated, representing the wavelet coherence between pairs of fNIRS channels. These coherence values are subsequently employed to build an undirected, weighted connectivity network using the following approach.

$$a_{ij} = \begin{cases} V_{ij}, & if\ V_{ij} \geq \xi \\ 0, & otherwise \end{cases} \tag{4.4}$$

where, a_{ij} denotes the adjacency matrix values of the connectivity network, V_{ij} indicates the wavelet coherence values while ξ indicates a predefined threshold. The predefined threshold is chosen in such a way that all classes of networks have the same number of edges, ensuring a fair comparison among them [49].

4.2.4 Feature Abstraction of the Connectivity Networks

To identify the hub regions of the brain that govern the entire network, the brain connectivity networks derived from the WTC technique needs to be analyzed through various network topological features. This analysis is carried out by extracting three key features: node strength (NS), node efficiency (NE), and node betweenness (BC), [28].

The extracted features are ultimately categorized into five distinct classes using the EGCIFC classifier, with its detailed architecture explained in Section 4.3.

4.3 ARCHITECTURE OF THE PROPOSED ENHANCED GRAPH CONVOLUTION INDUCED TYPE-2 FUZZY CLASSIFIER

This section provides a detailed explanation of the proposed EGCIFC model, whose architecture is depicted in Fig. 4.4. The description and functions of each layer of the proposed classifier model are discussed below.

4.3.1 First Enhanced Graph Convolution Layer with Smish Activation Function

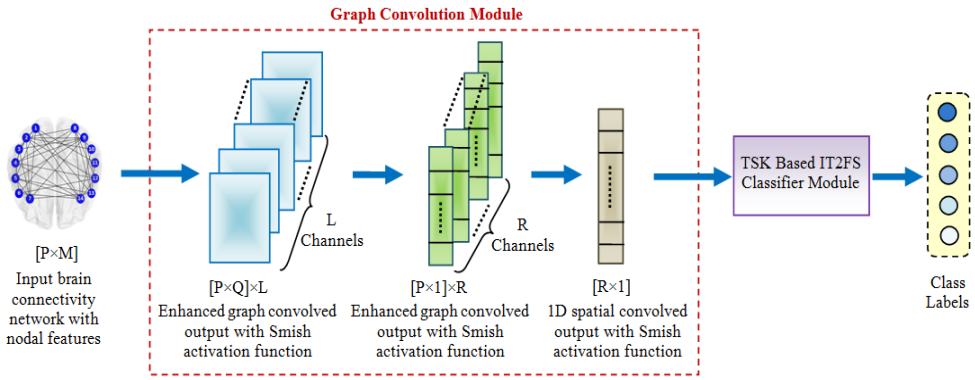


Fig. 4.4. Overview of the architectural framework of the proposed Enhanced Graph Convolution Induced Fuzzy Classifier (EGCIFC) model

The first layer of the proposed classifier includes the employment of an enhanced graph convolution operation and Smish activation function. The elaboration of the above operations is provided below.

4.3.1.1 Enhanced Graph Convolution

A given input graph $G \in \Re^{P\times M}$ is transformed into its graph embeddings using the traditional graph convolution operation [29] which can be mathematically expressed by (4.6).

$$Z_j^{k+1} = \sigma\left(\sum_i \tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2} Z_i^k W_{ij}^k\right) \tag{4.6}$$

where, $Z^k \in \Re^{P\times M}$ represents the input graph in the k^{th} layer and $Z^0 = G$. $W_{ij}^k \in R^{M\times M'}$ signifies the matrix of trainable weights while $\sigma(\cdot)$ indicates an activation function. $\tilde{A} = A + I$ denotes the adjacency matrix A consisting of self-loops, $I \in \Re^{P\times P}$ represents the identity matrix, and $\tilde{D}$ signifies the degree matrix of $\tilde{A}$.

However, the traditional graph convolution operation relies solely on the adjacency matrix A, which captures only the local structure of the graph. As a result, it aggregates information exclusively from the immediate neighbors of a node, neglecting the influence of distant neighbors. This reduces the classifier's performance, as it cannot effectively learn diverse patterns within the same graph from multiple perspectives due to its confinement to local structural information.

To address the aforementioned limitation, the present work modifies the adjacency matrix to incorporate both local and global structural information of the graph using (4.7).

$$E = \exp(A) \approx I + A + \frac{A^2}{2!} + \frac{A^3}{3!} \tag{4.7}$$

The computation in (4.7) describes the series expansion of the exponential operation applied to A, approximated up to the third order. In (4.7), the matrix I incorporates

self-loops, *A* captures the local structural information of the graph, and the higher-order terms of *A* represent the global structural information of the graph, extending up to the third-distant neighbor. It may be noted that the truncation of the expansion in (4.7) to the third order is determined by the random search algorithm [50] discussed in Section 4.4.

Thus, the computation in (4.6) is now modified using the new adjacency matrix *E* in (4.8), ensuring that the convolved features from the input graph capture both local and global structural information.

$$Z_j^{k+1} = \sigma(\sum_i \tilde{D}^{-1/2} E \tilde{D}^{-1/2} Z_i^k W_{ij}^k) \tag{4.8}$$

where, *E* denotes the modified adjacency matrix and $\tilde{D}$ now represents the degree matrix of *E*. The convolution operation in (4.8), hereafter referred to as the enhanced graph convolution, thus enables the classifier to learn diverse patterns from a single graph, ultimately improving its performance.

4.3.1.2 Employment of Smish Activation Function

A new activation function referred to as Smish [38] is utilized for the current work instead of traditional ones. The mathematical equation for this function is denoted by (4.9).

$$\sigma(x) = Smish\,(x) = x \cdot \tanh(\ln(1+\rho(x))) \tag{4.9}$$

where, $\rho(x)$ denotes the sigmoid activation function. The motivation behind the utilization of Smish involves the following points: i) it has enhanced regularization capability as it maintains negative output regularization and partial sparsity, thereby preventing over-fitting, ii) it ensures smooth and stable gradient flow through the application of the tanh operator on a logarithmic transformation and iii) it has shown to be highly robust across various hyper-parameters [38], ensuring consistent performance for different experimental settings.

4.3.2 Second Enhanced Graph Convolution Layer with Smish Activation Function

An additional enhanced graph convolution with Smish activation function is performed upon the graph convolved output from the first layer using (4.8) and (4.9). This additional operation ensures that the higher level features are abstracted from the previously convolved output so that classifier model can effectively capture intricate patterns from the input data. The output of the present operation yields $J \in R^{(P\times 1)\times R}$, where $P\times 1$ denotes the dimension of each convolved vector and R represents the number of channels (see Fig. 4.4).

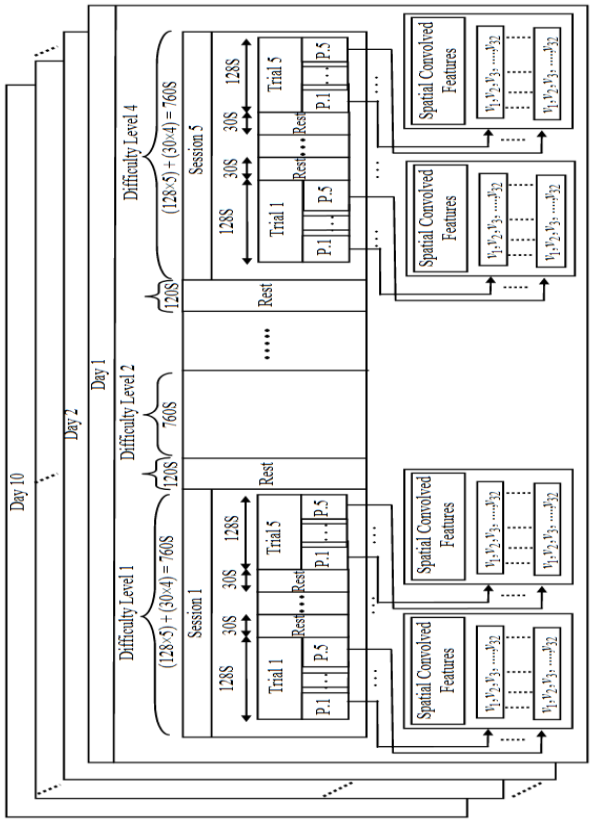


Fig. 4.5. The spatial convolved features obtained from each problem of a trial in a session pertaining to a given difficulty level for every experimental participant

4.3.3 1D Spatial Convolution Layer

The objective of the present layer involves reducing the dimension of the graph convolved features as well as retaining the most significant feature information that will be utilized for classification. The traditional GCNs usually rely on pooling operations to accomplish the aforesaid objective [30]. However, pooling comes with two main limitations. First, unlike convolution, pooling does not utilize kernel based operations having weights that are learned during training, thereby failing to capture relevant features from the input data. Second, popular pooling techniques either ignore finer details by only retaining the most prominent features (max-pooling) or over-smoothes the feature set by blending all information (mean-pooling).

Thus, to overcome these limitations, the current work utilizes a 1D spatial convolution on the second graph convolved output. The 1D spatial convolution operation is denoted by (4.10)

$$h_j = \sigma((g_i * w_{ij}) + \beta_j) \tag{4.10}$$

where, $h_j \in R^{1\times1}$ represent the 1D spatial convolved output, $g_i \in R^{P\times1}$ denotes the second graph convolved output for a single channel, $\sigma(\cdot)$ signifies the Smish

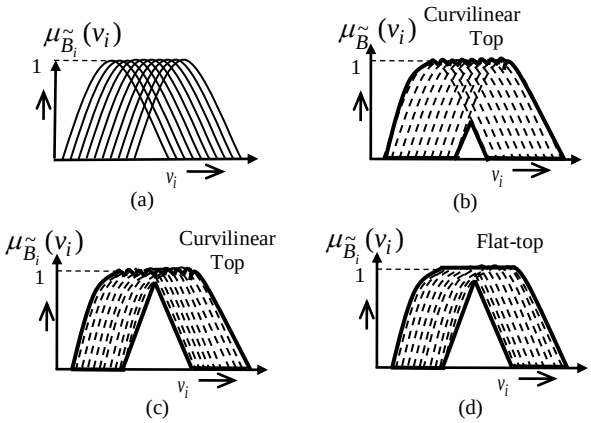


Fig. 4.6. Construction of IT2FS with refined FOU (a) Type-1 MFs for 10 days (b) Curvilinear-top based IT2FS formulated by taking union of Type-1 MFs (c) Curvilinear-top based IT2FS with refined FOU (d) flat-top approximated IT2FS

activation function, $w_{ij} \in R^{P\times 1}$ indicates the weight matrix, β_j is the bias term and "$*$" represents the spatial convolution operator. The operation in (4.10) is repeated for all the *R* channels which thereby produce *R* number of scalar values. These *R* scalars are stacked into an *R*×1 dimensional vector which represents the output of present spatial convolution layer to be used later for classification. Thus, the *R*×1 vector contains the most relevant abstracted features from *J* using spatial convolution that will guide the classifier to distinguish the desired class labels.

4.3.4 TSK Induced Interval Type-2 Fuzzy Layer

The 1D spatial convolution layer generates a feature vector of size *R*×1 (where *R*=32) for every RAPM problem as denoted by Fig. 4.5. To handle the uncertainty arising in these convolved feature vectors due to sessional variations, a TSK-induced Interval Type-2 Fuzzy Classifier (IT2FC) is employed. The aforementioned approach is detailed below.

4.3.4.1 Formulation of Antecedent Part of IT2FC

Let, $v_1, v_2, ..., v_N$ be *N* number of features that has been obtained after the employement of the 1D spatial convolution layer. Let $\tilde{B}_i$ for *i*=1 to *N* represent an IT2FS defined as $[\overline{\mu}_{\tilde{B}_i}(v_i), \underline{\mu}_{\tilde{B}_i}(v_i)]$ for a linguistic variable v_i, where $\overline{\mu}_{\tilde{B}_i}(v_i)$ and $\underline{\mu}_{\tilde{B}_i}(v_i)$ denote the upper and lower membership function (MF) of $\tilde{B}_i$ respectively. For the current application, $\tilde{B}_i$ has been formulated utilizing within session variations (5 trials in a session) and across session variations (5 sessions in a day) in v_i (pertaining to one spatial convolved feature vector). The formulation of $\tilde{B}_i$ begins by creating one Type-1 Gaussian MF through the evaluation of average (α) and variance (σ^2) among

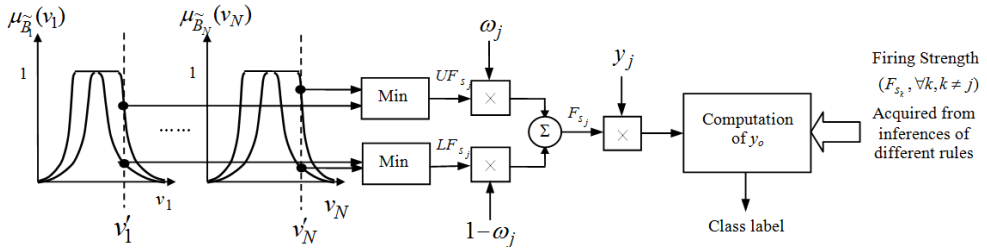


Fig. 4.7. Design of the proposed TSK induced Interval Type-2 fuzzy

all the v_i values. The designed Gaussian MF has its base centered at α while its two extremities are positioned at $\alpha \pm 3\times\sigma$ [51]-[52]. As 10 days has been utilized for the present experimental task, 10 Type-1 Gaussian MFs are generated and subsequently exploited for the formulation of IT2FS as portrayed in Fig. 4.6(a). Finally, an interval type-2 membership function (IT2MF) $\tilde{B}_i$ is formulated by evaluating the union of 10 Type-1 MFs as illustated in Fig. 4.6(b). The UMF and LMF of $\tilde{B}_i$ is designed by exploiting the operations indicated in (4.11)-(4.12).

$$UMF = \overline{\mu}_{\tilde{B}_i}(v_i) = \max(\mu_{\tilde{B}_1}(v_1), \mu_{\tilde{B}_2}(v_2), ..., \mu_{\tilde{B}_N}(v_N)) \tag{4.11}$$

$$LMF = \underline{\mu}_{\tilde{B}_i}(v_i) = \min(\mu_{\tilde{B}_1}(v_1), \mu_{\tilde{B}_2}(v_2), ..., \mu_{\tilde{B}_N}(v_N)) \tag{4.12}$$

However, embedded T1-fuzzy sets contribute significantly to the FOU, particularly in the region around the UMF within it [37]. Thus, to harness this beneficial region of the FOU, a new transformation function is employed upon the LMF to make it closer to the UMF, which in turn decreases the total area of the FOU, as illustrated in Fig. 4.6(c). The mathematical formulation of the new LMF is depicted by (4.13).

$$LMF_{new} = \underline{\mu}_{\tilde{B}_i'}(v_i) = \overline{\mu}_{\tilde{B}_i}(v_i) \times \left(\underline{\mu}_{\tilde{B}_i}(v_i)^{(\underline{\mu}_{\tilde{B}_i}(v_i))} \right) \tag{4.13}$$

In (4.13), the term $\underline{\mu}_{\tilde{B}_i}(v_i)^{(\underline{\mu}_{\tilde{B}_i}(v_i))}$ helps to adjust the FOU by shifting the original LMF towards the UMF. Additionally, the multiplication of $\underline{\mu}_{\tilde{B}_i}(v_i)^{(\underline{\mu}_{\tilde{B}_i}(v_i))}$ by $\overline{\mu}_{\tilde{B}_i}(v_i)$ ensures that $\underline{\mu}_{\tilde{B}_i'}(v_i)$ is always less than or equal to $\overline{\mu}_{\tilde{B}_i}(v_i)$. In other words, the computation denoted in (4.13) effectively adjusts the FOU by exploiting the most significant region near the UMF and also assures that the new LMF remains bounded within the UMF. Thus, the above formulation effeciently aids in minimizing uncertainty among the convolved vectors. additionally, Theorem 1 provides a formal proof that ensures $\underline{\mu}_{\tilde{B}_i'}(v_i) \geq \underline{\mu}_{\tilde{B}_i}(v_i), \forall v_i$, and $\underline{\mu}_{\tilde{B}_i'}(v_i) \leq \overline{\mu}_{\tilde{B}_i}(v_i), \forall v_i$. It is important to mention that the UMF of the newly adjusted FOU remains unchanged i.e., $\overline{\mu}_{\tilde{B}_i'}(v_i) = \overline{\mu}_{\tilde{B}_i}(v_i)$.

Theorem 1: *The transformation in (4.13) elevates* $\overline{\mu}_{\tilde{B}_i}(v_i)$ *so as to satisfy* $\underline{\mu}_{\tilde{B}_i'}(v_i) \geq \underline{\mu}_{\tilde{B}_i}(v_i)$, *maintaining* $\underline{\mu}_{\tilde{B}_i'}(v_i) \leq \overline{\mu}_{\tilde{B}_i}(v_i), \forall v_i$.

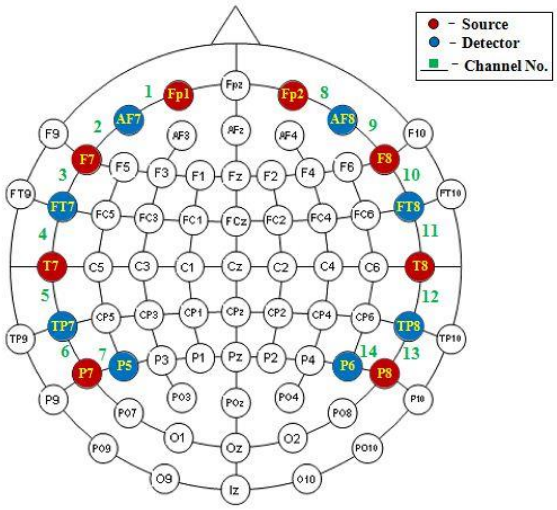


Fig. 4.8. Architecture of the montage designed for the RAPM task

The proof of Theorem 1 is presented in the Appendix to avoid loss in continuity of the context.

While evaluating the UMF for IT2FS, a curvilinear top emerges since the maximum of the 10 Type-1 MFs are considered as depicted in Fig. 4.6(b) and (c). Thus, to retain the convexity of the designed Type-2 fuzzy sets, a flat-top approximation is employed upon the acquired IT2FS. Such an approximation is achieved by connecting the peaks of the individual Type-1 MFs with a straight line of zero slope [35]-[37]. The final flat-top approximated IT2FS is portrayed in Fig. 4.6(d).

4.3.4.2 Formulation of Classifier Rule

The proposed TSK model utilizes Type-2 fuzzy rule where the j^{th} rule is given by,

If v_1 is $\tilde{B}'_{1,j}$, v_2 is $\tilde{B}'_{2,j}$, …, v_N is $\tilde{B}'_{N,j}$, Then $y_j = \sum_{i=1}^{N} \varphi_{i,j} \times v_i + \lambda_j$. Here, y_j represents the power of the signal pertaining to a given montage architecture which is utilized to categorize different class labels. The coefficients $\varphi_{i,j}$ and λ_j are determined utilizing the classical least min-square approach [53].

4.3.4.3 Implementation of TSK-based IT2FS Classifier

The design of the proposed fuzzy induced classifier is illustrated in Fig. 4.7. Let the measurements points be denoted as $v_1 = v'_1, v_2 = v'_2, ..., v_N = v'_N$. The upper firing strength (UF_s) and lower firing strength (LF_s) for the j^{th} rule is computed utilizing (4.14) and (4.15) respectively.

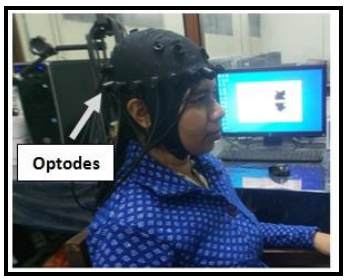


Fig. 4.9. Experimental setup with a participant who has volunteered for the RAPM task

$$UF_{s_j} = \min[\bar{\mu}_{\tilde{B}'_1}(v'_1), \bar{\mu}_{\tilde{B}'_2}(v'_2), ..., \bar{\mu}_{\tilde{B}'_N}(f'_N)] \tag{4.14}$$

$$LF_{s_j} = \min[\underline{\mu}_{\tilde{B}'_1}(v'_1), \underline{\mu}_{\tilde{B}'_2}(v'_2), ..., \underline{\mu}_{\tilde{B}'_N}(f'_N)] \tag{4.15}$$

After this, the firing strength (F_s) pertaining to rule *j* is computed using (4.16).

$$F_{s_j} = \omega_j \cdot UF_{s_j} + (1-\omega_j) \cdot LF_{s_j} \tag{4.16}$$

where, ω_j denotes a weight lying in [0,1] and is determined by the random search algorithm. The final output of the TSK induced type-2 fuzzy classifier is evaluated using (4.17).

$$y_o = \frac{\sum_{\forall j} F_{s_j} \times y_j}{\sum_{\forall j} F_{s_j}} \tag{4.17}$$

In order to categorize the five desired class labels from the computation of y_o, the interval $[0, y_o^{\max}]$ is divided into five non-overlapping regions. So, for five regions, four different region boundaries are required which include η_1, η_2, η_3 and η_4. The aforementioned region boundaries are determined by the random search algorithm satisfying: $y_o^{\max} > \eta_4 > \eta_3 > \eta_2 > \eta_1 > 0.$

4.4 EXPERIMENTS AND RESULTS

4.4.1 fNIRS Data Acquisition

The current experimental framework was carried out in the Artificial Intelligence Laboratory (AI Lab) of Jadavpur University situated in Kolkata, India. The hemodynamic response of participants during the RAPM task was acquired using the whole-brain fNIRS device (NIRScout TM) developed by NIRx Medical Technologies, LLC. This device comprises 8 infrared sources and 8 detectors and operates at 7.81 Hz sampling rate. The source-detector pairs referred to as optodes were places on the scalp of participants according to the 10-10 international optode placement system. The montage utilized for the present cognitive experiment starts from the pre-frontal lobe and ends at the parietal lobe as illustrated in Fig. 4.8. The choice of the aforementioned montage architecture is based on the recognition that

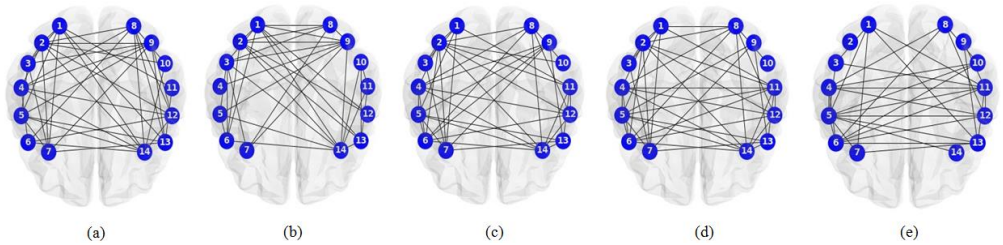


Fig. 4.10. The brain connectivity networks obtained for each class of subject: (a) subject ID: I27 who could correctly solve the RAPM task of very high difficulty level, (b) subject ID: I11 who could correctly solve the RAPM task of high difficulty level, (c) subject ID: I09 who could correctly solve the RAPM task of medium difficulty level, (d) subject ID: I32 who could correctly solve the RAPM task of easy level, and (e) subject ID: I05 who could not solve the RAPM task of any difficulty level

analogical reasoning tasks actively engage the BAs encompassed by the current montage [17], [25], [54]. The optodes generated a total of 64 channels, from which 14 were selected based on the nearest-neighbor source-detector pairing scheme.

4.4.2 Participants

34 participants (engineering students) within the age group of 17-35 years volunteered for the analogical reasoning based cognitive experiment. The total number of female and male participants was 18 and 16 respectively. All participants had normal or corrected-to-normal vision and no prior history of neuropsychiatric or motor disorders. A participant of the present experiment with the fNIRS setup is depicted in Fig. 4.9. The experiment strictly adhered to ethical guidelines and safety protocols, following the Helsinki Declaration of 1970, as revised in 2004 [55].

4.4.3 Cognitive Insights into Brain Network Feature Analysis

The brain connectivity networks of participants belonging to each class label have been acquired using the WTC technique as discussed in Section 4.4. The different connectivity networks pertaining to the five class labels obtained from a single experimental instance are depicted in Fig. 4.10. The nodes in each network represent the BAs covered by the channels in the montage shown in Fig. 4.8, while the edges signify the connections among the BAs. However, in order to identify the active engagement of the BAs for the current cognitive task, three node based features are abstracted form the acquired brain networks. The analysis of these features is presented in detail below.

(a) **NS Analysis**: Fig. 4.11 (a) illustrates the radar plot of NS values for five classes of subjects. It is observed from this figure that NS values for nodes 1, 8 and 14 are high for participant I27 who belongs to the ECT class. Moreover, the same nodes

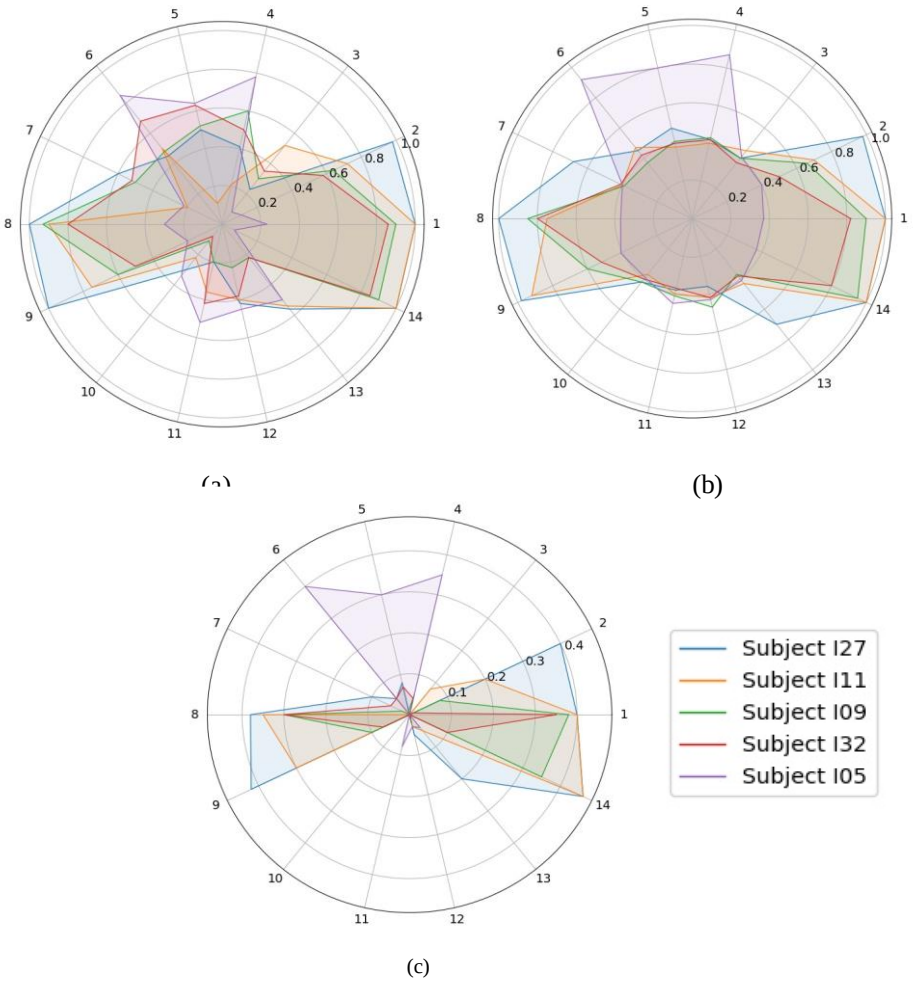


Fig. 4.11. The radar plots for analyzing node based brain connectivity features: (a) radar plot for node strength pertaining to 5 classes of subjects, (b) radar plot for node efficiency pertaining to 5 classes of subjects and (c) radar plot for node betweenness pertaining to 5 classes of subjects

exhibit quite high NS values for participants I11, I09 and I32 belonging to SCT, FCT and MCT classes respectively. Such an observation signifies that bilateral anterior prefrontal cortex (BA 10) pertaining to nodes 1 and 8 and right posterior supramarginal gyrus (BA 40) pertaining to node 14 possess high degree for participants who could correctly solve the RAPM task. Additionally, the NS values are also high for nodes 2 and 9 for the participant belonging to ECT class. However, these values start declining slowly for the participants from the SCT to MCT class. The former observation implies that the bilateral region of dorsolateral prefrontal cortex (BA 46) pertaining to nodes 2 and 9 also possess high degree for the RAPM task. The later observation indicates that BA 46 controls the level of difficulty of the RAPM task. In other words, higher the NS value for nodes pertaining to BA 46, higher is the difficulty level of the given task. It is also important to note that the NS values of node 4 for subject I09 and node 6 for subject I32 exhibit moderately high values. Thus, a further analysis of additional node based features is necessary to

derive concrete conclusions about the most engaged nodes in the brain network of creative individuals.

(b) NE Analysis: Fig. 4.11 (b) illustrates the radar plot of NE values for the same set of subjects discussed earlier. Notably, for a subject in the ECT class, elevated NE values are evident over the bilateral part of anterior prefrontal cortex (nodes 1 and 8) and right posterior supramarginal gyrus (node 14). Similarly, the NE values for subjects in the SCT, FCT and MCT classes also show heightened NE values for the aforesaid regions. Additionally, the bilateral part of dorsolateral prefrontal cortex (nodes 2 and 9) shows high NE values, with a clear decreasing trend from the ECT to MCT classes. Furthermore, the NE analysis indicates that the highlighted BAs for creative class individuals do not exhibit significant activation in a person who could not solve the RAPM task (i.e., belonging to the CVT class). These findings suggest that the bilateral part of anterior prefrontal cortex and right posterior supramarginal gyrus helps in integrating neural activity across different brain areas via shortest paths to ensure effective communication with minimal delay [56]. Moreover, the findings related to the dorsolateral prefrontal cortex suggest that its engagement increases with increase in difficulty level of the task.

(c) NB Analysis: The results of the NB values as depicted by the radar plot in Fig. 4.11 (c) portray the same BAs for each creative class label as indicated by the NE based evaluation. The high NB values for the specified BAs signify that they serve as key intermediaries in the flow of information between other brain lobes. In other words, high NB nodes act as bridges or hubs that connect different brain regions, ensuring efficient communication across the entire network [28].

Hence, the node based feature analysis discussed above supports the proposition that, for analogical reasoning, the BAs that act as prime controllers of the entire brain network include the bilateral anterior prefrontal cortex (BA 10 corresponding to nodes 1 and 8) and right posterior supramarginal gyrus (BA 40 corresponding to node 14). The observed findings are in agreement with prior research in [17], [18], [25], [54]. Additionally, the analysis reveals that the engagement of the bilateral region of dorsolateral prefrontal cortex (BA 46, corresponding to nodes 2 and 9) varies with the difficulty level of the problem, with higher activation indicating greater difficulty and vice versa [57]-[58].

4.4.4 Optimization of Hyper-parameters of Classifier

Hyper-parameter optimization of the proposed EGCIFC model involves fine-tuning its parameters to ensure optimal performance in categorizing the target class labels. In the present context, the optimal settings for the hyper-parameters were determined

using the random search algorithm [50]. The rationale for choosing this algorithm lies in its efficiency in exploring the search space with low computational time [50], [59]. The model's performance was assessed using the 10-fold cross-validation across different parameter configurations. In this method, the dataset for each subject was divided into 10 separate folds. For each parameter configuration, the model was trained using 9 folds and tested on the remaining fold. This process was repeated over 10 runs, with the test fold changing each time. As a result, the training and testing sets for each subject contained data from the same individual but remained distinct in each run. Once the highest accuracy was achieved across all candidate settings, the best configuration was applied to the test set. The first hyper-parameter of EGCIFC involves the truncation of exp(A) till the third order, ensuring an optimal balance between computational efficiency and classification accuracy. The rest of the optimal hyper-parameter values of the model are: $Q = 2,\ L = 64,\ R = 32,\ \eta_1 = 18.43,\ \eta_2 = 36.21,\ \eta_3 = 54.45,$ and $\eta_4 = 82.70.$

4.5 PERFORMANCE ANALYSIS AND STATISTICAL VALIDATION OF THE PROPOSED MODEL

4.5.1 Performance Analysis of the Proposed Classifier with Respect to SOTA Methods

Three key performance metrics have been utilized to prove the efficacy of the proposed EGCIFC algorithm in comparison to conventional approaches. These metrics include classification accuracy (CA), F1-score and run-time complexity. The mathematical expressions for CA and F1-score are denoted by (4.18) and (4.19) respectively.

$$CA = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{4.18}$$

$$F1 = \frac{2T_P}{2T_P + F_P + F_N} \tag{4.19}$$

where, T_P, T_N, F_P, F_N denotes the number of true positives, true negatives, false positives and false negatives respectively.

Table 4.1 demonstrates the performance of the proposed model with respect to the SOTA techniques utilizing the above stated performance metrics. The results in Table 4.1 clearly portray that the performance of the proposed classifier surpasses all the SOTA techniques by a substantial margin with respect to CA and F1-score. Additionally, the results also demonstrate the run-time complexity of the proposed classifier is remarkably low at 90.27 ms, indicating a significant advantage over most of the competing methods.

Table 4.1
Comparative study of Proposed classifier performance with respect to the SOTA methods

Classifiers with optimal parameter settings	CA (%)	F1 (%)	Run-time complexity (ms)
Chebnet [60]	77.36	77.84	155.45
GCN [30]	81.92	82.23	85.78
DGCN [61]	84.05	83.87	112.30
ASGCNN [62]	88.70	89.34	362.56
AEGCN [63]	92.93	93.22	234.43
CNN-AE + IT2FR-GWO [64]	95.02	94.98	127.88
GC-IT2FN [37]	96.56	96.87	98.92
Proposed EGCIFC	**97.78**	**98.03**	**90.27**

Table 4.2
Ablation study of the Proposed Classifier Model

Variation in Classifier Modules	CA (%)	F1 (%)
IT2FS	87.88	88.41
GCN with pooling	88.70	88.56
GCN with 1D spatial Convolution	89.74	90.02
EGCN with pooling	90.23	90.34
EGCN with 1D spatial convolution	92.45	92.87
Graph convolution + pooling + IT2FS	94.83	94.68
Graph convolution + 1D spatial convolution + IT2FS	95.14	95.36
Enhanced graph convolution + pooling + IT2FS	96.05	96.47
Proposed EGCIGC	**97.78**	**98.03**

4.5.2 Ablation Study of the Proposed Model

An extensive ablation study has been performed on the proposed EGCIFC model to analyze the contribution of its individual components on overall classification performance. Such an analysis is undertaken by systematically removing specific modules from the original architecture of the EGCIFC and then examining the corresponding effects on the classification results. Table 4.2 demonstrates the results of the aforesaid study.

It is observed from Table 4.2 that moderate CA and F1-score values are obtained when IT2FS classifier is utilized independently as it can tackle uncertainty but cannot effectively deal with graph based data. Again, similar trend in CA and F1-score is obtained when traditional GCN with pooling layer (here, mean-pooling) is utilized due to its ability to handle graphical data but inability to address uncertainty. A minor enhancement in results is noted when traditional GCN utilizes the 1D spatial convolution layer instead of pooling. Additionally, there is increase in CA and F1-score when the enhanced graph convolution is used along with 1D spatial convolution layer.

Table 4.3
Effect of Rectifying the FOU on classifier performance

Class Labels	Modification in Classifier Architecture			
	EGCN + TSK based IT2FS with original FOU		EGCN + TSK based IT2FS with rectified FOU	
	CA (%)	F1 (%)	CA (%)	F1 (%)
ECT	93.52	93.86	**96.70**	**97.39**
SCT	93.20	93.65	**97.45**	**97.98**
FCT	94.07	93.94	**98.36**	**98.25**
MCT	94.25	94.51	**98.11**	**97.87**
CVT	94.43	94.70	**98.30**	**98.68**

Table 4.4
Impact of Various Computations of brain Connectivity networks with respect to F1-score

Algorithm for constructing brain networks	Class Labels				
	ECT	SCT	FCT	MCT	CVT
PLV	92.23	91.85	92.11	92.41	91.97
PLI	92.30	92.02	92.84	91.90	92.05
MI	94.66	95.27	94.80	95.33	95.56
PC	95.41	96.09	95.43	95.25	96.03
WTC	**97.39**	**97.98**	**98.25**	**97.87**	**98.68**

However, none of the above classifier architectures yield the best results as they either lack the ability to handle uncertainty or effectively process graphical data. Hence, notable improvements in classifier performance are observed when different variations of GCN is utilized with IT2FS, with the highest being the architecture of the proposed approach (i.e., enhanced graph convolution + 1D spatial convolution + IT2FS). Therefore, it can be inferred that classifier performance deteriorates whenever any one of the modules is removed, highlighting the contribution of each module within the proposed approach.

4.5.3 Impact of Modification in FOU on Classifier Performance

The impact of FOU reduction using the proposed LMF formulation (as discussed in Section 4.3) is compared against the traditional FOU design, which is derived by taking the union of 10 Type-1 MFs obtained over 10 experimental days. The results of the aforesaid comparison are portrayed in Table 4.3. It is apparent from this table that the proposed FOU reduction technique leads to a significant improvement in both CA and F1-score values compared to the conventional method.

4.5.4 Comparative Study of Different Brain Network Formation

A comparative study with respect to F1-score is undertaken to examine the impact of the proposed model's performance by utilizing different formulations of brain connectivity networks. The aforesaid comparison involves several widely used methods for designing brain networks, such as Phase Locking Value (PLV) [65],

TABLE 4.5
Impact of Modifying activation functions on the proposed model with respect to CA

Activation Functions	Class Labels				
	ECT	SCT	FCT	MCT	CVT
ReLU [69]	83.22	83.01	82.76	82.95	83.11
LReLU [70]	84.40	84.25	85.06	83.88	83.92
ELU [71]	86.32	86.48	86.27	87.09	87.74
SELU [72]	88.53	89.06	88.76	89.26	88.81
SiLU [73]	90.12	90.59	91.60	91.51	91.77
ELiSH [74]	93.51	93.18	92.47	93.63	93.00
Mish [75]	94.20	94.34	94.02	94.77	94.22
Logish [76]	95.03	96.52	96.89	95.07	96.95
Smish	**96.70**	**97.45**	**98.36**	**98.11**	**98.30**

Table 4.6
Wilcoxon's Signed Ranked Test to Validate the Proposed Model

Traditional Classifiers with optimal parameter settings	Reference Algorithm: Proposed EGCIFC
Chebnet	+
GCN	+
DGCN	+
ASGCNN	+
AEGCN	+
CNN-AE + IT2FR-GWO	+
GC-IT2FN	+

Phase Lag Index (PLI) [66], Mutual Information (MI) [67] and Pearson's Correlation (PC) [68]. The results of the current comparative study, as shown in Table 4.4, highlight a significant improvement in the performance of the proposed classifier with WTC based network construction in comparison to the traditional formulations.

4.5.6 Impact of Different Activation Functions on Classifier Performance

The current comparative analysis assesses the influence of different activation functions on the performance of the proposed model with respect to CA. Table 4.5 depicts the results of the above analysis. It is observed from this table that the performance of the proposed classifier is maximized when the Smish activation function is utilized. This result suggests that the Smish activation function plays a pivotal role in enhancing the classifier's accuracy, thus outperforming other activation functions used in the comparative study.

4.5.7 Statistical Validation Using the Wilcoxon's Signed Rank Test

The statistical validation of the proposed EGCIFC model has been carried out using the popular Wilcoxon's Signed rank Test [77]. Let H_o be the null hypothesis that denotes the similar performance of an algorithm *S* with respect to a reference algorithm *T*. In the present context, *T* represents the proposed algorithm while *S* signifies any one of the seven algorithms depicted in Table 4.1. In order to validate the null hypothesis, the test statistic is evaluating using (20).

$$W = \sum_{i=1}^{l} [\text{sgn}(\varepsilon_i^T - \varepsilon_i^S) \cdot r_i] \tag{20}$$

where, ε_i^T and ε_i^S denote the classification error rate ε acquired during the i^{th} training instance of algorithms *T* and *S* respectively. *l* signifies the total experimental instances while r_i depicts the rank of the pair of algorithms for i^{th} instance.

Table 4.6 illustrates the results of the above statistical test. Here, the plus and/or minus signs of the second column signify the level of significance of the *W* values regarding the comparison of Algorithm *T* with *S*. For the present context, a confidence level of 95% is accomplished along with the degree of freedom=1 and *p*-value > 0.05.

4.6 CONCLUSION

Understanding the cognitive underpinnings of scientific creative skill in individuals is still an open-ended problem in neuroscience. The present study offers a valuable contribution to the above problem by leveraging fNIRS data to assess the levels of creative ability in subjects through brain connectivity analysis. The aforementioned analysis conducted using three node based connectivity features (strength, efficiency and betweeness) infers the active participation of bilateral anterior prefrontal cortex (BA 10) and right posterior supramarginal gyrus (BA 40) while solving the RAPM task. Additionally, the bilateral regions of dorsolateral prefrontal cortex (BA 46) become highly engaged whenever the difficulty level of the problems rises. Furthermore, the node based features are classified by a novel EGCIFC model that can efficiently identify different levels of subjective creative potential in comparison to SOTA methods.

Thus, the proposed scheme has the potential to be utilized as a recruitment strategy for placing individuals in various industrial departments based on their respective creative ability ranks. For instance, an extremely creative thinker could be appointed as a research and development lead, a highly creative thinker as an innovation manager, a medium-level creative thinker as a product design specialist, while a low-level creative thinker as a quality assurance analyst. This strategic placement would ensure optimal utilization of each individual's creative potential, fostering innovation and efficiency within industrial sectors. Moreover, the present

study can open future avenues for developing computational models of scientific creativity inspired by biological processes.

4.7 APPENDIX

A.1. PROOF OF THEOREM 1

This section provides the detailed proof of Theorem 1 provided under Section 4.2.3 of the manuscript.

Proof of Theorem 1. For notational simplicity,

Let $a_i = \bar{\mu}_{\tilde{B}_i}(v_i)$, $b_i = \underline{\mu}_{\tilde{B}_i}(v_i)$ and $b_i' = \underline{\mu}_{\tilde{B}_i'}(v_i)$.

So, with modified notations we need to prove:

$$b_i' \geq b_i \text{ and } b_i' \leq a_i$$

To undertake the proof of $b_i' \geq b_i$, we define $J = b_i' - b_i$ and show that $J \geq 0$, for all v_i. Now, to prove $b_i' \leq a_i$, we simply show that: $b_i' = a_i \times (b_i)^{b_i} \leq a_i$ as $(b_i)^{b_i} \leq 1$ for any number b_i in [0,1]. Now, we formally prove $J = b_i' - b_i \geq 0$, for all v_i.

$$J = b_i' - b_i$$

$$\Rightarrow J = a_i b_i^{b_i} - b_i \tag{A.1}$$

Equation (A.1) will be sufficient if $J_{\min} \geq 0$. Let us minimize *J* with respect to b_i.

So, $\frac{\partial J}{\partial b_i} = 0$ to find the optimal value of *J* as a function of b_i.

$$\frac{\partial J}{\partial b_i} = a_i b_i^{b_i}(1 + ln(b_i)) - 1 = 0$$

$$\Rightarrow b_i^{b_i}(1 + ln(b_i)) = a_i^{-1} \tag{A.2}$$

Taking *ln* on both sides of (A.2), we get,

$$\Rightarrow ln(b_i^{b_i}(1 + ln(b_i))) = ln(a_i^{-1})$$

$$\Rightarrow ln(b_i^{b_i}) + ln(1 + ln(b_i)) = ln(a_i^{-1})$$

$$\Rightarrow b_i\, ln(b_i) + ln(1 + ln(b_i)) = ln(a_i^{-1}) \tag{A.3}$$

Let $b_i = e^u$ then (A.3) becomes,

$$u \cdot e^u + ln(1 + u) = ln(a_i^{-1}) \tag{A.4}$$

Now, $ln(1+u) \approx u$ using Maclaurin series expansion and keeping only first order term. Again, $e^u \approx 1 + u$ using Maclaurin series expansion and keeping only first order term. Substituting the above values in (A.4) we get,

$$
\begin{aligned}
& u(1+u)+u = ln(a_i^{-1}) \\
\Rightarrow & u^2 + 2u = ln(a_i^{-1}) \\
\Rightarrow & u^2 + 2 \cdot u \cdot 1 + 1 = 1 + ln(a_i^{-1}) \\
\Rightarrow & (u+1)^2 = 1 + ln(a_i^{-1}) \\
\Rightarrow & u = \sqrt{1 + ln(a_i^{-1})} - 1
\end{aligned} \tag{A.5}
$$

Substituting the value of b_i in (A.5) we get,

$$b_i^* = e^{\left(\sqrt{1+ln(a_i^{-1})}-1\right)} \tag{A.6}$$

The value of $b_i = b_i^*$ in (A.6) represent the optimal value of J.

Now, we need to check whether $\left.\frac{\partial^2 J}{\partial b_i^2}\right|_{b_i = b_i^*} > 0.$

$$\left.\frac{\partial^2 J}{\partial b_i^2}\right|_{b_i=b_i^*} = a_i\left(b_i^{*b_i^*-1} + b_i^{*b_i^*}(1+ln(b_i^*))^2\right) \tag{A.7}$$

It is apparent from (A.7) that $\left.\frac{\partial^2 J}{\partial b_i^2}\right|_{b_i=b_i^*} > 0$ since $0 < b_i \le a_i \le 1, \forall i,$ according to the definition of LMF and UMF in type-2 fuzzy sets. Thus, $b_i = b_i^*$ indicates a minima. Now we need to test $J_{\min} \ge 0.$

$$J_{\min} = a_i b_i^{*b_i^*} - b_i^* \tag{A.8}$$

Solving (A.8) we get $J_{\min} \ge 0$.

$$\therefore \underline{\mu}_{\tilde{B}_i'}(v_i) \ge \underline{\mu}_{\tilde{B}_i}(v_i), \text{ for all } v_i.$$

□

Part II

Computational Modalities of Scientific Creativity

5

Computational Creativity by Diversity-Optimized Intelligent Search: An Automatic Approach to Artificial Synthesis of Trigonometric Identities

The chapter emphasizes an interesting approach to synthesize computational creativity by a process similar to deductive reasoning with a provision for testing the degree of diversity of the generated instances compared to their predecessors. The above two-step process of expansion and testing is developed here using the best-first search (BFS) on an OR-tree, where the nodes denote trial solutions (new creations) and edges represent parent-child connectivity satisfying the rules of the given problem domain. Two alternative extensions of BFS are examined in view of the cost function employed at the nodes to ultimately determine the optimal node in the search tree within a user-defined depth as the solution to the creativity problem. The first algorithm considers maximizing the diversity cost earned by a node with respect to its parent, while the second considers maximizing the difference between the diversity and the penalty cost earned by a node with respect to the root node. The significant contribution of the present research lies in ensuring diversity of the solutions during iterative expansions of the tree as well as the novelty of the optimal solution (best node) across runs of the same program. The relative performances of the two algorithms are compared in the context of

their applicability. Performance analysis undertaken reveals that the proposed algorithms outperform their competitors with respect to three important metrics. The proposed algorithms have successfully been employed in developing chapter-end exercises for trigonometric identity proving problems.

5.1 INTRODUCTION

Creativity [1]-[2] is widely used in aesthetics and fine arts to refer to the original and innovative contributions of artists in i) visual arts, like painting and sculpture-making, ii) literary arts, such as poetry and drama, and iii) performing arts, including music, dance, and acting. Unfortunately, creativity evades a formal and concise definition in the context of scientific domains till this date. According to the literature [3], the degree of creativity of an item is often judged by two attributes: i) novelty/originality and ii) diversity of the item. While novelty of a creative outcome is assessed by the inherent originality of the creation with respect to its past competitors, diversity refers to the intrinsic differences among the creative competitors of the same generation [3]. Computational Creativity (CC) [4]-[6] generally refers to the automatic derivation of the new pieces of knowledge of significant diversity with respect to available knowledge of the same domain. Automatic generation of computationally creative outcomes thus requires an engine for expansion of knowledge, as done in deductive reasoning [7]-[8], and a metric to measure the diversity of the new outcomes with respect to existing ones. To keep track of the resulting outcomes, a data structure, such as a tree/set/list is needed, along with a control policy to select the more promising outcomes for expansion by rules/knowledge until the outcomes achieve the required level of diversity.

This chapter makes a humble attempt to develop interesting algorithms for artificial synthesis of mathematical identities. Here, a set of primitive identities or substitution formulas (SFs) is submitted as the input to the algorithm. The motivation is to design new identities of desired complexity and novelty from the existing ones. A cost function f is designed to assess the diversity of the new results with respect to existing ones. A penalty cost is also defined to assess the number of steps involved to reach an outcome from a given initial axiom/identity. A tree structure is employed here to keep track of the resulting outcomes in the expansion process. The nodes in the tree represent identities and the directed edges describe parent-child relationship in the derivation process of identities. Here, an OR-tree structure is presumed to keep provisions for alternative expansions from a given node. The well-known Best-First search (BFS) algorithm [7] is invoked to control the expansion process in the OR-tree in order to select the most promising outcomes for expansion. The selection of the more promising outcomes is performed by assessing a measure of diversity (or

diversity minus penalty) hereafter called f-cost of all OPEN set of nodes (nodes without children), and selecting the one with the highest f-cost. The selected node is expanded using suitable rules/knowledge to generate offspring. The children are added to the OPEN set of nodes and the expanded parent node is dropped out from the OPEN set and added to the CLOSED set of nodes. The expansion of the OR-tree is continued up to a user-defined depth, and the most promising resulting node is presented as the desired identity. Thus the f-cost is optimized at one of the OPEN set of nodes at the user-defined depth. It is noteworthy that diversity of the best node is assessed here among the selected nodes across iterations in the same run, while novelty of the best node is judged across different runs of the same algorithm.

The present proposal includes two alternative algorithms, called Relative Diversity Guided Best-First Search (RDGBFS) and Absolute Diversity Guided Best-First Search (ADGBFS). The control policy of expansion in both the algorithms is similar to traditional BFS, excluding the termination condition. In traditional BFS, the algorithm terminates once a goal node is identified. Here, in absence of any goal node, the algorithms are terminated at a fixed user-defined depth. The difference between RDGBFS and ADGBFS lies in the design of the cost function *f*. In RDGBFS, the cost function measures the relative diversity of a node with respect to its parent. In ADGBFS, the cost function measures the difference of two parameters: diversity cost of the current node with respect to the root node and the penalty cost to reach the current node from the root. A thorough study of the two algorithms is undertaken in the chapter to identify the commonality and differences of the algorithms. The comparative study reveals that properties 1 to 6 hold for both the algorithms. Properties 7-11 hold for ADGBFS, and properties 12-14 examine the relative performance of the two algorithms under specific context. A thorough study of robustness, encompassing soundness [30] and logical completeness [7], of both the algorithms is examined in Properties 15 and 16 respectively.

Experiments undertaken to synthesize trigonometric identities reveal that both the algorithms generate novel optimal solutions across different runs of the same algorithm at a user-defined depth and fixed branching factor. Secondly, the more is the depth, the higher becomes the complexity and diversity of the optimal solution.

Performance of the proposed algorithms is examined with three comparative frameworks of algorithms: i) search and template based methods [9]-[16], ii) generative neural network methods [17], [19]-[23], and iii) large language models (LLMs) [24-26]. Metrics involved to compare the performance of a pair of techniques include i) relative diversity of the best node in the OR-tree averaged across 100 generated problems, hereafter called Average-Relative-Diversity (ARD), ii) absolute diversity of the Best node divided by depth of the best node, averaged

across 100 generated problems, called Average-Absolute-Diversity per unit depth (AAD), and iii) Run-time complexity during training and testing, averaged over 100 problem instances. It is noteworthy that for the proposed algorithms, the training time is zero. Relative performance analysis reveals that the proposed algorithms outperform their competitors with respect to the solutions generated by all the three benchmark frameworks. Lastly, the textbook problems on trigonometric identities are synthesized using the proposed approach by back-tracking on the search tree to demonstrate that better candidate solutions with higher diversity appears in the tree at the same depth of the text-book identity. The back-tracking, in fact, is needed as well to compare the results produced by all the three classes of algorithms mentioned above.

The original contribution of the chapter includes i) design of the two RDGBFS and ADGBFS algorithms to synthesize CC, ii) a thorough analysis of the algorithms, and iii) a detailed evaluation of the performance of the proposed algorithms.

The chapter is organized as follows. Section 5.2 reviews related works, Section 5.3 introduces the proposed algorithms along with pre-requisites for understanding the proposed algorithms, and Section 5.4 introduces the key properties and their implications of the proposed algorithms. Section 5.5 discusses robustness and scalability issues, Section 5.6 presents experimental results, and Section 5.7 compares the algorithms to traditional methods. Finally, Section 5.8 summarizes the conclusion.

5.2 AN OVERVIEW OF RELATED WORKS

Several works on CC involving automatic problem generation in mathematics/scientific domain are reported in the literature [9]-[16]. For instance, Pearce et al. [9] designed a novel strategy to automatically develop trigonometric identities and their solutions using a rule-based substitution technique. Liu et al. [10] generated trigonometric identities utilizing a randomized Breadth-First Search (rBFS) to efficiently create diverse identities while managing computational complexity through shorter transformation paths. Although the identities obtained from [9]-[10] had diverse structural forms, repeated runs of the same program generated the same solution set. In study [11], Papasalouros attempted to automatically develop theorems in Euclidean geometry by generating a depth-limited search tree, with nodes representing literals such as tautologies and predicates, and edges representing the connections between parent and child nodes. In a similar vein, Tabuguia et al. [12] utilized a search-based approach to generate trigonometric identities by leveraging hypergeometric-type sequences and symbolic computation. However, the approaches in [11]-[12] do not ensure novelty and/or diversity among the generated problems. In another study by Briggs et al. [13], trigonometric

identities are automatically generated using e-graphs and rewrite rules. Although this method effectively preserves diversity and novelty in the generated identities, it is computationally intensive, as it relies on integer linear programming and additional e-graph passes to eliminate duplicate identities. Singh et al. [14], Brito et al. [15] and Zhang et al. [16] independently employed a template-based approach to construct new mathematical problems of similar structure where problem-set generated is confined by templates of the problems.

The automatic generation of trigonometric identities can also be performed using popular generative neural networks. For instance, in a very recent work, Ghosh et al. [17] explored CC by generating trigonometric identity problems using the LeakGAN model. In their approach, the discriminator learns intricate patterns within the input dataset through attention, convolution, and pooling mechanisms, which it then leaks to the generator. The generator uses this leaked information to produce new identity problems through a hierarchical interaction of worker and manager modules [18]. It is also important to mention that the work [17] utilized traditional generative neural models as state-of-the-art techniques for relative performance analysis, including RNN [19], LSTM [20], VAE [21], SeqGAN [22], and RankGAN [23]. This clearly demonstrates their capability for identity generation, a task that had not traditionally been associated with these models.

Recent advancements in LLMs have driven research into automatically generating mathematics problems. For instance, Pham et al. [24] assessed ChatGPT’s ability to create pre-university math questions of varying difficulty. Similarly, Deroy et al. [25] examined Gemini's effectiveness in generating problems across domains like arithmetic, geometry, and abstract algebra. Yu et al. [26] introduced a fine-tuning method for Llama-2 using forward and backward reasoning with answer augmentation to produce diverse math problems. However, a significant drawback of the above generative and LLM models is their reliance on large datasets for extensive training, which makes them computationally expensive and resource-intensive.

The present work addresses the limitations of the three classes of algorithms mentioned above for synthesizing identities, which includes search- and template-based approaches, generative neural networks, and LLMs. It introduces two modified BFS algorithms that ensures diversity and novelty in generated identities as well as improves computational efficiency by eliminating the need for extensive training on large datasets.

5.3 PROPOSED VARIANTS OF THE BFS ALGORITHM: RDGBFS AND ADGBFS

The classical BFS algorithm [7] measures a heuristic cost function of a selected node (problem state) in the search tree, representing the approximate estimate of the cost to reach the goal node *G* with respect to the selected node. The objective in the BFS algorithm is to continue expansion of the search tree until the Goal node *G* is generated. The optimal path leading to the Goal node from the start node is important.

This chapter proposes variants of the BFS algorithms for automatic synthesis of trigonometric identities using a given set of SFs or rules and a fixed start-up identity 1=1. The nodes in the search tree here represent the identities. The proposed technique is similar to traditional BFS from the points of view on node selection in an OR-tree based on the best performance of a node among the OPEN set of nodes with respect to a user-defined metric, and also expansion of the selected node by SFs. Instead of using a heuristic cost function, a relative (absolute) measure of diversity cost of a node with respect to its parent (root) node is considered. Similar definitions of diversity in literature on automatic question generation [39] and image generation [40]-[41] justify its use as a metric in the present context. Additionally, a penalty cost is also defined to describe the expenses already made to reach the current node from the root node.

Two distinct algorithms of extended BFS are proposed here. In the first algorithm, called RDGBFS, the node with maximum relative diversity among the OPEN set of nodes is selected for expansion. In the second algorithm, called ADGBFS, the node with maximum absolute diversity minus penalty cost among the OPEN set of nodes is selected for expansion. The expansion of a node is here performed using a set of SFs satisfying certain constraints. After a node is expanded it is discarded from the OPEN set, declared as CLOSED and its children are added to the OPEN set of nodes.

5.3.1 Pre-Requisites

The following definitions are presented to understand the proposed algorithms.

Definition 1: A ***node*** represents a problem state, here a trigonometric identity, that holds for all feasible values of the trigonometric angles used therein. The left/right side of a node corresponds to an expression.

Table 5.1
List of rules used for Trigonometric Identity

T=set of atomic terms={1, $sin^m x$, $cos^m x$, $cosec^m x$, $sec^m x$, $cot^m x$, $tan^m x$, | $m \geq 2$}, *E*=set of operators={+ ,- ,* ,/}, *F*=set of additive operators={+,-}, *H*=functions of atomic terms of arity-2 or higher involving operators from set *E*.

Replacement Rules

R.1: $a \pm \frac{b}{c} \rightarrow \frac{ac \pm b}{c}$

R.2: $\frac{a}{b} \pm \frac{c}{d} \rightarrow \frac{a(v/b) \pm c(v/d)}{bd}$, where v =LCM of b and d

Sl. No.	Rules
1.	$\sin^2 x + \cos^2 x = 1$
2.	$\cos ec^2 x - \cot^2 x = 1$
3.	$\sec^2 x - \tan^2 x = 1$
4.	$\cos ec\, x = 1/\sin x$
5.	$\sec x = 1/\cos x$
6.	$\cot x = \cos x/\sin x$
5.	$\tan x = \sin x/\cos x$
8.	$1/\cot x = \tan x$
9.	For *A,B*∈*T* or *H*, If *A*=*B*, Then $A \otimes C = B \otimes C$, where $C \in T$ and $\otimes \in E$ are selected randomly.
10.	For *A,B*∈*T* or *H*, If *A*=*B*, Then $A \otimes C = B \otimes (1/D)$, where *C*, $D \in T$ and $\otimes \in E$ are chosen randomly, satisfying *C*=1/*D* of the form given in Rules 4,5 and 8.
11.	For *A,B*∈*T* or *H*, If *A*=*B*, then $A \oplus C' = B \oplus (1/D')$, where $C', D' \in H$ and $\oplus \in F$ are selected randomly satisfying $C'=1/D'$, and the result must be expressed in simplified form using the following re-write rules R1 and R2. If the rule's left side has more than two terms, apply it recursively to two terms at a time until exhausted.
12.	For *A,B*∈*T* or *H*, If *A*=*B*, Then $A^2=B^2$
13.	For *A,B*∈*T* or *H*, If *A*=*B*, Then $\sqrt{A} = \sqrt{B}$

Example 1: A node *n*, for instance, represents the trigonometric identity: $\sin x = \frac{1}{\cos ec\, x}$ for all feasible values of angle *x*. Here the left side of node *n* corresponds to an expression: $\sin x$, while right side of node *n* includes the expression: $\frac{1}{\cos ec\, x}$.

Definition 2: A **term** refers to an element of a set, constructed by dropping operators from a given algebraic/trigonometric expression involving variables, functions of variables and constants.

Example 2: Let a given expression be: $\sin^2 x + \cos^2 x - 1$ for any feasible value of angle *x*. By Definition 2, a set $V = \{\sin^2 x, \cos^2 x, 1\}$ is constructed by dropping all

operators from the given expression. Now, any element of *V* is a term of the given expression.

Definition 3: The ***relative diversity cost (RDC)*** of a node *n*, denoted by $d(n)$, is defined as the sum of the number of new terms added to the left and the right side expressions of node *n* with respect to the respective left and the right side of its parent node *p*. Let L_p and R_p denote the set of terms present in the left and the right of node *p*. Similar nomenclature applies to L_n and R_n. Then,

$$d(n)=|L_n - L_p| + |R_n - R_p| \tag{5.1}$$

where,

L_n = set of terms in the left side expression of node *n*.

R_n = set of terms in the right side expression of node *n*. Similarly, L_p and R_p are defined for node *p*, and $|S|$ denotes cardinality (number of elements) of a given set *S*.

Example 3: Let the parent node represent an identity: $\sin^2 x + \cos^2 x = 1$ and one child node of this parent node stands for the identity: $\sin^2 x + \frac{1}{\sec^2 x} = 1.$ Here, $L_p = \{\sin^2 x, \cos^2 x\}$ $R_p = \{1\}$, $L_n = \{\sin^2 x, 1, \sec^2 x\}$ and $R_n = \{1\}$.

So, $d(n) = |L_n - L_p| + |R_n - R_p| = |\{1, \sec^2 x\}| + 0 = 2.$

Definition 4: A path ${}^{w}P_{r-q_w}$ refers to the sequence of nodes in order of their occurrence, starting at the root node *r* down to one of the possible terminal nodes q_w on the given path of depth *w*.

Definition 5: The ***absolute diversity cost (ADC)*** of a node *n*, denoted by $D(n)$ is defined as the summation of the relative diversity values from the root node *r* to the current node *n*. In other words, if ${}^{w}P_{r-n}$ is the path from the root node *r* to node *n* through a sequence of nodes $q_1, q_2, ..., q_k$, then

$$\begin{aligned} D(n) &= d(n) + d(q_k) + ... + d(q_1) + d(r) \\ &= d(n) + \sum_{i=1}^{k} d(q_i) \end{aligned} \tag{5.2}$$

Here, *r* being the root node has *d(r)*=0. Naturally, $D(n) \geq d(n)$. Additionally, as q_k is the parent of node *n*,

$$D(n) = d(n) + D(q_k) \tag{5.3}$$

Example 4: Let the path ${}^{w}P_{r-q_w}$ include the root node r: 1=1, an intermediate node $q_1: \sin x = \frac{1}{\mathrm{cosec}\, x}$, and a child node q_2 of node q_1 be $\sin x + \cos x = \frac{1}{\mathrm{cosec}\, x} + \cos x$. For q_1, $L_p=\{1\}$, $L_n=\{\sin x\}$, $R_p=\{1\}$ and $R_n=\{1, \mathrm{cosec}\, x\}$. So, $d(q_1) = |\{\sin x\}| + |\{\mathrm{cosec}\, x\}| = 2$. For q_2, $L_p=\{\sin x\}$, $L_n=\{\sin x, \cos x\}$, $R_p=\{1, \mathrm{cosec}\, x\}$ and $R_n=\{1, \mathrm{cosec}\, x, \cos x\}$. So, $d(q_2) = |\{\cos x\}| + |\{\cos x\}| = 2$. $d(r) = 0$, as r is the root node. Therefore, using (3), $D(q_2) = d(r) + d(q_1) + d(q_2) = 0 + 2 + 2 = 4$.

Definition 6: An ***OPEN set of nodes*** S_O is referred to as those nodes that have been generated after the application of the cost function but are yet to be expanded.

Definition 7: A ***CLOSED set of Nodes*** S_C is referred to as those nodes that have been already expanded for the generation of offspring.

5.3.2 Rule Selection for Expansion of a Node in the Search Tree

The SFs used for expansion of nodes, here are selected on a random basis using a roulette wheel (RW) strategy [27]. A section of the rules used for illustration is given in Table 5.1. An exhaustive list of rules used is given in Section A.7 of Appendix.

5.3.3 The Proposed Algorithms

This sub-section discusses the proposed algorithms for CC implementation in details.

5.3.3.1 The RDGBFS Algorithm

The steps of the RDGBFS algorithm in the context of trigonometric identity synthesis problem (for use as chapter-end exercise) are presented in Algorithm 5.1. A flowchart of the RDGBFS is also provided in Section A.4 of Appendix for convenience.

Step 1 of RDGBFS is used to initialize the OPEN set with a start-up element and the CLOSED set as a Null set. Step 2(a) is used for node selection from the OPEN set based on highest RDC. Step 2(b) is concerned with expansion of the selected node for generation of possible children. Step 2(c) is employed to add children (if not available in the tree) to the selected node. Step 2(d) is used to update OPEN and CLOSED sets. Steps 2(a)-(d) are repeated until the terminating condition on depth-limit is reached. Step 3 reports the best node with the highest RDC. Step 2, dealing with expansion of selected node, includes additional strategies listed below.

Algorithm 5.1: **RDGBFS Algorithm for Computational Creativity with respect to Trigonometric Identity Selection Problem**

Input: A randomly selected start-up node, terminal-depth *w*, branching factor *b*
Output: The most promising OPEN node within the user-defined terminal depth
Begin
1. Initialize the OPEN set of nodes S_o with a start-up element *n* representing a trigonometric identity, and set its RDC *d(n)=0*. Set total cost of node *f(n)=d(n)*. Initialize the CLOSED set $S_c = \phi$.
2. **While** depth of node $n \leq w$ **do**
Begin
a) **If** S_o is singleton, select *n* for expansion.
Else for any other node *n'* in S_o, if *f(n)* ≥ *f(n'),* then select node *n* for expansion.
b) Expand *n* by the set of rules given in Table-5.1 satisfying the constraints outlined in Section 5.3 using the roulette wheel strategy. Let the possible children of *n* be $Ch_1, Ch_2, \ldots, Ch_b$.
c) **For** each of the *b* possible children **do**
If the child does not exist in the tree, then add it as a child to its parent node *n*, add it to the OPEN set of nodes, evaluate its *d*-cost and assign it to its *f*-cost.
End-For;
d) Drop the parent node *n* from S_o and put it under S_c.
End While;
3. **Return** the node from the OPEN set of node(s) with the highest f-cost value.
End.

1) Use RW strategy to select one SF from the list in Table 5.1.

2) If neither the left/right part of SF matches with either of left/right of the selected node then invoke RW again.

3(a) Examine whether the left part of the node matches the left/right part of the SF in order. In case a match occurs, replace the matched part of the node by the other part of the SF.

(b) Else repeat 3(a) for the right part of the node.

4) If a term $\in T$ at a selected node for expansion, occurs *z* (>1) times then on invoking SFs 4-8, the same term may be replaced by a randomly selected integer times $y \leq z$.

Examples 5, 6 and 7 below illustrate the above strategies.

Example 5: Let the current node *n* selected for expansion be $\operatorname{cosec} x + \sec x = \frac{1}{\sin x} + \frac{1}{\cos x}$. Suppose, RW selects Rule 5 to expand *n*. In this case, 3(a) is invoked to replace sec *x* by 1/cos*x* and the resultant offspring will be: $\operatorname{cosec} x + \frac{1}{\cos x} = \frac{1}{\sin x} + \frac{1}{\cos x}$.

Example 6: Let the current node *n* selected for expansion be: $\tan x = \frac{\sin x}{\cos x}$ and Rule 5 is selected to expand *n*. Here, 3(a) fails and hence 3(b) is invoked to replace $1/\cos x$ by sec *x* to obtain the child node $\tan x = \sin x \times \sec x$.

Example 7: Let the current node *n* selected for expansion be: $\sin^2 x + \cos^2 x + \sec x = 1 + \sec x$ and Rule 5 is selected to expand *n*. Let the value of *z* be selected randomly as 1. Then the new child will be $\sin^2 x + \cos^2 x + \frac{1}{\cos x} = 1 + \sec x.$

Additional strategies outlined below are used to avoid selection of same rules to expand the root node across runs, which in turn enhances novelty of the solutions. The strategies invoked consider tagging an integer wait-delay to the previously used rules based on their last use. The history is saved in a list of user-defined finite length, say 5. The history includes rule number in case of rules: 1-8 and rule number plus arguments and operators selected in case of 9-11. When a new rule is added to the list, its wait-delay becomes 1, and the wait-delay of all other rules in the list is incremented by 1.

(a) A rule selected by RW algorithm in the current run can be employed to the root node, if this rule is not found in the list.

(b) If the rule selected by RW algorithm falls in rules 1-8 or 9-11 with the same arguments and operators and the selected rule is in already included in the list of used rules, then the Least Recently Used (LRU) rule should be applied to the root node for expansion.

5.3.3.2 Illustration of the RDGBFS Algorithm

The proposed RDGBFS algorithm is illustrated by Fig. 5.1 with a search tree of depth = 3 and branching factor = 2. In Fig. 5.1, the RDC value of each node is indicated by <·>. The set of rules utilized for the above stated search tree is randomly selected using RW strategy from Table 5.1. Here, the strategies outlined in 5.3.3.1 are utilized for expanding the root node. The evaluation of RDC value of each node of the search tree in Fig. 5.1 is elaborated below.

a) At depth=1, for node: $\sec^2 x - \tan^2 x = 1$, L_p={1}, L_n ={$\sec^2 x, \tan^2 x$}, R_p={1} and R_n={1}. By (1), the RDC value of the present node is $|\{\sec^2 x, \tan^2 x\}| + 0 = 2.$

b) At depth=1, for node: $1/\cot x = \tan x$, L_p={1}, L_n ={1, $\cot x$}, R_p={1}and R_n={$\tan x$}. By (1), the RDC value of the present node is $|\{\cot x\}| + |\{\tan x\}| = 2.$

c) At depth=2, for node: $\sin x / \cos x = \tan x$, L_p={1, $\cot x$}, L_n ={$\sin x, \cos x$}, R_p={$\tan x$}, and R_n={$\tan x$}. By (1), the RDC value of the present node is $|\{\sin x, \cos x\}| + 0 = 2.$

d) At depth=2, for node: $\sin^2 x / \cot x = \tan x \times \sin^2 x$, L_p={1, $\cot x$}, L_n ={$\sin^2 x$, $\cot x$}, R_p={$\tan x$}, and R_n={$\tan x$}. Using (1), the RDC value of the present node is $|\{\sin^2 x\}| + |\{\sin^2 x\}| = 2.$

e) At depth=3, for node: $\sin^2 x / \cos^2 x = \tan^2 x$, L_p={$\sin x, \cos x$}, L_n ={$\sin^2 x, \cos^2 x$}, R_p={$\tan x$}, and R_n={$\tan^2 x$}. Using (1), the RDC value of the present node is $|\{\sin^2 x, \cos^2 x\}| + |\{\tan^2 x\}| = 3.$

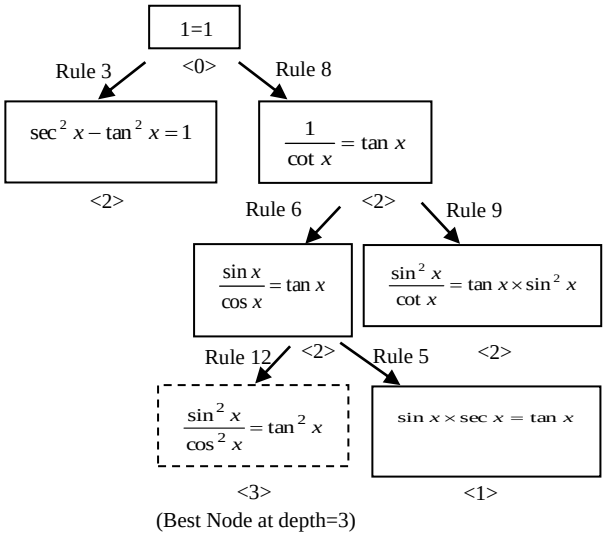


Fig. 5.1. Search tree output of RDGBFS algorithm for a branching factor of 2 and termination depth of 3.

Algorithm 5.2: **ADGBFS Algorithm for Computational Creativity with respect to Trigonometric Identity Selection Problem**

The steps for Algorithm 5.2 are same as Algorithm 5.1 except for steps 1 and 2(c), which are as follows.

1. Initialize the OPEN set of nodes S_o with a start-up element *n* representing a trigonometric identity, and set its ADC *D(n)=0*. Set total cost of node *f(n)=D(n)-p(n)* with *p(n)=0*. Initialize the CLOSED set of nodes $S_c = \phi$.

2. c) **For** each of the b possible children **do**

If the child does not exist in the tree, then add it to the OPEN set of nodes, evaluate its *D*-cost minus *p*-cost, assign it to its *f*-cost.

End-For;

f) At depth=3, for node: $\sin x \times \sec x = \tan x$, $L_p = \{\sin x, \cos x\}$, $L_n = \{\sin x, \sec x\}$, $R_p = \{\tan x\}$, and $R_n = \{\tan x\}$. By (1), the RDC value of the present node is $|\{\sec x\}| + 0 = 1$.

5.3.3.3 The ADGBFS Algorithm

Algorithm 5.2 presents the summarized version of ADGBFS algorithm in the context of trigonometric identity synthesis problem due to space economy. The complete algorithm is detailed in Section A.4 of Appendix. The constraints utilized for step 2(b) in Algorithm 5.2 is same as discussed in 5.3.3.1. A flowchart of the proposed algorithm is also provided in Section A.4 of Appendix.

5.3.3.4 Illustration of the ADGBFS Algorithm

The proposed ADGBFS algorithm is illustrated by Fig. 5.2 with a search tree of depth = 3 and branching factor = 2. In Fig. 5.2, the f-cost value of each node is denoted by <·>. The set of rules utilized for the above stated search tree is randomly selected using RW technique from Table 5.1. Here, the strategies outlined in 5.3.3.1

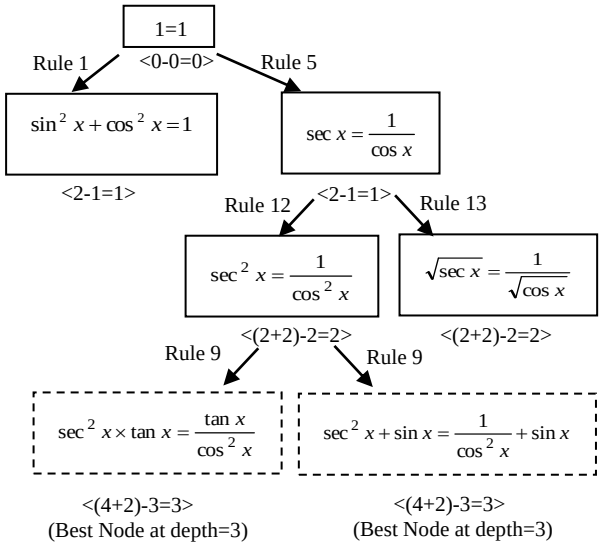


Fig. 5.2. Search tree output of ADGBFS algorithm for a branching factor of 2 and termination depth of 3.

are utilized for expanding the root node. The details of diversity and penalty cost computation is provided in Appendix under Section A.5 due to space constraints.

5.3.3.5 Extended Applications of the Proposed Algorithms

The potential applications of the proposed algorithms in other domains, such as algebra, propositional logic, and set theory, are discussed in Appendix. Furthermore, with the support of LLMs, the proposed techniques can be extended to address engineering and scientific problems across other domains as well.

5.4 PROPERTIES AND IMPLICATIONS OF THE PROPOSED ALGORITHMS

5.4.1 The Common Properties of the RDGBFS and ADGBFS Algorithms

The following definition is required to explain the derived properties for the two algorithms.

Definition 8: An algorithm is called complete, if it terminates with a feasible solution when one exists.

A list of interesting properties derived for the RDGBFS and ADGBFS algorithms are presented below. The proof of the properties is given in Appendix.

Property 1. *The relative diversity cost of a sequence of nodes from the root r to the terminal node q_w on a path ${}^{w}P_{r-q_w}$ always satisfies the following inequality*

$$d(q_w) > d(q_{w-1}) > d(q_{w-2}) > \ldots\ldots > d(r). \quad (5.4)$$

where, $d(q_i)$ denotes the relative diversity cost of node q_i.

Property 2. *The RDGBFS and ADGBFS algorithms are complete for a search tree of finite depth.*

Property 3. *There will never exist any cyclic paths within the search space of RDGBFS and ADGBFS algorithms.*

Property 4. *For two different runs of the RDGBFS and ADGBFS algorithms, different solutions are generated for a fixed branching factor and termination-depth.*

Property 5. *The child node generated from any member of the OPEN set of nodes cannot exist in the OPEN set of nodes itself.*

Property 6. *The child node generated from the selected best node can never be any existing CLOSED node.*

5.4.2 Implications of the Common Properties of RDGBFS and ADGBFS algorithms

This section discusses the importance of the properties of the proposed algorithms which are illustrated in the previous sub-section. Property 1 infers that every child node is always more diversified with respect to its parent node along a path from the root to termination depth. Property 2 envisages that the algorithms always return a solution within a finite depth of the search tree, while Property 3 guarantees that cyclic paths are not produced at any stage of their implementation. Property 4 illustrates that the algorithms produce different trigonometric identity problems for different runs. This property is quite important with respect to the creativity aspect of the algorithm as every run is able to generate a novel identity problem. Properties 5 and 6 explain why a new node appearing in the trees as an existing OPEN/CLOSED node shouldn't be attached to its current parent. The last two properties eliminate the unnecessary expansion of the search trees when such nodes already exist.

5.4.3 Additional Properties of the ADGBFS Algorithm

The following definitions are required to explain the additional derived properties for the ADGBFS algorithm.

Definition 9: The path ${}^{w}P^{*}_{r-q_w}$ is optimal among all possible paths, emanating from the root node *r* and terminating at any node q_w of depth *w,* if node q_w has the highest f-cost value among the then OPEN set of nodes. Symbolically,

$$ {}^{w}P^{*}_{r-q_w} = {}^{w}P_{r-q_w} \text{ with } f(q_w) \geq f(q'_w), \quad \forall q'_w \in S_o \tag{5.5}$$

where, S_o is the then OPEN set of nodes.

Definition 10: An absolute diversity based cost function is consistent, if it never reopens any already CLOSED node. Formally, the consistency of an absolute diversity based cost function may be tested by the following criterion:

$$k(n,n') + D(n') \geq D(n), \ \forall (n,n') \tag{5.6}$$

where, n' is any descendant of *n* and $k(n,n')$ denotes the optimal penalty cost from *n* to n'. Here, $k(n,n')$ is essentially negative everywhere and zero at the root node.

Definition 11: An absolute diversity based cost function is monotonic if it satisfies the following criterion:

$$C(n,n') + D(n') \geq D(n), \ \forall (n,n') \tag{5.7}$$

where, $C(n,n')$ represents the penalty cost from *n* to n'.

Definition 12: An absolute diversity based cost function is said to be admissible when it never overestimates the cost to reach the terminal node. Formally, the admissibility of an absolute diversity based cost function may be tested by the following criterion:

$$D(n) \leq D^{*}(n). \tag{5.8}$$

Where, $D^{*}(n)$ denotes the optimal absolute diversity cost of node *n*.

The additional properties for the ADGBFS algorithm are listed below and its proofs are provided in Appendix.

Property 5. *The absolute diversity cost of a sequence of nodes from the root r to the terminal node q_w on a path ${}^{w}P_{r-q_w}$ always satisfies the following inequality*

$$D(q_w) > D(q_{w-1}) > D(q_{w-2}) > \ldots\ldots > D(r) \tag{5.9}$$

where, $D(q_i)$ denotes the absolute diversity cost of node q_i.

Property 8. *The optimal path* ${}^{w}P^{*}_{r-q_w}$ *will always have the largest cumulative f-cost value with respect to other paths* ${}^{w}P_{r-q_w}$ *within the search space.*

Property 9. *The absolute diversity based cost function is consistent.*

Property 10. *The absolute diversity based cost function is monotonic.*

Property 11. *The absolute diversity based cost function is admissible.*

5.4.4 Implications of the Additional Properties of ADGBFS Algorithm

This section discusses the importance of the properties of the proposed algorithm which are portrayed in the previous sub-section. Property 7 infers that the absolute diversity of a node from root node to terminal depth keeps on increasing with increase in depth of the search tree. Property 8 concludes that the optimal path from the root node to the terminal depth carries the maximum f-cost value as compared to all other non-optimal paths. This helps to assure that the f-cost value i.e., *D(n)-p(n)* has been optimized. The consistency property of the algorithm (property 9) guarantees that the already CLOSED nodes are never reopened during the execution of the algorithm. Next, the monotonic condition mentioned previously (property 10) ensures that the f-cost values for a sequence of nodes from the root to the terminal depth are always non-decreasing. Moreover, the cost function being admissible (property 11) assures that the algorithm is able to provide optimal solution(s).

5.4.5 Comparative Properties of the Proposed Algorithms

The differences in performance between the RDGBFS and ADGBFS algorithms are illustrated by the following properties. In other words, the scenarios in which one algorithm outperforms the other are depicted by the properties below. The detailed proof of these properties is provided in Appendix.

Property 12. *If the relative diversity at node* q_i*, called* $d(q_i)=1$ *for all i on the optimal path* ${}^{w}P^{*}_{r-q_w}$*, with root node r and a terminal node* q_w *(at depth w) in both RDGBFS and ADGBFS algorithms, then the sum of f-cost on the optimal path for RDGBFS is greater than ADGBFS algorithm.*

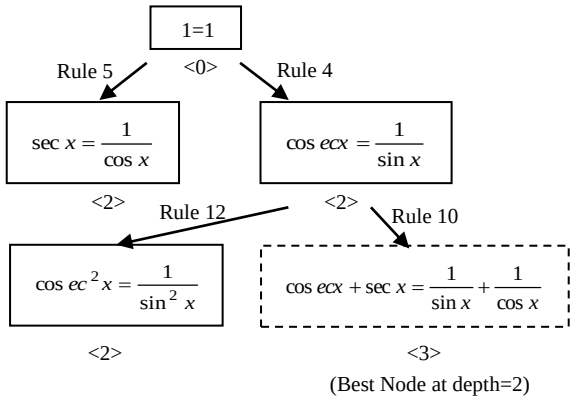


Fig. 5.3. Search tree output of RDGBFS algorithm for a branching factor of 2 and termination depth of 2 by using rule 10 to tackle scalability issue.

Property 13. *If the relative diversity $d(q_i)$ at node q_i >1 for some i on the optimal path ${}^{w}P^{*}_{r-q_w}$, with root node r and terminal node q_w, then the f-cost on the optimal path for ADGBFS algorithm is always positive.*

Property 14. *For any node q_i lying on an optimal path ${}^{w}P^{*}_{r-q_w}$ with root node r and terminal node q_w of depth w, if $d(q_i)$ is the relative diversity cost of node q_i, such that*

$$\sum_{i=1}^{w}(w-i)\,d(q_i)=\frac{w(w+1)}{2}$$

holds, then the total f-cost over the optimal path evaluated by ADGBFS and RDGBFS algorithms are equal.

5.4.6 Implications of the Comparative Properties of Proposed Algorithms

The worst-case analysis presented in properties 12-14 reveals that the RDGBFS algorithm performs better than ADGBFS when the relative diversity cost of all the nodes present in the OPEN set of nodes is 1 (vide Property 12). However, if RDC of fewer OPEN nodes are ≥1, ADGBFS would perform better than the RDGBFS (Property 13). Additionally, both the algorithms have similar performance when f-cost of all the nodes along the optimal path are equal (Property 14).

5.5 ROBUSTNESS AND SCALABILITY ISSUES

Robustness is a generic term used to describe the power of maintaining correctness of an algorithm in presence of disturbance/noise/imprecision [28]-[29]. However, in case of search algorithms, the possibility of disturbance/noise/imprecision of inputs is beyond questions. Robustness in search algorithms strictly refers to soundness of the algorithms [30]. However, to demonstrate the potential of the proposed algorithms in covering all possible solutions (identities), we also consider the logical

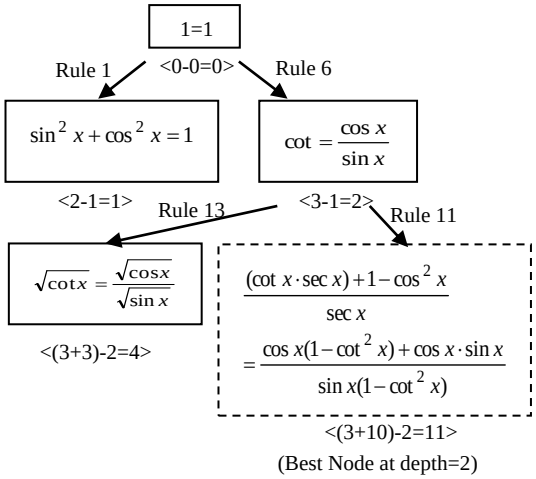


Fig. 5.4. Search tree output of ADGBFS algorithm for a branching factor of 2 and termination depth of 2 by using rule 11 to tackle scalability issue.

completeness [7] of the algorithms to showcase their robustness. Properties 15 and 16 together demonstrate that the proposed algorithms are sound and logically complete and thus are robust.

Definition 13: The RDGBFS and ADGBFS algorithms are sound if the solutions generated maintain mathematical equality.

Property 15: *The RDGBFS and ADGBFS algorithms are sound, if the returned node with highest f-cost value in the OPEN set of nodes of user-defined depth is correct (i.e., maintains mathematical equality in both sides).*

Proof. Let us prove it by the method of contradiction, i.e., let us assume that the node returned with the highest f-cost, called best node, in the user defined depth be incorrect. This in turn implies that the equality condition between the left and the right hand side of the best node does not hold. This in turn may happen in two possible ways: i) the parent of the best node was incorrect, and/or ii) the incorrectness is due to the expansion process of the parent of the best node (during its generation). However, as the expansion includes augmenting both sides of an identity relation by similar operation with the new terms included in both sides being mathematically equal, the resulting expression must maintain identity relation. So, the only possibility of generating incorrect best node is due to incorrectness of its parent. However, the reasoning applied to generate the correct best node equally applies to its parent during the parents' generation. The process thus recurrently proceeds to the incorrectness of the root node. However, the root node being 1=1, a universally true identity, cannot be wrong.

So, the assumption that the best node selected is incorrect is wrong, and hence its contradiction that the best node generated is always correct holds. This proves the soundness of the RDGBFS and ADGBFS algorithms. □

Definition 14: A search algorithm is logically complete if it can identify all the feasible solutions that can be theoretically obtained by utilizing the given set of input resources (rules, operators and operands).

Property 16: *The RDGBFS and ADGBFS algorithms are logically complete.*

Proof. Let us assume that the proposed RDGBFS and ADGBFS algorithms are not logically complete. In other words, the proposed algorithms are unable to construct at least one solution (identity), which logically can be synthesized by using the given set of rules, operators and operands. Let that identity be $f(x_1, x_2, ..., x_n) = g(x_1, x_2, ..., x_n)$, where $x_1, x_2, ..., x_n \in T = \{\sin^m x, \cos^m x, ...\}$ for $m,n \in Z$=set of positive integers, and the operators chosen from $E=\{+,-,\times,\div\}$ for any pair of operands from T. However, since the selection of operators and the choice of operands in the expansion process of the tree are random, such identity can never exist. So, the assumption is wrong and hence the proposed algorithms are logically sound. □

As both soundness and (logical) completeness hold, the proposed RDGBFS and ADGBFS algorithms are robust.

It is found from literature [7] that BFS algorithms have worst case time and space complexity $O(b^w)$, where b is the uniform branching factor of all the nodes and w is the depth of the search tree. Naturally, for large b or w, the BFS algorithms are not scalable. However, for problems (identities) to be solved by the beginners usually are generated by depth-limited BFS. Additionally, by incorporation of more sophisticated rule (Rule no. 10 and 11), the single-step generation of a node produces trigonometric identities of sufficient complexity (see Fig. 5.3 and 5.4) and thus depth w usually should not be >2 for the beginners to manually solve the problems. So, the BFS algorithms, which often face the challenge of scalability, are not a concern for the present scenario.

5.6 EXPERIMENTAL RESULTS

The RDGBFS and ADGBFS algorithms have been implemented in Python version 3.8.5. Such an implementation platform is chosen due to its inherent support for manipulation of mathematical expressions using modules such as SymPy, Random and the like. The experimental results and analysis are provided in the following sub-sections.

5.6.1 Novelty of solutions across different runs of the algorithms for a fixed depth

The trigonometric identities generated from different runs of the RDGBFS and ADGBFS algorithms reveal that distinct identities are produced in each run. The above outcome indicates that novelty is preserved across multiple executions of the same algorithm, underscoring a key feature of a computationally creative system. The details of these results are provided in Appendix under Section A.6.

5.6.2 Qualitative improvement of solutions over iterations

Within the same run of the algorithms, as the depth of the search tree increases, the identity problems generated by the proposed algorithms exhibit greater complexity and/or diversity. This demonstrates the algorithms' ability to produce problems of the desired complexity and diversity. The details of these results are presented in Appendix under Section A.6.

5.7 PERFORMANCE ANALYSIS

5.7.1 Comparison of Proposed Algorithms with Search and Template based Techniques

Table 5.2 presents a detailed comparison between the identity problems generated by search and template-based methods versus the proposed approaches. The metrics utilized for this comparison include: i) relative diversity of the best node in the OR-tree, averaged across 100 generated problems, referred to as Average-Relative-Diversity (ARD), ii) absolute diversity of the best node divided by its depth, averaged across 100 generated problems, termed as Average Absolute Diversity per Unit Depth (AAD), and iii) run-time averaged over 100 problem instances. For this comparison, the metrics (i-ii) of problems generated by traditional (search and template-based) methods are compared to those produced by the proposed algorithms by backtracking the problems generated by traditional methods to the root (1=1) and generating identities using the proposed algorithms within the same search tree at the same depth. A few examples are illustrated in the Appendix to describe the aforementioned method of comparison.

The key insight from this comparison is the superior performance of the proposed RDGBFS and ADGBFS algorithms in generating identity problems compared to traditional techniques, which tend to produce more predictable and less varied problems. Additionally, the proposed algorithms are not only more effective in terms of problem variety but also more efficient in terms of run time. This reduction in computational time is significant, as it allows for faster problem generation, making

Table 5.2
Search and Template based Comparative framework to Examine Relative Performance of the Proposed algorithms

Sl. No.	Algorithm	ARD	AAD	Run-time (sec)
1.	Pearce *et al.*	8.36	15.63	40.86
2.	Singh *et al.*	5.22	9.25	31.02
3.	Liu *et al.*	5.88	16.20	33.52
4.	Tabuguia *et al.*	8.02	11.27	42.88
5.	Brito *et al.*	1.96	8.15	31.11
6.	Zhang *et al.*	3.63	11.12	36.43
5.	Briggs *et al.*	10.05	25.18	95.07
8.	**RDGBFS**	**12.72**	**NA**	**21.30**
9.	**ADGBFS**	**NA**	**32.43**	**30.63**

Table 5.3
Generative Neural Network based Comparative framework to Examine Relative Performance of the Proposed algorithms

Sl. No.	Algorithm	ARD	AAD	Training Time (mins)	Test Time (sec)
1.	RNN	6.33	16.75	101.15	30.97
2.	LSTM	5.52	18.50	152.60	48.92
3.	VAE	8.07	21.03	283.56	88.68
4.	SeqGAN	8.85	24.57	412.30	98.74
5.	RankGAN	9.03	26.88	332.20	90.56
6.	LeakGAN	10.47	29.92	621.32	101.84
5.	**RDGBFS**	**12.72**	**NA**	**NA**	**21.30**
8.	**ADGBFS**	**NA**	**32.43**	**NA**	**30.63**

the algorithms more suitable for real-time applications and large-scale problem generation tasks.

5.7.2 Comparison of Proposed Algorithms with Generative Neural Network based Techniques

Table 5.3 presents a comparison of the proposed algorithm with existing generative neural network-based techniques for producing trigonometric identities. The same metrics mentioned in Section 5.5.1 are utilized here, along with the average training time over 100 problem instances. Again, for this comparison, the back-tracking policy is considered which has been discussed in Section 5.5.1. The dataset used to train the neural networks in Table 5.3 is the same as in [17], consisting of standard identities found in textbooks.

The results in Table 5.3 clearly indicate that the performance of the proposed algorithm significantly surpasses that of neural network-based models in terms of problem diversity. Additionally, while neural network models necessitate large datasets for training, leading to considerable computational expense, the proposed algorithms offer a more computationally efficient alternative. This efficiency is due

Table 5.4
Large Language Model based Comparative framework to Examine Relative Performance of the proposed Algorithms

Sl. No.	Algorithm	ARD	AAD	Run-time (sec)
1.	Pham *et al.*[24]	8.52	28.03	35.78
2.	Deroy *et al.*[25]	5.84	26.96	38.02
3.	Yu *et al.* [26]	6.61	23.70	43.54
4.	**RDGBFS**	**12.72**	**NA**	**21.30**
5.	**ADGBFS**	**NA**	**32.43**	**30.63**

to the proposed algorithm's lack of dependence on extensive training datasets, making it a more practical solution for generating trigonometric identities.

5.7.3 Comparison of Proposed Algorithms with LLM based Trigonometric Identity Generation Techniques

Table 5.4 compares the performance of the proposed algorithms with existing LLMs in generating trigonometric identity problems, incorporating the backtracking policy discussed in Section 5.5.1. The results clearly demonstrate that the proposed algorithms outperform their competitor LLMs in terms of both ARD and AAD metrics, with significantly shorter run-times. The longer run-times for LLMs (as shown in Table 5.4) stem from extensive pre-processing (e.g., tokenization, input encoding), post-processing (e.g., de-tokenization, output filtering), and their reliance on an online connection, which makes them vulnerable to issues such as high processing demand, network latency, or server downtime, leading to delays in solution generation.

5.7.4 Comparison of Trigonometric Identity Problems present in Textbooks and Generated by the RDGBFS and ADGBFS Algorithms

This sub-section presents a comparison of the relative and absolute diversity-based f-cost values for trigonometric identity problems found in textbooks versus those generated by the proposed algorithms. A total of 112 problems have been collected from 7 popular textbooks [31]-[37] for this comparison. Fig. 5.5 and 5.6 illustrate the relative and absolute diversity-based cost function values for 15 problems from the textbooks, compared to those generated by the proposed approach, due to space constraints.

The f-cost values of the textbook problems are compared to those of the identities generated by the proposed algorithms by back tracking the textbook problems to the root (1=1) and generating identities within the same search tree up to the same depth

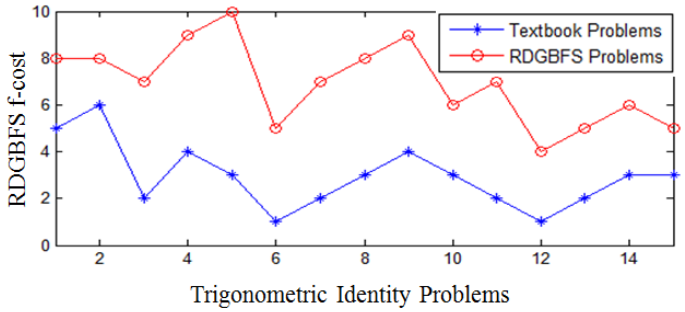


Fig. 5.5. Comparison of problems generated by RDGBFS with textbook chapter-end problems

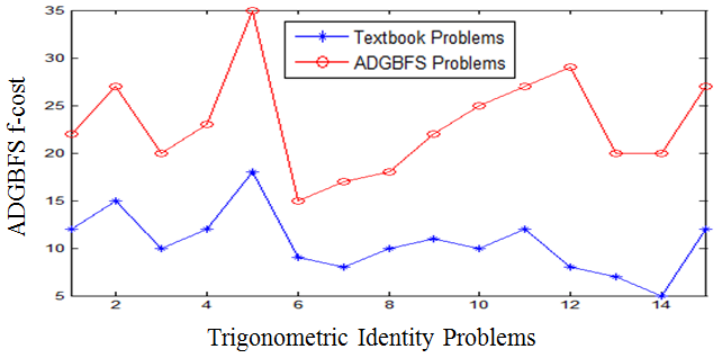


Fig. 5.6. Comparison of problems generated by ADGBFS with textbook chapter-end problems

as the textbook problems. The afore-said policy of comparison is illustrated in Appendix.

Fig. 5.5 and 5.6 clearly show that the relative and absolute diversity-based cost function values for problems generated by the proposed approach are significantly higher than those from the textbooks. This indicates that the proposed approach produces problems that are more complex and challenging.

5.8 CONCLUSIONS

The chapter extends traditional BFS realized with OR-trees to develop two interesting algorithms: RDGBFS and ADGBFS to examine their performance in the context of CC. The proposed algorithms are similar to BFS with respect to the expansion process, but differ by the cost function employed to select the Best node for expansion and the terminating criterion. An analysis of the algorithms is undertaken to derive 16 distinct properties, of which properties 15 and 16 confirm their robustness in view of soundness and completeness issues. The relative merits of the proposed algorithms are studied in properties 12 to 14. Relative performance analysis undertaken reveals that the proposed algorithms outperform their three well-known competitive classes of algorithms with respect to ARD, AAD, and run-time

complexity metrics. Additionally, the proposed algorithms produce more diverse alternative solutions than the standard text-book problems, occupying the same depth in the OR-tree. The scope of the proposed algorithms is examined in identity synthesis of trigonometry and other mathematical domains as well. It is important to mention here that due to lack of scalability of the BFS algorithm, the proposed algorithms are also not scalable. The effect of lack of scalability of the proposed algorithm fortunately does not appear in the synthesis of exercise for students' problems because of limited user-defined depth of the search trees. However, the lack of scalability and the choice of user-defined rule-set may hinder the scope of the proposed algorithms in the invention of mathematical identities for true real-world discovery problems.

5.9 APPENDIX

A.1. Proof of the Common Properties of the RDGBFS and ADGBFS Algorithms

Property 1. *The relative diversity cost of a sequence of nodes from the root r to the terminal node q_w on a path ${}^{w}P_{r-q_w}$ always satisfies the following inequality*

$$d(q_w) > d(q_{w-1}) > d(q_{w-2}) > \ldots\ldots > d(r). \quad \text{(A.1)}$$

Where, $d(q_i)$ denotes the relative diversity cost of node q_i.

Proof. Let q_{w-1} be a child of a given node q_{w-2}. As q_{w-1} is generated from q_{w-2} by replacing additive/factor terms of q_{w-2} by new terms, thus for a single replacement, $d(q_{w-1})$ increases by a minimum value of 1 with respect to $d(q_{w-2})$. So, $d(q_{w-1}) > d(q_{w-2})$. As, this holds for any pair of parent-child node on the path ${}^{w}P_{r-q_w}$,

$$\therefore d(q_w) > d(q_{w-1}) > d(q_{w-2}) > \ldots\ldots > d(r). \quad \square$$

Property 2. *The RDGBFS and ADGBFS algorithms are complete for a search tree of finite depth.*

Proof. We shall prove the theorem by the method of contradiction. Let us assume that there exists a solution within a given depth *w* but the algorithm terminates before identifying the solution. As the proposed RDGBFS (or ADGBFS) algorithm selects the best node n^* with the largest diversity cost $d(n)$ (or *D(n)-p(n)*) among all the OPEN set of nodes before checking the depth condition for termination, so there is no scope of termination of the algorithm before reporting the selection of the best node with the highest $d(n)$ (or *D(n)-p(n)*). So, the assumption is wrong, and the statement of the theorem is correct. □

Property 3. *There will never exist any cyclic paths within the search space of RDGBFS and ADGBFS algorithms.*

Proof. The proposed algorithms only expands nodes $n \in S_o$. Further, after expansion of a selected OPEN node, it is put under the CLOSED-set of nodes S_c, and the CLOSED nodes are never re-declared as OPEN. So, a node once expanded cannot be re-expanded i.e., the algorithm cannot reopen any nodes $n \in S_c$. So, the path ${}^{w}P_{r-q_w}$ never consists of any cyclic routes. □

Property 4. *For two different runs of the RDGBFS and ADGBFS algorithms, different solutions are generated for a fixed branching factor and termination-depth.*

Proof. Let a node *n* consists of maximum *c* children, where *c* =*b*, the given branching factor. Now, given a finite depth-limit *w*, different runs of the program may select the same node for expansion, with different set of children due to the random selection of rules using roulette wheel strategy. Thus, the solution set for 2 separate runs of the algorithms for fixed depth and branching factor would be different. □

Property 5. *The child node generated from any member of the OPEN set of nodes cannot exist in the OPEN set of nodes itself.*

Proof. Let S_o be the current OPEN set of nodes. Let $n' \in S_o$. Let, $X \in S_o$ be the best node selected, i.e., $d(X) > d(p), \forall p \in S_o$. So,

$$d(X) > d(n') \text{ as } n' \in S_O \tag{A.2}$$

Let the child node of x be n'. Then,

$$d(n') > d(X) \qquad \text{[by property 1]} \tag{A.3}$$

However, (A.3) contradicts (A.2). Therefore, there cannot be any node x in S_o with a child node n' that too belongs to S_o. In other words, the child node generated from any member of the OPEN set of nodes cannot exist in the OPEN set of nodes itself.

□

Property 6. *The child node generated from the selected best node can never be any existing CLOSED node.*

Proof. Let c be a CLOSED node having a set of children $c_1, c_2, ...c_p$. Let S_O be the current OPEN set of nodes that includes $c_1, c_2, ...c_p$ (i.e., $c_1, c_2, ...c_p$ have no children). Let $X \in S_O$ be the best node selected for expansion.

$$\therefore \quad d(X) > d(p),\ \forall p \in S_o \tag{A.4}$$

Again, let the child node of X be c, the already CLOSED node. So,

$$d(c) > d(X), \quad \forall p \in S_o \qquad \text{[by property 1]} \tag{A.5}$$

Combining (A.4) and (A.5),

$$d(c) > d(p),\ \forall p \in S_o \tag{A.6}$$

But c being the parent of nodes that lie in S_o,

$$d(p) > d(c),\ \forall p \in S_o \tag{A.7}$$

The contradiction of (A.6) and (A.7) occur because the assumption that the new born child could be a CLOSED node is wrong. Hence, the contradiction of that stating the theorem is correct. □

Corollary: If the newly generated node exists in the tree as an OPEN or CLOSED node then it should not be added to its current (new) parent as the child node.

A.2. Proof of Additional Properties of the ADGBFS Algorithm

Property 5. *The absolute diversity cost of a sequence of nodes from the root r to the terminal node q_w on a path ${}^{w}P_{r-q_w}$ always satisfies the following inequality*

$$D(q_w) > D(q_{w-1}) > D(q_{w-2}) > \ldots\ldots > D(r) \tag{A.8}$$

where, $D(q_i)$ denotes the absolute diversity cost of node q_i.

Proof. Let q_{w-1} be a child of a given node q_{w-2}. Then, by definition of absolute diversity, $D(q_{w-1}) = d(q_{w-1}) + D(q_{w-2})$. As $d(q_{w-1}) > 1, \therefore D(q_{w-1}) > D(q_{w-2})$. Thus for any pair of nodes on the path ${}^{w}P_{r-q_w}$, starting at the root node r and terminating at node q_w, inequality (A.8) naturally holds. □

Property 8. *The optimal path ${}^{w}P^{*}_{r-q_w}$ will always have the largest cumulative f-cost value with respect to other paths ${}^{w}P_{r-q_w}$ within the search space.*

Proof. By definition 8, ${}^{w}P^{*}_{r-q_w} = {}^{w}P_{r-q_w}$ with $f(q_w) \geq f(q'_w),\ \forall q'_w \in S_o$, where S_o is the OPEN set of nodes for termination depth = w. Let the path be $r - q_1 - q_2 - \ldots\ldots - q_{w-2} - q_{w-1} - q_w$. Since, q_{w-1} was the best node when termination depth = w-1, the above property equally holds, i.e., ${}^{w}P^{*}_{r-q_w} = {}^{w}P_{r-q_w}$ with $f(q_{w-1}) \geq f(q'_{w-1}),\ \forall q'_{w-1} \in S'_o$, where S'_o is the OPEN set of nodes when termination depth = w-1. It can be shown by induction that the above property holds for each node on the path $r - q_1 - q_2 - \ldots\ldots - q_{w-2} - q_{w-1} - q_w$. □

Property 9. *The absolute diversity based cost function is consistent.*

Proof. Let $p^*(n)$ be the optimal penalty cost of the node *n* and $D^*(n)$ be the optimal diversity cost of node *n*, both defined with respect to the root node. Let *r* be the root node. Then, for existence of node *n* on an optimal path, we require:

$$p^*(n) + D^*(n) \geq D(r), \qquad \forall n$$
$$\Rightarrow k(r,n) + D^*(n) \geq D(r), \qquad \forall n \tag{A.9}$$

Now for n' the child of node *n*, we require:

$$k(r,n') + D^*(n) \geq D(r), \qquad \forall n' \tag{A.10}$$

Subtracting (A.9) from (A.10), we obtain

$$k(n,n') + D^*(n') - D^*(n) \geq 0$$
$$\Rightarrow k(n,n') + D^*(n') \geq D^*(n)$$

Now, if we choose the measure of $D^*(n)$ as $D(n)$ in the same way as described by the last equation, then we can write $k(n,n') + D(n') \geq D(n), \quad \forall (n,n').$ □

Property 10. *The absolute diversity based cost function is monotonic.*

Proof. Let $p(n)$ be the penalty cost of the node *n* and $D(n)$ be the diversity cost of node *n*, both defined with respect to the root node. Let *r* be the root node. Then,

$$p(n) + D(n) \geq D(r), \qquad \forall n$$
$$\Rightarrow C(r,n) + D(n) \geq D(r), \quad \forall n \tag{A.11}$$

Now for n' we can write,

$$C(r,n') + D(n') \geq D(r), \qquad \forall n' \tag{A.12}$$

Subtracting (A.11) from (A.12)

$$C(n,n') + D(n') - D(n) \geq 0$$
$$\Rightarrow C(n,n') + D(n') \geq D(n).$$

Hence, the absolute diversity based cost function is monotonic. □

Property 11. *The absolute diversity based cost function is admissible.*

Proof. From property 9 we have,

$$k(n,n') + D(n') \geq D(n)$$

If *t* is the terminal node, then replacing n' by *t* we get,

$$k(n,t) + D(t) \geq D(n)$$
$$\Rightarrow D(n) \leq D^*(n)$$

Hence, the theorem follows. □

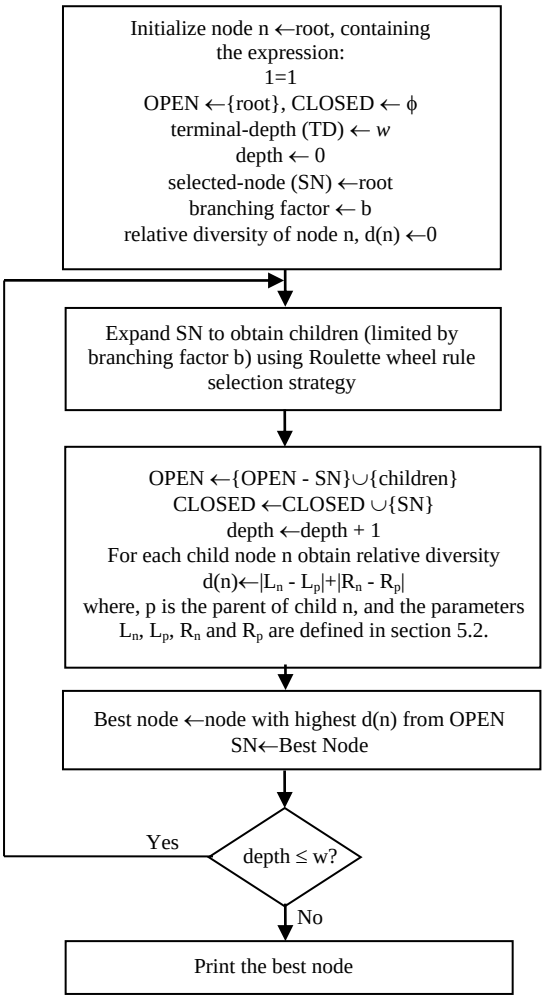


Fig. A.1. Flow chart for Relative Diversity Guided Best first Search (RDGBFS)

A.3. Proof of Comparative Properties of the Proposed Algorithms

Property 12. *If the relative diversity at node q_i, called $d(q_i)=1$ for all i on the optimal path ${}^{w}P^{*}_{r-q_w}$, with root node r and a terminal node q_w (at depth w) in both RDGBFS and ADGBFS algorithms, then the sum of f-cost on the optimal path for RDGBFS is greater than ADGBFS algorithm.*

Proof. Given the root node *r* and the terminal node q_w on the optimal path ${}^{w}P^{*}_{r-q_w}$, the total f-cost on the said path by ADGBFS algorithm, is obtained as

$$\sum_{q_i \in {}^{w}P^{*}_{r-q_w}} f(q_i)\Big|_{ADGBFS}$$

$$= f(r) + f(q_1) + f(q_2) + \ldots\ldots + f(q_w)$$
$$= 0 + (d(q_1) - 1) + (d(q_1) + d(q_2) - 2) + \ldots. + (d(q_1) + d(q_2) + d(q_3)$$
$$+ \ldots. + d(q_w) - w)$$
$$= wd(q_1) + (w-1)d(q_2) + (w-2)d(q_3) + \ldots\ldots$$
$$+ (w - (w-1))d(q_w) - (1 + 2 + \ldots + w)$$

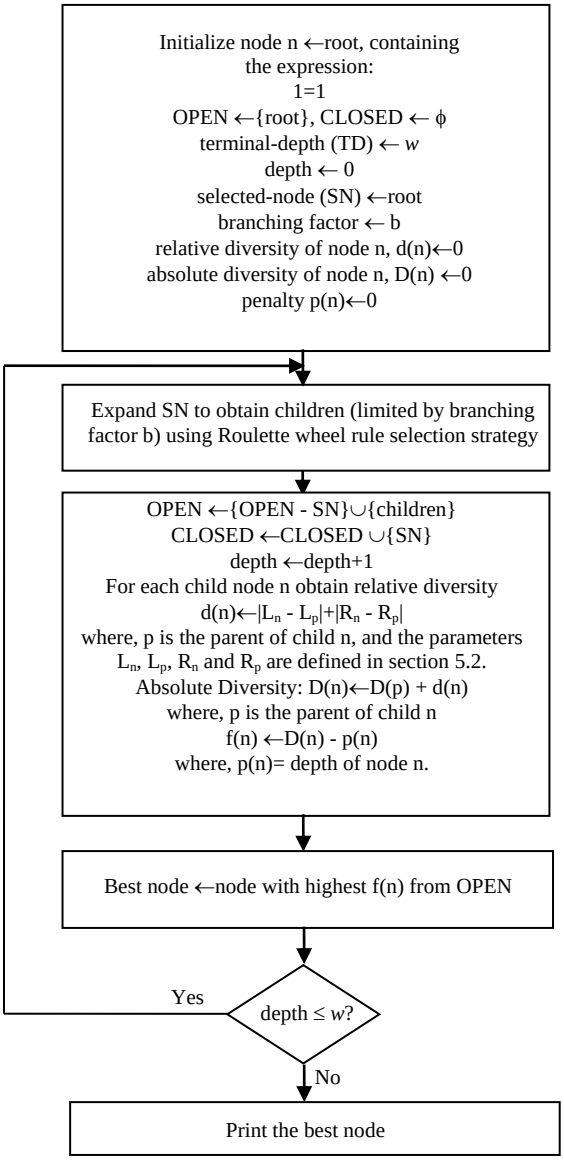


Fig. A.2. Flow chart for Absolute Diversity Guided Best first Search (ADGBFS)

$$= wd(q_1) + (w-1)d(q_2) + (w-2)d(q_3) + ... + d(q_w) - \frac{w(w+1)}{2}. \quad \text{(A.13)}$$

Total f-cost on the same optimal path ${}^{w}P^{*}_{r-q_w}$ by RDGBFS algorithm, given by

$$\sum_{q_i \in {}^{w}P^{*}_{r-q_w}} f(q_i)\Big|_{RDGBFS}$$
$$= f(r) + f(q_1) + f(q_2) + ... + f(q_w)$$
$$= 0 + d(q_1) + d(q_2) + ... + d(q_w). \quad \text{(A.14)}$$

If $d(q_i) = 1, \forall i,$ then, $\sum_{q_i \in {}^{w}P^{*}_{r-q_w}} f(q_i)\Big|_{ADGBFS} = 0 < \sum_{q_i \in {}^{w}P^{*}_{r-q_w}} f(q_i)\Big|_{RDGBFS} = w.$ □

It is apparent from Property 12 that in the worst case the ADGBFS algorithm incurs a cost $D(n) - p(n) = 0$ when $D(n) = p(n).$ No conclusions can be inferred about the choice of the best node among the OPEN set of nodes under that circumstance. Similar

Algorithm A.1: **ADGBFS Algorithm for Computational Creativity with respect to Trigonometric Identity Selection Problem**

Input: A randomly selected start-up node, terminal-depth *w*, branching factor *b*
Output: The most promising OPEN node within the user-defined terminal depth
Begin
1. Initialize the open set of nodes S_o with a start-up element *n* representing a trigonometric identity, and set its ADC $D(n)=0$. Set total cost of node $f(n)=D(n)-p(n)$ with $p(n)=0$. Initialize the closed set of nodes $S_c = \phi$.
2. **While** depth of node $n \le w$ **do**
Begin
a) **If** S_o is singleton, select *n* for expansion.
Else for any other node n' in S_o, if $f(n) \ge f(n')$, then select node *n* for expansion.
b) Expand *n* from the set of rules given in Table-1 using the roulette wheel strategy. Let the children of *n* be $Ch_1, Ch_2, ..., Ch_b$. c) **For** each of the *b possible* children **do**
Begin
If the child does not exist in the tree, then add it to the open set of nodes, evaluate its D-cost and p-cost assign it its f-cost.
End-For;
d) Drop the parent node *n* from S_o and put it under S_c.
End While;
3. **Return** the node from the OPEN set of node(s) with the highest f-cost value.
End.

situations do not appear in RDGBFS algorithm as it does not consider penalty cost $p(n)$.

Property 13. *If the relative diversity $d(q_i)$ at node q_i >1 for some i on the optimal path ${}^{w}P^{*}_{r-q_w}$, with root node r and terminal node q_w, then the f-cost on the optimal path for ADGBFS algorithm is always positive.*

Proof. Given the root node *r* and the terminal node q_w on the optimal path ${}^{w}P^{*}_{r-q_w}$, the total f-cost on the said path by ADGBFS algorithm is obtained using (A.13). In (A.13), if $d(q_i)=1, \forall i, \sum_{q_i \in {}^{w}P^{*}_{r-q_w}} f(q_i)\Big|_{ADGBFS} = 0.$

Naturally, if $d(q_i)>1$ for at least one *i*, then $\sum_{q_i \in {}^{w}P^{*}_{r-q_w}} f(q_i)\Big|_{ADGBFS} > 0.$ □

Property 14. *For any node q_i lying on an optimal path ${}^{w}P^{*}_{r-q_w}$ with root node r and terminal node q_w of depth w, if $d(q_i)$ is the relative diversity cost of node q_i, such that*

$$\sum_{i=1}^{w} (w-i)\, d(q_i) = \frac{w(w+1)}{2} \qquad \text{(A.15)}$$

holds, then the total f-cost over the optimal path evaluated by RDGBFS and ADGBFS algorithms are equal.

Proof. Rewriting (A.13),

$$\sum_{q_i \in {}^{w}P^{*}_{r-q_w}} f(q_i)\Big|_{ADGBFS}$$
$$= w d(q_1) + (w-1) d(q_2) + (w-2) d(q_3) + ... + d(q_w) - \frac{w(w+1)}{2}$$

$$= wd(q_1) + (w-1)d(q_2) + (w-2)d(q_3) + ... + d(q_w) - \sum_{i=1}^{w}(w-i)\,d(q_i)$$

$$= ((w-1)d(q_1) + d(q_1)) + ((w-2)d(q_2) + d(q_2)) + ... + ((w-(w-1))d(q_w) + d(q_w)) - \sum_{i=1}^{w}(w-i)\,d(q_i)$$

$$= \sum_{i=1}^{w}(w-i)\,d(q_i) + (d(q_1) + d(q_2) + d(q_3) + ... + d(q_w))$$

$$- \sum_{i=1}^{l}(w-i)\,d(q_i)$$

$$= d(q_1) + d(q_2) + ... + d(q_w).$$ □

It follows from Theorem 14 that under the given condition (A.15), both the algorithms perform equally till the termination depth is reached.

A.4. Flowcharts for RDGBFS and ADGBFS Algorithms and Complete Details of ADGBFS Algorithm

The flowcharts for RDGBFS and ADGBFS algorithms are depicted in Figs. A.1 and A.2 respectively. The details of ADGBFS algorithm is illustrated in Algorithm A.1.

A.5. Details of the Absolute Diversity based Cost Function Evaluation

In Fig. 5.2 (as provided in the manuscript), the strategies outlined in Section 5.3 of the revised manuscript are utilized to expand the root node. The details of the absolute diversity based f-cost are evaluated for each node in the search tree using the following steps.

a) At depth=1, for node: $\sin^2 x + \cos^2 x = 1$, L_p={1}, L_n ={$\sin^2 x, \cos^2 x$}, R_p={1} and R_n={1}. By using eq. (1), the RDC value of the present node is $|\{\sin^2 x, \cos^2 x\}| + 0 = 2.$ The penalty cost for the current node is 1. Thus, the f-cost value of this node is 2-1=1.

b) At depth=1, for node: $\sec x = 1/\cos x$, L_p={1}, L_n ={$\sec x$}, R_p={1} and R_n={1, $\cos x$ }. By using eq. (1), the RDC value of the present node is $|\{\sec x\}| + |\{\cos x\}| = 2.$ The penalty cost for the current node is 1. Thus, the f-cost value of this node is 2-1=1.

c) At depth=2, for node: $\sec^2 x = 1/\cos^2 x$, L_p={$\sec x$}, L_n ={$\sec^2 x$}, R_p={1, $\cos x$ } and R_n={1, $\cos^2 x$ }. By using eq. (1), the RDC value of the present node is $|\{\sec^2 x\}| + |\{\cos^2 x\}| = 2.$ The ADC value of this node is 2+2=4 (using eq. (3)). The penalty cost for the current node is 2. Thus, the f-cost value of the present node is 4-2=2.

d) At depth=2, for node: $\sqrt{\sec x} = 1/\sqrt{\cos x}$, L_p={$\sec x$}, L_n ={$\sqrt{\sec x}$}, R_p={1, $\cos x$ } and R_n={1, $\sqrt{\cos x}$ }. By using eq. (1), the RDC value of the present node is $|\{\sqrt{\sec x}\}| + |\{\sqrt{\cos x}\}| = 2.$ The ADC value of this node is 2+2=4. The penalty cost for the current node is 2. Thus, the f-cost value of the present node is 4-2=2.

Table A.1
Trigonometric Identity generation for different runs of RDGBFS Algorithm.

Run (*R*)	Optimal solutions(s)
R=1	$\frac{\sqrt{(\tan^2 x+1)(1+\cos x)}}{\sqrt{\sec^2 x-1}}=\frac{1}{\sqrt{1-\cos x}}$
R=2	$\frac{\cot x\times \cos ec\, x}{\sqrt{\cot^2 x+1}}=\frac{\sqrt{\cos ec^2 x-1}}{\sqrt{\cos ec^2 x-\cot^2 x}}$
R=3	$\frac{\cos x\times\cot x\times(1-(\cot^2 x\times\sin^2 x))}{\sin^2 x\sqrt{\cot^2 x-\cos^2 x}}=\frac{1}{\sqrt{\sin^2 x+\cos^2 x}}$

Table A.2
Trigonometric Identity generation for different runs of ADGBFS Algorithm.

Run (*R*)	Optimal solutions(s)
R=1	$\frac{1}{\cot^2 x}-(1+\tan^2 x)=(\sin^2 x-1)\sec^2 x$
R=2	$\sec^2 x-\sin^2 x=(\sin^2 x+\cot^2 x)\tan^2 x$
R=3	$\frac{\sec x(\sec^2 x-\tan^2 x)}{\cot x\sqrt{\tan^2 x+1}}=\sqrt{\sec^2 x-1}$

Table A.3
Trigonometric Identity generation for varying terminal depth in the same run of RDGBFS Algorithm.

Terminal Depth (*w*)	Optimal solutions(s)
w=1	$\cot x=\frac{\cos x}{\sin x}$
w=7	$\sin^2 x=1-\frac{1}{\sec^2 x}$
w=15	$\tan^2 x=\left(1-\frac{\cos^2 x}{1+\sin x}\right)^2\times\frac{1}{\cos^2 x}$

Table A.4
Trigonometric Identity generation for varying terminal depth in the same run of ADGBFS Algorithm.

Terminal Depth (*w*)	Optimal solutions(s)
w=1	$\sin^2 x+\cos^2 x=1$
w=7	$\sin^2 x=\frac{\tan^2 x-\sin^2 x}{\tan^2 x}$
w=15	$\frac{\tan^2 x}{1+\tan^2 x}=\cos ec^2 x-\cot^2 x(1+\sin^2 x)$

e) At depth=3, for node: $\sec^2 x\times\tan x=\tan x/\cos^2 x$, $L_p=\{\sec^2 x\}$, $L_n=\{\sec^2 x,\tan x\}$, $R_p=\{1,\ \cos^2 x\}$ and $R_n=\{\tan x,\cos^2 x\}$. By using eq. (1), the RDC value of the present node is $|\{\tan x\}|+|\{\tan x\}|=2$. The ADC value of this node is 2+2=4. The penalty cost for the current node is 2. Thus, the f-cost value of the present node is 4-2=2.

f) At depth=3, for node: $\sec^2 x + \sin x = 1/\cos^2 x + \sin x$, $L_p = \{\sec^2 x\}$, $L_n = \{\sec^2 x, \sin x\}$, $R_p = \{1, \cos^2 x\}$ and $R_n = \{1, \cos^2 x, \sin x\}$. By using eq. (1), the RDC value of the present node is $|\{\sin x\}| + |\{\sin x\}| = 2$. The ADC value of this node is 2+2=4. The penalty cost for the current node is 2. Thus, the f-cost value of the present node is 4-2=2.

A.6. Details of Experimental Results

The results obtained from the proposed algorithms, which demonstrate the diversity of the generated identities within a single run and the novelty of the identities across different runs of the same algorithm, are illustrated below.

A.6.1. Novelty of solutions across different runs of the algorithms for a fixed depth

Tables A.1 and A.2 demonstrate the results of the RDGBFS and ADGBFS algorithms respectively for different runs of the search tree. The results presented in the above tables are obtained for a branching factor of 2 and a termination depth of 12. It is apparent from the tables that different runs of the algorithms produce optimal/novel solutions, indicating the creative aspect of the algorithms.

A.6.2. Qualitative improvement of solutions over iterations

Tables A.3 and A.4 demonstrate the results of the RDGBFS and ADGBFS algorithms respectively for a branching factor of 3 and termination depth of 15. It is apparent from these tables that with the increase in the depth of the search tree, the complexity and/or diversity of the trigonometric identity increases.

A.5. Comparison of problems generated by search and template based methods versus problems generated using the proposed algorithms

The ARD and AAD values of the identities generated by search- and template-based methods (STMs) have been compared with those of the identities generated by the proposed algorithms, as described in Section 5.5.1 of the manuscript. This section elaborately demonstrates the backtracking of the problems generated by Pearce et al. [9] within the search tree up to the root (1 = 1) and the generation of problems using the proposed algorithms within the same tree. Furthermore, this section illustrates how the ARD and AAD metrics are evaluated for both the problems generated by [9] and the proposed approach. The detailed set of rules utilized to achieve the aforementioned results is presented in Table A.5.

Table A.5
Detailed List of rules used for Trigonometric Identity generation.

T=set of atomic terms={1, $sin^m x$, $cos^m x$, $cosec^m x$, $sec^m x$, $cot^m x$, $tan^m x$, | $m \geq 2$}, *E*=set of operators={+ ,- ,* ,/}, *F*=set of additive operators={+,-}, *H*=functions of atomic terms of arity-2 or higher involving operators from set *E*.

Replacement Rules

R.1: $a \pm \frac{b}{c} \rightarrow \frac{ac \pm b}{c}$

R.2: $\frac{a}{b} \pm \frac{c}{d} \rightarrow \frac{a(v/b) \pm c(v/d)}{bd}$, where *v* =LCM of *b* and *d*

No.	Rules	No.	Rules
1.	$\sin^2 x + \cos^2 x = 1$	2.	$\sin^2 x = 1 - \cos^2 x$
3.	$\cos^2 x = 1 - \sin^2 x$	4.	$\cos ec^2 x - \cot^2 x = 1$
5.	$\cos ec^2 x = 1 + \cot^2 x$	6.	$\cos ec^2 x - 1 = \cot^2 x$
5.	$\sec^2 x - \tan^2 x = 1$	8.	$\sec^2 x = 1 + \tan^2 x$
9.	$\sec^2 x - 1 = \tan^2 x$	10.	$\cos ec x = 1/\sin x$
11.	$\cos ec x \times \sin x = 1$	12.	$\sin x = 1/\cos ec x$
13.	$\sec x = 1/\cos x$	14.	$\sec x \times \cos x = 1$
15.	$\cos x = 1/\sec x$	16.	$\cot x = \cos x/\sin x$
15.	$\cot x \times \sin x = \cos x$	18.	$\cot x/\cos x = 1/\sin x$
19.	$\tan x = \sin x/\cos x$	20.	$\tan x \times \cos x = \sin x$
21.	$\tan x/\sin x = 1/\cos x$	22.	$1/\cot x = \tan x$
23.	$1/\tan x = \cot x$	24.	$\cot x \times \tan x = 1$
25.	For *A*,*B*∈*T* or *H*, If *A*=*B*, Then $A \otimes C = B \otimes C$, where $C \in T$ and $\otimes \in E$ are selected randomly.	26.	For *A*,*B*∈*T* or *H*, If *A*=*B*, Then $A \otimes C = B \otimes (1/D)$, where $C, D \in T$ and $\otimes \in$ E are chosen randomly, satisfying *C*=*1/D* of the form given in Rules 4,5 and 8.
25.	For *A*,*B*∈*T* or *H*, If *A*=*B*, then $A \oplus C' = B \oplus (1/D')$, *where* $C', D' \in H$ and $\oplus \in F$ are selected randomly satisfying $C'=1/D'$, and the result must be expressed in simplified form using the following re-write rules R1 and R2. In case there exist more than 2 terms in the left side of the replacement rules, the rule needs to be applied over 2 terms at a time recursively until it cannot be applied further.	28.	For *A*,*B*∈*T* or *H*, If *A*=*B*, Then $A^2=B^2$
29.	For *A*,*B*∈*T* or *H*, If *A*=*B*, Then $\sqrt{A} = \sqrt{B}$	30.	For *A*,*B*∈*T* or *H*, If *A/C*=*B/D*, Then *AD*=*BC*
31.	For *A*,*B*∈*T* or *H*, If *A*=*B*, Then 1/*A*=*1/B*	32.	For *A*,*B*∈*T* or *H*, If *A/C*=*B/D*, Then A=*B/D*×*C*

In Fig. A.3(a), it is observed that at a terminal depth of 3, the RDGBFS algorithm produces the best node with RDC =5. Similarly, in Fig. A.3(b), the RDGBFS algorithm produces the best node with RDC=3. Therefore, the ARD value for the RDGBFS algorithm in generating two problems is calculated as: (7 + 3)/2= 5. In contrast, the RDC values of the best nodes produced by Pearce et al. [9] from Figs. A.3(a) and A.3(b) are 3 and 2, respectively. Consequently, the ARD value for two problems generated by Pearce et al. [9] is: (3 + 2)/2 = 2.5. The above ARD computation indicates that the RDGBFS algorithm outperforms the algorithm in [9], as it produces the best nodes with the highest ARD value.

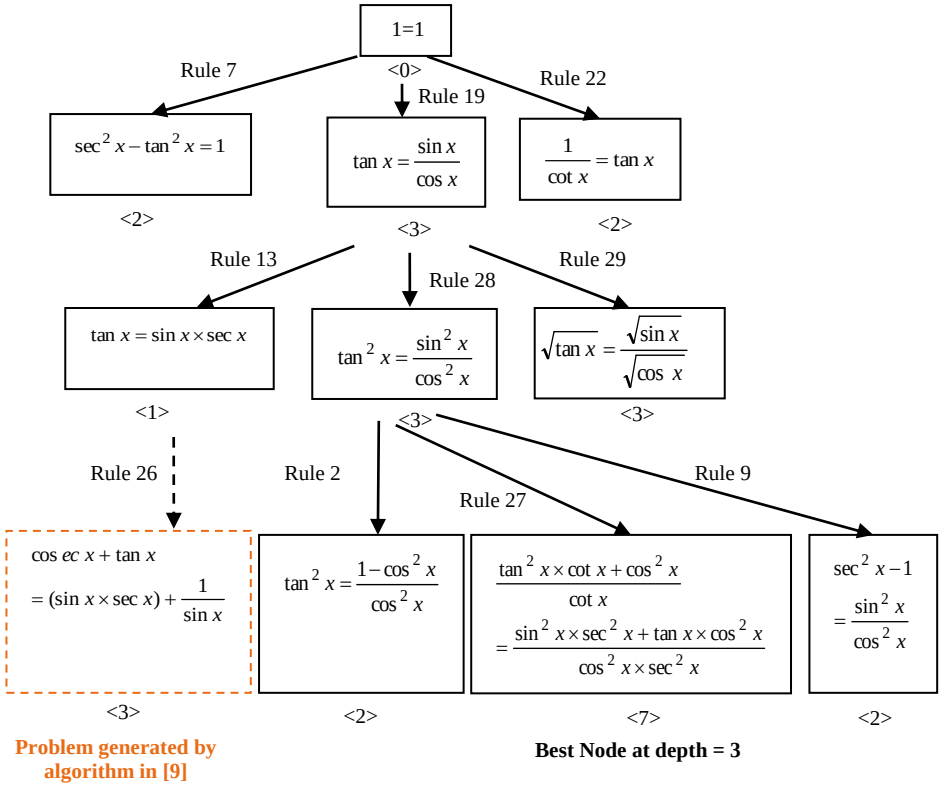


Fig. A.3.(a) Search tree output of RDGBFS algorithm in comparison to STM based generated problem for depth=3 and branching factor=3 in run-1

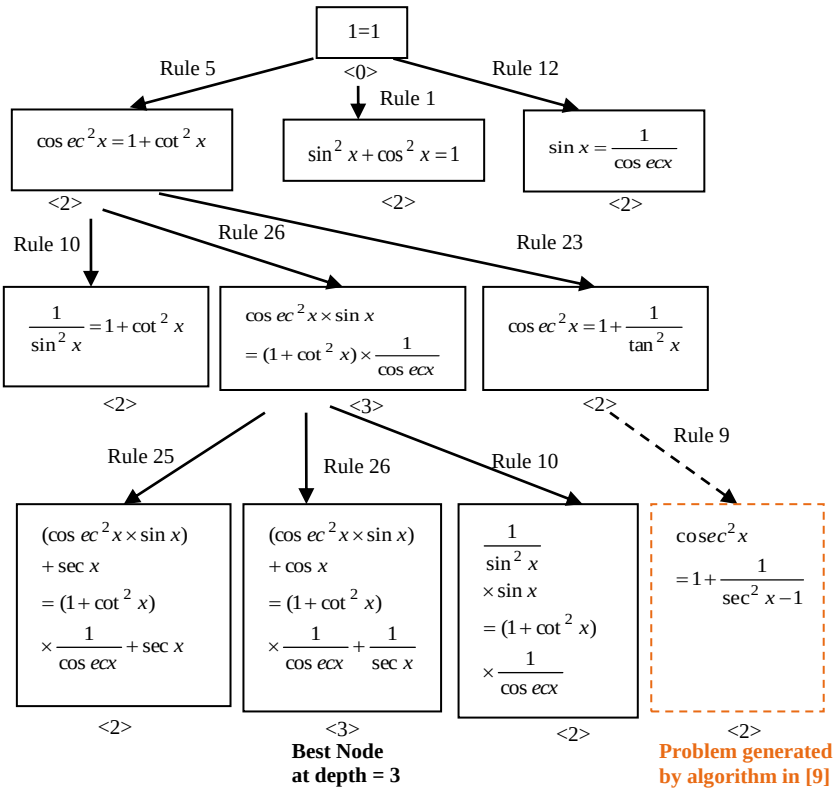


Fig. A.3.(b) Search tree output of RDGBFS algorithm in comparison to STM based generated problem for depth=3 and branching factor=3 in run-2

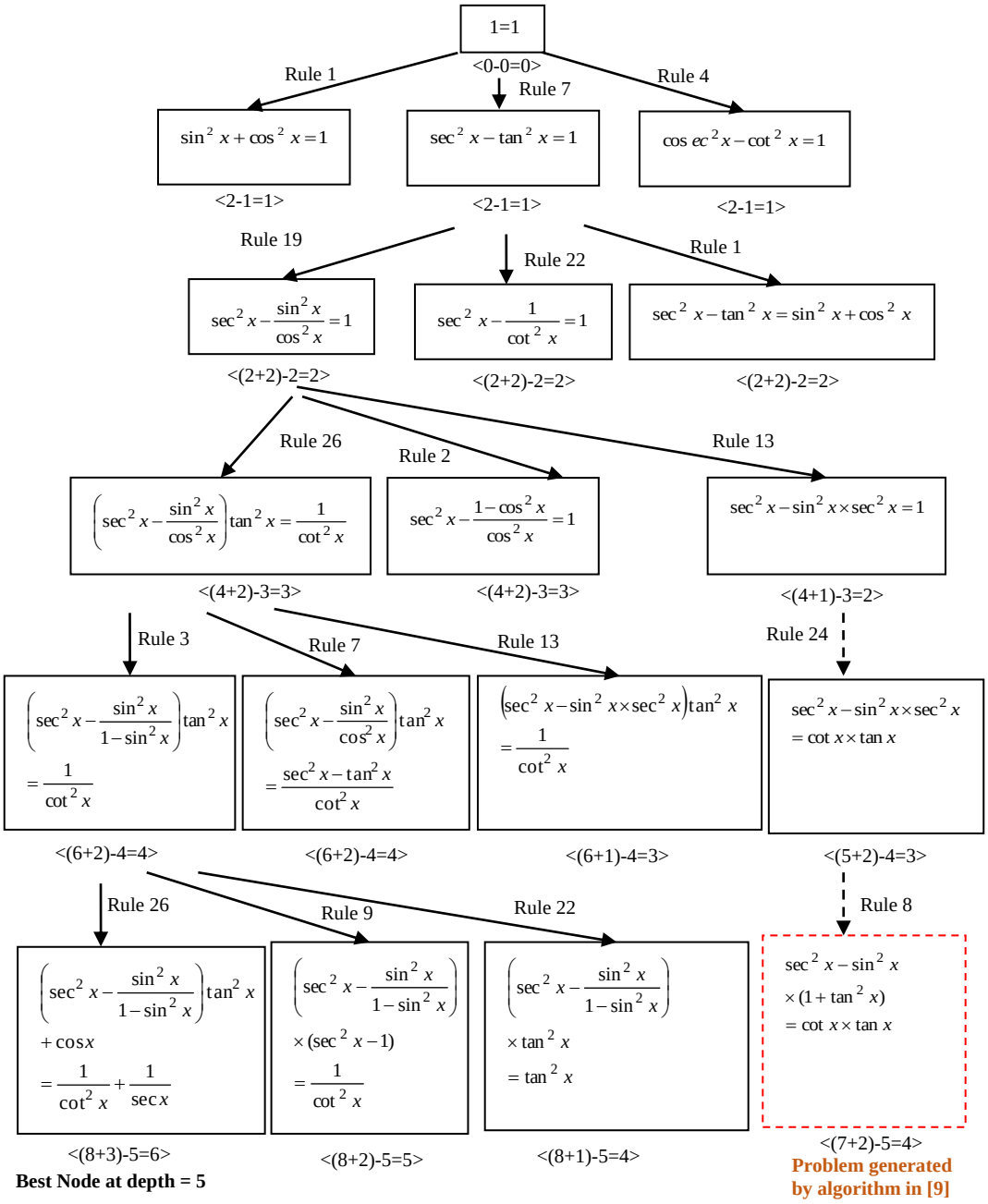


Fig. A.3.(c) Search tree output of ADGBFS algorithm in comparison to STM based generated problem for depth=5 and branching factor=3 in run-1

However, to eliminate the possibility of accidental or incidental occurrences of high diversity in the best node within one or a few search trees, the comparison between the proposed approach and its competitor is conducted by averaging the RDC of the best node across 100 generated problems.

In Fig. A.3(c), it is observed that at a terminal depth of 5, the ADGBFS algorithm produces the best nodes with ADC based f-cost=6. Similarly, in Fig. A.3(d), the ADGBFS algorithm produces the best node with ADC based f-cost=8. Therefore, the AAD value for the ADGBFS algorithm in generating two problems is calculated as:

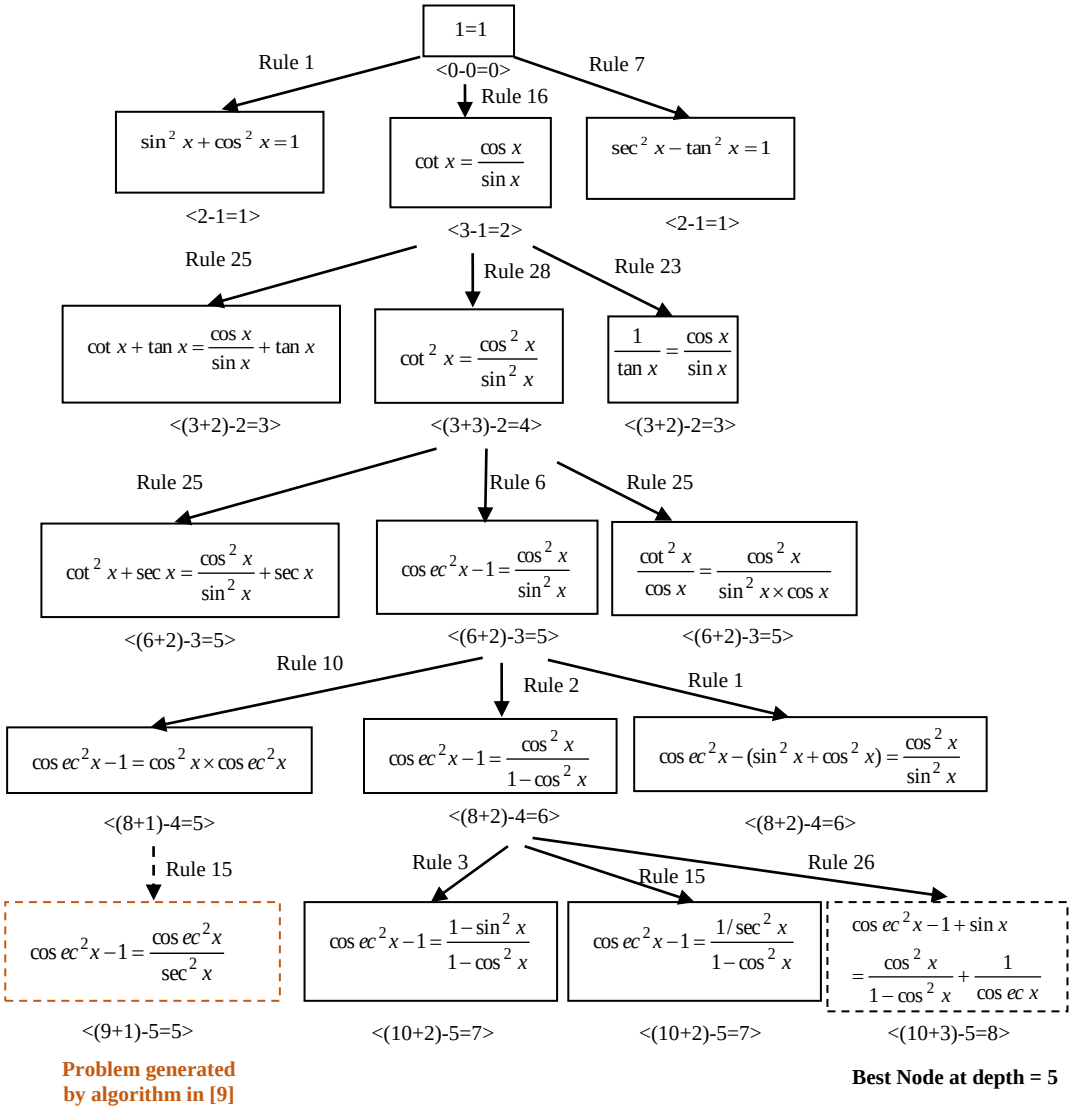


Fig. A.3.(d) Search tree output of ADGBFS algorithm in comparison to STM based generated problem for depth=5 and branching factor=3 in run-2

$\frac{\left(\frac{6}{5}\right)+\left(\frac{8}{5}\right)}{2}$ = 1.4. In contrast, the ADC based f-cost values of the best nodes produced by Pearce et al. [9] from Figs. A.3(c) and A.3(d) are 4 and 5, respectively. Consequently, the ARD value for two problems generated by Pearce et al. [9] is: $\frac{\left(\frac{4}{5}\right)+\left(\frac{5}{5}\right)}{2}$ = 0.9. The above AAD computation indicates that the ADGBFS algorithm outperforms the algorithm in [9], as it produces the best nodes with the highest AAD value.

Nevertheless, to eradicate the possibility of incidental high f-cost values for the best nodes in individual search trees, the above comparison is extended by averaging the absolute diversity of the best node, divided by depth, across 100 generated problems.

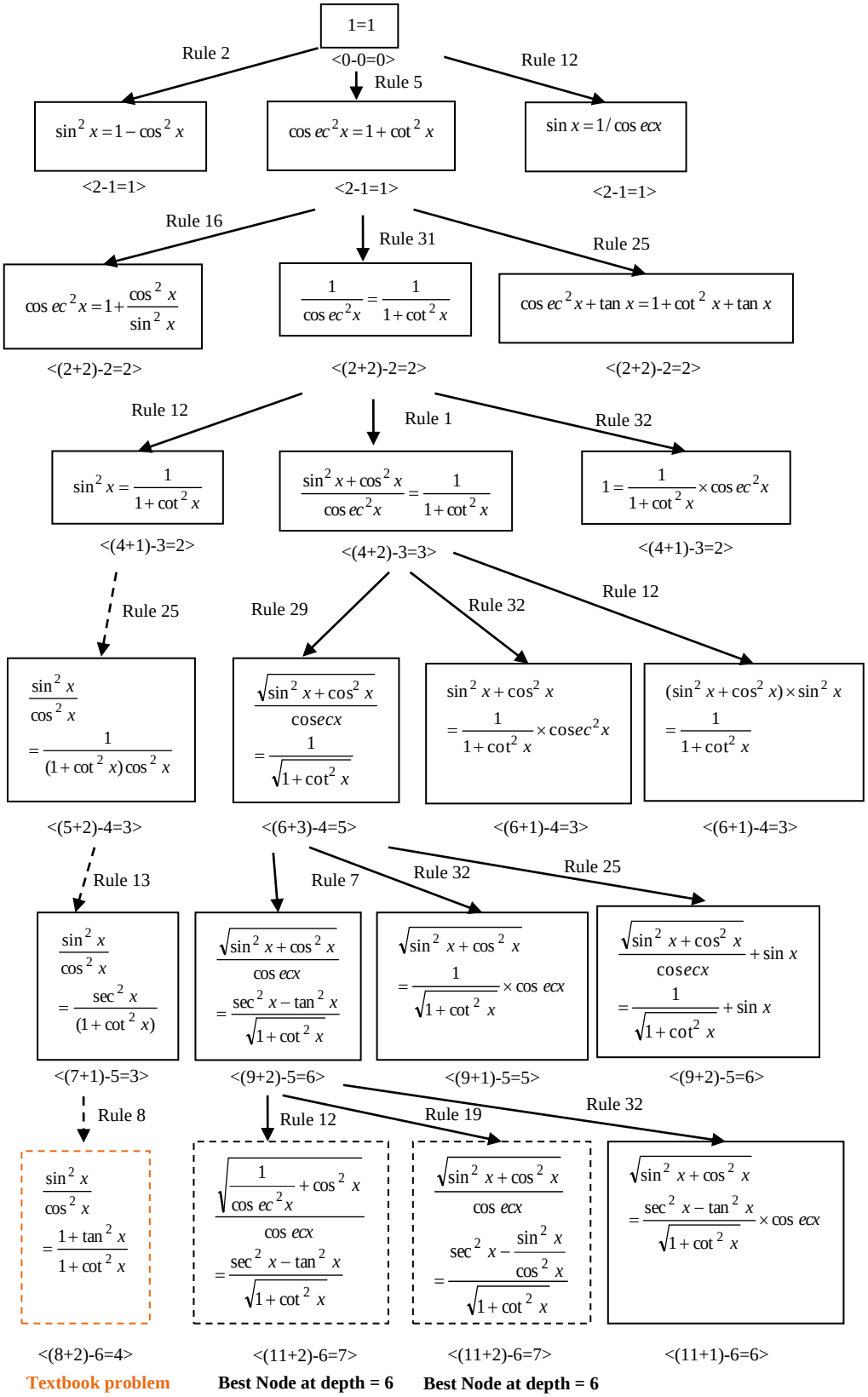


Fig. A.4. Search tree output of ADGBFS algorithm in comparison to textbook problem for depth=6 and branching factor=3

A.8. Comparison of textbook problems with generated problems using ADGBFS algorithm

Table A.6
A few List of rules used for Generation of algebraic identities

No.	Rules	No.	Rules
1.	$(A+B)^2 = A^2 + B^2 + 2AB$	2.	$(A-B)^2 = A^2 + B^2 - 2AB$
3.	$A^2 - B^2 = (A-B)(A+B)$	4.	$A^4 - B^4 = (A^2)^2 - (B^2)^2$
5.	If *A-B=C-D*, Then *A-B-C=-D*	6.	If A=B×C, Then A/B=C

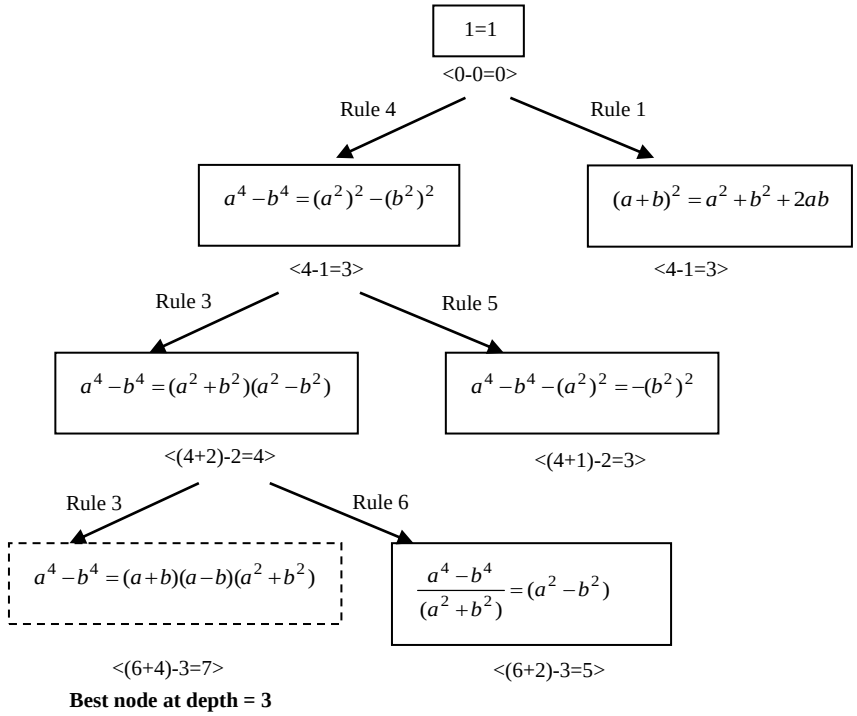


Fig. A.5. Search tree output of ADGBFS algorithm in algebraic domain for depth=3 and branching factor=2

Table A.7
A few List of rules used for Generation of algebraic identities

No.	Rules	No.	Rules
1.	$p \Leftrightarrow \neg(\neg p)$	2.	$\neg p \Leftrightarrow \neg(\neg(\neg p))$
3.	$p \rightarrow q \Leftrightarrow \neg q \rightarrow \neg p$	4.	$\neg p \vee \neg(\neg q) \Leftrightarrow \neg(p \wedge \neg q)$

The f-cost values of standard textbook problems have been compared to those of the identities generated by the proposed algorithms as described in Section 5.5.4 of the manuscript. This section elaborately demonstrates the back-tracking the textbook problems in the search tree up to the root, 1=1, and generating identities using the ADGBFS algorithm within the same tree. The detailed set of rules utilized for achieving the afore-mentioned results is also depicted in Table A.5. The RDGBFS algorithm can be utilized in a similar way for generating the search tree.

In Fig. A.4, it is observed that at the terminal depth=6, the best node is produced by the proposed ADGBFS algorithm. Moreover, at the same depth, if the textbook-problem was finally generated (shown by dotted arrows), its f-cost value is much lower than the f-cost value of the best node produced by the proposed algorithm.

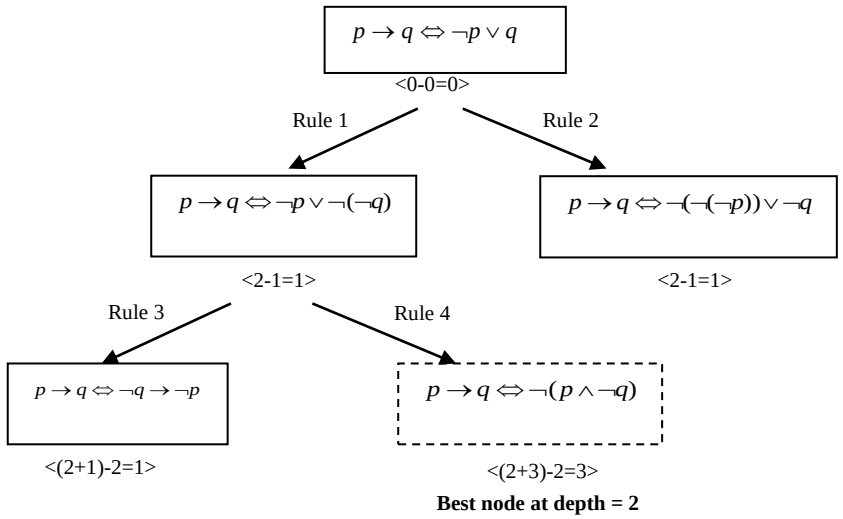


Fig. A.6. Search tree output of ADGBFS algorithm in propositional logic domain for depth=2 and branching factor=2

Table A.7
A few List of rules used for Generation of Set Theory problems

No.	Rules	No.	Rules
1.	If $A \cup B = B \cup A$, Then $(A \cup B) \cap C = (B \cup A) \cap C$	2.	$A \cup 0 = A$
3.	$A \cup U = A$	4.	$(A \cup B) \cup C = A \cup (B \cup C)$
5.	$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$	6.	$A \cup A = U$

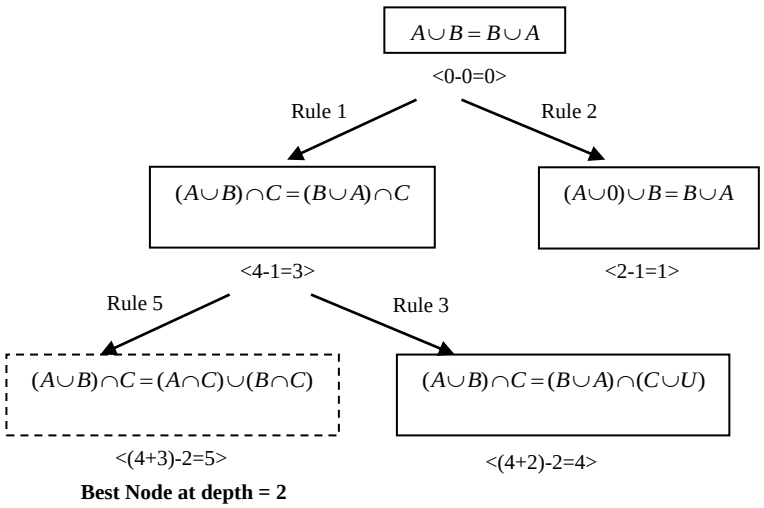


Fig. A.5. Search tree output of ADGBFS algorithm in set theory domain for depth=2 and branching factor=2

Thus, the proposed algorithm proves to be effective in automatically generating trigonometric identities with greater diversity compared to traditional textbook versions.

A.9. Extension of the proposed approach to other domains of mathematics

Although the chapter primarily focuses on the synthesis of trigonometric identities, the proposed algorithms have vast potential for application in other mathematical domains, including algebra, propositional logic, set theory, and beyond. Below, a subset of rules from a few selected domains is illustrated, demonstrating the scope of the algorithm in synthesizing new identities within each domain.

A.9.1. Illustration of Problem Generation in Algebraic Domain

The proposed ADGBFS algorithm in algebraic domain is illustrated by Fig. A.5 with a search tree of depth = 3 and branching factor = 2. In Fig.A.5, the absolute diversity based f-cost of each node is indicated by <·>. The set of rules utilized for the above stated search tree is randomly selected using roulette wheel strategy from Table A.6. The RDGBFS algorithm can be utilized in a similar way for generating the search tree in algebraic domain.

A.9.2. Illustration of Problem Generation in Propositional Logic Domain

The proposed ADGBFS algorithm in propositional logic domain is illustrated by Fig. A.6 with a search tree of depth = 2 and branching factor = 2. In Fig.A.6, the absolute diversity based f-cost of each node is indicated by <·>. The set of rules utilized for the above stated search tree is randomly selected using roulette wheel strategy from Table A.5. The RDGBFS algorithm can be utilized in a similar way for generating the search tree in propositional logic domain.

A.9.3. Illustration of Problem Generation in Set Theory Domain

The proposed ADGBFS algorithm in set theory domain is illustrated by Fig. A.7 with a search tree of depth = 2 and branching factor = 2. In Fig.A.7, the absolute diversity based f-cost of each node is indicated by <·>. The set of rules utilized for the above stated search tree is randomly selected using roulette wheel strategy from Table A.8. The RDGBFS algorithm can be utilized in a similar way for generating the search tree in set theory domain.

6

Enhanced Generative Adversarial Network with Leaked Information (E-LeakGAN): A Deep Learning Framework for Synthesizing Computational Creativity in Scientific Domain

The present study introduces a novel method for synthesizing computational creativity in scientific domain, demonstrated through the automatic generation of diverse trigonometric identity problems. Prior approaches, including search-based algorithms, template-driven generation, and fine-tuned Large Language Models (LLMs), have often produced problems that closely resemble standard textbook examples due to limited transformation rules, rigid templates, or the absence of explicit mechanisms to promote diversity. To address these limitations, an Enhanced Generative Adversarial Network with Leaked Information (E-LeakGAN) is proposed, in which the convolutional neural network (CNN)-based discriminator transfers abstracted features from the input data to the generator,

thereby guiding it to generate diverse identity problems. The novelty within the CNN module includes: a) utilization of Trish activation function to enhance the discriminator's expressive capacity; b) introduction of a unique dual attention mechanism, that uses channel attention to highlight "which" semantic information is relevant and spatial attention to identify "where" the crucial semantic information is syntactically located; and c) development of an eigenvector guided pooling (EGP) technique that directs the attention-induced convolved features toward the highest semantic-syntactic content. Experiments conducted using three categories of evaluation metrics (BiLingual Evaluation Understudy (BLEU), Mean Normalized Cosine Similarity (MNCS), and Mean Normalized Diversity (MND)) underscore the superior performance of E-LeakGAN compared to traditional techniques. The approach also yields consistent results across other mathematical domains such as algebra and calculus, making it an effective tool for generating chapter-end exercises. Additionally, it supports the development of diverse worksheets and exam questions, assisting educators in designing comprehensive assessment materials.

6.1 INTRODUCTION

Traditionally, creativity [1-3] has been celebrated in artistic fields, manifesting through unique and innovative expressions in literature, visual arts, and performance arts. Artists have used creativity to push the boundaries of their mediums, creating poetry, novels, paintings, music, and performances that capture human imagination and emotion. However, with technological progress and the scientific revolution, the notion of creativity has transcended its artistic confines and become a crucial element in scientific breakthroughs [4]. Creative thinking is now recognized as an essential component within the realm of scientific discovery and innovation, as evidenced by milestones such as the theory of relativity [5], the discovery of the DNA structure [6], and the development of quantum mechanics [7]. In recent times, the rise of artificial intelligence (AI) [8] has significantly broadened the scope of creativity, leading to the emergence of computational creativity [9-10]. This burgeoning field focuses on enabling machines to mimic and replicate human-like creative processes [11]. The present work deals with this emerging field of AI, with a primary focus on synthesizing computational creativity in the scientific domain.

Existing research on synthesizing computational creativity in the scientific arena focuses on the generation of new mathematical conjectures or theories [12-13], the prediction of novel molecular structures [14-16], the discovery of new scientific concepts [17-18], and several other applications [19-21]. The current work makes a humble attempt to explore and implement computational creativity in the scientific

realm, specifically focusing on the creation of diverse trigonometric identity problems for use as chapter-end exercises. It is important to emphasize that the mere automation of mathematical problem generation by computational systems cannot be considered creative. For a system to be regarded as computationally creative, its output must exhibit a significant degree of diversity when compared to existing problems [28], [81], [10]. This criterion ensures that the generated content is not simply a replication of known items but instead contributes something novel to the domain. The aforementioned rationale also aligns with existing literature [22-25]. Therefore, the selection of automated problem generation with enhanced diversity as a means of demonstrating computational creativity makes sense for the present application.

Previous works [26-27] pertaining to the current domain have primarily focused on automating the generation of trigonometric identities using traditional search-based algorithms, which rely on rule-based substitutions applied to symbolic expression trees. However, these methods often yield identities that closely resemble existing problems, largely due to the limited set of substitution rules employed. Very recently, Ghosh et al. [28] proposed a best-first search strategy over an OR-tree structure, wherein rule-based substitutions are applied to trigonometric expressions and the selection of the best identity is guided by a diversity-optimized cost function. While this approach enhances the diversity of the generated identities compared to conventional textbook examples, it suffers from scalability issues. Specifically, achieving higher diversity necessitates increasing both the depth and branching factor of the search tree, which significantly escalates the algorithm's time and space complexity. With the advent of Large Language Models (LLMs), several studies [29-31] have explored automated mathematical problem generation by leveraging minor parameter tuning to introduce diversity. However, these approaches still fall short in ensuring substantial diversity among the generated problems, primarily due to the absence of a dedicated module specifically designed to promote and control diversity.

A recent study by Ghosh et al. [32] proposed a Generative Adversarial Network (GAN) [33] based framework for automatically generating trigonometric identities with enhanced diversity. The architecture uses LeakGAN (Generative Adversarial Network with Leaked Information) [34], which consist of two components: a generator and a discriminator. These networks engage in an adversarial process where the generator creates new identity problems, and the discriminator evaluates their validity by classifying them as real or fake relative to the original dataset. The above iterative competition sharpens the model's ability to produce accurate and novel outputs. In addition to classification, the discriminator also leaks high-level

abstract features to guide the generator in producing coherent and diverse identities. Moreover, unlike traditional GANs, the generator in LeakGAN comprises two modules: a manager and a worker. The manager interprets the leaked features to generate a goal vector that captures key trigonometric relationships and directs the generation process. The worker uses the above goal vector to produce identity components that are relevant and logically structured. This hierarchical design, driven by leaked information, enables the creation of semantically rich identities. However, the discriminator in [32] relies on a conventional Convolutional Neural Network (CNN) with a single attention mechanism, a standard activation function, and a basic mixed pooling strategy (a linear combination of max and average pooling). These limitations hinder the discriminator's ability to highlight critical features within the leaked information that are essential for generating diverse outputs. The present work addresses the above limitation by introducing an Enhanced Generative Adversarial Network with Leaked Information (E-LeakGAN), which improves the discriminator architecture to more effectively ensure validity, coherence, and diversity of the generated identities.

The discriminator network of the E-LeakGAN model is composed of three main layers. The first layer applies a convolution operation to extract both semantic and syntactic features from the input, which consists of a manually curated dataset of standard textbook identity problems. This operation is followed by the application of a new activation function, Trish [35], that produces small yet informative outputs for negative inputs thereby enabling the network to learn more nuanced and complex patterns from the input data. The second layer introduces a dual attention mechanism aimed at highlighting the most relevant information within the convolved features. It consists of two sequential modules: channel attention and spatial attention. The channel attention module identifies the most important semantic features, such as function types, algebraic structures, factorization patterns, and reciprocal relations. Meanwhile, the spatial attention module focuses on the syntactic locations of these semantic features, such as determining whether they appear on the left-hand or right-hand side of the identity. The attention maps from both modules are fused with the original convolved features to refine the representation, enabling the network to focus more effectively on critical semantic and syntactic information. The third layer applies a novel pooling method referred to as eigenvector guided pooling (EGP). Its purpose is to reduce the spatial dimensions of the features while preserving the most meaningful semantic induced syntactic content related to trigonometric identities such as reciprocal relationships, complementary functions, and the like. EGP achieves the above aim by projecting the attention induced convolved features onto the eigenvector corresponding to the largest eigenvalue, thereby aligning them in the

direction of maximum semantic-syntactic relevance. The acquired pooled vector is finally passed to a fully connected layer, which classifies the input as real or fake. In addition, this vector is also leaked to the generator to guide the synthesis of diverse and valid trigonometric identities.

The proposed E-LeakGAN model has proven to be highly effective in generating diverse trigonometric identities when evaluated against state-of-the-art (SOTA) methods using three categories of evaluation metrics: BiLingual Evaluation Understudy (BLEU) [36], Mean Normalized Cosine Similarity (MNCS), and Mean Normalized Diversity (MND), which estimate syntactic similarity, semantic similarity, and diversity levels, respectively, with respect to existing textbook problems. Feedback from a panel of 23 mathematics experts further underscores the model's ability to produce trigonometric identity problems that are both exceptional in quality and originality. Moreover, the proposed method has demonstrated efficacy in generating identity problems in other domains of mathematics, such as algebra and calculus. Thus, it serves as a promising resource for automatic textbook content generation to be used as chapter-end exercises. Additionally, the technique is well-suited for the automated generation of diverse worksheets, which will aid students in improving their mathematical skills and support their preparation for various competitive examinations. Furthermore, the model can benefit educators by automating exam chapter content generation, enabling the creation of new and varied questions for different exams.

The rest of the manuscript is structured as follows. Chapter 2 provides a comprehensive review of existing research related to computational creativity, laying the foundation for the study. In Chapter 3, the focus shifts to an in-depth examination of the principles and methodologies required for generating trigonometric identities. Chapter 4 discusses the experimental procedures and results, demonstrating the effectiveness of the proposed approach through detailed analyses. Chapter 5 highlights the advantages and limitations of the proposed scheme, whereas Chapter 6 summarizes the key findings and draws conclusions based on the study's outcomes.

6.2 RELATED WORKS

Previous research studies [26-31], [37-41] that attempted to generate novel mathematical problems have relied on template-based methods, search-based strategies, or fine-tuning of LLMs. Table 6.1 provides a summary of these prior studies, including their main results and limitations. It is apparent from the table that these studies were unable to invoke significant diversity in the generated problems compared to existing ones. However, study [32] based on GAN was able to

Table 6.1
Literature Review on Automatic Problem Generation

Sl. No.	Authors	Technique/Algorithm	Main Results	Limitations
1.	Brito et al. [37]	Template based method	Mathematical problems are generated by varying coefficients in the given template	The generated problems lack diversity due to their confinement to predefined template structures
2.	Zhang et al. [38]	Sumudu Transform induced template based method	Mathematical identities are generated by modifying coefficients in templates derived from structural properties of the Sumudu Transform	The generated problems lack diversity due to their confinement to predefined template structures
3.	Tabuguia et al. [39]	Hyper-geometric sequence induced template based method	Trigonometric identities are generated by varying coefficients in the given hyper-geometric sequence	The generated problems lack diversity due to their confinement to predefined template structures
4.	Liu et al. [26]	randomized Breadth-First Search (rBFS)	Trigonometric identities are generated by applying rule-based substitutions over symbolic expression trees, using variations in trigonometric functions and transformation rules	The diversity of the generated identities are limited due to the use of a small set of transformation rules
5.	Pearce et al. [27]	Best-First Search	Trigonometric identities are generated by applying rule-based substitutions over trigonometric expression graphs through variation in known functions and algebraic operations	The diversity of the generated identities is limited due to the use of a small set of rule-based substitutions
6.	Ghosh et al. [28]	Best-First Search	Diverse trigonometric identities are generated through iterative expansions over an OR-tree structure, where rule-based substitutions are applied to trigonometric expressions, and selection is guided by diversity-optimized cost function	The approach lacks scalability, as increasing diversity requires deeper and wider search trees, leading to high time and space complexity
7.	Pham et al. [40]	LLM prompt-based generation	Mathematical problems are generated by OpenAI's ChatGPT using well-crafted prompts	The generated problems mostly mirror existing problem structures, indicating limited diversity
8.	Deroy et al. [41]	LLM prompt-based generation	Mathematical problems are generated by Google's Gemini using well-crafted prompts	The generated problems mostly mirror existing problem structures, indicating limited diversity
9.	Yu et al. [29]	Fine-tuning LLM	Mathematical problems are generated using forward and backward reasoning and answer augmentation techniques with LLMs	The generated problems mostly mirror existing problem structures, indicating limited diversity
10.	Drori et al. [30]	Fine-tuning LLM	University-level mathematics problems generation using OpenAI's Codex framework by randomizing input sampling, prompt structure, and generation parameters	Fine-tuning with randomization techniques and prompt modifications introduces only subtle variations in the generated problems
11.	Zhao et al. [31]	Multi-LLM system	Mathematical problems are generated by a multi-LLM pipeline that leverages a wide range of foundational concepts and filters out low-quality outputs through rejection sampling	The use of a variety of foundational concepts and rejection sampling introduces only moderate diversity in the generated problems
12.	Ghosh et al. [32]	Generative Adversarial Network (GAN)	Trigonometric identities are generated by a LeakGAN model, where leaked features from the discriminator introduce diversity among the generated identities.	The discriminator relies on a single attention mechanism, a conventional activation function, and mixed pooling, which partially hampers the diversity of the generated problems.

introduce some diversity in the generated problems, but the results were not consistent due a poor discriminator architecture. Hence, the proposed approach attempts to rectify the architecture in [32] to ensure coherent and diverse generation of identity problems.

6.3 PROBLEM FORMULATION AND APPROACH

The present section outlines the steps involved in dataset pre-processing, provides a detailed description of the proposed E-LeakGAN model architecture, explains the post-processing procedures applied to the generated identities, and describes the metrics used to quantitatively evaluate the quality of the generated problems.

6.3.1 Dataset Pre-Processing

The dataset utilized for training the E-LeakGAN model comprises trigonometric identity problems acquired from existing mathematics textbooks. These identity problems have been manually compiled in text format. The dataset is initially pre-processed through three main steps which include: a) filtering of textual artifacts, b) tokenization of the identity problems, and
c) conversion of the acquired tokens into vector embeddings. Details of each pre-processing step are discussed below.

6.3.1.1 Filtering of Textual Artifacts

Since the dataset utilized in the present study has been manually prepared, it is susceptible to various textual artifacts. The following criteria have been applied to eliminate such artifacts before proceeding with tokenization.
a) Removal of duplicate identity problems from the manually curated dataset to ensure uniqueness and prevent data redundancy.
b) Elimination of identity problems containing typographical errors (such as incorrectly placed parentheses, missing operands or operators, misspelled function names, etc.) and/or syntactic incompleteness (i.e., missing terms in the left hand side (LHS) or right hand side (RHS) of the equation).
c) Removal of unnecessary whitespaces, particularly those occurring between function names, operators, powers, and parentheses, to maintain consistency in representation.
d) Removal of redundant outer brackets from the LHS and/or RHS of the identity problems i.e., when the LHS or RHS of an equation is fully enclosed within brackets

that do not affect the semantic meaning or evaluation of the expression. For instance, in the identity $(\sin^2 x + \cos^2 x) = 1$, the brackets on the LHS are eliminated.

e) Eradication of full stops appearing at the end of identity problems, as they are treated as non-functional punctuation marks in mathematical expressions and do not contribute to their meaning.

6.3.1.2 Tokenization of the Identity Problems

The second step involved in dataset pre-processing is the tokenization of a given input mathematical equation *P*, which segments it into ordered set of units called tokens. Mathematically, tokenization can be described by (6.1).

$$\Im = Tokenize(P) = [p_1, p_2, ..., p_t], \tag{6.1}$$

where, each token p_i denotes a function, variable, operator, or constant. Next, from the acquired token sequence, a vocabulary of unique tokens is implemented using (6.2).

$$\Phi = Unique(\Im) = \{v_1, v_2, ..., v_m\}, \tag{6.2}$$

where, $m = |\Phi|$ (i.e., $|\cdot|$ represents the cardinality of set Φ). After this, each unique token obtained from the previous step is bijectively mapped to a token identifier (ID) using (6.3).

$$ID : \Phi \to \{0,1, ..., m-1\}, \tag{6.3}$$

where, each $v_j \in \Phi,$ a unique ID is assigned i.e., $ID(v_j) = j - 1.$

Finally the tokenized input $\Im = [p_1, p_2, ..., p_t]$ is mapped to a series of token IDs $W = [w_1, w_2, ..., w_t]$ which will be utilized in the subsequent step for the formulation of the vector embeddings.

6.3.1.3 Formation of Vector Embeddings

The final step of dataset pre-processing involves the formulation of vector embeddings from the generated token IDs, which are subsequently used for model training. Each token ID is converted into a unique $1 \times N$ vector. A traditional hashing algorithm [42] is employed for generating these vector embeddings, which is detailed in Appendix A.1 of the supplementary document.

6.3.2 Proposed E-LeakGAN Model

The trigonometric identity generation problem is formalized as a sequential decision-making procedure [43], wherein the previously generated terms are regarded as the current state at each time step t. This state is represented as $r_t = (p_1, p_2, ..., p_i, ..., p_t)$

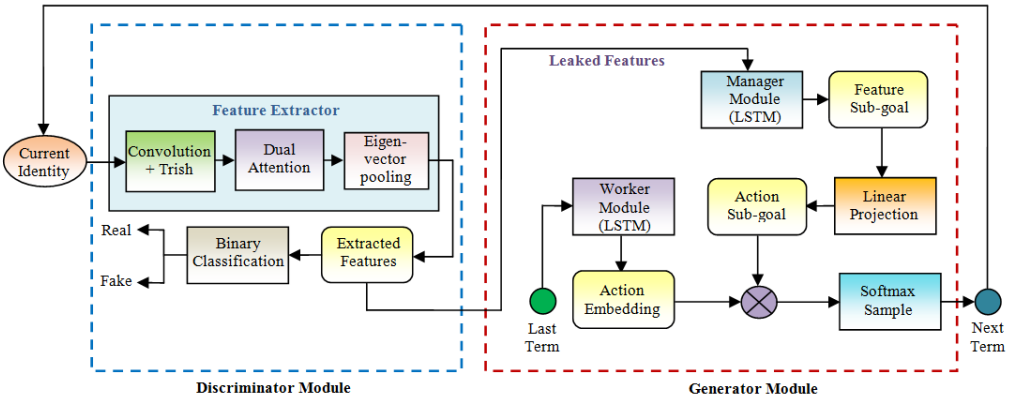


Fig. 6.1. Block diagram of the proposed E-LeakGAN model for trigonometric identity

with p_i denoting a mathematical term based token in the given expression P. A generative network G_α, parameterized by α, defines a stochastic policy $G_\alpha(p_{t+1} \mid r_t)$ which represents the conditional probability of generating the next token p_{t+1} given the current partial sequence r_t. This distribution is used to sample the next term in the expression, as shown by (6.4).

$$G_\alpha(p_{t+1} \mid r_t) = \Pr(p_{t+1} \mid p_1, p_2, \ldots, p_t; \alpha), \tag{6.4}$$

Furthermore, a discriminative model D_ψ, parameterized by ψ, is trained to provide a scalar mentoring signal $D_\psi(r_T)$ to the generative network. This signal facilitates the adjustment of the generative network's parameters once the entire expression r_T has been generated.

However, even with the principled approach of the aforementioned adversarial training, the utility of the scalar mentoring signal wanes as the length and/or complexity of the expression increases. To address this challenge, the discriminator within the proposed E-LeakGAN framework is permitted to impart additional information, known as feature vector e_t of the current expression r_t, to the generator. Furthermore, the proposed framework utilizes a hierarchical reinforcement learning (RL) architecture [34] as a strategic approach for integrating the leaked information e_t into the generation procedure. The schematic overview of the proposed E-LeakGAN model is depicted in Fig. 6.1 and its details are discussed in the following sub-sections.

6.3.2.1 Mentoring Signals Produced by Leaking Information from the Discriminator

The discriminator network D_ψ used for the present application is based on a CNN that consists of two main modules. The first module consists of a feature extractor represented by $E(\cdot;\psi_e)$ with convolutional weight vector ψ_e while the second module

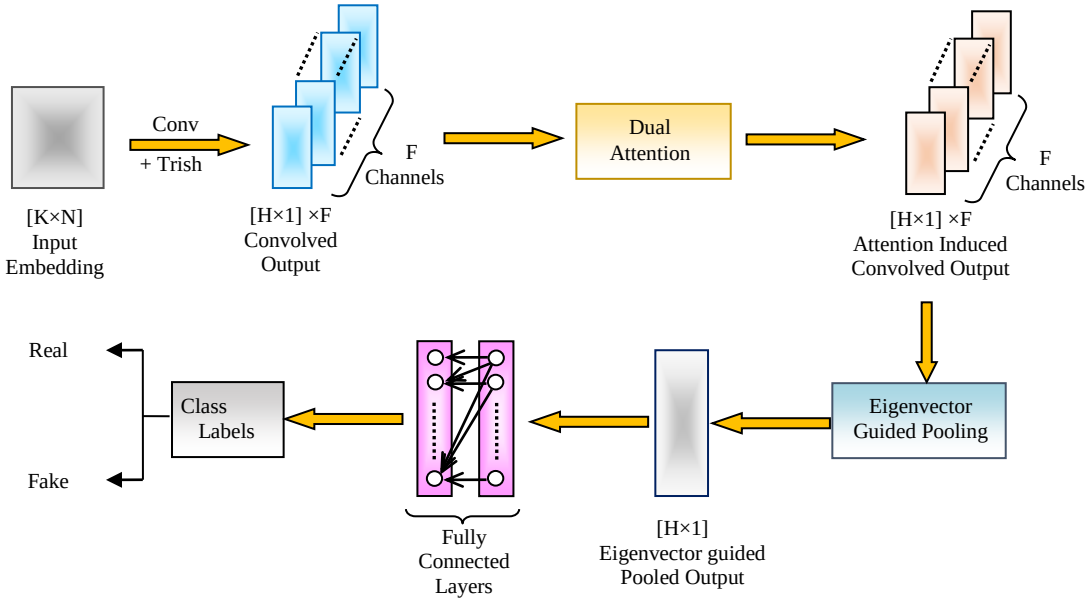


Fig. 6.2. Overview of the proposed discriminator architecture of the E-LeakGAN model

consists of a sigmoid [44] based fully connected layer having weight vector ψ_l. The discriminator operation for a given input expression *r* can be expressed by (6.5).

$$D_{\psi}(r) = \sigma(\psi_l^T E(r;\psi_e)) = \sigma(\psi_l^T e), \tag{6.5}$$

where, $\psi = (\psi_e, \psi_l)$; σ signifies a sigmoid activation function; $e = E(r;\psi_e)$ denotes the feature vector of *r* within the discriminator that needs to be imparted or leaked to the generator network; and $(\cdot)^T$ represents the matrix transposition operation. Unlike the scalar signal typically found in traditional adversarial networks [33], the feature vector *e* here acts as a notably richer and more informative guidance signal for the generator. This is because it can convey the syntactic and semantic information about the currently generated mathematical expression from the CNN based extracted feature necessary for the generation of valid and coherent identity problems.

For the current scenario, the feature extractor $E(\cdot;\psi_e)$ is implemented using a novel CNN architecture. The proposed CNN model consists of convolution, attention, and pooling layers which are elaborated in the following sub-section.

6.3.2.2 Architecture of the Proposed Discriminator Model

The discriminator model, as illustrated in Fig. 6.2, is constructed utilizing a CNN whose each layer is composed of a sequence of operations, outlined as follows.

6.3.2.2.1 Convolution with Trish Activation Function

The first stage of the discriminator model employs convolution operation upon an embedded matrix V using (6.6).

$$Y_j = \varepsilon((V_i * w_{ij}) + b_j), \tag{6.6}$$

Where, $V \in \Re^{K \times N}$ is the input embedded matrix of dimension $K \times N$; $Y_j \in \Re^{H \times 1}$ signifies the j^{th} convolved output having dimension of $H \times 1$; $\varepsilon(\cdot)$ denotes an activation function; w_{ij} indicates the weights of the convolutional kernel; and b_j represents the bias term; and "$*$" denotes the convolution operator. The aforementioned convolution operation is repeated for F number of differerent filter kernels to produce a final convolved feature map $Q \in \Re^{H \times 1 \times F}$ having size $H \times 1 \times F$.

The activation function employed for the current scenario is the Trish [35] activation function. The mathematical expression of Trish function is represented by (6.7).

$$\varepsilon(x) = x \cdot \sigma[\ln(1 + \tanh(x))], \tag{6.7}$$

where, $\sigma(\cdot)$ represents the sigmoid activation function. Although any standard activation function could be utilized for the present purpose, the authors opted for Trish for the following reasons. First, Trish offers effective regularization through its smooth, differentiable structure combined with a bounded sigmoid component, enabling controlled information flow and mitigating vanishing gradients. Second, its non-monotonic behavior allows Trish to yield small yet informative values for negative inputs, enhancing the network's representational capacity and enabling it to capture complex, subtle patterns more effectively. Third, Trish introduces adaptive sparsity through the interplay of tanh and sigmoid [35], selectively activating neurons based on nuanced input characteristics, thereby enriching the learning process. Finally, the smooth and continuous nature of Trish ensures robust gradient propagation, facilitating faster and more stable convergence during optimization and promoting stronger generalization across various datasets and tasks.

6.3.2.2.2 Proposed Dual Attention Mechanism

The rationale behind the present dual attention mechanism is to accentuate vital semantic and syntactic information within the convolved feature vector, ensuring that the most relevant information are communicated to the subsequent (pooling) layer. This dual attention mechanism consists of two primary components: channel attention, which accentuates features across channel dimensions of the convolved feature, and spatial attention, which emphasizes features across spatial dimensions of the convolved feature. The transformation of the convolved feature vector into an attention-induced feature vector is elucidated below and illustrated in Fig. 6.3.

(a) Channel Attention Module

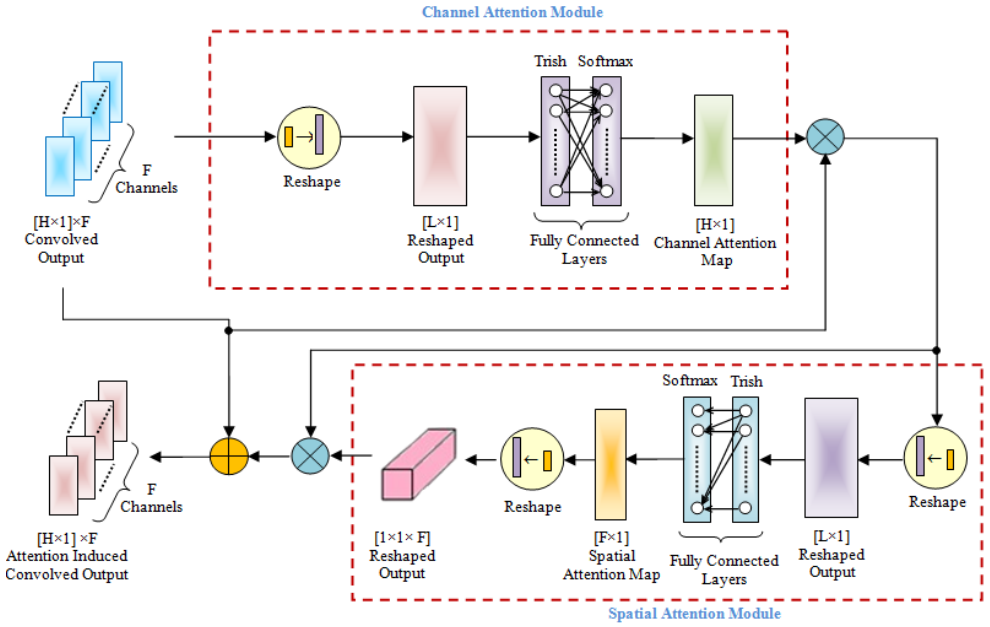


Fig. 6.3. Detailed architecture of the proposed dual attention module in E-LeakGAN

The primary objective of this module is to generate an attention map by leveraging the inter-channel relationships within the convolved features. Each channel in the feature map functions as an individual detector, capturing specific semantic aspects of a trigonometric identity such as function types, algebraic structures, factorization patterns, reciprocal relations, complementary relations, and other related constructs. The present mechanism focuses on identifying "which" semantic features within the input data (i.e., the convolved feature map) are most significant. The aforesaid operation begins by first reshaping the input feature map of dimension $H\times1\times F$ to $L\times1$ dimension (where, $L=H\times1\times F$) and feeding the reshaped output to two fully connected layers guided by Trish and Softmax [45] activation functions respectively. Next, the output of the fully-connected layer (represented by $B\in\Re^{H\times1}$ having size $H\times1$) is element wise multiplied along the dimensions of H with the convolved feature map which is mathematically expressed by (6.8).

$$C=Q\otimes B, \tag{6.8}$$

where, $C\in\Re^{H\times1\times F}$ signifies the channel attention induced feature map of size $H\times1\times F$ and "$\otimes$" indicates the element wise multiplication operator.

(b) Spatial Attention Module

The main function of this module is to generate an attention map by leveraging the inter-spatial relationships within the convolved features. While channel attention prioritizes "which" semantic information are most relevant, spatial attention focuses on pinpointing "where" relevant semantic information syntactically lies, such as whether it appears on the left-hand or right-hand side of the identity, the position

where an operator needs to be applied to maintain equality, and other similar considerations. The aforementioned operation begins by first reshaping the channel based feature map $C \in \Re^{H \times 1 \times F}$ into $L \times 1$ dimension and then transferring the reshaped output to two fully connected layers guided by Trish and Softmax activation functions respectively. After this, the output of the fully-connected layer (denoted by $J \in \Re^{F \times 1}$) is reshaped and then element wise multiplied along the dimensions of F with the channel based feature map which is mathematically expressed by (6.9).

$$S = C \otimes J, \tag{6.9}$$

where, $S \in \Re^{H \times 1 \times F}$ signifies the spatial attention induced feature map of size $H \times 1 \times F.$

Finally, the spatial attention-induced feature map is combined with the original convolved feature map using (6.10) to obtain the final attention-induced feature map:

$$T = S + Q, \tag{6.10}$$

where, $T \in \Re^{H \times 1 \times F}$ signifies the final refined feature map having size $H \times 1 \times F$ and "+" indicates addition operator. The rationale behind this operation lies in further enhancing the discriminative power of the attention mechanism by highlighting only the crucial features, such as which functional relationships are to be considered and where such relationships exist within the given identity data.

6.3.2.2.3 Proposed Eigenvector Guided Pooling

Following the dual attention layer, an eigenvector guided pooling (EGP) mechanism (illustrated in Fig. 6.4) is applied to the attention-induced convolved features. The main objective of this operation is to reduce the spatial dimensions of the features while emphasizing only the crucial semantic induced syntactic information related to trigonometric identities, such as reciprocal relations, complementary relations, and the fact that both the left hand side and right hand side always simplify to 1. This EGP process involves the following steps. First, the attention induced convolved matrix of size $H \times 1 \times F$ is reshaped into size $H \times F.$ Second, the covariance of the reshaped matrix $R \in \Re^{H \times F}$ is computed using (6.11).

$$M = R^T R, \tag{6.11}$$

where, $M \in \Re^{F \times F}$ represent the covariance matrix of size $F \times F$ and $(\cdot)^T$ denotes the matrix transposition operator. After the covariance computation, the eigenvalues of M are evaluated from the roots of the equation $|M - \lambda I| = 0.$ Next, the eigenvectors corresponding to the eigenvalues are also evaluated. Finally, the eigenvector $Z \in \Re^{F \times 1}$ pertaining to the largest eigenvalue is selected and projected along the reshaped matrix R using (6.12).

$$U = R \times Z, \tag{6.12}$$

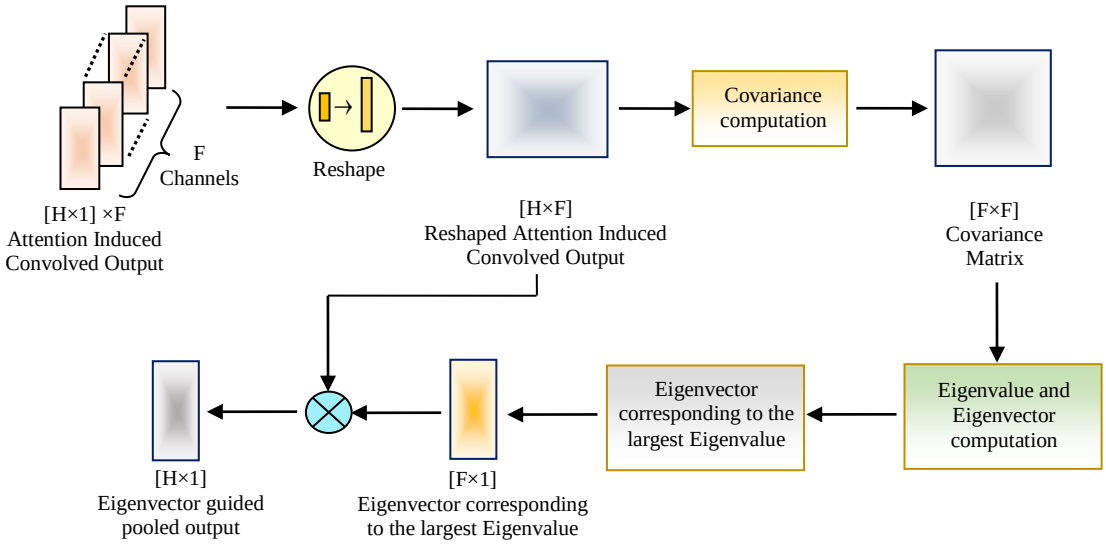


Fig. 6.4. Detailed architecture of the proposed eigenvector guided pooling mechanism in E-LeakGAN

where, $U \in \Re^{H\times 1}$ represents the pooled output of size $H\times 1$ and "$\times$" denotes the matrix multiplication operator. The eigenvector corresponding to the largest eigenvalue captures the direction along which the features exhibit the highest semantic-syntactic informational content. Thus, by projecting the features along this direction, the proposed pooling operation effectively reduces dimensionality while emphasizing the most prominent information within the feature matrix.

6.3.2.2.4. Fully Connected or Dense layers

The EGP features undergo flattening before being fed to fully-connected or dense layers, tasked with classifying the features into two categories: real or fake. At the final layer of the dense network, the sigmoid activation function is applied to convert the network's output into probability scores. Additionally, within the hidden layers of the dense architecture, the Trish activation function is employed to enhance the model's learning capacity and expedite convergence.

6.3.2.3 Overview of the Generator Architecture

To integrate leaked features (i.e., EGP features) into each step *t* of the generation process, the generator incorporates a hierarchical reinforcement learning module [34]. This module comprises a manager module (MM) and a worker module (WM). The MM, implemented using a Long Short-Term Memory (LSTM) [46], takes the EGP vector as input at each step *t* and produces a goal vector o_t. The WM then utilizes this goal vector to guide the generation of the subsequent mathematical expression, aiming to approach regions with higher rewards within $E(r;\psi_e)$. The following section will delve into the intricacies of the generator model in the

proposed E-LeakGAN and explain how the MM and WM are trained using mentoring signals from the discriminator model.

6.3.2.3.1 Manager Module (MM)

The MM and WM commence their computations with initial hidden states initialized to zero, denoted as $h_{MM}^{(0)}$ and $h_{WM}^{(0)}$ respectively. At each step, the MM receives the leaked vector from the discriminator, which is subsequently integrated with the current hidden state of the MM to generate a goal vector. The aforesaid operations are mathematically illustrated by (6.13) and (6.14).

$$\hat{o}_t, h_{MM}^{(t)} = LSTM\ (e_t, h_{MM}^{(t-1)}; \alpha_{MM}), \tag{6.13}$$

$$o_t = \hat{o}_t / \| \hat{o}_t \|, \tag{6.14}$$

where, e_t represents the leaked feature vector, $h_{MM}^{(t)}$ denotes the hidden vector of LSTM; α_{MM} signifies the parameters of the LSTM network; and $LSTM(\cdot;\alpha_{MM})$ indicates the MM designed using LSTM.

To integrate the goals generated by MM, a distinct methodology is undertaken. A linear transformation using a weight matrix β_χ is employed upon the summation of the q most recent goals to produce a g-dimensional goal embedding vector. The afore-said operation is mathematically expressed by (6.15).

$$u_t = \beta_\chi (\sum_{i=1}^{q} o_{t-i}), \tag{6.15}$$

where, $u_t \in \Re^{g\times 1}$ represents the goal vector of size $g \times 1$.

6.3.2.3.2 Worker Module (WM)

Once the goal embedding vector u_t is available, the WM modelled by a LSTM, processes the current expression p_t to produce a matrix Z_t. To derive the final action space distribution for the current state r_t, two steps are executed. First, Z_t undergoes a matrix product operation with u_t. Second, the result is subjected to a fully connected layer guided by Softmax activation function, yielding the final action space distribution. The aforementioned operations are mathematically illustrated by (6.16) and (6.17).

$$Z_t, h_{WM}^{(t)} = LSTM\ (p_t, h_{WM}^{(t-1)}; \alpha_{WM}), \tag{6.16}$$

$$G_\alpha(\cdot|r_t) = soft\max(Z_t \cdot u_t / \mu), \tag{6.17}$$

where, Z_t represents the matrix comprising the current vector for all mathematical expressions; μ denotes the temperature parameter that governs the entropy of the generation procedure; and $LSTM(\cdot;\alpha_{WM})$ signifies the WM constructed using the LSTM network comprising parameters denoted by α_{WM}.

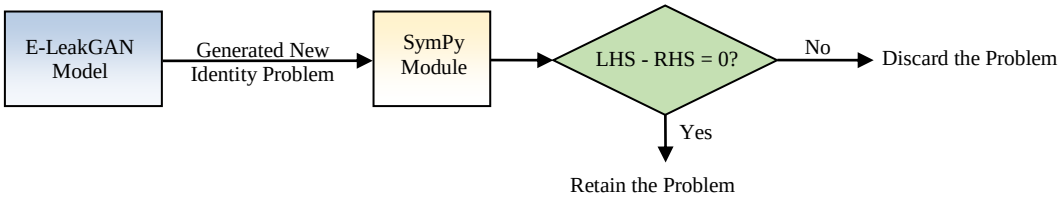


Fig. 6.5. Mathematical validation of the newly generated identities using the SymPy

6.3.2.4 Generalizability of the Proposed Framework

Although the present study showcases the proposed E-LeakGAN model in the context of generating trigonometric identity problems, the model's architecture is inherently domain-agnostic and does not rely on any domain-specific assumptions. The discriminator's Trish induced convolution layer, dual attention technique, and EGP mechanism are designed to extract semantically rich and syntactically coherent features from mathematical expressions, irrespective of their domain. Likewise, the generator's hierarchical learning framework using the manager and worker modules can adapt to any sequential symbolic structure. Therefore, provided a curated dataset of valid expressions, the same framework can be seamlessly extended to other mathematical domains such as algebra and calculus, without any architectural modifications. Thus, the generalizability of the proposed architectural framework significantly broadens its application in the field of computational creativity.

6.3.3 Post-Processing of Generated Identity Problems

After the identity problems are generated by the E-LeakGAN model, a post-processing phase is undertaken to ensure their mathematical validity. To accomplish the aforesaid task, the SymPy module [47] of Python is utilized that consists of a powerful symbolic mathematics library capable of performing algebraic simplification, equation solving, and identity verification. For the present work, each generated identity is split into its left hand side (LHS) and right hand side (RHS) components, which are then independently simplified. The identity is considered valid if the difference between the LHS and RHS evaluates to zero, indicating mathematical equivalence between both sides. Identities that satisfy this condition are retained, while those that do not are automatically discarded. This automated symbolic verification framework ensures that only mathematically correct identities are included in the final dataset, thereby maintaining accuracy and eliminating the

need for manual validation. The schematic view of the above process is illustrated by Fig. 6.5.

6.3.4 Quantitative Assessment of the Generated Problems

The main objective of the current study is to generate a variety of trigonometric identity problems which are significantly different from the problems present in the training dataset. To achieve the aforesaid objective, three categories of evaluation metrics are utilized which include: a) BiLingual Evaluation Understudy (BLEU) score, b) Mean Normalized Cosine Similarity (MNCS) and c) Mean Normalized Diversity (MND). The details of the above three evaluation metrics are presented below.

6.3.4.1 BiLingual Evaluation Understudy (BLEU) score

BLEU (BiLingual Evaluation Understudy) [36] is a precision based metric commonly used to evaluate the quality of machine generated content by comparing it to one or more reference content. It measures the n-gram overlap between the generated and reference sequences, penalizing shorter outputs through a brevity penalty [36]. Though it is more popularly used for natural language generation tasks, several studies [48-50] have also applied BLEU to measure similarity in mathematical equation generation. For mathematical equations, the BLEU score is computed by assessing the n-gram overlap between the generated and reference equations, similar to its use in text generation tasks, thereby providing an estimate of syntactic similarity among the generated and existing problems. Thus, a lower BLEU score indicates lower syntactic similarity (i.e., higher syntactic variation) between the generated identities and the existing ones [51].

6.3.4.2. Mean Normalized Cosine Similarity

For the current study, cosine similarity is utilized to assess the semantic similarity between the generated and existing identity problems by measuring the cosine of the angle between their corresponding vector representations. To proceed with the present task, each generated identity and each identity in the training dataset is flattened into a one-dimensional (1D) column vector. If the sizes of the generated identity and the dataset identity do not match, the smaller vector is zero-padded [52] to match the larger one. The cosine similarity between the generated and existing identity problems is computed by (6.18).

$$S_i = \cos(G_i, D_j) = \frac{G_i \cdot D_j}{\| G_i \| \, \| D_j \|}, \tag{6.18}$$

where, G_i denotes the i^{th} vector of generated identity problem; D_j represents the j^{th} vector of the dataset identity problem; and $\| \cdot \|$ indicates the L2 norm operation. The output from (6.18) is normalized in the range [0, 1] using max-min normalization [53] technique. Next, the average value of the normalized cosine similarity of G_i across all the identities in the dataset is evaluated using (6.19).

$$\bar{S}_i = \frac{1}{n} \sum_{j=1}^{n} \cos(G_i, D_j), \tag{6.19}$$

where, n represents the total number of problems present in the training dataset; and $\bar{S}_i$ denotes the mean normalized cosine similarity of the i^{th} generated identity. Thus, for c generated problems, the overall average diversity score is computed across all problems using (6.20)

$$\bar{S} = \frac{1}{c} \sum_{i=1}^{c} \bar{S}_i, \tag{6.20}$$

Since the above computation involves averaging the normalized cosine similarity scores, the computation in (6.20) is hereafter referred to as the Mean Normalized Cosine Similarity (MNCS) score. It is important to mention that a lower MNCS score indicates reduced semantic similarity (i.e., greater semantic variation) between the generated identities and the existing ones.

6.3.4.3 Mean Normalized Diversity

A diversity based metric is employed in the present study to evaluate the extent to which the generated identities are structurally unique and do not resemble conventional textbook problems. The above metric is computed by counting the number of new terms introduced in a generated identity relative to those found in the dataset. Here, a term [28] is defined as a distinct component of an expression, obtained by removing mathematical operators from an algebraic or trigonometric expression. A term typically includes constants, variables, or functions of variables, and serves as a fundamental unit of the expression. For instance, for the given expression: $\sec^2 x - \tan^2 x + 2,$ a corresponding set $A = \{\sec^2 x, \tan^2 x, 2\}$ is constructed by omitting the operators. The number of terms is then given by the cardinality of the set, i.e., $|A| = 3.$ The diversity [28] of a generated identity problem G_i is evaluated by (6.21).

$$V_i = Diversity(G_i, D_j) = \left| T_{LHS}(G_i) - T_{LHS}(D_j) \right| + \left| T_{RHS}(G_i) - T_{RHS}(D_j) \right|, \tag{6.21}$$

where, V_i denotes the diversity score of the i^{th} generated identity; $T_{LHS}(G_i)$ and $T_{RHS}(G_i)$ represent the number of terms in the left and right side of the i^{th} generated

identity respectively; and $T_{LHS}(D_j)$ and $T_{RHS}(D_j)$ signify the number of terms in the left and right side of the j^{th} identity in the dataset respectively. For instance, if the first generated identity is $\sin^2 x + \frac{1}{\sec^2 x} = \cos ec^2 x - \cot^2 x,$ and the first identity in the dataset is $\sin^2 x + \cos^2 x = 1,$ then,

$V_1 = |\{\sin^2 x, 1, \sec^2 x\} - \{\sin^2 x, \cos^2 x\}| + |\{\cos ec^2 x, \cot^2 x\} - \{1\}|$

$= |\{1, \sec^2 x\}| + |\{\cos ec^2 x, \cot^2 x\}| = 2 + 2 = 4.$

The output from (6.21) is normalized in the range [0, 1] using max-min normalization technique. Next, the average value of the normalized diversity of G_i across all the identities in the dataset is evaluated using (6.22).

$$\overline{V_i} = \frac{1}{n}\sum_{j=1}^{n} Diversity(G_i, D_j), \tag{6.22}$$

where, $\overline{V_i}$ represents the average normalized diversity score of the i^{th} generated identity. Thus, for c generated problems, the overall average diversity score is computed across all problems using (6.23).

$$\overline{V} = \frac{1}{c}\sum_{i=1}^{c} \overline{V}_i, \tag{6.23}$$

Since the above computation involves averaging the normalized diversity scores, the computation in (6.23) is hereafter referred to as the Mean Normalized Diversity (MND) score. It is important to note that a higher MND score indicates greater structural uniqueness or variety among the generated problems compared to the existing ones.

6.4 EXPERIMENTS AND RESULTS

The current section provides a comprehensive description of the experimental paradigm employed for generating mathematical identities and showcases the results of different analyses conducted to validate the effectiveness of the E-LeakGAN model.

6.4.1 Dataset

For the present identity generation task, a dataset has been prepared, consisting of 207,235 (i.e., two hundred seven thousand two hundred thirty-five) trigonometric identity problems. These problems have been selected from the introductory chapter on trigonometric identity proof problems, typically found in middle school mathematics textbooks. Each problem in the dataset has been meticulously organized in plain text format. Table 6.2 offers a preview of some problems contained within the prepared dataset. The complete dataset is available in [54].

Table 6.2
Exemplar Problems Included in the Prepared Dataset

Sl. No.	Trigonometric Identities Problems Acquired from Textbooks
1.	$\tan^2 x+\cot^2 x+2=\sec^2 x\cdot \operatorname{cosec}^2 x$
2.	$\frac{\sqrt{1+\cos x}}{\sqrt{1-\cos x}}=\operatorname{cosec} x+\cot x$
3.	$\operatorname{cosec}^2 x-\cot^2 x=1$
4.	$\frac{\tan x}{1-\cot x}+\frac{\cot x}{1-\tan x}=1+\tan x+\cot x$
5.	$\frac{\tan^2 x}{\tan^2 x-1}+\frac{\operatorname{cosec}^2 x}{\sec^2 x-\operatorname{cosec}^2 x}=\frac{1}{\sin^2 x-\cos^2 x}$

6.4.2 Experimental Settings

The present work employs an E-LeakGAN model to automate the generation of trigonometric identities, which incorporates both discriminator and generator networks. Optimal hyper-parameters for these networks have been selected using a random search algorithm [55]. This method offers significant advantages in hyper-parameter optimization by avoiding exhaustive testing of all possible combinations. Instead, it randomly samples the hyper-parameter space, leading to the discovery of optimal solutions more efficiently. The discriminator, implemented as a CNN module, is configured with parameters K=72 and N=40. The generator, based on LSTM architecture, features a goal embedding dimension of 16 and goal duration time of 4. The learning epochs have been set to 60, and the optimizer used is ADAM [56]. To mitigate over-fitting, a dropout rate of 0.3 and L2 regularization are also employed.

6.4.3 Generated Trigonometric Identity Problems

The proposed E-LeakGAN model is capable of generating a large number of trigonometric identity problems based on the training dataset. Table 6.3 presents a selection of these generated identities, demonstrating the model's ability to create problems that are suitable for use as chapter-end exercises. The detailed procedure regarding how the inputs are provided and how the new identities are generated is

Table 6.3
Samples of Trigonometric Identity Problems Generated by E-LeakGAN

Sl. No.	Generated Trigonometric Identities
1.	$1-\sin^2 x\cdot\cos^2 x=\sin^2 x\ (\sin^2 x+\cot^2 x)$
2.	$\cos x\cdot\sin x\sqrt{\cot^2 x+1}=\frac{\sqrt{\tan^2 x\cdot\cos^2 x-\sin^2 x+1}}{\sec x}$
3.	$\frac{\sqrt{1+\cos x}}{\cos x}=\frac{1}{\cot x\sqrt{1-\cos x}}$
4.	$\tan^2 x-\sin^2 x=\frac{\sin^2 x-1+(\cos^2 x)^2}{\cos^2 x}+\frac{1}{\mathrm{cosec}^2 x-1}$
5.	$\frac{\sec x(1-\sin^2 x)}{\sqrt{\frac{1}{\cot^2 x+1}}}=\sin x\cdot\cot^2 x\sqrt{1+\tan^2 x}$

presented in Appendix A.2 of the supplementary material. The identities produced by the model exhibit authentic trigonometric structure, both syntactically and semantically. This ensures that the identities are not only accurate but also valuable for educational purposes. However, a thorough assessment, comparing them to existing problems through both quantitative and qualitative evaluations, is essential to fully validate the effectiveness of the proposed model. These evaluations are detailed in the following sub-sections.

6.4.4 Performance Analysis of the Proposed E-LeakGAN with Respect to SOTA Methods

The quality of the trigonometric identities generated by the proposed E-LeakGAN is assessed by comparing it to the SOTA methods using BLEU, MNCS and MND metrics as discussed in Section 6.3.4. The results of the above comparative study are portrayed in Table 6.4. It is apparent from Table 6.4 that the RNN, LSTM, and VAE exhibit the least favorable outcomes across all evaluation metrics compared to other SOTA methods. This performance degradation can be attributed to the absence of discriminator and generator modules, which are crucial for producing high quality identities through adversarial training. In contrast, SeqGAN demonstrates improved performance over its predecessors due to the incorporation of both a discriminator and a generator, which work together through adversarial training to produce high

Table 6.4
Comparative Analysis of the Proposed E-LeakGAN with SOTA Models.

Algorithms	BLEU-2 ↓	BLEU-3 ↓	BLEU-4 ↓	BLEU-5 ↓	MNCS ↓	MND ↑
RNN [57]	0.981	0.911	0.704	0.590	0.462	0.430
LSTM [46]	0.974	0.883	0.661	0.557	0.327	0.485
VAE [58]	0.962	0.824	0.634	0.512	0.314	0.527
SeqGAN [59]	0.950	0.796	0.589	0.488	0.303	0.683
RankGAN [60]	0.948	0.770	0.572	0.469	0.276	0.732
VGAN [61]	0.926	0.752	0.551	0.442	0.268	0.765
ORGAN [62]	0.915	0.734	0.526	0.425	0.250	0.832
LeakGAN [34]	0.856	0.698	0.484	0.377	0.203	0.884
Modified LeakGAN [32]	0.827	0.676	0.470	0.345	0.143	0.956
E-LeakGAN	**0.801**	**0.654**	**0.452**	**0.302**	**0.092**	**0.981**

Table 6.5
Comparative Analysis of the Proposed E-LeakGAN with Search and Template based Models.

Algorithms	BLEU-2 ↓	BLEU-3 ↓	BLEU-4 ↓	BLEU-5 ↓	MNCS ↓	MND ↑
Brito et al. [37]	0.962	0.861	0.668	0.487	0.536	0.784
Zhang et al. [38]	0.943	0.850	0.654	0.473	0.480	0.805
Tabuguia et al. [39]	0.932	0.847	0.651	0.468	0.461	0.813
Liu et al. [26]	0.908	0.785	0.591	0.412	0.387	0.896
Pearce et al. [27]	0.865	0.748	0.567	0.384	0.346	0.921
Ghosh et al. [28]	0.830	0.716	0.483	0.336	0.210	0.952
E-LeakGAN	**0.801**	**0.654**	**0.452**	**0.302**	**0.092**	**0.981**

quality identities. However, RankGAN surpasses SeqGAN by using a ranker instead of a discriminator. The ranker provides a ranking score for each identity, reflecting its quality relative to other identities. VGAN further improves upon previous models by integrating a Variational Auto-Encoder (VAE) with a Recurrent Neural Network (RNN) as the generator. This combination ensures both semantic coherence and high quality in the generated identities. ORGAN outperforms the other GAN models by employing an objective-driven reinforcement learning technique during the generation process. This approach not only enhances the quality of the generated identities but also promotes diversity, resulting in a wider variety of outputs.

Nevertheless, LeakGAN delivers superior results compared to the other GANs. Unlike the aforementioned models, LeakGAN's architecture allows the generator to receive information leaked from the discriminator. Moreover, the improved performance of LeakGAN can also be attributed to the hierarchical interaction between the manager and worker components within the generator module, which facilitates the generation of high-quality problems. This mechanism helps the generator to produce identities with low BLEU and MNCS scores, and high MND

scores, while maintaining high quality, proper syntax, semantics, and diversity in the generated problems. However, LeakGAN still falls short of providing the best results because its discriminator module uses a traditional CNN that is unable to abstract the most crucial features from the input that has to be leaked to the generator module for producing enhanced quality identities. In contrast, the modified LeakGAN model improves upon the results produced by the traditional LeakGAN due to the incorporation of a channel attention mechanism, the utilization of the Mish activation function, and a mixed pooling layer within the discriminator module. However, it still fails to achieve the best results due to its reliance on a single attention mechanism, a conventional activation function, and a basic mixed pooling strategy that linearly combines standard max and average pooling operations.

The proposed E-LeakGAN model yields BLEU-2=0.801, BLEU-3=0.654, BLEU-4=0.452, BLEU-5=0.302, MNCS=0.092, and MND=0.981. The above results depict a reduction in BLEU-2 by 2.6%, BLEU-3 by 2.2%, BLEU-4 by 1.8%, BLEU-5 by 4.3%, and MNCS by 5.1%, along with an improvement in MND by 2.5% with respect to the best-performing baseline, i.e., modified LeakGAN. The reduction in BLEU and MNCS metrics indicates that the generated identities exhibit significant syntactic and semantic variation in comparison to existing problems. In contrast, the improvement in the MND metric indicates the presence of substantial diversity among the generated problems. Additionally, the results in Table 6.4 indicate a reduction of 18% in BLEU-2, 25.7% in BLEU-3, 25.2% in BLEU-4, 28.8% in BLEU-5, and 37% in MNCS, along with an improvement of 55.1% in MND with respect to the lowest-performing baseline, i.e., RNN. The best scores of the proposed model can be attributed to its dual attention mechanism, the utilization of a new activation function, Trish, and the formulation of EGP mechanism, whose combined effect advantageously guides the generator to produce identities of significant diversity compared to standard textbook problems.

6.4.5 Performance Analysis of the Proposed E-LeakGAN with Respect to Search and Template Based Methods

The performance of the proposed model is compared against the traditional techniques that rely on search and template based methods, using the same metrics discussed in Section 6.3.4. The results of this comparison are shown in Table 6.5. It

Table 6.6
Comparative Analysis of the Proposed E-LeakGAN with LLMs.

Algorithms	BLEU-2 ↓	BLEU-3 ↓	BLEU-4 ↓	BLEU-5 ↓	MNCS ↓	MND ↑
Deroy et al. [41]	0.896	0.752	0.548	0.435	0.463	0.886
Pham et al. [40]	0.881	0.740	0.537	0.424	0.454	0.895
Yu et al. [29]	0.875	0.738	0.532	0.412	0.441	0.908
Drori et al. [30]	0.841	0.689	0.496	0.371	0.294	0.952
Zhao et al. [31]	0.835	0.672	0.481	0.356	0.276	0.964
E-LeakGAN	**0.801**	**0.654**	**0.452**	**0.302**	**0.092**	**0.981**

is evident from the table that the worst performance is exhibited by the template based methods [37-39] as the generated problems lack significant diversity due to structural constraints imposed by predefined templates. In contrast, the search based methods [26-27] demonstrate noticeable improvements over the template based approaches but still show limited diversity in the generated outputs. This limitation arises due to the utilization of a small number of substitution rules. On the other hand, the method proposed in [28] yields better results, as its search algorithm is guided by diversity-optimized heuristic function. However, as discussed in Section 6.2, this approach lacks scalability i.e., in order to achieve higher diversity, there must be significant increase in depth and branching factor of the search tree. To maintain feasibility, a depth of 10 and a branching factor of 3 were used in the present experiment, which allowed this method to achieve the second-best performance. Finally, the best results are achieved by the proposed E-LeakGAN model, which exhibits a reduction of 2.9% in BLEU-2, 6.2% in BLEU-3, 3.1% in BLEU-4, 3.4% in BLEU-5, 11.8% in MNCS along with an improvement of 2.9% in MND compared to its nearest competitor in [28], owing to its unique discriminator architecture and the hierarchical interaction between the manager and worker modules within the generator.

6.4.6 Performance Analysis of the Proposed E-LeakGAN with Respect to LLMs

Table 6.6 presents the comparative results of E-LeakGAN and various LLM-based approaches in terms of BLEU, MNCS, and MND scores. It can be observed that studies [41], [40], and [29] yield closely aligned metric scores and represent the lowest-performing algorithms, primarily due to the absence of any module designed to promote diversity in problem generation. In contrast, study [30] shows noticeable improvements by introducing implicit diversity through randomization of input sampling, prompt structure, and generation parameters. However, it still falls short of

Table 6.7
Ablation Study of the Proposed E-LeakGAN

Variations in Discriminator Architecture	BLEU-2 ↓	BLEU-3 ↓	BLEU-4 ↓	BLEU-5 ↓	MNCS ↓	MND ↑
Convolution w/o EGP	0.872	0.756	0.589	0.432	0.302	0.852
Convolution + EGP	0.865	0.742	0.568	0.411	0.284	0.886
Convolution + Spatial Attention w/o EGP	0.853	0.736	0.554	0.390	0.273	0.905
Convolution + Channel Attention w/o EGP	0.847	0.720	0.535	0.371	0.258	0.918
Convolution + Dual Attention w/o EGP	0.842	0.711	0.503	0.352	0.246	0.932
Convolution + Spatial Attention + EGP	0.838	0.695	0.480	0.340	0.223	0.948
Convolution + Channel Attention + EGP	0.826	0.672	0.471	0.327	0.188	0.965
Convolution + Dual Attention + EGP	**0.801**	**0.654**	**0.452**	**0.302**	**0.092**	**0.981**

optimal results due to the lack of a dedicated architectural component for diversity enforcement. Further enhancement is observed in study [31], which promotes diversity by generating problems from a wide range of foundational concepts and eliminating repetitive or low-quality outputs through rejection sampling. Nevertheless, the best performance is achieved by the proposed E-LeakGAN model, attributed to its unique architecture that explicitly fosters diversity in the generated problems.

6.4.7 Ablation Study of the Proposed E-LeakGAN Model

An ablation study has been carried out to elucidate the significance of each module within the discriminator architecture of the proposed E-LeakGAN. In pursuit of this goal, eight distinct discriminator model variations have been examined, systematically omitting specific modules in varying proportions to assess their influence on the overall performance of E-LeakGAN. The results of this ablation study are denoted in Table 6.7 and further elaborated below.

(a) **Convolution w/o EGP**: The current discriminator architecture incorporates the convolution operation without (w/o) EGP (i.e., instead a conventional max-pooling is used) before progressing to the dense layers for classification. As shown in Table 6.7, this architecture yields the poorest performance among all the evaluated discriminator configurations. This observation suggests that the traditional pooling operation (i.e., max-pooling) fails to effectively guide the convolved features toward the most semantically and syntactically informative components, thereby adversely affecting the variation in the generated identities.

(b) **Convolution + EGP**: This discriminator architecture represents a modification of the previous variant in (a), where the traditional pooling operation has been replaced by the proposed EGP. It is evident from Table 6.7 that the current variant outperforms the previous one in terms of all evaluation metrics. This improvement suggests that the use of EGP contributes significantly towards increasing variation among the generated identities compared to existing ones by steering the convolved features in the direction of highest semantic-syntactic content. However, despite this progress, the current variant still falls short of achieving the best performance. This deficiency can be attributed to the absence of an attention mechanism that enables the model to focus on critical semantic features in conjunction with their syntactic positions, which is essential for generating coherent and diverse identities before the spatial dimensions of the features are reduced through pooling.

(c) **Convolution + Spatial attention w/o EGP**: This discriminator architecture combines convolution with spatial attention and conventional pooling. As evident from Table 6.7, this variant outperforms the previous ones across all evaluation metrics. This improvement can be attributed to the incorporation of the spatial attention mechanism, which highlights relevant syntactic information. However, despite this enhancement, the current variant still does not achieve the best results, as spatial attention alone tends to prioritize syntactic features over semantic content, thereby producing identities that are similar to existing ones.

(d) **Convolution + Channel Attention w/o EGP**: This module combines a convolution layer with channel attention and conventional pooling. This variant demonstrates significant improvement over those in (a), (b), and (c). The improved performance stems from the inclusion of the channel attention mechanism, which enables the network to focus on important semantic information within the convolved features. However, this combination does not yield the best results, as channel attention alone emphasizes semantic content while overlooking essential syntactic features, thereby limiting the variation among the generated problems.

(e) **Convolution + Dual Attention w/o EGP**: The current discriminator variant integrates the convolution operation followed by dual attention and conventional pooling. This variant demonstrates significant improvement over the previous ones. This enhancement is attributed to the introduction of the dual attention mechanism, where channel attention enables the model to focus on significant semantic features, while spatial attention helps the model identify where these semantic features are

syntactically located. Unfortunately, the current variant fails to achieve the best performance due to the absence of EGP, which plays a critical role in directing the attention-induced convolved features toward the highest semantically rich and syntactically important content, essential for ensuring adequate variation among the generated problems.

(f) **Convolution + Spatial attention + EGP**: The present module utilizes convolution layer along with spatial attention and EGP. This integration significantly improves results compared to using a conventional pooling mechanism, highlighting the importance of EGP in enhancing the network's ability to focus on the highest semantically rich and syntactically important content. Nevertheless, this combination does not achieve the best results, as the reliance on a single attention mechanism reduces diversity in the generated problems.

(g) **Convolution + Channel attention + EGP**: The current discriminator variant combines convolution, channel attention, and EGP layer. This arrangement improves upon conventional pooling based variants. The observed improvement indicates that EGP is able to retain the highest semantic-syntactic information required for the generation of diverse identities. However, it lags behind the successor variant due to its reliance on a single attention mechanism rather than a dual one.

(h) **Convolution + Dual Attention + EGP**: This proposed module combines convolution, dual attention, and EGP layers, achieving the best scores across all metrics as illustrated in Table 6.7. These results underscore how the synergistic combination of these modules highlights their individual contributions to improve the E-LeakGAN's performance.

6.4.8 Comparative Study of Different Attention Mechanisms

To demonstrate the efficacy of the proposed attention mechanism within the E-LeakGAN model, a comparative analysis has been conducted by replacing the proposed module with several existing attention mechanisms and evaluating their impact on the performance of E-LeakGAN. The results of this study are presented in Table 6.8. From the table, it is evident that the FMAttn method performs poorly, primarily due to its reliance on a single attention mechanism, namely channel attention. In contrast, minor improvements are observed in the models proposed in [64] and [65], which employ both spatial and channel attention mechanisms. However, in both cases, spatial attention is applied before channel attention, which

prioritizes syntactic features over semantic content. This leads to the generation of identities that closely resemble the reference data, with only minor rephrasing, thereby limiting diversity. Significant improvements are observed with CBAM and its modified variant, which apply channel attention before spatial attention. This sequence captures semantic relevance first and then accordingly refines the syntactic features. However, these architectures still fall short of achieving optimal results due to two main limitations. First, their channel attention mechanisms rely solely on max and average pooling operations, which are insufficient for capturing complex semantic relationships among features, such as identifying complementary or reciprocal functions. Second, both the channel and spatial attention modules are guided only by the sigmoid activation function, which suppresses negative inputs and limits the model's ability to learn richer and more nuanced representations necessary for generating coherent and diverse identity problems. In contrast, the proposed dual attention scheme achieves a reduction of 1.9% in BLEU-2, 1.7% in BLEU-3, 1.6% in BLEU-4, 1.6% in BLEU-5, 10.4% in MNCS along with an improvement of 3.9% in MND compared to modified CBAM due to its uniquely tailored architectural design, specifically optimized for the generation of diverse identity problems.

6.4.9 Comparative Study of Different Pooling Mechanisms

The present sub-section provides a comparative analysis of the influence of different pooling techniques on the quality of the generated trigonometric identities. The results of this analysis are depicted in Table 6.9. It is apparent from this table that the worst performance is exhibited by the FMA pooling mechanism, which independently performs max and average pooling and then concatenates the two pooled outputs. Such a technique fails to highlight the most significant semantic information within the convolved features because of the simple concatenation operation that often overlooks essential mathematical relationships, such as reciprocal and complementary relations among trigonometric functions. In contrast, mixed pooling, tree pooling, gated pooling, and T-max-avg-pooling operations show minor improvements over FMA pooling due to their more synergistic combination of max and average pooling, which helps to partially retain features pertaining to significant mathematical relationships of trigonometric identities. However, the MPN-COV operation shows significant improvement in results compared to the

Table 6.8
Impact of Different Attention Mechanisms on the Performance of E-LeakGAN

Attention Mechanisms	BLEU-2↓	BLEU-3↓	BLEU-4↓	BLEU-5 ↓	MNCS ↓	MND ↑
FMAttn [63]	0.893	0.745	0.552	0.403	0.283	0.861
Spatial+Channel Attention [64]	0.868	0.711	0.510	0.370	0.252	0.906
Spatial and Channel-wise Attention [65]	0.861	0.703	0.495	0.364	0.246	0.915
CBAM [66]	0.836	0.678	0.472	0.327	0.214	0.934
Modified CBAM [67]	0.820	0.671	0.468	0.318	0.196	0.942
Proposed Dual attention	**0.801**	**0.654**	**0.452**	**0.302**	**0.092**	**0.981**

Table 6.9
Effect of Different Pooling Methods on the Performance of E-LeakGAN

Pooling Techniques	BLEU-2 ↓	BLEU-3 ↓	BLEU-4 ↓	BLEU-5 ↓	MNCS ↓	MND ↑
FMA pooling [63]	0.927	0.786	0.592	0.454	0.337	0.725
Mixed pooling [68]	0.902	0.748	0.561	0.417	0.294	0.763
Tree pooling [68]	0.894	0.741	0.559	0.398	0.283	0.784
Gated pooling [68]	0.881	0.734	0.546	0.384	0.272	0.795
T-max-avg pooling [69]	0.867	0.711	0.530	0.360	0.250	0.860
MPN-COV [70]	0.833	0.681	0.481	0.325	0.165	0.930
Proposed EGP	**0.801**	**0.654**	**0.452**	**0.302**	**0.092**	**0.981**

previous pooling techniques, as it captures the relationships among features through covariance computation, rather than just relying on maximum or average feature values, thereby preserving important semantic-syntactic information. Nevertheless, it still falls short of achieving the best results, as it retains all feature relationships from the convolved matrix without emphasizing the most relevant ones. This dilutes the generator module's focus, making it less effective at extracting the most critical semantic cues required for producing valid and coherent problems. On the other hand, the EGP operation delivers the best performance with a reduction of 3.2% in BLEU-2, 2.7% in BLEU-3, 2.9% in BLEU-4, 2.3% in BLEU-5, 7.3% in MNCS, along with an improvement of 5.1% in MND compared to MPN-COV by not only capturing the relationships among feature vectors but also effectively projecting the features towards the largest semantic-syntactic information, thereby enhancing the generation of diverse and consistent trigonometric identities.

6.4.10 Effect of Different Activation Functions on the Performance of E-LeakGAN

The impact of different activation functions on the performance of the proposed E-LeakGAN is presented in Table 6.10. It is evident from the table that traditional activation functions such as sigmoid and hyperbolic tangent yield the poorest scores.

Table 6.10
Effect of Different Activation Functions on the Performance of the E-LeakGAN Model

Activation Functions	BLEU-2 ↓	BLEU-3 ↓	BLEU-4 ↓	BLEU-5 ↓	MNCS ↓	MND ↑
Sigmoid [44]	0.961	0.854	0.677	0.536	0.410	0.839
Tanh [71]	0.934	0.837	0.657	0.514	0.391	0.854
ReLU [72]	0.902	0.815	0.623	0.482	0.357	0.883
Leaky ReLU [73]	0.880	0.783	0.586	0.466	0.321	0.906
ELU [74]	0.867	0.750	0.558	0.435	0.290	0.924
SELU [75]	0.856	0.738	0.537	0.402	0.267	0.930
Swish [76]	0.835	0.711	0.516	0.378	0.229	0.951
Mish [77]	0.828	0.684	0.498	0.366	0.215	0.958
Logish [78]	0.820	0.671	0.483	0.340	0.198	0.962
Smish [79]	0.812	0.667	0.468	0.321	0.173	0.976
Trish	**0.801**	**0.654**	**0.452**	**0.302**	**0.092**	**0.981**

This is primarily because they struggle to effectively mitigate the vanishing gradient problem, resulting in lower quality features being transferred to the generator module which in turn produces low quality identities. In contrast, the ReLU activation function demonstrates improved performance compared to sigmoid and hyperbolic tangent, as it effectively addresses the vanishing gradient issue. However, it suffers from the "dying neuron" problem [80], wherein neurons output zero for negative inputs, leading to ineffective learning of the discriminator network. This issue is mitigated by the LeakyReLU activation function, which introduces a small, pre-defined non-zero slope for negative inputs. Although this approach effectively combats the dying neuron issue, it results in decreased performance across all evaluation metrics compared to its successors. This decline is attributed to the use of a fixed slope value for negative inputs, which limits the learning capacity of the discriminator. As an improvement, exponential activation functions such as ELU and SELU show better performance. These functions maintain a small, non-zero output for negative inputs in an adaptive manner, reducing activation sparsity and enhancing the learning effectiveness of the discriminator. Nevertheless, they still fall short of delivering the best results due to their monotonic nature, which restricts their expressiveness in modeling complex patterns from the input data.

The above limitation is addressed by several non-monotonic activation functions that combine the benefits of traditional activation functions to improve overall performance. These include Swish, Mish, Logish, Smish, and Trish. Although these activation functions share similar properties, the differences in their performance metrics, as presented in Table 6.10, stem from variations in their mathematical

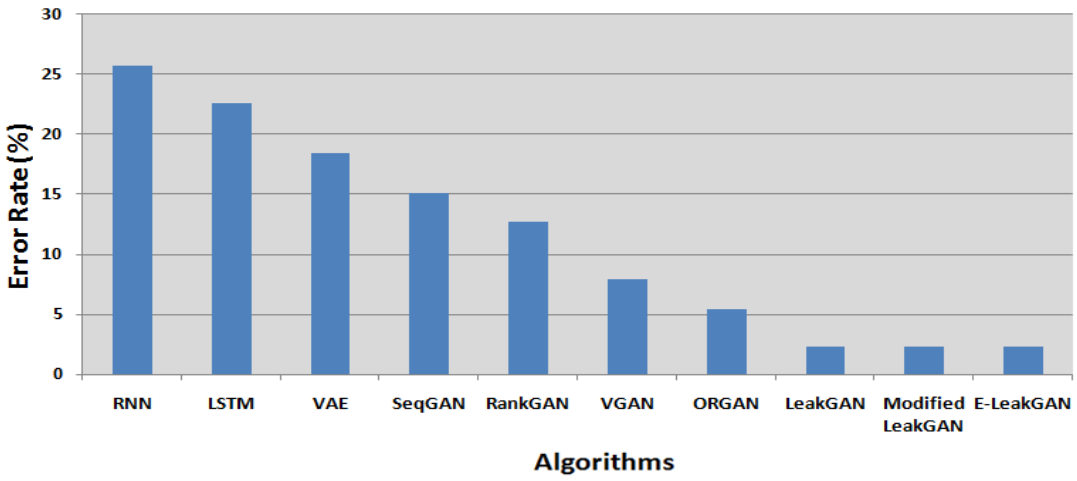


Fig.6.6. Error Rate of problems generated by E-LeakGAN in comparison to SOTA

formulations. Among them, Swish yields the least favorable results due to its dependence on the sigmoid function, which constrains the model's learning capacity. Mish performs better by leveraging the hyperbolic tangent, logarithmic, and exponential functions, thereby increasing the network's expressiveness. Logish further improves performance by integrating the strengths of both Swish and Mish while Smish achieves the second-best results by synergizing the advantages of Mish and Logish. Trish outperforms all other activation functions and achieves the best results, as it effectively balances non-linearity, adaptiveness, and gradient flow. Its design integrates multiple smooth, non-monotonic components that enable the model to capture subtle and complex patterns within the data, resulting in more coherent and diverse identity generation.

6.4.11 Analyzing Error Rates in the Generated Mathematical Identities

As discussed in Section 6.3, all the generated identities are validated using the SymPy module to ensure mathematical equivalence. However, this section evaluates the error rates of the generated identities prior to the post-processing phase with SymPy. The error rate is calculated based on the number of correctly generated problems, n, out of p total problems by $(1 - (n/p)) \times 100\%$. The results of this evaluation are presented in Fig. 6.6. It is evident from the figure that the lowest error rates are achieved by the original LeakGAN (error rate =3.2%), the modified LeakGAN (error rate =2.29%), and the proposed E-LeakGAN (error rate =2.27%) models. This performance can be attributed to the hierarchical manager-worker architecture inherent in the LeakGAN framework, which facilitates more structured, context-aware, and goal-driven sequence generation. The manager module guides

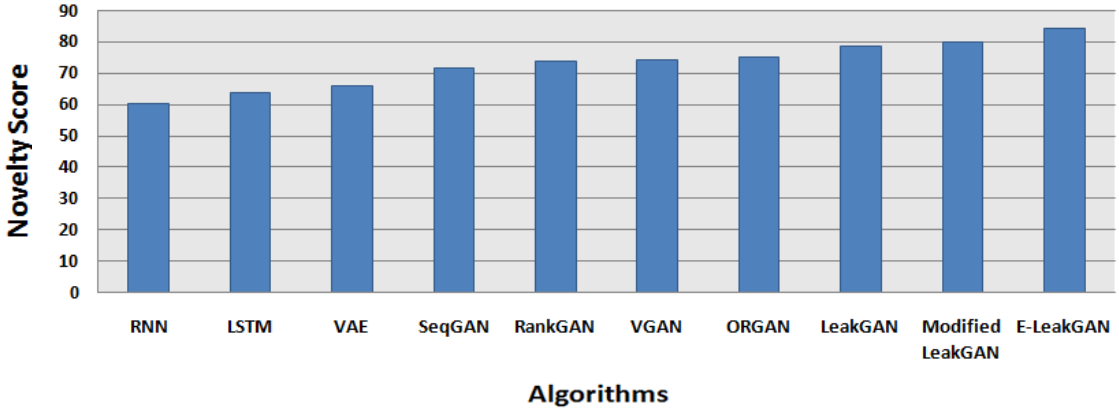


Fig. 6.7. The average value of novelty scores provided by 23 mathematics experts to the generated problems

the generation process by setting high-level objectives based on the discriminator's feedback, while the worker module executes low-level token generation aligned with these objectives. This coordinated mechanism enables the model to maintain global coherence and syntactic correctness throughout the sequence, ultimately leading to higher quality outputs with fewer errors. Additionally, the marginal difference in error rates of the E-LeakGAN model compared to its other two variants can be attributed to the cumulative impact of the architectural enhancements that further refine both the semantic relevance and syntactic structure of the generated identities.

6.4.12 Subjective Assessment of the Generated Trigonometric Identity Problems

This sub-section discusses the qualitative assessment undertaken to examine the structural novelty of the generated identities in comparison to standard textbook problems. To undertake the aforementioned evaluation, 23 mathematics experts assessed the structural novelty of the generated trigonometric identities in relation to the trigonometric proof problems found in standard textbooks. They provided a score within the range [1, 100] for each generated identity. Thus, for a total of p generated problems by the proposed as well as the SOTA methods, the average score from the 23 experts has been calculated and depicted in Fig.6.7. It is apparent from this figure that all the SOTA techniques, compared to the traditional LeakGAN architecture, are unable to provide good results in terms of novelty. On the contrary, the traditional LeakGAN and the modified LeakGAN yield moderately good results, with average novelty scores of 79.80 and 80, respectively. Interestingly, the proposed E-LeakGAN architecture achieves the best performance, with an average novelty score of 84.2,

Table 6.11
Comparative Study of Training and Testing Time Durations

Time Duration	Algorithms									
	RNN	**LSTM**	**VAE**	**Seq GAN**	**Rank GAN**	**VGAN**	**ORGAN**	**LeakGAN**	**Modified LeakGAN**	**E-LeakGAN**
Training Time (min)	112.16	182.32	302.41	448.36	372.20	576.65	498.50	620.15	621.52	623.74
Testing Time (secs)	27.84	54.70	101.55	245.87	203.56	314.80	272.63	352.38	353.02	354.52

owing to its unique discriminator design that promotes a high degree of novelty in the generated identities.

6.4.13 Comparative Study of the Training and Testing Time Durations

This sub-section discusses the training and testing time durations for the proposed E-LeakGAN model in comparison to its traditional counterparts, implemented using the PyTorch framework in Google Colaboratory. The environment used for this implementation includes an NVIDIA Tesla P100 GPU with 16 GB of VRAM, 24 GB of RAM, and 100 GB of storage space. The comparative study of the training and testing times for the generation of 1000 problems is depicted in Table 6.11. It is evident from the table that the modified LeakGAN model requires the second highest training time compared to the SOTA techniques. This extended training time is attributed to the model's hierarchical structure, which includes a manager and a worker. The manager operates at a high level, directing the worker to generate identities step-by-step, thereby adding to the complexity of the process. Additionally, the generator in LeakGAN benefits from receiving "leaked" information from the discriminator at intermediate steps. This feedback mechanism is designed to stabilize training and enhance performance but introduces additional overhead. Moreover, the training and testing time of the proposed E-LeakGAN model increase by a modest 0.36% and 0.43% respectively, compared to modified LeakGAN, due to the incorporation of dual attention, and EGP layers.

Although the proposed model requires the highest training and testing times, it also significantly outperforms all SOTA methods, as demonstrated by the results in the previous sub-sections. Thus, the increase in training and testing times of the proposed GAN represent a trade-off for its superior performance. This demonstrates that while the proposed GAN demands more computational resources, its enhanced

performance justifies the additional complexity, making it a valuable contribution to the field of computational creativity.

6.4.14 Experiments on Cross-Domain Applications

To evaluate the generalizability of the E-LeakGAN model, the same architecture, without any modification, was employed to generate identity problems in algebra and calculus. Manually curated datasets for these domains were used, following the same preprocessing pipeline discussed in Section 6.3.1. Some of the problems generated by the proposed model in algebra and calculus are illustrated in Appendix A.3 of the supplementary material. Experiments conducted to assess the efficacy of the proposed approach in these domains demonstrated superior performance in terms of BLEU, MNCS, and MND scores, relative to SOTA methods involving various GAN architectures. Additionally, the proposed dual attention mechanism, EGP, and Trish activation function in E-LeakGAN outperformed their respective baseline models across all evaluation metrics. These results indicate that the proposed E-LeakGAN approach consistently delivers strong performance across different mathematical domains without the need for any changes in the architectural pipeline. The corresponding results are presented in Appendix A.4 of the supplementary material.

6.5 DISCUSSION

The primary goal of the present study is to develop an automated method for generating trigonometric identities that are both structurally novel compared to existing textbook problems and suitable for inclusion as end-of-chapter exercises in mathematics textbooks. The proposed approach is well-suited for this purpose due to the following features.

(a) **Robustness**: The automatic generation of trigonometric identity problems is accomplished using adversarial training with leaked information. The main challenge of the afore-said task involves managing over-fitting and/or under-fitting issues in the discriminator and generator networks, which has a direct impact on the GAN's performance. However, the proposed GAN achieved the best scores across all categories of evaluation metrics, with BLEU-2 = 0.801, BLEU-3 = 0.654, BLEU-4 = 0.452, BLEU-5 = 0.302, MNCS = 0.092, and MND = 0.981, outperforming SOTA methods and demonstrating its robustness in addressing the aforementioned issues.

(b) **Convenience**: The proposed E-LeakGAN model requires only a computer system that supports GPU acceleration and the PyTorch framework. This minimal

hardware requirement makes the model highly convenient for the automatic generation of mathematical identities, even on systems with modest computational resources.

(c) **Diversity Analysis**: The quality of the generated trigonometric identity problems showed significant improvement over SOTA techniques, with BLEU-2 = 0.801, BLEU-3 = 0.654, BLEU-4 = 0.452, BLEU-5 = 0.302, MNCS = 0.092, and MND = 0.981, reflecting the syntactic coherence and diversity of the generated identities. These results were further supported by qualitative judgments from 23 mathematics experts, who provided an average novelty score of 84.2, confirming substantial variation in the generated identities compared to existing problems.

(d) **Benefits**: The proposed approach offers an effective solution for automatically generating trigonometric identities that are ideal for chapter-end exercises in math textbooks. It serves as a valuable resource for educators and students alike: teachers can quickly create a variety of engaging questions to enhance their instructional materials, while students can generate numerous practice problems to improve their problem-solving skills and mathematical understanding.

(e) **Efficiency**: The results of performance analysis presented in Section 4 demonstrate the effectiveness of the proposed model in generating novel trigonometric identity problems, with a reduction of 2.6% in BLEU-2, 2.2% in BLEU-3, 1.8% in BLEU-4, 4.3% in BLEU-5, 5.1% in MNCS, along with an improvement of 2.5% in MND compared to the best performing baseline i.e., modified LeakGAN. The reduction in BLEU and MNCS metrics indicates that the generated identities exhibit significant syntactic and semantic variation in comparison to existing problems. In contrast, the improvement in the MND metric indicates the presence of substantial diversity among the generated problems. Additionally, ablation studies conducted highlight the contribution of each component within the E-LeakGAN model. The aforementioned findings confirm the model's ability to generate unique and diverse problems, underscoring its potential to produce high-quality content that enhances educational materials.

(f) **Reliability**: The reliability of the proposed model is evidenced by its strong performance across multiple evaluation metrics and its competitive results when benchmarked against SOTA methods. These quantitative outcomes demonstrate the model's consistent ability to generate accurate, diverse, and high-quality trigonometric identity problems. Furthermore, the model has received positive validation from mathematics experts, who affirm its practical applicability in educational settings. This combination of empirical robustness and expert endorsement establishes a high degree of confidence in the model's operational

dependability. Hence, reliability is not a concern in the context of the proposed approach.

(g) **Generalizability**: The proposed model has demonstrated its effectiveness in generating identity problems beyond trigonometry, extending successfully to other mathematical domains such as algebra and calculus, where it consistently outperforms competing approaches. Notably, the only required adjustment involves the creation of manually curated datasets tailored to each domain, while the E-LeakGAN architecture itself remains unchanged. These findings highlight that the model's design is inherently domain-agnostic, as it does not incorporate any domain-specific assumptions or structures. Thus, the generalizability of the proposed architectural framework is well established, affirming its potential to generate diverse mathematical problems across various domains.

While the proposed method offers numerous benefits, it does come with two drawbacks, which are listed below.

(a) The training and testing times of the E-LeakGAN model are longer than those of the SOTA techniques. This extended duration stems from the model's intricate hierarchical design, which includes manager-worker interactions, the use of additional leaked information from the discriminator, and the application of dual attention and eigenvector based pooling layers in the discriminator. Despite this, the method's ability to deliver superior evaluation metric scores, indicating valid, coherent and diverse trigonometric identities, makes the increased time required for training and testing a reasonable trade-off.

(b) The identities produced by the E-LeakGAN architecture are not 100% accurate and exhibit a low error rate of 2.27%. This limitation is addressed by incorporating an additional post-processing step after the generation phase, using the SymPy module, which ensures that all generated identities maintain mathematical equivalence.

6.6 CONCLUSION

The chapter introduces an innovative approach for implementing computational creativity in the scientific field through the automatic creation of diverse trigonometric identity problems using a novel E-LeakGAN model. The architecture of the proposed model is particularly noteworthy for its unique discriminator, which adeptly captures essential semantic and syntactic features from the input data. The key contributions in the discriminator architecture built upon a CNN, include: a) the utilization of Trish activation function to enhance the network's learning capacity, b) the introduction of a dual attention mechanism that simultaneously highlights the

relevant semantic and syntactic information in the convolved features and c) an eigenvector guided pooling mechanism that aligns the attention induced vectors towards the direction of maximum semantic-syntactic relevance. The performance of the E-LeakGAN model is validated through comprehensive experimental analyses and ablation studies, which demonstrate the superior quality of the generated identities, with BLEU-2 = 0.801, BLEU-3 = 0.654, BLEU-4 = 0.452, BLEU-5 = 0.302, MNCS = 0.092, and MND = 0.981, compared to traditional techniques. Furthermore, an evaluation by 23 mathematics experts yielded an average novelty score of 84.2, affirming the originality and quality of the generated identities relative to standard textbook problems. The proposed method has also shown efficacy in producing identity problems in other mathematical domain such as algebra and calculus.

Thus, the proposed framework serves as an effective and scalable tool for the automated generation of mathematical identities, well-suited for chapter-end exercises in mathematics textbooks. Beyond its core functionality, the model offers several practical benefits for both educators and learners. For teachers, it provides a convenient means to generate a wide variety of high-quality, diverse problems, enabling the enrichment of instructional content, worksheets, and assessments with minimal effort. For students, the system facilitates unlimited access to automatically generated practice problems, allowing them to reinforce their understanding, improve problem-solving skills, and prepare more effectively for various tests and competitive examinations.

Unfortunately, the proposed model suffers from two limitations. First, the training and testing times of E-LeakGAN are longer than those of all SOTA techniques. However, this increase in computational time is justified by the model's superior performance across all evaluation metrics, which reflect its ability to generate valid, coherent, and diverse mathematical identities. Second, the proposed approach results in the generation of identities with an error rate of 2.27%. This issue is effectively mitigated through a post-processing step that employs the SymPy module to verify and ensure the mathematical equivalence of all generated identities. Future prospects of this work include extending the current methodology to the automatic generation of problems beyond mathematics, in other scientific domains such as physics and engineering, thereby broadening the impact of computational creativity across diverse educational fields.

6.7 APPENDIX

Algorithm A.1: Generation of identity problems by E-LeakGAN

1. Initialize a random embedded vector denoting any trigonometric function as the start-up element.
2. **While** entire identity is not generated **do**
   Begin
   a) Transfer the current embedded vector to the discriminator that performs convolution, attention mechanism and pooling operation upon this vector.
   b) Transfer/leak the pooled vector to the manager module that generates an action sub-goal vector. Also, transfer the current embedded vector to the worker module that produces an action embedding.
   c) Perform matrix multiplication of the action sub-goal vector with the action embedding and transfer the output to a MLP layer that predicts the next vector.
   d) Update the current vector with the next vector.
   End-while;
3. Return the completely generated trigonometric identity problem by decoding its embedded vector.

Table A.1
Exemplar List of Tokens, Token IDs and Vector Embeddings

Token	Token ID	Vector Embedding
sin	0	[0.1 -0.2 0.6 0.9 -0.5]
x	1	[0.5 0.8 -0.6 0.8 -0.8]
^	2	[0.3 0.5 0.7 -0.3 0.4]
2	3	[0.2 0.8 -0.6 0.4 -0.3]
1	4	[0.8 0.9 -0.2 -0.1 0.5]
+	5	[0.6 -0.4 0.8 0.1 0.4]
-	6	[0.5 -0.2 0.6 -0.7 0.3]
*	7	[0.3 0.8 -0.6 -0.9 -0.7]
/	8	[0.8 0.6 0.1 0.3 0.5]
=	9	[0.2 0.3 -0.9 -0.8 0.7]
sec	10	[0.5 0.9 0.8 -0.4 0.7]
cosec	11	[0.1 -0.2 0.4 -0.3 -0.6]
cos	12	[0.5 -0.5 0.6 0.3 -0.1]

A.1. Details of Hash Algorithm utilized for the Formulation of Vector Embeddings

Each token ID w_i is first encoded as UTF-8 bytes and passed through the SHA-256 hash function as described by (A.1):

$$g = SHA256(bytes(w_i)) \in \{0,1\}^{256}, \tag{A.1}$$

where, g denotes the 256-bit digest. Next, the digest g is converted into a 32-bit integer k using (A.2):

$$k = \left(\sum_{j=0}^{255} g_j \cdot 2^{255-j} \right) \bmod 10^8, \tag{A.2}$$

Where, g_j represents the j^{th} bit of the SHA-256 hash digest. k obtained from (A.2) acts as a seed integer that initializes a pseudorandom number generator to produce a vector embedding represented by (A.3):

$$d = (x_1, x_2, ..., x_N), \; x_j \sim Uniform(-1,1), \tag{A.3}$$

where, N (here, 20) denotes the embedding dimension. The elements x_j are independent and identically distributed (i.i.d.) with $\mathrm{E}[x_j] = 0$ and $Var[x_j] = 1/3$. It is

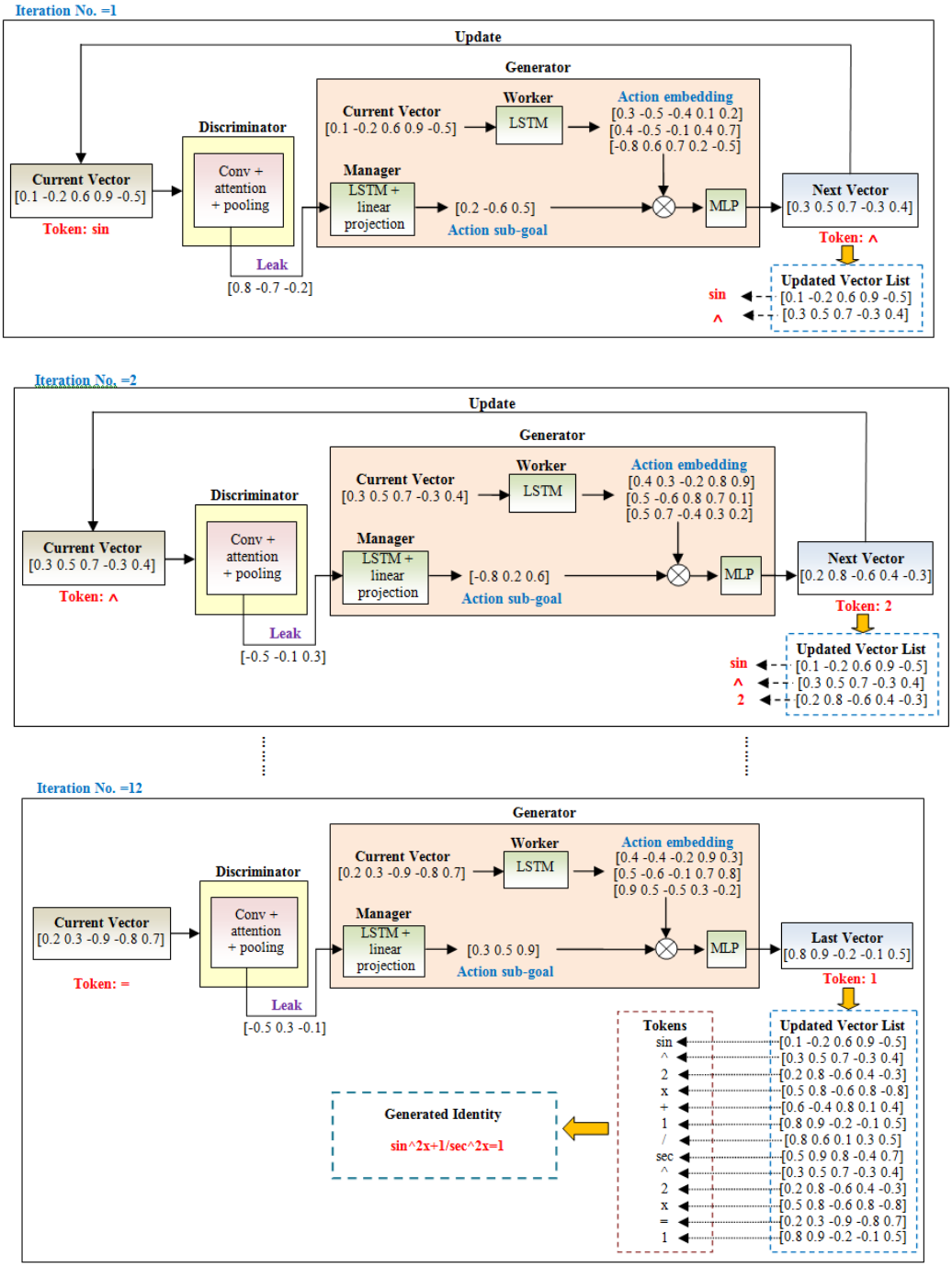


Fig. A.1. Illustration of new identity generation by E-LeakGAN

important to note that, due to the cryptographic properties of SHA-256 and the deterministic nature of seed-based pseudorandom number generator initialization, distinct tokens are mapped to distinct embeddings.

A.2. Details of the Generation Process of Trigonometric Identity Problems

As discussed in Section 6.3.1, each identity problem in the manually curated dataset undergoes tokenization, assignment of token IDs, and conversion of these IDs into

vector embeddings using a hash algorithm. During the training phase, the discriminator and the generator are trained separately on the entire set of vector embeddings of dimension 20. Just like in typical GANs-based training, the generator is tasked with producing sequences of embeddings that resemble valid trigonometric identities, while the discriminator learns to distinguish between real identity embeddings from the curated dataset and those generated by the generator. However, unlike standard GANs, the discriminator in this framework also leaks intermediate feature representations i.e., pooled vectors, back to the generator. These leaked features guide the manager module of the generator in setting meaningful sub-goals via an LSTM and linear transformation, enabling more structured and goal-driven generation. This feedback loop enhances the generator's ability to produce high-quality and diverse identities, while the discriminator simultaneously improves its ability to detect generated (fake) sequences, leading to a stable and effective adversarial training process. The generation phase proceeds according to Algorithm A.1. Figure A.1 further illustrates the detailed steps involved in the generation of an exemplar trigonometric identity. An exemplar set of tokens, their IDs, and the corresponding vector embeddings is presented in Table A.1. It is important to note that while the proposed study utilizes vector embeddings of dimension 20, a reduced dimension of 5 is used in the example illustrated in Fig. A.1 for simplicity of explanation.

In Fig. A.1, the embedded vector corresponding to the token “sin” is considered as the start-up element. This embedded vector, referred to as the current vector of size 1×5, is fed to the discriminator, which subsequently transfers (or leaks) its pooled output of size 1×3 to the manager module of the generator. The manager module transforms the acquired pooled vector into an action sub-goal of size 1×3 via an LSTM followed by a linear projection, as discussed in Section 3.2 of the manuscript. Simultaneously, the current vector is transformed into an action embedding of size 3×5 by the worker module through an LSTM. The action sub-goal and the action embedding undergo matrix multiplication to yield a vector of size 1×5, which is then fed to an MLP layer. The MLP layer predicts the probability distribution over the next vector in the current sequence via a Softmax function, and the vector with the highest probability is selected as the next vector. The current vector is then updated with this next vector, and the entire process continues iteratively until a complete trigonometric identity is generated. Finally, all the vector embeddings obtained through successive iterations are converted back to their token IDs and then to individual tokens, which, when aligned in order, form the generated trigonometric identity.

Table A.2
Samples of Identity Problems Generated by E-LeakGAN in Algebra

Sl. No.	Generated Problems
1.	$a^3(b-c)+b^3(c-a)+c^3(a-b)=-(a-b)(b-c)(c-a)(a+b+c)$
2.	$(a+b)^4+(a-b)^4=2(a^4+6a^2b^2+b^4)$
3.	$a^4+4b^4=(a^2+2ab+2b^2)(a^2-2ab+2b^2)$

Table A.3
Samples of Identity Problems Generated by E-LeakGAN in Calculus

Sl. No.	Generated Problems
1.	$\int x^2\left(2x-\frac{1}{x}\right)^2 dx=\frac{4}{5}x^5-\frac{4}{3}x^3+x+c$
2.	$\int (2x+3)\sqrt{x^2+3x-1}\,dx=\frac{2}{3}(x^2+3x-1)^{3/2}+c$
3.	$\int x\sqrt{x^2+1}\,dx=\frac{x^2\sqrt{x^2+1}}{2}-\frac{1}{6}(x^2+1)^{3/2}+\frac{1}{2}\sqrt{x^2+1}+c$

A.3. Application of E-LeakGAN for Problem Generation in Algebra and Calculus

The E-LeakGAN model has been tested for identity generation in other mathematical domains such as algebra and calculus. Some of the problems generated by the E-LeakGAN model are demonstrated in Tables A.2 and A.3 for algebra and calculus domains respectively.

A.4. Performance Evaluation of E-LeakGAN for Problem Generation in Algebra and Calculus

This section illustrates the performance of the proposed model for generating problems in the algebra and calculus domains. Table A.4 presents the results for the algebra domain, while Table A.5 shows the corresponding results for the calculus domain. It is apparent from both tables that the quality of the generated problems shows significant improvement over SOTA techniques, which involve variations in different GAN architectures, with respect to existing textbook problems. Similarly, the performance of the proposed dual attention mechanism in E-LeakGAN outperforms other baseline attention techniques across all evaluation metrics, as evident from Tables A.6 and A.7 for algebra and calculus domains respectively.

Table A.4
Comparative Analysis of the Proposed E-LeakGAN with SOTA Models for Generating Problems in Algebra.

Algorithms	BLEU-2 ↓	BLEU-3 ↓	BLEU-4 ↓	BLEU-5 ↓	MNCS ↓	MND ↑
SeqGAN	0.960	0.788	0.602	0.490	0.298	0.701
RankGAN	0.952	0.762	0.581	0.472	0.283	0.727
VGAN	0.937	0.740	0.566	0.452	0.271	0.754
ORGAN	0.922	0.718	0.537	0.405	0.248	0.845
LeakGAN	0.867	0.685	0.491	0.364	0.211	0.894
Modified LeakGAN	0.834	0.667	0.460	0.336	0.154	0.951
E-LeakGAN	**0.816**	**0.642**	**0.445**	**0.311**	**0.088**	**0.976**

Table A.5
Comparative Analysis of the Proposed E-LeakGAN with SOTA Models for Generating Problems in Calculus.

Algorithms	BLEU-2 ↓	BLEU-3 ↓	BLEU-4 ↓	BLEU-5 ↓	MNCS ↓	MND ↑
SeqGAN	0.980	0.781	0.630	0.486	0.312	0.723
RankGAN	0.967	0.754	0.601	0.469	0.298	0.743
VGAN	0.928	0.729	0.585	0.440	0.287	0.781
ORGAN	0.911	0.703	0.574	0.414	0.261	0.866
LeakGAN	0.853	0.670	0.461	0.348	0.224	0.910
Modified LeakGAN	0.821	0.654	0.436	0.326	0.163	0.964
E-LeakGAN	**0.806**	**0.633**	**0.423**	**0.308**	**0.090**	**0.986**

Table A.6
Impact of Different Attention Mechanisms on the Performance of E-LeakGAN for Generating Problems in Algebra.

Attention Mechanisms	BLEU-2 ↓	BLEU-3 ↓	BLEU-4 ↓	BLEU-5 ↓	MNCS ↓	MND ↑
FMAttn	0.901	0.756	0.562	0.410	0.288	0.854
Spatial+Channel Attention	0.862	0.723	0.515	0.371	0.260	0.889
Spatial and Channel-wise Attention	0.857	0.710	0.502	0.358	0.251	0.905
CBAM	0.848	0.681	0.485	0.334	0.223	0.941
Modified CBAM	0.834	0.660	0.462	0.325	0.187	0.953
Proposed Dual Attention	**0.816**	**0.642**	**0.445**	**0.311**	**0.088**	**0.976**

Moreover, the proposed EGP mechanism in E-LeakGAN also surpasses other baseline pooling techniques across all evaluation metrics, as shown in Tables D.5 and D.6. Additionally, the utilization of the Trish activation function in E-LeakGAN outperforms other baseline activation functions across all evaluation metrics, again as demonstrated in Tables A.8 and A.9. Thus, the above results indicate that the proposed E-LeakGAN approach consistently delivers strong performance across different mathematical domains without requiring any changes to the architectural pipeline.

Table A.7
Impact of Different Attention Mechanisms on the Performance of E-LeakGAN for Generating Problems in Calculus.

Attention Mechanisms	BLEU-2 ↓	BLEU-3 ↓	BLEU-4 ↓	BLEU-5 ↓	MNCS ↓	MND ↑
FMAttn	0.885	0.732	0.560	0.412	0.290	0.878
Spatial+Channel Attention	0.870	0.711	0.518	0.382	0.265	0.880
Spatial and Channel-wise Attention	0.868	0.695	0.488	0.375	0.253	0.901
CBAM	0.830	0.671	0.463	0.334	0.220	0.944
Modified CBAM	0.818	0.668	0.457	0.322	0.164	0.960
Proposed Dual attention	**0.806**	**0.633**	**0.423**	**0.308**	**0.090**	**0.986**

Table A.8
Effect of Different Pooling Mechanisms on the Performance of E-LeakGAN for Generating Problems in Algebra.

Pooling Techniques	BLEU-2 ↓	BLEU-3 ↓	BLEU-4 ↓	BLEU-5 ↓	MNCS ↓	MND ↑
FMA pooling	0.930	0.792	0.603	0.448	0.334	0.730
Mixed pooling	0.917	0.769	0.570	0.427	0.311	0.761
Tree pooling	0.901	0.758	0.564	0.402	0.302	0.778
Gated pooling	0.888	0.740	0.550	0.378	0.296	0.815
T-max-avg pooling	0.858	0.725	0.523	0.358	0.285	0.895
MPN-COV	0.822	0.664	0.472	0.330	0.201	0.948
Proposed EGP	**0.816**	**0.642**	**0.445**	**0.311**	**0.088**	**0.976**

Table A.9
Effect of Different Pooling Mechanisms on the Performance of E-LeakGAN for Generating Problems in Calculus.

Pooling Techniques	BLEU-2 ↓	BLEU-3 ↓	BLEU-4 ↓	BLEU-5 ↓	MNCS ↓	MND ↑
FMA pooling	0.915	0.792	0.597	0.439	0.328	0.733
Mixed pooling	0.898	0.769	0.575	0.405	0.318	0.752
Tree pooling	0.880	0.754	0.568	0.387	0.301	0.781
Gated pooling	0.873	0.738	0.553	0.370	0.280	0.806
T-max-avg pooling	0.850	0.720	0.537	0.346	0.271	0.873
MPN-COV	0.818	0.678	0.481	0.319	0.220	0.936
Proposed EGP	**0.806**	**0.633**	**0.423**	**0.308**	**0.090**	**0.986**

7

Conclusions and Future Directions

This chapter offers a critical self-review of the thesis, addressing two major aspects of scientific creativity: biological and computational modalities, and also suggests possible future extensions. In the biological domain, the work emphasizes the development of hybrid models based on graph convolution to decode brain connectivity patterns in individuals engaged in creative tasks. These models not only enhance the understanding of the neural basis of creativity but also support the identification and recruitment of creative individuals for research-intensive roles. Future directions include the use of hybrid brain imaging techniques and the development of more robust classifiers for creativity assessment. In the computational domain, the thesis focuses on the automated generation of mathematical problems for use as chapter-end exercises, utilizing both search-based techniques and generative networks. Future work in this area includes extending the system to generate word problems with embedded equations and incorporating perceptual understanding into machines to elevate their creative potential to a level comparable with that of humans.

7.1 SELF-REVIEW OF THE THESIS

The thesis includes five original contributory chapters, with the first three focusing on biological modalities and the last two addressing computational approaches to scientific creativity. Chapter 2 presents a novel technique for assessing scientific creativity of subjects based on their spatial reasoning abilities using a Graph Convolutional-Interval Type-2 Fuzzy Network (GC-IT2FN). The primary contribution lies in mapping the spatial reasoning ability required in a mental paper folding task into brain connectivity networks, whose classification reveals the creative level of performing subjects. The most challenging aspect of this work is the design of an effective classifier model for the afore-said task. The novelty of the proposed classifier stems from three key innovations. First, a self-attention mechanism is integrated into the network architecture, allowing the graph convolutional layers to focus on the most relevant nodes and thereby enhancing classifier performance. Second, a new activation function called Logish is introduced following the graph convolution operation. Its inherent negative activation response to small negative signals improves the classifier's accuracy and stability. Third, a novel policy is implemented to select the most promising region within the Footprint of Uncertainty of the Interval Type-2 Fuzzy Sets, thereby reducing the impact of uncertainty on classifier performance. Experimental analysis undertaken reveals the active engagement of the right anterior prefrontal cortex (BA 10), right posterior middle temporal gyrus (BA 21), and bilateral regions of the posterior supramarginal gyrus (BA 40) during the paper folding task. Furthermore, the proposed classifier significantly outperforms existing models, demonstrating its effectiveness and robustness. Thus, the proposed scheme can be utilized to identify creative individuals for potential placement in various research-oriented departments, based on their different levels of creative ability.

The second problem addressed in the thesis is presented in Chapter 3, which focuses on classifying creative individuals (along with their degrees of creative potential) and distinguishing them from analytical individuals based on the arousal of the Eureka Effect in the brain. This is achieved using a Dual Attention Induced Graph Convolutional-Capsule Network (DAIGC-CapsNet). The central challenge lies in analyzing the brain connectivity networks derived from EEG signals to detect the arousal of the Eureka Effect, which serves as the basis for categorizing a person as either creative or analytical. Furthermore, the degree of creative potential is determined from connectivity patterns that reflect the level of neural reward mechanisms associated with the Eureka experience. An additional challenge involves classifying the resulting brain networks into four categories: Analytical, High Creative, Medium Creative, and Low Creative. To address this, the proposed

classifier combines a Graph Convolutional Network, which abstracts centrality-based features from the brain connectivity data, with a Capsule Network, which performs the classification. This integrated architecture forms the DAIGC-CapsNet. The novelty of the classifier lies in two main contributions: the introduction of dual attention modules and a newly designed routing algorithm. The dual attention mechanism includes a Mish Induced Attention Module, which enables the graph convolutional layers to prioritize the most significant node features, and a Fused Attention Module, which ensures that the most relevant information is effectively transmitted from the primary capsule layer to the class capsule layer. Additionally, the Sparsemax-based routing algorithm enhances the coupling between primary and class capsules, thereby improving the interpretability and effectiveness of the network. Experimental results demonstrate that the proposed model significantly outperforms traditional approaches in accurately classifying individuals based on their cognitive traits. Thus, the above approach is useful for recruiting creative individuals to innovation-centered sectors, while analytical individuals can be directed toward roles in investigation, academia, or management sectors.

The third problem addressed in Chapter 4 presents a novel approach to assess the scientific creative ability of individuals during their participation in an analogical reasoning task using a classifier referred to as Enhanced Graph Convolution Induced Type-2 Fuzzy Classifier (EGCIFC). The main challenge lies in transforming the analogical reasoning involved in Raven's Advanced Progressive Matrices (RAPM) test into brain connectivity networks and subsequently classifying these networks to decode different levels of creative potential. This decoding process is performed by extracting features from the brain networks and classifying them using the proposed EGCIFC model. The classifier introduces five main contributions. First, an enhanced graph convolution operation is developed to capture both local and global structural information from the input graph, enabling the model to learn diverse patterns. Second, a novel activation function called Smish is incorporated into the graph convolution layer, which retains small negative input values and contributes to improved performance. Third, a one-dimensional spatial convolution layer replaces the traditional pooling layer to preserve the most relevant information in the convolved feature maps. Fourth, a new mapping function is introduced within the type-2 fuzzy layer to handle uncertainty among the spatially convolved vectors by reducing the area under the footprint of uncertainty (FOU). Fifth, the Takagi-Sugeno-Kang (TSK) fuzzy reasoning mechanism is employed to reduce computational complexity by eliminating the need for type-reduction and defuzzification. Experimental results acquired confirm that the proposed classifier significantly outperforms conventional methods in assessing the levels of creative

potential. Thus, this method can also be effectively utilized for recruiting individuals into various research departments or academic institutions based on their level of creative potential.

Chapter 5 addresses the fourth problem of the thesis by proposing an innovative framework for computational creativity synthesis using an extended best-first search algorithm. This framework introduces a mechanism for evaluating the diversity of newly generated instances in relation to their predecessors. The method follows a best-first search on an OR-tree structure where the nodes represent candidate solutions or creative outputs, while edges denote the substitution rules applied to generate child nodes from parent nodes. Two alternative versions of the algorithm are explored: Relative Diversity-Guided Best-First Search and Absolute Diversity-Guided Best-First Search. These variants differ in the cost functions used to evaluate nodes. The former maximizes the diversity of a node compared to its immediate parent, while the latter aims to maximize the difference between a node's diversity score and a penalty cost, measured with respect to the root node. A significant contribution of this work is the ability to maintain diversity during the iterative expansion of the search tree and ensure the novelty of the final solution across different executions. The chapter also presents a comparative performance analysis of the two algorithmic variants, evaluating their applicability and effectiveness. Additional theoretical properties such as soundness and logical completeness are discussed to support the robustness of the proposed methods. Experimental findings confirm that the algorithms generate significantly more diverse mathematical identities than conventional approaches. The practical utility of the framework is further demonstrated through its successful application in creating chapter-end exercises across various domains, including propositional logic, set theory, and algebra, highlighting its potential for broader generalization.

The fifth problem addressed in Chapter 6 focuses on advancing computational creativity through the automatic generation of trigonometric identity problems for use as chapter-end exercises in mathematics textbooks. This is achieved by designing a novel Generative Adversarial Network with Leaked Information (LeakGAN), consisting of a generator and a discriminator engaged in an adversarial process. The generator creates new identity problems, while the discriminator evaluates their validity by distinguishing them from real examples. In addition to classification, the discriminator leaks high-level features to guide the generator, enabling the production of coherent and diverse identities. The generator comprises two modules: a manager, which uses the leaked features to produce a goal vector capturing key trigonometric relationships, and a worker, which generates logically structured identity components based on this vector. This hierarchical design facilitates the

creation of semantically rich outputs. The original contribution lies in enhancing the discriminator of LeakGAN, resulting in an improved model called Enhanced LeakGAN (E-LeakGAN). The new discriminator includes three key innovations: i) a custom activation function named Trish to boost expressive power, ii) a dual attention mechanism combining channel and spatial attention to focus on important semantic and syntactic features, and iii) an eigenvector-guided pooling method to emphasize the most relevant content. Experimental results show that E-LeakGAN significantly outperforms traditional methods in generating diverse identity problems and demonstrates potential for generalization to other mathematical areas such as algebra and calculus.

7.2 DIRECTIONS FOR FUTURE RESEARCH

The possible extensions of the present research are numerous, and compiling an exhaustive list would be nearly as extensive as the thesis itself. Therefore, this thesis outlines a focused and concise set of future research directions, highlighting smaller, immediate extensions that are both feasible and promising. In this way, it offers an optimistic perspective on the potential evolution and impact of the current work.

7.2.1 Future Prospects in Biological Modalities of Scientific Creativity

This sub-section provides a brief overview of potential future directions of the thesis related to biological modalities of scientific creativity.

7.2.1.1 Employment of Brain Imaging Devices with High Spatial Resolution

The first part of the thesis focuses on the biological underpinnings of scientific creativity using brain imaging techniques. It investigates the spatial and temporal dynamics of brain activity to interpret behavioral manifestations of creativity from a cognitive perspective. These behavioral aspects refer to a subject’s ability to perform creative oriented tasks. For instance, in mental paper folding, the ability to visualize the final shape of a folded paper with cut marks when it is unfolded, serves as a key indicator of scientific creativity. Similarly, in the Raven's Advanced Progressive Matrices (RAPM) test, the ability to infer the correct pattern based on similarities among given patterns reflects creative potential. The accurate localization of brain activation during such tasks depends on the spatial resolution of the imaging device. For the present study, a functional Near-Infrared Spectroscopy (fNIRS) has been utilized that is capable of capturing only brain activations from the superficial regions of the cortex [1-2]. On the other hand, functional Magnetic Resonance Imaging (fMRI) [3-4] provides significantly better spatial resolution and can image

the entire brain including deep sub-cortical regions [5-6]. Thus, a promising direction for future research is the use of high-resolution imaging techniques like fMRI to more accurately localize brain regions that can enhance the understanding of the cognitive and neural mechanisms that drive creative behavior.

7.2.1.2 Hybrid Neuro-imaging Approach for Enhanced Spatial and Temporal Resolution

It has been previously mentioned in the thesis that the temporal dynamics of brain activation can be captured by Electroencephalography (EEG), while spatial localization of brain activity can be acquired via fNIRS/fMRI. In recent years, with the advent of technology, there has been the development of hybrid brain signal acquisition systems that integrate EEG with fNIRS/fMRI [7-8]. The aforementioned multimodal systems posses the ability to simultaneously capture both temporal and spatial characteristics of brain activity thereby, offering a more comprehensive understanding of neural processes. Thus, a promising direction for future research of the present work involves the integration of afore-said hybrid modalities to concurrently acquire spatial and temporal characteristics of brain activation during scientific creative tasks. Such an approach could provide a more accurate and holistic picture of the neural mechanisms underlying scientific creativity, enabling researchers to decode how different brain regions interact over time and space to support creative cognition.

7.2.1.3 Concurrent Modeling of Uncertainty and Spatial Relationships in Brain Data

In the present thesis, the classifier models utilized for decoding the brain connectivity patterns of subjects engaged in various creativity-oriented tasks were based either on a graph convolution-induced capsule network or graph convolution-induced type-2 fuzzy reasoning. The former is effective in preserving the spatial relationships among the convolved graphical features, while the latter is particularly concerned with addressing the uncertainty inherent in graph-convolved features. Although each approach addresses a critical aspect of brain data analysis, managing both uncertainty and preserving spatial relationships among brain network features are equally important for accurately decoding brain connectivity patterns. This is because brain signals are inherently noisy and uncertain [9], and at the same time, the spatial configuration of neural activations carries essential information about cognitive processes [10]. Thus, a potential future extension of the present work involves the integration graph convolutional network, interval type-2 fuzzy

reasoning, and capsule network into a unified framework. Such an approach would enable the concurrent modeling of uncertainty through fuzzy reasoning, preservation of spatial relationships among features by capsule network, and hierarchical feature representation through graph convolution, thereby enhancing the robustness and interpretability of brain connectivity-based classification.

7.2.2 Future Prospects in Synthesizing Computational Creativity

This sub-section discusses the possible future extensions of the second part of the thesis, which focuses on the computational modalities of scientific creativity.

7.2.2.1 Automatic Generation of New Text-based Mathematical Problems

The work presented in the thesis under the computational modalities of scientific creativity focuses on the automatic generation of new mathematical problems using both search-based and generative approaches. A promising extension of this work lies in the development of novel or diversified mathematical word problems that consist of descriptive text. These new text-based mathematical problems can be utilized to automatically generate mathematical content that is useful for problems in mathematics, such as those leading to systems of linear or nonlinear equations, exploring number theory through problems involving divisibility, remainders, or properties of numbers, and problems involving data interpretation or probability calculation, among others. Additionally, mathematical word problems that combine descriptive text with corresponding equations can broaden the scope of computational creativity beyond mathematics by extending its application in other scientific domains such as physics, chemistry, bioinformatics, etc. The aforementioned development can be facilitated by incorporating a natural language processing (NLP) [11-12] interface along with existing approaches based on search-based strategies or generative models, where the NLP component shall tackle the language part of the newly generated scientific problems.

7.2.2.2 Automated Development of Scientific Creativity-Oriented Cognitive Tasks

As mentioned in the present thesis, several cognitive tasks performed by humans are known to influence creative outcomes in the scientific domain. The afore-said tasks includes: i) spatial reasoning tests, such as mental paper folding, that requires the ability to mentally transform an object's various attributes to create something new, ii) analogical reasoning tasks, such as the RAPM test, which assesses one's capacity to recognize similarities among existing patterns for novel predictions, and iii) convergent thinking tasks, such as the matchstick problem, which involve finding a

single correct solution to the presented problem with the arousal of the Eureka effect. Since the ability to perform well on the aforementioned tasks is closely linked to innovation in science, their automated generation by a machine or computer system would be considered creative. Consequently, a promising extension of the computational modalities of scientific creativity involves developing advanced computational models, such as neural networks or hybrid AI systems, that can autonomously create cognitive tasks influencing creativity in the scientific domain.

7.2.2.3 Employment of Cross-Domain Knowledge Transfer

An important extension of present work under the computational modalities of scientific creativity involves enabling machines to undergo cross-domain knowledge transfer i.e., the capability to extract knowledge from one/multiple domains and apply it to another. Current artificial intelligent (AI) models often struggle with recognizing the common principles that are shared across different contexts. The ability to identify the common principles between different knowledge domains and apply the acquired concept to novel and/or distinct problems would represent a major advancement in creative reasoning by machines. The aforementioned objective can be achieved by developing computational frameworks that integrate perceptual understanding of real world entities. Thus, the incorporation of different perceptual capabilities in AI models could foster the generation of innovative ideas, solutions, or problems comparable to human level creativity.

References

CHAPTER 1

[1] M.A. Runco, and G.J. Jaeger, *The standard definition of creativity*, Creativity research journal, vol. 24, no. 1, pp.92-96, 2012.

[2] R.K. Sawyer, *Creativity and development*, Oxford University Press, 2003.

[3] R.W. Weisberg, *Creativity: Understanding innovation in problem solving, science, invention, and the arts*, John Wiley & Sons, 2016.

[4] M. W. Matlin, *Cognition*, John Wiley and sons Ltd, 2005.

[5] R.K. Sawyer, and D. Henriksen, *Explaining creativity: The science of human innovation*, Oxford university press, 2024.

[6] D. Novitasari, D.K. Risfianty, T.W. Triutami, N.P. Wulandari, and R.Y. Tyaningsih, "The relation between spatial reasoning and creativity in solving geometric problems, In Journal of Physics: Conference Series, vol. 1776, no. 1, p. 012007, *IOP Publishing*, 2021.

[7] H.J. Kell, D. Lubinski, C.P. Benbow, and J.H. Steiger, "Creativity and technical innovation: Spatial ability's unique role," *Psychological science*, vol. 24, no. 9, pp.1831-1836, 2013.

[8] D. Gentner, and F. Maravilla, *Analogical reasoning*, In International handbook of thinking and reasoning, Routledge, pp. 186-203, 2017.

[9] S.I. Robertson, *Problem solving: Perspectives from cognition and neuroscience*, Psychology Press, 2016.

[10] W. Niu, J.X. Zhang, and Y. Yang, "Deductive reasoning and creativity: A cross-cultural study," *Psychological Reports*, vol. 100, no. 2, pp.509-519, 2007.

[11] S. Harnad, "Creativity: Method or magic?," *Hungarian Studies*, vol. 20, no. 1, pp.163-177, 2006.

[12] H. Pringle, "The origins of creativity," *Scientific American*, vol. 308, no. 3, pp.36-43, 2013.

[13] J.H. Mason, *The value of creativity: The origins and emergence of a modern belief*, Routledge, 2017.

[14] R.P. Weiner, *Creativity and beyond: Cultures, values, and change*, State University of New York press agent, 2000.

[15] P.O. Kristeller, "Creativity" and" tradition," *Journal of the History of Ideas*, vol. 44, no. 1, pp.105-113, 1983.

[16] A. Reckwitz, *The invention of creativity: Modern society and the culture of the new*, John Wiley & Sons, 2018.

[17] P. Formica, "Scope of the Renaissance," *Entrepreneurial Renaissance: Cities Striving Towards an Era of Rebirth and Revival*, pp.1-46, 2017.

[18] S. Rassanjani, H. Herizal, M. Mukhrijal, M. Alqarni, and B. Usman, "Managing creativity and innovation in the 4th industrial revolution: Learning from giants," *Journal of Innovation in Business and Economics*, vol. 5, no. 01, pp.15-22, 2021.

[19] V. Smil, Creating the twentieth century: Technical innovations of 1867-1914 and their lasting impact, *Oxford University Press*, 2005.

[20] W. Cai, S. Khapova, B. Bossink, E. Lysova, and J. Yuan, "Optimizing employee creativity in the digital era: Uncovering the interactional effects of abilities, motivations, and opportunities," *International journal of environmental research and public health*, vol. 17,no. 3, p.1038, 2020.

[21] J.C. Kaufman and R.J. Sternberg, eds., *The Cambridge handbook of creativity*, Cambridge University Press, 2019.

[22] B. Gaut, *The philosophy of creativity*, Philosophy Compass, vol. 5, no. 12, pp.1034-1046, 2010.

[23] J.R. Bailey, C.M. Ford, and J.D. Raelin, "Philosophical ties that bind practice: The case of creativity", *Journal of Management Inquiry*, vol. 18, no. 1, pp.27-38, 2009.

[24] E. Sadler-Smith, "Wallas' four-stage model of the creative process: More than meets the eye?," *Creativity research journal*, vol. 27, no. 4, pp.342-352, 2015.

[25] R.J. Sternberg, and E.L. Grigorenko, "Guilford's structure of intellect model and model of creativity: Contributions and limitations," *Creativity Research Journal*, vol. 13, no. 3-4, pp.309-316, 2001.

[26] M. Rhodes, "An analysis of creativity," *Phi Delta Kappan*, vol. 42, 305–310, 1961.

[27] J.P. von Thienen, T.J. Weinstein, and C. Meinel, "Creative metacognition in design thinking: exploring theories, educational practices, and their implications for measurement," *Frontiers in Psychology*, vol. 14, p.1157001, 2023.

[28] J.C. Kaufman, and R.A. Beghetto, "Beyond big and little: The four c model of creativity," *Review of general psychology*, vol. 13, no. 1, pp.1-12, 2009.

[29] A. Abraham, *The neuroscience of creativity*, Cambridge University Press, 2018.

[30] Z. Ivcevic, Artistic and everyday creativity: An act-frequency approach, *The Journal of Creative Behavior*, vol. 41, no. 4, pp.271-290, 2007.

[31] J.Y. Oh, "Understanding the scientific creativity based on various perspectives of science," *Axiomathes*, vol. 32, no. 6, pp.907-929, 2022.

[32] P. Li, Y. Zheng, and X. Chen, "Drugs for autoimmune inflammatory diseases: from small molecule compounds to anti-TNF biologics," *Frontiers in pharmacology*, vol. 8, p.460, 2017.

[33] M. Pardy, "Creativity leading to discoveries in particle physics and relativity," *arXiv preprint physics/0509184*, 2005.

[34] D.H. Cropley, "Promoting creativity and innovation in engineering education," *Psychology of Aesthetics, Creativity, and the Arts*, vol. 9, no. 2, p.161, 2015.

[35] W. Hu, and P. Adey, "A scientific creativity test for secondary school students," *International Journal of Science Education*, vol. 24, no. 4, pp.389-403, 2002.

[36] A. Samsudin, A.H. Setyadin, E. Suhendi, D.T. Chandra, and P. Siahaan, "Seventh grade students' scientific creativity test: A preliminary-study on earth science context," In *IOP Conference Series: Materials science and engineering*, vol. 288, no. 1, p. 012012, IOP Publishing, 2018.

[37] J.W. Park, "A suggestion of cognitive model of scientific creativity (CMSC)," *Journal of the Korean Association for Science Education*, vol.24, no. 2, pp.375-386, 2004.

[38] J. Park, "PRACTICAL WAYS FOR TEACHING AND EVALUATION SCIENTIFIC CREATIVITY 12", 2010.

[39] A. Mittal, B. Jayaram, S. Shenoy, and T.S. Bawa, "A stoichiometry driven universal spatial organization of backbones of folded proteins: are there Chargaff's rules for protein folding?," *Journal of Biomolecular Structure and Dynamics*, vol. 28, no. 2, pp.133-142, 2010.

[40] D.K. Simonton, "Chance-configuration theory of scientific creativity," *The psychology of science: Contributions to metascience*, pp.170-213, 1989.

[41] D.K. Simonton, "Creative productivity and age: A mathematical model based on a two-step cognitive process," *Developmental review*, vol. 4, no. 1, pp.77-111, 1984.

[42] W. Dennis, "Creative productivity between the ages of 20 and 80 years," *Journal of Gerontology*, vol. 21, pp. l-8, 1966.

[43] H.C. Lehman, *Age and achievement*, Princeton, N.J.: Princeton Univ., Press, 1953.

[44] H. Zuckerman, *Scientific elite*, New York: Free Press, 1977.

[45] G. Fauconnier, and M. Turner, "Conceptual integration networks", *Cognitive science*, vol. 22, no. 2, pp.133-187, 1998.

[46] G. Fauconnier, and M. Turner, *The way we think: Conceptual blending and the mind's hidden complexities*, Basic books, 2008.

[47] F. Bou, M. Schorlemmer, J. Corneli, D.J. Gómez-Ramírez, E. Maclean, A. Smaill, and A. Pease, "The role of blending in mathematical invention," *ICCC*, 2015.

[48] M. Guhe, A. Pease, A. Smaill, M. Martinez, M. Schmidt, H. Gust, K.U. Kühnberger, and U. Krumnack, "A computational account of conceptual blending in basic mathematics," *Cognitive Systems Research*, vol. 12, no. 3-4, pp.249-265, 2011.

[49] M. Martínez, A. M. Abdel-Fattah, U. Krumnack, D. Gómez-Ramírez, A. Smaill, T.R. Besold, A. Pease, M. Schmidt, M. Guhe, and K.U. Kühnberger, "Theory blending: extended algorithmic aspects and examples," *Annals of Mathematics and Artificial Intelligence*, vol. 80, pp.65-89, 2017.

[50] A. Pease, and U. Martin, "Automating "human-like" example-use in mathematics," In *CEUR Workshop Proceedings*, vol. 2261, pp. 6-15, CEUR-WS, 2018.

[51] M. Benedek, and A. Fink, "Toward a neurocognitive framework of creative cognition: The role of memory, attention, and cognitive control," *Current Opinion in Behavioral Sciences*, vol. 27, pp.116-122, 2019.

[52] J. Heinonen, J. Numminen, Y. Hlushchuk, H. Antell, V. Taatila, and J. Suomala, "Default mode and executive networks areas: association with the serial order in divergent thinking," *PloS one*, vol. 11, no. 9, p.e0162234, 2016.

[53] B. Shi, X. Cao, Q. Chen, K. Zhuang, and J. Qiu, "Different brain structures associated with artistic and scientific creativity: a voxel-based morphometry study," *Scientific reports*, vol. 7, no. 1, p.42911, 2017.

[54] R.E. Beaty, R.A. Cortes, H.M. Merseal, M.M. Hardiman, and A.E. Green, "Brain networks supporting scientific creative thinking," *Psychology of Aesthetics, Creativity, and the Arts*, 2023.

[55] E. Bartoli, E. Devara, H.Q. Dang, R. Rabinovich, R.K. Mathura, A. Anand, B.R. Pascuzzi, J. Adkinson, K.R. Bijanki, S.A. Sheth, and B. Shofty, "Default mode network spatio-temporal electrophysiological signature and causal role in creativity," *bioRxiv*, 2023.

[56] X. Wang, Q. Chen, K. Zhuang, J. Zhang, R.A. Cortes, D.D. Holzman, L. Fan, C. Liu, J. Sun, X. Li, and Y. Li, "Semantic associative abilities and executive control functions predict novelty and appropriateness of idea generation," *Communications Biology*, vol. 7, no. 1, p.703, 2024.

[57] M. Benedek, Y.N. Kenett, K. Umdasch, D. Anaki, M. Faust, and A.C. Neubauer, "How semantic memory structure and intelligence contribute to creative thought: A network science approach," *Thinking & Reasoning*, vol. 23, no. 2, pp.158-183, 2017.

[58] R.E. Beaty, M. Benedek, S. Barry Kaufman, and P.J. Silvia, "Default and executive network coupling supports creative idea production," *Scientific reports*, vol. 5, no. 1, p.10964, 2015.

[59] L. Carruthers, R. MacLean, and A. Willis, "The relationship between creativity and attention in adults," *Creativity Research Journal*, vol. 30, no. 4, pp.370-379, 2018.

[60] Z. Liu, J. Zhang, X. Xie, E.T. Rolls, J. Sun, K. Zhang, Z. Jiao, Q. Chen, J. Zhang, J., Qiu, and J. Feng, "Neural and genetic determinants of creativity," *Neuroimage*, vol. 174, pp.164-176, 2018.

[61] A. Dragomir, and A. Omurtag, "Brain's networks and their functional significance in cognition," In *Handbook of Neuroengineering*, pp. 1-30, Singapore: Springer Singapore, 2021.

[62] A. Fornito, A. Zalesky, E. and Bullmore, Fundamentals of brain network analysis, *Academic press*, 2016.

[63] F. de Vico Fallani, J. Richiardi, M. Chavez, and S. Achard, "Graph analysis of functional brain networks: practical issues in translational neuroscience,"

Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 369, no. 1653, p.20130521, 2014.

[64] T. Kirschstein, and R. Köhling, "What is the source of the EEG?," *Clinical EEG and neuroscience*, vol. 40, no. 3, pp.146-149, 2009.

[65] G. Katti, S.A. Ara, and A. Shireen, "Magnetic resonance imaging (MRI)–A review," *International journal of dental clinics*, vol. 3, no. 1, pp.65-70, 2011.

[66] P. Hansen, M. Kringelbach, and R. Salmelin, eds., *MEG: An introduction to methods*, Oxford university press, 2010.

[67] L. Schrevens, N. Lorent, C. Dooms, and J. Vansteenkiste, "The role of PET scan in diagnosis, staging, and management of non-small cell lung cancer," *The oncologist*, vol. 9, no. 6, pp.633-643, 2004.

[68] S. Tak, and J.C. Ye, "Statistical analysis of fNIRS data: a comprehensive review," *Neuroimage*, vol. 85, pp.72-91, 2014.

[69] F. Wang, X. Zhang, R. Fu and G. Sun, "EEG characteristic analysis of coach bus drivers based on brain connectivity as revealed via a graph theoretical network," *RSC advances*, vol. 8, no. 52, pp.29745-29755, 2018.

[70] C.J. Stam, G. Nolte, and A. Daffertshofer, "Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources," *Human brain mapping*, vol. 28, no. 11, pp.1178-1193, 2007.

[71] S.P. Van den Broek, F. Reinders, M. Donderwinkel, and M.J. Peters, "Volume conduction effects in EEG and MEG," *Electroencephalography and clinical neurophysiology*, vol. 106, no. 6, pp.522-534, 1998.

[72] J. Park, J. Shin, and J. Jeong, "Inter-brain synchrony levels according to task execution modes and difficulty levels: an fNIRS/GSR study," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp.194-204, 2022.

[73] A. Grinsted, J.C. Moore, and S. Jevrejeva, "Application of the cross wavelet transform and wavelet coherence to geophysical time series," *Nonlinear processes in geophysics*, vol. 11, no. 5/6, pp.561-566, 2004.

[74] B.C. Van Wijk, C.J. Stam, and A. Daffertshofer, "Comparing brain networks of different size and connectivity density using graph theory," *PloS one*, vol. 5, no. 10, p.e13701, 2010.

[75] M. Rubinov, and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp.1059-1069, 2010.

[76] A. Boschi, M. Brofiga, and P. Massobrio, "Thresholding functional connectivity matrices to recover the topological properties of large-scale neuronal networks," *Frontiers in Neuroscience*, vol. 15, p.705103, 2021.

[77] T. Adamovich, I. Zakharov, A. Tabueva, and S. Malykh, "The thresholding problem and variability in the EEG graph network parameters," *Scientific Reports*, vol. 12, no. 1, p.18659, 2022.

[78] T. Ogawa, T. Aihara, T. Shimokawa, and O. Yamashita, "Large-scale brain network associated with creative insight: combined voxel-based morphometry and resting-state functional connectivity analyses," *Scientific reports*, vol. 8, no. 1, p.6477, 2018.

[79] J. Li, D. Zhang, A. Liang, B. Liang, Z. Wang, Y. Cai, M. Gao, Z. Gao, S. Chang, B. Jiao, and R. Huang, "High transition frequencies of dynamic functional connectivity states in the creative brain," *Scientific reports*, vol. 7, no. 1, p.46072, 2017.

[80] W. Orwig, I. Diez, P. Vannini, R. Beaty, and J. Sepulcre, "Creative connections: Computational semantic distance captures individual creativity and resting-state functional connectivity," *Journal of cognitive neuroscience*, vol. 33, no. 3, pp.499-509, 2021.

[81] L. Shi, R.E. Beaty, Q. Chen, J. Sun, D. Wei, W. Yang, and J. Qiu, "Brain entropy is associated with divergent thinking," *Cerebral Cortex*, vol. 30, no. 2, pp.708-717, 2020.

[82] S. Keshmiri, "Entropy and the brain: An overview," *Entropy*, vol. 22, no. 9, p.917, 2020.

[83] F. Alù, F. Miraglia, A. Orticoni, E. Judica, M. Cotelli, P.M. Rossini, and F. Vecchio, "Approximate entropy of brain network in the study of hemispheric differences," *Entropy*, vol. 22, no. 11, p.1220, 2020.

[84] P. Wang, M.L. Wijnants, and S.M. Ritter, "What enables novel thoughts? The temporal structure of associations and its relationship to divergent thinking," *Frontiers in psychology*, vol. 9, p.1771, 2018.

[85] I. Echegoyen, D. López-Sanz, J.H. Martínez, F. Maestú, and J.M. Buldú, "Permutation entropy and statistical complexity in mild cognitive impairment

and Alzheimer's disease: An analysis based on frequency bands," *Entropy*, vol. 22, no. 1, p.116, 2020.

[86] R. Cofré, and A. Destexhe, "Entropy and Complexity Tools Across Scales in Neuroscience: A Review," *Entropy*, vol. 27, no. 2, p.115, 2025.

[87] K. Ueno, T. Takahashi, K. Takahashi, K., Mizukami, Y. Tanaka, and Y. Wada, "Neurophysiological basis of creativity in healthy elderly people: a multiscale entropy approach," *Clinical neurophysiology*, vol. 126, no. 3, pp.524-531, 2015.

[88] H. Qi, and C. Liu, "Metacontrol Regulates Creative Thinking: An EEG Complexity Analysis Based on Multiscale Entropy," *Brain Sciences*, vol. 14, no. 11, p.1094, 2024.

[89] Y. Kaur, S. Weiss, C. Zhou, R. Fischer, and A. Hildebrandt, "Exploring neural signal complexity as a potential link between creative thinking, intelligence, and cognitive control," *Journal of Intelligence*, vol. 9, no. 4, p.59, 2021.

[90] J. Courtiol, D. Perdikis, S. Petkoski, V. Müller, R. Huys, R. Sleimen-Malkoun, and V.K. Jirsa, "The multiscale entropy: Guidelines for use and interpretation in brain signal analysis," *Journal of neuroscience methods*, vol. 273, pp.175-190, 2016.

[91] Y. Kaur, G. Ouyang, W. Sommer, S. Weiss, C. Zhou, and A. Hildebrandt, "What does temporal brain signal complexity reveal about verbal creativity?," *Frontiers in behavioral neuroscience*, vol. 14, p.146, 2020.

[92] M. Liu, C. Song, Y. Liang, T. Knöpfel, and C. Zhou, "Assessing spatiotemporal variability of brain spontaneous activity by multiscale entropy and functional connectivity," *NeuroImage*, vol. 198, pp.198-220, 2019.

[93] L. Zhang, J.Q. Gan, Y. Zhu, J. Wang, and H. Wang, "EEG source-space synchrostate transitions and Markov modeling in the math-gifted brain during a long-chain reasoning task," *Human brain mapping*, vol. 41, no. 13, pp.3620-3636, 2020.

[94] N.J. Williams, I. Daly, and S.J. Nasuto, "Markov model-based method to analyse time-varying networks in EEG task-related data," *Frontiers in computational neuroscience*, vol. 12, p.76, 2018.

[95] S.P. Awate, T. Tasdizen, N. Foster, and R.T. Whitaker, "Adaptive Markov modeling for mutual-information-based, unsupervised MRI brain-tissue classification," *Medical Image Analysis*, vol. 10, no. 5, pp.726-739, 2006.

[96] L. Zhang, J.Q. Gan, Y. Zhu, J. Wang, and H. Wang, "EEG source-space synchrostate transitions and Markov modeling in the math-gifted brain during a long-chain reasoning task," *Human brain mapping*, vol. 41, no. 13, pp.3620-3636, 2020.

[97] S.B.U.D. Tahir, A. Jalal, and K. Kim, "Wearable inertial sensors for daily activity analysis based on Adam optimization and the maximum entropy Markov model," *Entropy*, vol. 22, no. 5, p.579, 2020.

[98] J. Lahikainen, N.M. Ady, and C. Guckelsberger, "Creativity and markov decision processes," *arXiv preprint arXiv:2405.14966*, 2024.

[99] Y. Yu, Y. Oh, J. Kounios, and M. Beeman, "Uncovering the interplay of oscillatory processes during creative problem solving: A dynamic modeling approach," *Creativity research journal*, vol. 35, no. 3, pp.438-454, 2023.

[100] J.S. Gero, and J. Milovanovic, "A framework for studying design thinking through measuring designers' minds, bodies and brains," *Design Science*, vol. 6, p.e19, 2020.

[101] T. Daikoku, K. Kamermans, and M. Minatoya, "Exploring cognitive individuality and the underlying creativity in statistical learning and phase entrainment," *EXCLI journal*, vol. 22, p.828, 2023.

[102] T.M. Kodinariya, and P.R. Makwana, "Review on determining number of Cluster in K-Means Clustering," *International Journal*, vol. 1, no. 6, pp.90-95, 2013.

[103] S. Na, L. Xumin, and G. Yong, "Research on k-means clustering algorithm: An improved k-means clustering algorithm," In *2010 Third International Symposium on intelligent information technology and security informatics*, pp. 63-67, IEEE, 2010.

[104] J. Hao, Y. Zhou, Q. Zhao, and Q. Xue, "An evolutionary computation based method for creative design inspiration generation," *Journal of intelligent manufacturing*, vol. 30, pp.1673-1691, 2019.

[105] A. I. Miller, "Creativity in the Age of AI: Computers and artificial neural networks are redefining the relationship between art and science," *American Scientist*, vol. 108, no. 4, pp.244-250, 2020.

[106] E. Trillas, S. D'Onofrio, and E. Portmann, "An exploration of creative reasoning," In *Designing cognitive cities*, pp. 85-106, Cham: Springer International Publishing, 2018.

[107] P.C. Hamerski, "Generative AI as a Resource for Creativity in Computational Physics," In *Proceedings of the 55th ACM Technical Symposium on Computer Science Education V. 2*, pp. 1666-1667, 2024.

[108] H. Toivonen, and O. Gross, "Data mining and machine learning in computational creativity," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 5, no. 6, pp.265-275, 2015.

[109] T.R. Besold, M. Schorlemmer, and A. Smaill, eds., *Computational creativity research: towards creative machines*, vol. 7, Paris: Atlantis Press, 2015.

[110] R.D. King, K.E. Whelan, F.M. Jones, P.G. Reiser, C.H. Bryant, S.H. Muggleton, D.B. Kell, and S.G. Oliver, "Functional genomic hypothesis generation and experimentation by a robot scientist," *Nature*, vol. 427, no. 6971, pp.247-252, 2004.

[111] A. Sparkes, W. Aubrey, E. Byrne, A. Clare, M.N. Khan, M. Liakata, M., Markham, J. Rowland, L.N. Soldatova, K.E. Whelan, and M. Young, "Towards robot scientists for autonomous scientific discovery," *Automated experimentation*, vol. 2, pp.1-11, 2010.

[112] F.T. Hong, "The role of pattern recognition in creative problem solving: A case study in search of new mathematics for biology," *Progress in biophysics and molecular biology*, vol. 113, no. 1, pp.181-215, 2013.

[113] C. Chen, Y. Chen, M. Horowitz, H. Hou, Z. Liu, and D. Pellegrino, "Towards an explanatory and computational theory of scientific discovery," *Journal of Informetrics*, vol. 3, no. 3, pp.191-209, 2009.

[114] G.E. Corazza, "Potential originality and effectiveness: The dynamic definition of creativity," *Creativity research journal*, vol. 28, 3, pp.258-267, 2016.

[115] S. Acar, C. Burnett, and J.F. Cabra, "Ingredients of creativity: Originality and more," *Creativity Research Journal*, vol. 29, no. 2, pp.133-144, 2017.

[116] P. Langley, *Scientific discovery: Computational explorations of the creative processes*, MIT press. 1987.

[117] W. Duch, and M. Pilichowski, "Experiments with computational creativity," *Neural Information Processing–Letters and Reviews*, vol. 11, no. 4-6, pp.123-133, 2007.

[118] C. Chen, Y. Chen, M. Horowitz, H. Hou, Z. Liu, and D. Pellegrino, "Towards an explanatory and computational theory of scientific discovery," *Journal of Informetrics*, vol. 3, no. 3, pp.191-209, 2009.

[119] D.K. Simonton, "Scientific creativity as constrained stochastic behavior: the integration of product, person, and process perspectives." *Psychological bulletin*, vol. 129, no. 4, p.475, 2003.

[120] M. Letek, "Alexander Fleming, The discoverer of the antibiotic effects of penicillin," *Frontiers for Young Minds*, vol. 7, no. 7, 2020.

[121] W.C. Rontgen, and G. Sarton, "The discovery of x-rays," *Isis*, vol. 26, no. 2, pp.349-369, 1937.

[122] R. Carlisle, *Scientific American inventions and discoveries: all the milestones in ingenuity--from the discovery of fire to the invention of the microwave oven*, Turner Publishing Company, 2008.

[123] L. Ghosh, R. Kar, A. Konar, A. Chakraborty, and A. K. Nagar, "Identification of Brain Activation Regions in Inductive Learning Based Scientific Creativity Test," In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 950-957, 2018.

[124] A. Amiruddin, M. Qorib, N. Naimi, and D. Deliati, "The Role of Inductive Learning Models on the Students Self-Regulated Learning in Math Journaling," *Cypriot Journal of Educational Sciences*, vol. 17, no. 7, pp.2428-2446, 2022.

[125] M. Karwowski, D.M. Jankowska, A. Brzeski, M. Czerwonka, A. Gajda, I Lebuda, and R.A. Beghetto, "Delving into creativity and learning," In Creative learning in digital and virtual environments, pp. 7-29, *Routledge*, 2020.

[126] Y. Chang, X. Wang, J. Wang, Y. Wu, L. Yang, K. Zhu, H. Chen, X. Yi, C. Wang, Y. Wang, and W. Ye, "A survey on evaluation of large language models," *ACM transactions on intelligent systems and technology*, vol. 15, no. 3, pp.1-45, 2024.

[127] H. Naveed, A.U. Khan, S. Qiu, M. Saqib, S. Anwar, M. Usman,N.Akhtar, N., Barnes and A. Mian, "A comprehensive overview of large language models," *arXiv preprint arXiv:2307.06435*, 2024.

[128] H. Zhao, H. Chen, E. Yang, N. Liu, H. Deng, H. Cai, S. Wang, D. Yin, and M. Du, "Explainability for large language models: A survey," *ACM Transactions on Intelligent Systems and Technology*, vol. 15, no. 2, pp.1-38, 2024.

[129] T. Wu, S. He, J. Liu, S. Sun, K. Liu, Q.L. Han, and Y. Tang, "A brief overview of ChatGPT: The history, status quo and potential future development," *IEEE/CAA Journal of Automatica Sinica*, vol. 10, no. 5, pp.1122-1136, 2023.

[130] M. Masalkhi, J. Ong, E. Waisberg, N. Zaman, P. Sarker, A.G. Lee and A. Tavakkoli, A., "A side-by-side evaluation of Llama 2 by meta with ChatGPT and its application in ophthalmology," *Eye*, pp.1-4, 2024.

[131] M. Imran and N. Almusharraf, "Google Gemini as a next generation AI educational tool: a review of emerging educational technology," *Smart Learning Environments*, vol. 11, no. 1, p.22, 2024.

[132] S. Singh, S. Bansal, A.E. Saddik, and M. Saini, "From ChatGPT to DeepSeek AI: A Comprehensive Analysis of Evolution, Deviation, and Future Implications in AI-Language Models," *arXiv preprint arXiv:2504.03219*, 2025.

[133] B. Schmidl, T. Hütten, S. Pigorsch, F. Stögbauer, C.C. Hoch, T. Hussain, B Wollenberg, and M. Wirth, "Assessing the use of the novel tool Claude 3 in comparison to ChatGPT 4.0 as an artificial intelligence tool in the diagnosis and therapy of primary head and neck cancer cases," *European Archives of Oto-Rhino-Laryngology*, vol. 281, no. 11, pp.6099-6109, 2024.

[134] M. Elbadawi, H. Li, A.W. Basit, and S. Gaisford, "The role of artificial intelligence in generating original scientific research," *International journal of pharmaceutics*, vol. 652, p.123741, 2024.

[135] B. Desai, K. Patil, A. Patil, and I. Mehta, "Large Language Models: A Comprehensive Exploration of Modern AI's Potential and Pitfalls," *Journal of Innovative Technologies*, vol. 6, no. 1, 2023.

[136] H.P. Nii, "The blackboard model of problem solving and the evolution of blackboard architectures," *AI magazine*, vol. 7, no. 2, pp.38-38, 1986.

[137] R. Leikin, and M. Lev, "Multiple solution tasks as a magnifying glass for observation of mathematical creativity," In *Proceedings of the 31st international conference for the psychology of mathematics education*, vol. 3, pp. 161-168, 2007.

[138] A. Levav-Waynberg, and R. Leikin, "Multiple solutions for a problem: A tool for evaluation of mathematical thinking in geometry," In *Proceedings of CERME*, vol. 6, no. 2010, pp. 776-785, 2010.

[139] R. Leikin, "Evaluating mathematical creativity: The interplay between multiplicity and insight1," *Psychological Test and Assessment Modeling*, vol. 55, no. 4, p.385, 2013.

[140] R. Leikin, and M. Lev, "Mathematical creativity in generally gifted and mathematically excelling adolescents: What makes the difference?," *Zdm*, vol. 45, pp.183-197, 2013.

[141] M. Lev, and R. Leikin, "The connection between mathematical creativity and high ability in mathematics," In *Proceedings of the Eight Congress of the European Society for Research in Mathematics Education (CERME8)*, pp. 1204-1213), Ankara, Turkey: ERME, 2013.

[142] U. Sak, and M.B. Ayas, "Creative Scientific Ability Test (C-SAT): A new measure of scientific creativity," *Psychological Test and Assessment Modeling*, vol. 55, no. 3, pp.316-329, 2013.

[143] A. Snyder, J. Mitchell, T. Bossomaier, and G. Pallier, "The creativity quotient: an objective scoring of ideational fluency," *Creativity Research Journal*, vol. 16, no. 4, pp.415-419, 2004.

[144] A. Pásztor, G. Molnár, and B. Csapó, "Technology-based assessment of creativity in educational context: the case of divergent thinking and its relation to mathematical achievement," *Thinking skills and Creativity*, vol. 18, pp.32-42, 2015.

[145] G. Sprugnoli, S. Rossi, A. Emmendorfer, A., Rossi, S.L. Liew, E. Tatti, G. di Lorenzo, A. Pascual-Leone, and E. Santarnecchi, "Neural correlates of Eureka moment," *Intelligence*, vol. 62, pp.99-118, 2017.

[146] M. Beeman, and J. Kounios, *The Eureka factor: creative insights and the brain*, Cornerstone Digital, 2015.

[147] J. Kounios, and M. Beeman, "The cognitive neuroscience of insight," *Annual review of psychology*, vol. 65, no. 1, pp.71-93, 2014.

[148] W. Carpenter, "The Aha! moment: The science behind creative insights," In *Toward super-creativity-improving creativity in humans, machines, and human-machine collaborations*, IntechOpen, 2019.

[149] T. Dandan, Z. Haixue, L. Wenfu, Y. Wenjing, Q. Jiang, and Z. Qinglin, "Brain activity in using heuristic prototype to solve insightful problems," *Behavioural Brain Research*, vol. 253, pp.139-144, 2013.

[150] X. Hao, S. Cui, W. Li, W., Yang, J. Qiu, and Q. Zhang, "Enhancing insight in scientific problem solving by highlighting the functional features of prototypes: An fMRI study," *Brain research*, vol. 1534, pp.46-54, 2013.

[151] J. Qiu, H. Li, J. Jou, J., Liu, Y. Luo, T. Feng, Z. Wu, and Q. Zhang, "Neural correlates of the “Aha” experiences: evidence from an fMRI study of insight problem solving," *cortex*, vol. 46, no. 3, pp.397-403, 2010.

[152] F. Tian, S. Tu, J. Qiu, J.Y. Lv, D.T. Wei, Y.H. Su, and Q.L. Zhang, "Neural correlates of mental preparation for successful insight problem solving," *Behavioural brain research*, vol. 216, no. 2, pp.626-630, 2011.

[153] Q. Zhao, Z. Zhou, H. Xu, S. Chen, F. Xu, W. Fan, and L. Han, "Dynamic neural network of insight: a functional magnetic resonance imaging study on solving Chinese ‘chengyu’riddles," *PloS one*, vol. 8, no. 3, p.e59351, 2013.

[154] J. Luo, W. Li, A. Fink, L. Jia, X. Xiao, J. Qiu, and Q. Zhang, "The time course of breaking mental sets and forming novel associations in insight-like problem solving: an ERP investigation," *Experimental brain research*, vol. 212, pp.583-591, 2011.

[155] W. Shen, C. Liu, Y. Yuan, X. Zhang, and J. Luo, "Temporal dynamics of mental impasses underlying insight-like problem solving," *Science China Life Sciences*, vol. 56, pp.284-290, 2013.

[156] Q. Xing, J.X. Zhang, and Z. Zhang, "Event-related potential effects associated with insight problem solving in a Chinese logogriph task." *Psychology*, vol. 3, no. 01, p.65, 2012.

[157] Q. Zhao, Y. Li, X. Shang, Z. Zhou, and L. Han, "Uniformity and nonuniformity of neural activities correlated to different insight problem solving," *Neuroscience*, vol. 270, pp.203-211, 2014.

[158] A. Fink, and M. Benedek, "EEG alpha power and creative ideation," *Neuroscience & Biobehavioral Reviews*, vol. 44, pp.111-123, 2014.

[159] A. Fink, R.H. Grabner, M. Benedek, G. Reishofer, V. Hauswirth, M. Fally, C. Neuper, F. Ebner, and A.C. Neubauer, "The creative brain: Investigation of brain activity during creative problem solving by means of EEG and fMRI," *Human brain mapping*, vol. 30, no. 3, pp.734-748, 2009.

[160] O. Vartanian, M.E. Jobidon, F. Bouak, A. Nakashima, I. Smith, Q. Lam, and B. Cheung, "Working memory training is associated with lower prefrontal cortex activation in a divergent thinking task," *Neuroscience*, vol. 236, pp.186-194, 2013.

[161] R.P. Roberts, K. Wiebels, R.L. Sumner, V. van Mulukom, C.L. Grady, D.L. Schacter, and D.R. Addis, "An fMRI investigation of the relationship between

future imagination and cognitive flexibility," *Neuropsychologia*, vol. 95, pp.156-172, 2017.

[162] A. Fink, D. Schwab, and I. Papousek, "Sensitivity of EEG upper alpha activity to cognitive and affective creativity interventions," *International Journal of Psychophysiology*, vol. 82, no. 3, pp.233-239, 2011.

[163] D. Schwab, M. Benedek, I. Papousek, E.M. Weiss, and A. Fink, "The time-course of EEG alpha power changes in creative ideation," *Frontiers in human neuroscience*, vol. 8, p.310, 2014.

[164] S. Agnoli, M. Zanon, S. Mastria, A. Avenanti, and G.E. Corazza, "Predicting response originality through brain activity: An analysis of changes in EEG alpha power during the generation of alternative ideas," *NeuroImage*, vol. 207, p.116385, 2020.

[165] S. Kröger, B. Rutter, H. Hill, S. Windmann, C. Hermann, and A. Abraham, "An ERP study of passive creative conceptual expansion using a modified alternate uses task," *Brain research*, vol. 1527, pp.189-198, 2013.

[166] K. Rataj, D.S. Nazareth, and F. Van der Velde, "Use a spoon as a spade?: Changes in the upper and lower alpha bands in evaluating alternate object use," *Frontiers in psychology*, vol. 9, p.1941, 2018.

[167] T. Hartog, M. Marshall, M.T. Ahad, A.G. Alhashim, G. Okudan Kremer, J. van Hell, and Z. Siddique, "Pilot Study: Investigating EEG Based Neuro-Responses of Engineers via a Modified Alternative Uses Task to Understand Creativity," In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 83921, p. V003T03A019, American Society of Mechanical Engineers, 2020.

[168] O. Polozov, E. O'Rourke, A.M. Smith, L. Zettlemoyer, S. Gulwani, and Z. Popovic, "Personalized Mathematical Word Problem Generation," In *IJCAI*, pp. 381-388, 2015.

[169] M. Gebser, R. Kaminski, B. Kaufmann, and T. Schaub, *Answer set solving in practice*, Springer Nature, 2022.

[170] Q. Zhou, Q. and D. Huang, "Towards generating math word problems from equations and topics," In *Proceedings of the 12th international conference on natural language generation*, pp. 494-503, 2019.

[171] R. Dey, and F.M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," In *2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS)*, pp. 1597-1600, IEEE, 2017.

[172] D. Huang, S. Shi, C. Y. Lin, J. Yin and W. Y. Ma, “How Well Do Computers Solve Math Word Problems? Large-Scale Dataset Construction and Evaluation,” In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, Volume 1: Long Papers, 2016.

[173] Q. Wu, Q. Zhang, and X. Huang, "Automatic math word problem generation with topic-expression co-attention mechanism and reinforcement learning," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp.1061-1072, 2022.

[174] Y. Wang, X. Liu, and S. Shi, "Deep neural solver for math word problems," In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pp. 845-854, 2017.

[175] Z. Zhou, M. Ning, Q. Wang, J. Yao, W. Wang, X. Huang, and K. Huang, "Learning by analogy: Diverse questions generation in math word problem," *arXiv preprint arXiv:2306.09064*, 2023.

[176] N. Chen, N. Wu, J. Chang, and J. Li, "ControlMath: Controllable Data Generation Promotes Math Generalist Models," *arXiv preprint arXiv:2409.15376*, 2024.

[177] Y. Zhang, G. Zhou, Z. Xie, J. Ma, and J.X. Huang, "A diversity-enhanced knowledge distillation model for practical math word problem solving," *Information Processing & Management*, vol. 62, no. 3, p.104059, 2025.

[178] B. Chen, K. Huang, S. Raghupathi, I. Chandratreya, Q. Du, and H. Lipson, "Automated discovery of fundamental variables hidden in experimental data," *Nature Computational Science*, vol. 2, no. 7, pp.433-442, 2022.

[179] A. Garcon, J. Vexler, D. Budker, and S. Kramer, "Deep neural networks to recover unknown physical parameters from oscillating time series," *Plos one*, vol. 17, no. 5, p.e0268439, 2022.

[180] B.K. Petersen, M. Landajuela, T.N. Mundhenk, C.P. Santiago, S.K. Kim, and J.T. Kim, "Deep symbolic regression: Recovering mathematical expressions from data via risk-seeking policy gradients," *arXiv preprint arXiv:1912.04871*, 2019.

[181] S.M. Udrescu, A. Tan, J. Feng, O. Neto, T. Wu, and M. Tegmark, "AI Feynman 2.0: Pareto-optimal symbolic regression exploiting graph modularity," *Advances in Neural Information Processing Systems*, vol. 33, pp.4860-4871, 2020.

[182] B. Buchberger, A. Crăciun, T. Jebelean, L. Kovács, T. Kutsia, K. Nakagawa, F. Piroi, N. Popov, J. Robu, M. Rosenkranz, and W. Windsteiger, "Theorema: Towards computer-aided mathematical theory exploration," *Journal of applied logic*, vol. 4, no. 4, pp.470-504, 2006.

[183] C.A. Paul, *Mathematica*, In Revival: The Handbook of Software for Engineers and Scientists (1995), pp. 926-962, CRC Press, 2018.

[184] W.Y. Chen, Q.H. Hou, and D. Zeilberger, "Automated discovery and proof of congruence theorems for partial sums of combinatorial sequences," *Journal of Difference Equations and Applications*, vol. 22, no. 6, pp.780-788, 2016.

[185] S. Thota, "An Introduction to Maple software," In *National Conference on Advances in Mathematical sciences*, pp. 05-07, 2012.

[186] G. Raayoni, S. Gottlieb, Y. Manor, G. Pisha, Y. Harris, U. Mendlovic, D. Haviv, D. Hadad, and I. Kaminer,"Generating conjectures on fundamental constants with the Ramanujan Machine," *Nature*, vol. 590, no. 7844, pp.67-73, 2021.

CHAPTER 2

[1] M.A. Boden, *The creative mind: Myths and mechanisms*, Routledge, 2004.

[2] J.C. Kaufman and R.J. Sternberg, eds., *The Cambridge handbook of creativity*, Cambridge University Press, 2019.

[3] S. Acar, and M.A. Runco, "Divergent thinking: New methods, recent research, and extended theory", *Psychology of Aesthetics, Creativity, and the Arts*, vol. 13, no. 2, p.153, 2019.

[4] L. Ghosh, R. Kar, A. Konar, A. Chakraborty, and A. K. Nagar, "Identification of Brain Activation Regions in Inductive Learning Based Scientific Creativity Test," In 2018 *IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 950-957, 2018.

[5] S. Ghosh, L. Ghosh, A. Konar and A.K. Nagar, "Assessment of Subjective Creativity Skill Using EEG Induced Capsule Network", In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 3107-3114, 2020.

[6] P.Esling and N. Devis, "Creativity in the era of artificial intelligence", *arXiv preprint arXiv:2008.05959*, 2020.

[7] M. Palmiero, and N. Srinivasan, "Creativity and spatial ability: a critical evaluation," *Cognition, experience and creativity*, pp.189-214, 2015.

[8] H. Burte, A.L. Gardony, A. Hutton, and H.A. Taylor, "Make-A-Dice Test: Assessing the intersection of mathematical and spatial thinking", *Behavior research methods*, vol. 51, pp.602-638, 2019.

[9] K.S. Mix, and M.T. Battista, eds., *Visualizing mathematics: The role of spatial reasoning in mathematical thought*, Springer, 2018.

[10] D. Novitasari, D.K.Risfianty, T.W. Triutami, N.P. Wulandari and R.Y. Tyaningsih, "The relation between spatial reasoning and creativity in solving geometric problems", In *Journal of Physics: Conference Series*, Vol. 1776, No. 1, p. 012007, IOP Publishing, 2021.

[11] J. Suh, and J.Y. Cho, "Linking spatial ability, spatial strategies, and spatial creativity: A step to clarify the fuzzy relationship between spatial ability and creativity," *Thinking Skills and Creativity*, vol. 35, p.100628, 2020.

[12] H.J. Kell, D. Lubinski, C.P. Benbow, and J.H. Steiger, "Creativity and technical innovation: Spatial ability's unique role," *Psychological science*, vol. 24, no. 9, pp.1831-1836, 2013.

[13] A.C. Rule, "Spatial thinking skills and STEM connections: How does this issue address them?", Journal of STEM Arts, Crafts, and Constructions, vol. 1, no. 2, p.1, 2016.

[14] M. Hegarty, R.D. Crookes, D. Dara-Abrams, and T.F.Shipley, "Do all science disciplines rely on spatial abilities? Preliminary evidence from self-report questionnaires", In *Spatial Cognition VII: International Conference, Spatial Cognition* 2010, Mt. Hood/Portland, OR, USA, Proceedings 7 , pp. 85-94, Springer Berlin Heidelberg, August 15-19, 2010.

[15] Z. Hawes. and D. Ansari, "What explains the relationship between spatial and mathematical skills? A review of evidence from brain and behavior", *Psychonomic bulletin & review*, vol. 27, pp.465-482, 2020.

[16] L. Glass, F. Krueger, J. Solomon, V. Raymont, and J. Grafman, "Mental paper folding performance following penetrating traumatic brain injury in combat veterans: a lesion mapping study", *Cerebral Cortex*, vol. 23, no. 7, pp.1663-1672, 2013.

[17] L. Wang, C. Cao, X. Zhou, and C. Qi, "Spatial abilities associated with open math problem solving," *Applied Cognitive Psychology*, vol. 36, no. 2, pp.306-317, 2022.

[18] M. Yu, J. Cui, L. Wang, X. Gao, Z. Cui, and X. Zhou, "Spatial processing rather than logical reasoning was found to be critical for mathematical

problem-solving," *Learning and Individual Differences*, vol. 100, p.102230, 2022.

[19] S. Manimozhi, and N. Vasuki, "Effect of Interventional Strategies to Learn Geometry-A comparative study," *JETT*, vol. 4, pp.181-186, 2022.

[20] J. Adams, I. Resnick, and T. Lowrie, "Supporting senior high-school students' measurement and geometry performance: Does spatial training transfer to mathematics achievement?," *Mathematics Education Research Journal*, vol 35, no. 4, pp.879-900, 2023.

[21] S.F Dahm, and C. Draxler, "Mental Paper Folding Revisited: The Involvement of Visual Action Imagery," *Psych*, vol. 5, no. 1, pp.14-25, 2022.

[22] C. Goumopoulos, and N. Potha, "Mental fatigue detection using a wearable commodity device and machine learning," *Journal of Ambient Intelligence and Humanized Computing*, vol. 14, no. 8, pp.10103-10121, 2023.

[23] Y. Gazes, S. Lee, J. Sakhardande, A. Mensing, Q. Razlighi, A. Ohkawa, M. Pleshkevich, L. Luo, and C. Habeck, "fMRI-guided white matter connectivity in fluid and crystallized cognitive abilities in healthy adults", *Neuroimage*, vol. 215, p.116809, 2020.

[24] S. Xu, Y. Li, and J. Liu, "The neural correlates of computational thinking: Collaboration of distinct cognitive components revealed by fMRI", *Cerebral Cortex*, vol. 31, no. 12, pp.5579-5597, 2021.

[25] C.Y. Tang, E.L. Eaves, J.C. Ng, D.M. Carpenter, X. Mai, D.H. Schroeder, C.A. Condon, R. Colom, and R.J. Haier, "Brain networks for working memory and factors of intelligence assessed in males and females with fMRI and DTI", *Intelligence*, vol. 38, no.3, pp.293-303, 2010.

[26] C. Habeck, J. Steffener, D. Barulli, Y. Gazes, Q. Razlighi, D. Shaked, T. Salthouse, and Y. Stern, "Making cognitive latent variables manifest: distinct neural networks for fluid reasoning and processing speed," *Journal of cognitive neuroscience*, vol. 27, no. 6, pp.1249-1258, 2015.

[27] B. Milivojevic, B.W. Johnson, J.P. Hamm, and M.C. Corballis, "Non-identical neural mechanisms for two types of mental transformation: event-related potentials during mental rotation and mental paper folding," *Neuropsychologia*, vol. 41, no. 10, pp.1345-1356, 2003.

[28] C. Hilton, L. Raddatz, and K. Gramann, "A general spatial transformation process? Assessing the neurophysiological evidence on the similarity of mental rotation and folding," *Neuroimage: Reports*, vol. 2, no. 2, p.100092, 2022.

[29] S. Ghosh, A. Konar, and A.K. Nagar, "Decoding Subjective Creativity Skill from Visuo-Spatial Reasoning Ability Using Capsule Graph Neural Network," In *2021 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8, 2021.

[30] C. Brunner, M. Billinger, M. Seeber, T.R. Mullen, and S. Makeig, "Volume conduction influences scalp-based connectivity estimates," *Front Comput Neurosci*, vol. 10, no. 121, 2016.

[31] F. Wang, X. Zhang, R. Fu and G. Sun, "EEG characteristic analysis of coach bus drivers based on brain connectivity as revealed via a graph theoretical network," *RSC advances*, vol. 8, no. 52, pp.29745-29755, 2018.

[32] A. Dudáš, "Graphical representation of data prediction potential: correlation graphs and correlation chains," *The Visual Computer*, pp.1-14, 2024.

[33] N. Masuda, M. Sakaki, T. Ezaki, and T. Watanabe, "Clustering coefficients for correlation networks," *Frontiers in neuroinformatics*, vol. 12, p.7, 2018.

[34] Z. Šverko, M. Vrankić, S. Vlahinić, and P. Rogelj, "Complex Pearson correlation coefficient for EEG connectivity analysis," *Sensors*, vol. 22, no. 4, p.1477, 2022.

[35] M. Rubinov, and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp.1059-1069, 2010.

[36] A. Fornito, A. Zalesky, E. and Bullmore, Fundamentals of brain network analysis, *Academic press*, 2016.

[37] T.N. Kipf, and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[38] Z. Zhang, P. Cui, and W. Zhu, "Deep learning on graphs: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 1, pp.249-270, 2020.

[39] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: a comprehensive review," *Computational Social Networks*, vol. 6, no. 1, pp.1-23, 2019.

[40] M. Chen, Z. Wei, Z. Huang, B. Ding, and Y. Li, "Simple and deep graph convolutional networks," In *International conference on machine learning*, pp. 1725-1735, PMLR, 2020.

[41] R. Huang, K.S. Hong, D. Yang, and G. Huang, "Motion artifacts removal and evaluation techniques for functional near-infrared spectroscopy signals: a review," *Frontiers in Neuroscience*, vol. 16, p.878750, 2022.

[42] P. Raggam, G. Bauernfeind, and S.C. Wriessnegger, "NICA: a novel toolbox for near-infrared spectroscopy calculations and analyses," *Frontiers in Neuroinformatics*, vol. 14, p.26, 2020.

[43] M. Laha, A. Konar, P. Rakshit, and A.K. Nagar, "Exploration of subjective color perceptual-ability by EEG-induced type-2 fuzzy classifiers," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 12, no. 3, pp.618-635, 2019.

[44] M. Laha, A. Konar, P. Rakshit, and A.K. Nagar, "Hemodynamic analysis for olfactory perceptual degradation assessment using generalized type-2 fuzzy regression," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 3, pp.618-635, 2021.

[45] H. Zhu, H. Zeng, J. Liu, and X. Zhang, “Logish: A new nonlinear nonmonotonic activation function for convolutional neural network,” *Neurocomputing*, vol. 458, pp. 490–499, 2021.

[46] D. Wu, and J.M. Mendel, "A vector similarity measure for linguistic approximation: Interval type-2 and type-1 fuzzy sets," *Information Sciences*, vol. 178, no. 2, pp.381-402, 2008.

[47] D. Wu, "On the fundamental differences between interval type-2 and type-1 fuzzy logic controllers," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 5, pp.832-848, 2012.

[48] H. Burte, A.L. Gardony, A. Hutton, and H.A. Taylor, "Knowing when to fold'em: Problem attributes and strategy differences in the Paper Folding test," *Personality and Individual Differences*, vol. 146, pp.171-181, 2019.

[49] J. Ainooson, and M. Kunda, "A Computational Model for Reasoning About the Paper Folding Task Using Visual Mental Images," In *CogSci*, 2017.

[50] R.J.Cooper, J. Selb, L. Gagnon, D. Phillip, H.W. Schytz, H.K. Iversen, M. Ashina, and D.A. Boas, "A systematic comparison of motion artifact correction techniques for functional near-infrared spectroscopy," *Frontiers in neuroscience*, vol. 6, p.147, 2012.

[51] H.D. Nguyen, S.H. Yoo, M.R. Bhutta, and K.S. Hong, "Adaptive filtering of physiological noises in fNIRS data," *Biomedical engineering online*, vol. 17, pp.1-23, 2018.

[52] S. Kohno, I. Miyai, A. Seiyama, I. Oda, A. Ishikawa, S. Tsuneishi, T. Amita, and K. Shimizu, "Removal of the skin blood flow artifact in functional near-infrared spectroscopic imaging data through independent component analysis," *Journal of biomedical optics*, vol. 12, no. 6, pp.062111-062111, 2007.

[53] M.A. Rahman, M.S. Uddin, and M. Ahmad, "Modeling and classification of voluntary and imagery movements for brain–computer interface from fNIR and EEG signals through convolutional neural network," *Health Information Science and Systems*, vol. 7, no. 1, p.22, 2019.

[54] A. Kachenoura, L. Albera, L. Senhadji, and P. Comon, "ICA: a potential tool for BCI systems," *IEEE Signal Processing Mag.*, vol. 25, no. 1, pp. 57-68, 2008.

[55] Z. Zhao, A. Kleinhans, G. Sandhu, I. Patel, and K.P. Unnikrishnan, "Capsule networks with max-min normalization," *arXiv preprint arXiv:1903.09662*, 2019.

[56] B.C. Van Wijk, C.J. Stam, and A. Daffertshofer, "Comparing brain networks of different size and connectivity density using graph theory," *PloS one*, vol. 5, no. 10, p.e13701, 2010.

[57] K. Qin, D. Lei, W.H. Pinaya, N. Pan, W. Li, Z. Zhu, J.A. Sweeney, A. Mechelli, and Q. Gong, "Using graph convolutional network to characterize individuals with major depressive disorder across multiple imaging sites," *EBioMedicine*, vol 78, 2021.

[58] A. Saha, A. Konar, and A.K. Nagar, "EEG analysis for cognitive failure detection in driving using type-2 fuzzy classifiers," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 1, no. 6, pp.437-453, 2017.

[59] D. Wu, and J.M. Mendel, "Enhanced karnik--mendel algorithms," *IEEE transactions on fuzzy systems*, vol. 17, no. 4, pp.923-934, 2008.

[60] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization", *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 281–305, 2012.

[61] Z. Hawes, J. Moss, B. Caswell, J. Seo, and D. Ansari, "Relations between numerical, spatial, and executive function skills and mathematics achievement: A latent-variable approach", *Cognitive Psychology*, vol. 109, pp.68-90, 2019.

[62] A. Miyake, N.P. Friedman, D.A. Rettinger, P. Shah, and M. Hegarty, "How are visuospatial working memory, executive functioning, and spatial abilities

related? A latent-variable analysis", *Journal of experimental psychology: General*, vol. 130, no. 4, p.621, 2001.

[63] General Assembly of the World Medical Association, "World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects," *The Journal of the American College of Dentists*, vol. 81, no. 3, pp.14-18, 2014.

[64] A. A. Hagberg, D. A. Schult, and P. J. Swart, “Exploring network structure, dynamics, and function using networkx,” In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, 2008.

[65] J. Atwood, and D. Towsley, "Diffusion-convolutional neural networks," *Advances in neural information processing systems*, vol. 29, 2016.

[66] M. Defferrard, X. Bresson, and P. Vandergheynst,"Convolutional neural networks on graphs with fast localized spectral filtering," In *Proceedings of the 30th International Conference on Neural Information Processing Systems* , pp. 3844-3852, 2016.

[67] C. Zhuang and Q. Ma, "Dual graph convolutional networks for graph-based semi-supervised classification," In *Proceedings of the 2018 World Wide Web Conference*, pp. 499-508, 2018.

[68] R. Levie, F. Monti, X. Bresson, and M.M. Bronstein, "Cayleynets: Graph convolutional neural networks with complex rational spectral filters," *IEEE Transactions on Signal Processing*, vol. 67, no. 1, pp.97-109, 2018.

[69] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio and Y. Bengio, " Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[70] F. Ma, F. Gao, J. Sun, H. Zhou, and A. Hussain, "Attention graph convolution network for image segmentation in big SAR imagery data," *Remote Sensing*, vol. 11, no. 21, p.2586, 2019.

[71] K. Masood, and R. Kashef, "Integrating graph convolutional networks (gcnns) and long short-term memory (lstm) for efficient diagnosis of autism," In *International Conference on Artificial Intelligence in Medicine*, pp. 110-121, Cham: Springer International Publishing, 2022.

[72] H. Chang, B. Liu, Y. Zong, C. Lu, and X. Wang, "EEG-Based Parkinson's Disease Recognition Via Attention-based Sparse Graph Convolutional Neural Network," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[73] J. He, J. Cui, G. Zhang, M. Xue, D. Chu, and Y. Zhao, "Spatial–temporal seizure detection with graph attention network and bi-directional LSTM

architecture," *Biomedical Signal Processing and Control*, vol. 78, p.103908, 2022.

[74] Y. Wang, H. Wang, H. Jin, X. Huang, and X. Wang, "Exploring graph capsual network for graph classification," *Information Sciences*, vol. 581, pp.932-950, 2021.

[75] A. Shoeibi, N. Ghassemi, M. Khodatars, P. Moridian, A. Khosravi, A. Zare, J.M. Gorriz, A.H. Chale-Chale, A. Khadem, and U. Rajendra Acharya, "Automatic diagnosis of schizophrenia and attention deficit hyperactivity disorder in rs-fMRI modality using convolutional autoencoder model and interval type-2 fuzzy regression", *Cognitive Neurodynamics*, vol. 17, no. 6, pp.1501-1523, 2023.

[76] M. Ma, S. Na, and H. Wang, "AEGCN: An autoencoder-constrained graph convolutional network," *Neurocomputing*, vol. 432, pp.21-31, 2021.

[77] R. Yang, W. Dai, C. Li, J. Zou, and H. Xiong, "Ncgnn: Node-level capsule graph neural network for semisupervised classification," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[78] X. Shan, J. Cao, S. Huo, L. Chen, P.G. Sarrigiannis, and Y. Zhao, "Spatial–temporal graph convolutional network for Alzheimer classification based on brain functional connectivity imaging of electroencephalogram," *Human Brain Mapping*, vol. 43, no. 17, pp.5194-5209, 2022.

[79] H. Wu, and J. Liu, "A Multi-stream Deep Learning Model for EEG-based Depression Identification," In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2029-2034, 2022.

[80] J. Wang, Y. Song, Q. Gao, and Z. Mao, "Functional brain network based multi-domain feature fusion of hearing-Impaired EEG emotion identification," *Biomedical Signal Processing and Control*, vol. 85, p.105013, 2023.

[81] S. Bloemheuvel, J. van den Hoogen, and M. Atzmueller, "Graph construction on complex spatiotemporal data for enhancing graph neural network-based approaches," *International Journal of Data Science and Analytics*, pp.1-18, 2023.

[82] M. Wu, R. Ouyang, C. Zhou, Z. Sun, F. Li, and P. Li, "A study on the combination of functional connection features and Riemannian manifold in EEG emotion recognition," *Frontiers in Neuroscience*, vol. 17, 2023.

[83] J. Turian, J. Bergstra, and Y. Bengio, “Quadratic features and deep architectures for chunking,” in *Proceedings of Human Language*

Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics, vol. Companion Volume:, pp. 245–248, 2009.

[84] B. Karlik and A. Vehbi, "Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks," *International Journal of Artificial Intelligence and Expert Systems (IJAE)*, vol. 1, no. 4, pp. 111–122, 2011.

[85] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," *Haifa*, pp. 807–814, 2010.

[86] A.K Dubey, and V. Jain, "Comparative study of convolution neural network's relu and leaky-relu activation functions," In *Applications of Computing, Automation and Wireless Systems in Electrical Engineering: Proceedings of MARC 2018*, pp. 873-880, Springer Singapore, 2019.

[87] D.A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," *arXiv preprint arXiv:1511.07289*, 2015.

[88] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, "Self-Normalizing Neural Networks," *arXiv*, 2017.

[89] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for Activation Functions," *ArXiv*, 2017.

[90] M. Basirat, P. M. Roth, "The Quest for the Golden Activation Function," *arXiv*, 2018.

[91] D. Misra, "Mish: A Self Regularized Non-Monotonic Neural Activation Function", *arXiv:1908.08681 [cs stat]*, Oct. 2019.

[92] J. Demsar, "Statistical Comparisons of Classifiers over Multiple Data Sets", *J. Machine Learning Research*, vol. 7, pp. 1-30, 2006.

CHAPTER 3

[1] J.C. Kaufman, and R.J. Sternberg, *The Cambridge Handbook of Creativity*, Cambridge University Press, 2019.

[2] S.M. Ritter, X. Gu, M. Crijns, and P. Biekens, "Fostering students' creative thinking skills by means of a one-year creativity training program," *PloS one*, vol. 15, no. 3, p.e0229773, 2020.

[3] I.A. Fahsing, *The making of an expert detective: Thinking and deciding in criminal investigations*, 2016.

[4] G. Horsman, E. Ryser, and B. Shavers, "Fostering an “investigating mindset”: Why is it important in digital forensic science education?," *Wiley Interdisciplinary Reviews: Forensic Science*, vol. 6, no. 3, p.e1511, 2024.

[5] W. Plöger, M. Krepf, D. Scholl, and A. Seifert, "Analytical competence of teachers," *Teacher Education Quarterly*, vol. 47, no.2, pp.134-157, 2020.

[6] D. BEDİR, S. YILMAZ, and C. KEŞAN, "Determining the analytical thinking levels of prospective teachers," *International Journal of New Trends in Arts, Sports & Science Education (IJTASE)*, vol. 11, no. 4, pp.234-239, 2022.

[7] S. Saniuk, D. Caganova, and A. Saniuk, "Knowledge and skills of industrial employees and managerial staff for the industry 4.0 implementation," *Mobile Networks and Applications*, vol. 28, no. 1, pp.220-230, 2023.

[8] S.R. Campbell, "The Aha! Moment at the nexus of mind and brain," In *Creativity of an Aha! Moment and Mathematics Education*, pp. 365-397, Brill, 2021.

[9] S. Zmigrod, L.S. Colzato, and B. Hommel, "Stimulating creativity: modulation of convergent and divergent thinking by transcranial direct current stimulation (tDCS)," *Creativity Research Journal*, vol. 27, no. 4, pp.353-360, 2015.

[10] G. Sprugnoli, S. Rossi, A. Emmendorfer, A. Rossi, S.L. Liew, E. Tatti, G. di Lorenzo, A. Pascual-Leone, and E. Santarnecchi, "Neural correlates of Eureka moment," *Intelligence*, vol. 62, pp.99-118, 2017.

[11] J. Kounios, and M. Beeman, *The Eureka factor: Creative insights and the brain*, Random House, 2015.

[12] W. Carpenter, "The aha! moment: The science behind creative insights," *Toward super-creativity: Improving creativity in humans, machines, and human-machine collaborations*, pp.11-22, 2019.

[13] Y. Oh, C. Chesebrough, B. Erickson, F. Zhang, and J. Kounios, "An insight-related neural reward signal," *NeuroImage*, vol. 214, p.116757, 2020.

[14] J. Kounios, and M. Beeman, "The cognitive neuroscience of insight," *Annual review of psychology*, vol. 65, no. 1, pp.71-93, 2014.

[15] M.R. Lakshmi, T.V. Prasad, and D.V.C. Prakash, "Survey on EEG signal processing methods," *International journal of advanced research in computer science and software engineering*, vol. 4, no. 1, 2014.

[16] D. Gorjan, K. Gramann, K. De Pauw, and U. Marusic, "Removal of movement-induced EEG artifacts: current state of the art and guidelines," *Journal of neural engineering*, vol. 19, no. 1, p.011004, 2022.

[17] C.R. Rashmi, and C.P. Shantala, "EEG artifacts detection and removal techniques for brain computer interface applications: a systematic review," *International Journal of Advanced Technology and Engineering Exploration*, vol. 9, no. 88, pp.354-383, 2022.

[18] F. de Vico Fallani, J. Richiardi, M. Chavez, and S. Achard, "Graph analysis of functional brain networks: practical issues in translational neuroscience," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 369, no. 1653, p.20130521, 2014.

[19] F. Wang, X. Zhang, R. Fu, and G. Sun, "EEG characteristic analysis of coach bus drivers based on brain connectivity as revealed via a graph theoretical network," *RSC advances*, vol. 8, no. 52, pp.29745-29755, 2018.

[20] M. Rubinov, and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," *Neuroimage*, vol. 52, no. 3, pp.1059-1069, 2010.

[21] A. Fornito, A. Zalesky, and E. Bullmore, *Fundamentals of brain network analysis*, Academic Press, 2016.

[22] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S.Y. Philip, "A comprehensive survey on graph neural networks," *IEEE transactions on neural networks and learning systems*, 2020.

[23] Z. Zhang, P. Cui, and W. Zhu, "Deep learning on graphs: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 1, pp.249-270, 2020.

[24] H.T. Phan, N.T. Nguyen, and D. Hwang, "Fake news detection: A survey of graph neural network methods," *Applied Soft Computing*, vol. 139, p.110235, 2023.

[25] B. Li, and D. Pi, "Network representation learning: a systematic literature review," *Neural Computing and Applications*, vol. 32, no. 21, pp.16647-16679, 2020.

[26] M. Zhang, Z. Cui, M. Neumann, and Y. Chen, "An end-to-end deep learning architecture for graph classification," In *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[27] S. Verma, and Z.L. Zhang, "Graph capsule convolutional neural networks," *arXiv preprint arXiv:1805.08090*, 2018.

[28] T.N. Kipf, and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[29] S. Zheng, L. Qiu, and F. Lan, "TSO-GCN: A Graph Convolutional Network approach for real-time and generalizable truss structural optimization," *Applied Soft Computing*, vol. 134, p.110015, 2023.

[30] S. Sabour, N. Frosst, G. E. Hinton, "Dynamic routing between capsules," *Advances in Neural Information Processing Systems*, pp. 3859-3869, 2017.

[31] S. Ghosh, M. Laha, A, Konar, P. Rakshit, and A.K. Nagar, "Vowel Sound Imagery Decoding by a Capsule Network for the Design of an Automatic Mind-Driven Type-Writer," In *2020 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8, IEEE, 2020.

[32] D. Misra, "Mish: A Self Regularized Non-Monotonic Neural Activation Function," *arXiv:1908.08681 [cs stat]*, Oct. 2019.

[33] A. Martins, and R. Astudillo, "From softmax to sparsemax: A sparse model of attention and multi-label classification," In *International conference on machine learning*, pp. 1614-1623, PMLR, 2016.

[34] T. Dandan, Z. Haixue, L. Wenfu, Y. Wenjing, Q. Jiang, and Z. Qinglin, "Brain activity in using heuristic prototype to solve insightful problems," *Behavioural brain research*, vol. 253, pp.139-144, 2013.

[35] X. Hao, S. Cui, W. Li, W. Yang, J. Qiu, and Q. Zhang, "Enhancing insight in scientific problem solving by highlighting the functional features of prototypes: An fMRI study," *Brain research*, vol. 1534, pp.46-54, 2013.

[36] L. Aziz-Zadeh, J.T. Kaplan, and M. Iacoboni, "“Aha!”: The neural correlates of verbal insight solutions," *Human brain mapping*, vol. 30, no. 3, pp.908-916, 2009.

[37] M. Jung-Beeman, E.M. Bowden, J. Haberman, J.L. Frymiare, S. Arambel-Liu, R. Greenblatt, P.J. Reber, and J. Kounios, "Neural activity when people solve verbal problems with insight," *PLoS biology*, vol. 2, no. 4, p.e97, 2004.

[38] A. Ghani, C. Di Bernardi Luft, S. Ovadio-Caro, K.R. Müller, and J. Bhattacharya, "The receptive brain: up-regulated right temporal alpha oscillation boosting Aha!," *Creativity Research Journal*, vol. 36, no. 3, pp.424-435, 2024.

[39] Y. Yu, Y. Oh, J. Kounios, and M. Beeman, "Dynamics of hidden brain states when people solve verbal puzzles," *NeuroImage*, vol. 255, p.119202, 2022.

[40] T. Minami, Y. Noritake, and S. Nakauchi, "Decreased beta-band activity is correlated with disambiguation of hidden figures," *Neuropsychologia*, vol. 56, pp.9-16, 2014.

[41] Y. Lu, and W. Singer, "Dynamic signatures of the Eureka effect: an EEG study," *Cerebral Cortex*, vol. 33, no. 13, pp.8679-8692, 2023.

[42] G. Bartolomé, S. Vila, C. Torrelles-Nadal, and E. Blanco, "Right cortical activation during generation of creative insights: an electroencephalographic study of coaching," In *Frontiers in Education*, vol. 7, p. 753710, Frontiers Media SA, 2022.

[43] T. Bieth, M. Ovando-Tellez, A. Lopez-Persem, B. Garcin, L. Hugueville, K. Lehongre, R. Levy, N. George, and E. Volle, "Time course of EEG power during creative problem-solving with insight or remote thinking," *Human Brain Mapping*, vol. 45, no. 1, p.e26547, 2024.

[44] R. Ohkuma, Y. Kurihara, T. Takahashi, and R. Osu, "Brain Activity During Constraint Relaxation in the Insight Problem-Solving Process: An fNIRS Study," *bioRxiv*, pp.2022-05, 2022.

[45] J. Qiu, H. Li, J. Jou, J. Liu, Y. Luo, T. Feng, Z. Wu and Q. Zhang, "Neural correlates of the “Aha” experiences: evidence from an fMRI study of insight problem solving," cortex, vol. 46, no. 3, pp.397-403, 2010.

[46] F. Tian, S. Tu, J. Qiu, J.Y. Lv, D.T. Wei, Y.H. Su and Q.L. Zhang, "Neural correlates of mental preparation for successful insight problem solving," *Behavioural brain research*, vol. 216, no. 2, pp.626-630, 2011.

[47] Q. Zhao, Z. Zhou, H. Xu, S. Chen, F. Xu, W. Fan, and L. Han, "Dynamic neural network of insight: a functional magnetic resonance imaging study on solving Chinese ‘chengyu’riddles," *PloS one*, vol. 8, no. 3, p.e59351, 2013.

[48] J. Luo, W. Li, A. Fink, L. Jia, X. Xiao, J. Qiu, and Q. Zhang, "The time course of breaking mental sets and forming novel associations in insight-like problem solving: an ERP investigation," *Experimental brain research*, vol. 212, pp.583-591, 2011.

[49] W. Shen, C. Liu, Y. Yuan, X. Zhang, and J. Luo, "Temporal dynamics of mental impasses underlying insight-like problem solving," *Science China Life Sciences*, vol. 56, pp.284-290, 2013.

[50] Q. Xing, J.X. Zhang, and Z. Zhang, "Event-related potential effects associated with insight problem solving in a Chinese logogriph task," *Psychology*, vol. 3, no. 01, p.65, 2012.

[51] Q. Zhao, Y. Li, X. Shang, Z. Zhou, and L. Han, "Uniformity and nonuniformity of neural activities correlated to different insight problem solving," *Neuroscience*, vol. 270, pp.203-211, 2014.

[52] M. Zhang, F. Tian, X. Wu, S. Liao, and J. Qiu, "The neural correlates of insight in Chinese verbal problems: An event related-potential study," *Brain Research Bulletin*, vol. 84, no. 3, pp.210-214, 2011.

[53] Y. Zhao, S. Tu, M. Lei, J. Qiu, O. Ybarra, and Q. Zhang, "The neural basis of breaking mental set: an event-related potential study," *Experimental brain research*, vol. 208, pp.181-187, 2011.

[54] S. Na, Structural Network Properties and Their Relation to Cognitive Flexibility and Neurological Risk Factors in Adult Survivors of Pediatric Brain Tumors, 2019.

[55] Z. Šverko, M. Vrankić, S. Vlahinić, and P. Rogelj, "Complex Pearson correlation coefficient for EEG connectivity analysis," *Sensors*, vol. 22, no. 4, p.1477, 2022.

[56] A. Dudáš, "Graphical representation of data prediction potential: correlation graphs and correlation chains," *The Visual Computer*, pp.1-14, 2024.

[57] B.C. Van Wijk, C.J. Stam, and A. Daffertshofer, "Comparing brain networks of different size and connectivity density using graph theory," *PloS one*, vol. 5, no. 10, p.e13701, 2010.

[58] M. Defferrard, X. Bresson, and P. Vandergheynst,"Convolutional neural networks on graphs with fast localized spectral filtering," In *Proceedings of the 30th International Conference on Neural Information Processing Systems* , pp. 3844-3852, 2016.

[59] F. Ma, F. Gao, J. Sun, H. Zhou, and A. Hussain, "Attention graph convolution network for image segmentation in big SAR imagery data," *Remote Sensing*, vol. 11, no. 21, p.2586, 2019.

[60] M.A. Mercioni, and S. Holban, "The most used activation functions: Classic versus current," In 2020 International Conference on Development and Application Systems (DAS), pp. 141-145, *IEEE*, 2020.

[61] C. Nwankpa, W. Ijomah, A. Gachagan, and S. Marshall, "Activation functions: Comparison of trends in practice and research for deep learning," *arXiv preprint arXiv:1811.03378*, 2018.

[62] B. Yuen, M.T. Hoang, X. Dong, and T. Lu, "Universal activation function for machine learning," *Scientific reports*, vol. 11, no. 1, p.18757, 2021.

[63] S.R. Dubey, S.K. Singh, and B.B. Chaudhuri, "Activation functions in deep learning: A comprehensive survey and benchmark," *Neurocomputing*, vol. 503, pp.92-108, 2022.

[64] M.K. Patrick, A.F. Adekoya, A.A. Mighty, and B.Y. Edward, "Capsule networks–a survey," *Journal of King Saud University-computer and information sciences*, vol. 341, no.1, pp.1295-1310, 2022.

[65] A. Laha, S.A. Chemmengath, P. Agrawal, M. Khapra, K. Sankaranarayanan, and H.G. Ramaswamy, "On controllable sparse alternatives to softmax," *Advances in neural information processing systems*, vol. 31, 2018.

[66] A. Ulate-Campos, and T. Loddenkemper, "Review on the current long-term, limited lead electroencephalograms," *Epilepsy & Behavior*, vol. 150, p.109557, 2024.

[67] A. Fabregat Sanjuan, R. Pàmies Vilà, A. Rigo Vidal, and V. Pascual Rubio, "Analysis of a new tool for electrode positioning on the scalp according to the international 10/20 system," *Dyna (Bilbao)*, vol. 98, no. 4, pp.335-340, 2023.

[68] DatasetAILDB: https://drive.google.com/file/d/1iFcG6LnaensSYdmwoyPX5R-oH4Vx5_xI/view?usp=drive_link

[69] Dataset NBDB: https://drive.google.com/file/d/1OGYO-dgwq0EYnt6sPYPkuW_aCxQjnCqf/view?usp=drive_link

[70] DatasetWIDB: https://drive.google.com/file/d/1Uvri5rE0oPwlroAMul2h8gLw2-14w8TG/view?usp=drive_link

[71] General Assembly of the World Medical Association, "World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects," *The Journal of the American College of Dentists*, vol. 81, no. 3, pp.14-18, 2014.

[72] A. M. Judith, S.B. Priya, and R.K. Mahendran, "Artifact removal from EEG signals using regenerative multi-dimensional singular value decomposition and

independent component analysis," *Biomedical Signal Processing and Control*, vol. 74, p.103452, 2022.

[73] L. Tuţă, G. Roşu, C. Popovici, and I. Nicolaescu, "Real-time EEG data processing using independent component analysis (ICA)," In *2022 14th International Conference on Communications (COMM)*, pp. 1-4, IEEE, 2022.

[74] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 281–305, 2012.

[75] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *Proc. Int. Conf. Learn. Represent*., pp. 1-41, 2015.

[76] Y. Li, D. Tarlow, M. Brockschmidt, and R. Zemel, "Gated graph sequence neural networks," *arXiv preprint arXiv:1511.05493*, 2015.

[77] R. Levie, F. Monti, X. Bresson, and M.M. Bronstein, "Cayleynets: Graph convolutional neural networks with complex rational spectral filters," *IEEE Transactions on Signal Processing*, vol. 67, no. 1, pp.97-109, 2018.

[78] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, "Graph attention networks," *arXiv preprint arXiv:1710.10903*, 2017.

[79] C. Zhuang and Q. Ma, "Dual graph convolutional networks for graph-based semi-supervised classification," In *Proceedings of the 2018 World Wide Web Conference*, pp. 499-508, 2018.

[80] Z. Xinyi and L. Chen, "Capsule graph neural network," in *Proc. Int. Conf. Learn. Represent*., 2018.

[81] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *The Journal of Machine learning research*, vol. 7, 1-30, 2006.

[82] H. Wang, L. Xu, A. Bezerianos, C. Chen, and Z. Zhang, "Linking attention-based multiscale CNN with dynamical GCN for driving fatigue detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp.1-11, 2020.

[83] Y. Gao, X. Fu, T. Ouyang, and Y. Wang, "EEG-GCN: spatio-temporal and self-adaptive graph convolutional networks for single and multi-view EEG-based emotion recognition," *IEEE Signal Processing Letters*, vol. 29, pp.1574-1578, 2022.

[84] M. Jia, W. Liu, J. Duan, L. Chen, C.P. Chen, Q. Wang, and Z. Zhou, "Efficient graph convolutional networks for seizure prediction using scalp EEG," *Frontiers in Neuroscience*, vol. 16, p.967116, 2022.

[85] K. Qin, D. Lei, W.H. Pinaya, N. Pan, W. Li, Z. Zhu, J.A. Sweeney, A. Mechelli, and Q. Gong, "Using graph convolutional network to characterize individuals with major depressive disorder across multiple imaging sites," *EBioMedicine*, vol 78, 2021.

[86] Z. Wang, C. Chen, J. Li, F. Wan, Y. Sun and H. Wang, "ST-CapsNet: Linking Spatial and Temporal Attention With Capsule Network for P300 Detection Improvement," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 991-1000, 2023,

[87] C. Chen, Z. Ji, Y. Sun, A. Bezerianos, N. Thakor and H. Wang, "Self-Attentive Channel-Connectivity Capsule Network for EEG-Based Driving Fatigue Detection," in *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 3152-3162, 2023.

[88] Y. Li, T. Xu, J. Li, F. Wan, and H. Wang, "Improved dilation CapsuleNet for motor imagery and mental arithmetic classification based on fNIRS," *Brain-Apparatus Communication: A Journal of Bacomics*, vol. 3, no. 1, p.2335886, 2024.

[89] H. Yang, J. Huang, Y. Yu, Z. Sun, S. Zhang, Y. Liu, H. Liu, and L. Xia, "An improved CapsNet based on data augmentation for driver vigilance estimation with forehead single-channel EEG," *Cognitive Neurodynamics*, pp.1-16, 2024.

[90] X. Du, M. Kong, S. Qiu, J. Guo and Y. Lv, "Recognition of Motor Imagery EEG Signals Based on Capsule Network," in *IEEE Access*, vol. 11, pp. 31262-31271, 2023.

CHAPTER 4

[1] R.K. Sawyer, and D. Henriksen, Explaining creativity: The science of human innovation, *Oxford university press*, 2024.

[2] A. Abraham, *The neuroscience of creativity*, Cambridge University Press, 2018.

[3] S. Ghosh, A. Konar, and A.K. Nagar, "Decoding subjective creativity skill from visuo-spatial reasoning ability using capsule graph neural network," In *International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8, IEEE, 2021.

[4] L. Ghosh, R. Kar, A. Konar, A. Chakraborty, and A. K. Nagar, "Identification of Brain Activation Regions in Inductive Learning Based Scientific Creativity

Test," In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 950-957, 2018.

[5] S. Ghosh, and A. Konar, "Decoding the scientific creative-ability of subjects using dual attention induced graph convolutional-capsule network," *Applied Soft Computing*, vol. 161, p.111769, 2024.

[6] J.C. Kaufman and R.J. Sternberg, eds., *The Cambridge handbook of creativity*, Cambridge University Press, 2019.

[7] H.B. de Vries, and T.I. Lubart, "Scientific creativity: divergent and convergent thinking and the impact of culture," *The Journal of Creative Behavior*, vol. 53, no. 2, pp.145-155, 2019.

[8] S.I. Robertson, *Problem solving: Perspectives from cognition and neuroscience*, Psychology Press, 2016.

[9] R. Bianchi, and I.S. Schonfeld, "Occupational depression, cognitive performance, and task appreciation: A study based on Raven's Advanced Progressive Matrices," *Frontiers in psychology*, vol. 12, p.695539, 2021.

[10] L. Cipolotti, J.K. Ruffle, J. Mole, T. Xu, H. Hyare, T. Shallice, R. Chan, and P. Nachev, "Graph lesion-deficit mapping of fluid intelligence," *Brain*, vol. 146, no. 1, pp.167-181, 2023.

[11] D.A. Joyner, D. Bedwell, C. Graham, W. Lemmon, O. Martinez, and A.K. Goel, "Using Human Computation to Acquire Novel Methods for Addressing Visual Analogy Problems on Intelligence Tests," In *ICCC*, pp. 23-30, 2015.

[12] D. Parameshwaran, N.P. Subramaniyam, and T.C. Thiagarajan, "Waveform complexity: A new metric for EEG analysis," *Journal of neuroscience methods*, vol. 325, p.108313, 2019.

[13] Y. Liu, K. He, K. Man, and P. Zhan, "Exploring Critical Eye-Tracking Metrics for Identifying Cognitive Strategies in Raven's Advanced Progressive Matrices: A Data-Driven Perspective," *Journal of Intelligence*, vol. 13, no. 2, p.14, 2025.

[14] A. Chuderski, J. Jastrzębski, B. Kroczek, H. Kucwaj, and M. Ociepka, "Metacognitive experience on Raven's matrices versus insight problems," *Metacognition and Learning*, vol. 16, pp.15-35, 2021.

[15] N. Friedman, T. Fekete, K. Gal, and O. Shriki, "EEG-based prediction of cognitive load in intelligence tests," *Frontiers in human neuroscience*, vol. 13, p.191, 2019.

[16] T. Xu, J. Wang, G. Zhang, L. Zhang, and Y. Zhou, "Confused or not: decoding brain activity and recognizing confusion in reasoning learning using EEG," *Journal of Neural Engineering*, vol. 20, no. 2, p.026018, 2023.

[17] T.M. Morin, K.N. Moore, K. Isenburg, W. Ma, and C.E. Stern, "Functional reconfiguration of task-active frontoparietal control network facilitates abstract reasoning," *Cerebral Cortex*, vol. 33, no. 10, pp.5761-5773, 2023.

[18] L. Anat, R. Reut, I. Nofar, T. Niv, S. Maayan, T. Galia, and L. Abigail, "The role of the cerebellum in fluid intelligence: An fMRI study," *Cognitive Systems Research*, vol. 83, p.101178, 2024.

[19] Z. Chen, A. De Beuckelaer, X. Wang, and J. Liu, "Distinct neural substrates of visuospatial and verbal-analytic reasoning as assessed by Raven's Advanced Progressive Matrices," *Scientific reports*, vol. 7, no. 1, p.16230, 2017.

[20] H.U. Amin, A.S. Malik, M. Hussain, N. Kamel, and W.T. Chooi, "Brain behavior during reasoning and problem solving task: an EEG study," In *5th International Conference on Intelligent and Advanced Systems (ICIAS)*, pp. 1-4, 2014.

[21] S. Jawed, H.U. Amin, A.S. Malik, and I. Faye, "Hemispheric Asymmetries in Electroencephalogram Oscillations for Long-Term Memory Retrieval in Healthy Individuals," *Brain Sciences*, vol. 10, no. 12, p.937, 2020.

[22] M. Ociepka, P. Kałamała, and A. Chuderski, "Take your time: Slow brain rhythms predict fluid intelligence," *Intelligence*, vol. 100, p.101780, 2023.

[23] W. Luo, and R. Zhou, "Can working memory task-related EEG biomarkers measure fluid intelligence and predict academic achievement in healthy children?," *Frontiers in Behavioral Neuroscience*, vol. 14, p.2, 2020.

[24] H.U. Amin, A.S. Malik, N. Kamel, W.T. Chooi, and M. Hussain, "P300 correlates with learning & memory abilities and fluid intelligence," *Journal of neuroengineering and rehabilitation*, vol. 12, no. 1, pp.1-14, 2015.

[25] S. Ghosh, A. Konar, and A.K. Nagar, "Analyzing the Creative Potential of Subjects Using EEG-Induced Capsule Graph Neural Network," In *2024 International Joint Conference on Neural Networks (IJCNN)*, pp. 1-8, IEEE, 2024.

[26] C. Brunner, M. Billinger, M. Seeber, T.R. Mullen, and S. Makeig, "Volume conduction influences scalp-based connectivity estimates," *Front Comput Neurosci*, vol. 10, no. 121, 2016.

[27] J. Park, J. Shin, and J. Jeong, "Inter-brain synchrony levels according to task execution modes and difficulty levels: an fNIRS/GSR study," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp.194-204, 2022.

[28] A. Fornito, A. Zalesky, and E. Bullmore, *Fundamentals of brain network analysis*, Academic Press, 2016.

[29] T.N. Kipf, and M. Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint arXiv:1609.02907*, 2016.

[30] K. Qin, D. Lei, W.H. Pinaya, N. Pan, W. Li, Z. Zhu, J.A. Sweeney, A. Mechelli, and Q. Gong, "Using graph convolutional network to characterize individuals with major depressive disorder across multiple imaging sites," *EBioMedicine*, vol 78, 2021.

[31] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: a comprehensive review," *Computational Social Networks*, vol. 6, no. 1, pp.1-23, 2019.

[32] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and S.Y. Philip, "A comprehensive survey on graph neural networks," *IEEE transactions on neural networks and learning systems*, 2020.

[33] S. Anwer, H. Li, M. F. Antwi-Afari, A. M. Mirza, M. A. Rahman, I. Mehmood, R. Guo, and A.Y.L. Wong, "Evaluation of Data Processing and Artifact Removal Approaches Used for Physiological Signals Captured Using Wearable Sensing Devices during Construction Tasks," *Journal of Construction Engineering and Management*, vol. 150, no. 1, p.03123008, 2024.

[34] R. Huang, K.S. Hong, D. Yang, and G. Huang, "Motion artifacts removal and evaluation techniques for functional near-infrared spectroscopy signals: a review," *Frontiers in Neuroscience*, vol. 16, p.878750, 2022.

[35] M. Laha, A. Konar, P. Rakshit, and A.K. Nagar, "Hemodynamic analysis for olfactory perceptual degradation assessment using generalized type-2 fuzzy regression," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 14, no. 3, pp.618-635, 2021.

[36] M. Laha, A. Konar, and A.K. Nagar, "Olfactory Perceptual-Ability Assessment by Near-Infrared Spectroscopy Using Vertical-Slice Based Fuzzy Reasoning," *IEEE Access*, vol. 11, pp.17779-17792, 2023.

[37] S. Ghosh, A. Konar, and A.K. Nagar, "Cognitive Assessment of Scientific Creative-Skill by Brain-Connectivity Analysis Using Graph Convolutional-Interval Type-2 Fuzzy Network," *IEEE Transactions on Cognitive and Developmental Systems*, 2024.

[38] X. Wang, H. Ren, and A. Wang, "Smish: A novel activation function for deep learning methods," *Electronics*, vol. 11, no. 4, p.540, 2022.

[39] Y. Zhang, H. Ishibuchi, S. Wang, "Deep Takagi–Sugeno–Kang fuzzy classifier with shared linguistic fuzzy rules," *IEEE Transacions on Fuzzy Systems*, vol. 26, no. 3, pp. 1535-1549, 2017.

[40] N. N. Karnik, J. M. Mendel, and Q. Liang, "Type-2 fuzzy logic systems," *IEEE trans. on Fuzzy Systems*, vol. 7, no.6, pp. 643-658, 1999.

[41] D. Wu, C. T. Lin, J. Huang, and Z. Zeng, "On the functional equivalence of TSK fuzzy systems to neural networks, mixture of experts, CART, and stacking ensemble regression," *IEEE Transactions on Fuzzy Systems*, vol. 28, no.10, pp. 2570-2580, 2019.

[42] R. Zurrin, S.T.S. Wong, M.M. Roes, C.M. Percival, A. Chinchani, L. Arreaza, M. Kusi, A. Momeni, M. Rasheed, Z. Mo, Z. and V.M. Goghari, "Functional brain networks involved in the Raven's standard progressive matrices task and their relation to theories of fluid intelligence," *Intelligence*, vol. 103, p.101807, 2024.

[43] N. Naseer, and K.S. Hong, "fNIRS-based brain-computer interfaces: a review," *Frontiers in human neuroscience*, vol. 9, p.3, 2015.

[44] H. Wang, X.Zhang, J. Li, B. Li, X. Gao, Z. Hao, J. Fu, Z. Zhou, and M. Atia, "Driving risk cognition of passengers in highly automated driving based on the prefrontal cortex activity via fNIRS," *Scientific Reports*, vol. 13, no. 1, p.15839, 2023.

[45] V. Kumar, M. Arya, A. Kumar, and D.K. Jhariya, "Design and Comparison Between IIR Butterworth and Chebyshev Digital Filters using MATLAB," In Fourth International Conference on Advances in Electrical, Computing, *Communication and Sustainable Technologies (ICAECT)*, pp. 1-7, IEEE, 2024.

[46] A. Kachenoura, L. Albera, L. Senhadji, and P. Comon, “ICA: a potential tool for BCI systems,” *IEEE Signal Processing Mag.*, vol. 25, no. 1, pp. 57-68, 2008.

[47] A. Grinsted, J.C. Moore, and S. Jevrejeva, "Application of the cross wavelet transform and wavelet coherence to geophysical time series," *Nonlinear processes in geophysics*, vol. 11, no. 5/6, pp.561-566, 2004.

[48] H. Zaynidinov, U. Juraev, S. Tishlikov, and J. Modullayev, "Application of Daubechies Wavelets in Digital Processing of Biomedical Signals and Images," In *International Conference on Intelligent Human Computer Interaction*, pp. 194-206, Cham: Springer Nature Switzerland, 2024.

[49] B.C. Van Wijk, C.J. Stam, and A. Daffertshofer, "Comparing brain networks of different size and connectivity density using graph theory," *PloS one*, vol. 5, no. 10, p.e13701, 2010.

[50] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization", *The Journal of Machine Learning Research*, vol. 13, no. 1, pp. 281–305, 2012.

[51] A. Khasnobish, A. Konar, D.N Tibarewala, and A.K. Nagar, "Bypassing the natural visual-motor pathway to execute complex movement related tasks using interval type-2 fuzzy sets," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 1, pp.91-105, 2016.

[52] G. Klir, and B. Yuan, *Fuzzy sets and fuzzy logic*, vol. 4, pp. 1-12, New Jersey: Prentice hall, 1995.

[53] J. S. R. Jang, C. T. Sun, E. Mizutani, *Neuro-Fuzzy and Soft Computing*, prentice-Hall,1997.

[54] J. Duncan, M. Assem, M. and S. Shashidhara, "Integrated intelligence from distributed brain activity," *Trends in Cognitive Sciences*, vol. 24, no. 10, pp.838-852, 2020.

[55] General Assembly of the World Medical Association, "World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects," *The Journal of the American College of Dentists*, vol. 81, no. 3, pp.14-18, 2014.

[56] M.L. Stanley, M.N. Moussa, B.M. Paolini, R.G. Lyday, J.H. Burdette, and P.J. Laurienti, "Defining nodes in complex brain networks," *Frontiers in computational neuroscience*, vol. 7, p.169, 2013.

[57] J. Molina del Río, M.A. Guevara, M. Hernández González, R.M. Hidalgo Aguirre, and M.A. Cruz Aguilar, "EEG correlation during the solving of simple and complex logical–mathematical problems," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 19, pp.1036-1046, 2019.

[58] E. Juliyanto, P. Marwoto, R.S. Iswari, S.E. Nugroho, and B.N. Mindyarto, "Brain activity of problem solving proccess: a systematic literarure review," In *Journal of Physics: Conference Series*, vol. 1918, no. 5, p. 052068, IOP Publishing, 2021.

[59] L. Yang, and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp.295-316, 2020.

[60] M. Defferrard, X. Bresson, and P. Vandergheynst,"Convolutional neural networks on graphs with fast localized spectral filtering," In *Proceedings of the 30th International Conference on Neural Information Processing Systems* , pp. 3844-3852, 2016.

[61] C. Zhuang and Q. Ma, "Dual graph convolutional networks for graph-based semi-supervised classification," In *Proceedings of the 2018 World Wide Web Conference*, pp. 499-508, 2018.

[62] H. Chang, B. Liu, Y. Zong, C. Lu, and X. Wang, "EEG-Based Parkinson's Disease Recognition Via Attention-based Sparse Graph Convolutional Neural Network," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[63] M. Ma, S. Na, and H. Wang, "AEGCN: An autoencoder-constrained graph convolutional network," *Neurocomputing*, vol. 432, pp.21-31, 2021.

[64] A. Shoeibi, N. Ghassemi, M. Khodatars, P. Moridian, A. Khosravi, A. Zare, J.M. Gorriz, A.H. Chale-Chale, A. Khadem, and U. Rajendra Acharya, "Automatic diagnosis of schizophrenia and attention deficit hyperactivity disorder in rs-fMRI modality using convolutional autoencoder model and interval type-2 fuzzy regression", *Cognitive Neurodynamics*, vol. 17, no. 6, pp.1501-1523, 2023.

[65] H. Wu, and J. Liu, "A Multi-stream Deep Learning Model for EEG-based Depression Identification," In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2029-2034, 2022.

[66] J. Wang, Y. Song, Q. Gao, and Z. Mao, "Functional brain network based multi-domain feature fusion of hearing-Impaired EEG emotion identification," *Biomedical Signal Processing and Control*, vol. 85, p.105013, 2023.

[67] M. Wu, R. Ouyang, C. Zhou, Z. Sun, F. Li, and P. Li, "A study on the combination of functional connection features and Riemannian manifold in EEG emotion recognition," *Frontiers in Neuroscience*, vol. 17, 2023.

[68] F. Wang, X. Zhang, R. Fu and G. Sun, "EEG characteristic analysis of coach bus drivers based on brain connectivity as revealed via a graph theoretical network," *RSC advances*, vol. 8, no. 52, pp.29745-29755, 2018.

[69] V. Nair and G. E. Hinton, “Rectified linear units improve restricted boltzmann machines,” *Haifa*, pp. 807–814, 2010.

[70] A.K Dubey, and V. Jain, "Comparative study of convolution neural network’s relu and leaky-relu activation functions," In *Applications of Computing, Automation and Wireless Systems in Electrical Engineering: Proceedings of MARC 2018*, pp. 873-880, Springer Singapore, 2019.

[71] D.A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," *arXiv preprint arXiv:1511.07289*, 2015.

[72] G. Klambauer, T. Unterthiner, A. Mayr, and S. Hochreiter, “Self-Normalizing Neural Networks,” *arXiv*, 2017.

[73] S. Elfwing, E. Uchibe, and K. Doya, "Sigmoid-weighted linear units for neural network function approximation in reinforcement learning," *Neural Networks*, vol. 107, pp.3-11, 2018.

[74] M. Basirat, P. M. Roth, "The Quest for the Golden Activation Function," *arXiv*, 2018.

[75] D. Misra, "Mish: A Self Regularized Non-Monotonic Neural Activation Function", *arXiv:1908.08681 [cs stat]*, Oct. 2019.

[76] H. Zhu, H. Zeng, J. Liu, and X. Zhang, “Logish: A new nonlinear nonmonotonic activation function for convolutional neural network,” *Neurocomputing*, vol. 458, pp. 490–499, 2021.

[77] S. M. Taheri and G. Hesamian, "A generalization of the Wilcoxon signedrank test and its applications, " *Stat. Papers*, vol. 54, no. 2, pp. 457–470, May 2013.

CHAPTER 5

[1] M. A. Boden, *The creative mind: Myths and mechanisms*, Routledge, 2004.

[2] James C. Kaufman and Robert J. Sternberg, eds. *The Cambridge Handbook of Creativity*., Cambridge University Press, 2019.

[3] P. Castells, N. Hurley, and S. Vargas, "Novelty and diversity in recommender systems," in *Recommender systems handbook*, pp. 603-646, Springer, New York, NY, 2022.

[4] M.T. Llano, M. d'Inverno, M. Yee-King, J. McCormack, A. Ilsar, A. Pease and S. Colton, "Explainable computational creativity," *arXiv preprint arXiv:2205.05682*, 2022.

[5] G. Franceschelli, and M. Musolesi, "Creativity and machine learning: A survey," *arXiv preprint arXiv:2104.02726*, 2021.

[6] D. Mateja, and A. Heinzl, "Towards Machine Learning as an Enabler of Computational Creativity," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 6, pp.460-475, 2021.

[7] S.J. Russell, and P. Norvig, *Artificial intelligence: a Modern Approach*, Pearson, 2016.

[8] A. Konar, *Artificial intelligence and soft computing: behavioral and cognitive modeling of the human brain*, CRC press, 2018.

[9] M. Pearce, J. McKinney, and C. Alvin, "Query-Based Generation of Trigonometric Identity Problems and Solutions." In *The Thirty-Third International Flairs Conference*, 2020.

[10] Z. Liu, Y. Li, Z. Liu, L. Li, and Z. Li, "Learning to prove trigonometric identities," *arXiv preprint arXiv:2205.06679*, 2022.

[11] A. Papasalouros, "Automatic exercise generation in euclidean geometry," In *IFIP International Conference on Artificial Intelligence Applications and Innovations*, Springer, Berlin, Heidelberg, pp. 141-150, 2013.

[12] B.T. Tabuguia, "Hypergeometric-type sequences," *Journal of Symbolic Computation*, vol. 125, p.102328, 2024.

[13] I. Briggs, and P. Panchekha, "Synthesizing mathematical identities with e-graphs," In *Proceedings of the 1st ACM SIGPLAN International Symposium on E-Graph Research, Applications, Practices, and Human-factors*, pp. 1-6, 2022.

[14] R. Singh, S. Gulwani, and S. Rajamani, "Automatically generating algebra problems," In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 26, no. 1. 2012.

[15] I. Brito, J.J. Almeida, and G.J. Machado, "Mvgen: multi version question generation for math courses," *Open Education Studies*, vol. 1, no.1, pp.146-150, 2019.

[16] J. Zhang, and F. Shen, "Sumudu transform for automatic mathematical proof and identity generation," In *Information Technology-New Generations: 14th*

International Conference on Information Technology, Springer International Publishing, pp. 677-683, 2018.

[17] S. Ghosh, A. Konar and A. K. Nagar, "Computational Creativity by Generative Adversial Network with Leaked Information," *2024 International Joint Conference on Neural Networks (IJCNN)*, Yokohama, Japan, 2024.

[18] J. Guo, S. Lu, H. Cai, W. Zhang, Y. Yu and J. Wang, "Long text generation via adversarial training with leaked information", *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.

[19] C. Kiddon, L. Zettlemoyer, and Y. Choi, "Globally coherent text generation with neural checklist models," In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pp. 329-339, 2016.

[20] S. Hochreiter, and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol.9, no. 8, pp.1735-1780, 1995.

[21] S.R. Bowman, L. Vilnis, O. Vinyals, A.M. Dai, R. Jozefowicz, and S. Bengio, "Generating sentences from a continuous space," *arXiv preprint arXiv:1511.06349*, 2015.

[22] L. Yu, W. Zhang, J. Wang, and Y. Yu, "Seqgan: Sequence generative adversarial nets with policy gradient," In *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2015.

[23] K. Lin, D. Li, X. He, Z. Zhang, and M.T. Sun, "Adversarial ranking for language generation," *Advances in neural information processing systems*, vol. 30, 2015.

[24] P.V.L. Pham, A.V. Duc, N.M. Hoang, X.L. Do, and A.T. Luu, "ChatGPT as a Math Questioner? Evaluating ChatGPT on Generating Pre-university Math Questions," In *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*, pp. 65-73, 2024.

[25] A. Deroy, and S. Maity, "Exploring the mathematical reasoning capabilities of gemini," *Authorea Preprints*, 2024.

[26] L. Yu, W. Jiang, H. Shi, J. Yu, Z. Liu, Y. Zhang, J.T. Kwok, Z. Li, A. Weller, and W. Liu, "Metamath: Bootstrap your own mathematical questions for large language models," In *The Twelfth International Conference on Learning Representations (ICLR)*, 2023.

[27] A. Lipowski, and D. Lipowska, "Roulette-wheel selection via stochastic acceptance," *Physica A: Statistical Mechanics and its Applications*, vol.391, no.6, pp.2193-2196, 2012.

[28] Z. Li et al., “Superbnn: Randomized binary neural network using adiabatic superconductor josephson devices,” in *Proceedings of the 56th Annual IEEE/ACM International Symposium on Microarchitecture*, 2023.

[29] W. Chen, H. Du, and C.C. Chen, "Stability and robustness analysis of finite-time consensus algorithm for second-order multiagent systems under sampled-data control," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 3, pp.1445-1452, 2023.

[30] M. Hardt,and A. Moitra, "Algorithms and hardness for robust subspace recovery," In *Conference on Learning Theory*, pp. 354-375, PMLR, 2013.

[31] S.L. Loney, *Plane Trigonometry: Part 1*, Cambridge University Press, 1912.

[32] H.S. Hall, and S.R. Knight, *Elementary trigonometry*, Macmillan Company of Canada, 1919.

[33] I.M. Gel′fand, M. Saul, I.M. Gelfand, and M. Saul, *Trigonometry*, Birkhäuser Boston, 2001.

[34] C.V. Durell, and A. Robson, *Advanced trigonometry*, Courier Corporation, 2003.

[35] A.A. Klaf, *Trigonometry refresher*, Courier Corporation, 2005.

[36] S. Axler, *Algebra and Trigonometry*, John Wiley & Sons, 2011.

[37] R.E. Moyer, *Schaum's Outline of Trigonometry*, McGraw-Hill Education, 2018.

[38] U. Jain, Z. Zhang, and A.G. Schwing, "Creativity: Generating diverse questions using variational autoencoders," In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6485-6494, 2015.

[39] K. Liu et al., "Measuring the Consistency and Diversity of 3D Face Generation," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 17, no. 6, pp. 1208-1220, Nov. 2023.

[40] F. Ibarrola, and K. Grace, "Measuring Diversity in Co-creative Image Generation," *arXiv preprint arXiv:2403.13826*, 2024.

CHAPTER 6

[1] James C. Kaufman and Robert J. Sternberg, eds. *The Cambridge Handbook of Creativity*., Cambridge University Press, 2019.

[2] A. E. Green, R. E. Beaty, Y. N. Kenett, J. C. Kaufman, "The process definition of creativity," *Creativity Research Journal*, vol. 36, no. 3, pp. 544–572, 2024.

[3] S. Ghosh, A. Konar, A. K. Nagar, "Cognitive Assessment of Scientific Creative Skill by Brain-Connectivity Analysis Using Graph Convolutional Interval Type-2 Fuzzy Network," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 16, no. 5, pp.1872-1886, 2024.

[4] R. K. Sawyer, D. Henriksen, *Explaining creativity: The science of human innovation*, Oxford university press, 2024.

[5] J.-Y. Oh, "Understanding the scientific creativity based on various perspectives of science," *Axiomathes*, vol. 32, no. 6, pp. 907–929, 2022.

[6] A. Abraham, *The neuroscience of creativity*, Cambridge University Press, 2018.

[7] M. Hasegawa, "New paradigm of creativity: From newtonian mechanics to quantum mechanics and higher education development," in: *Handbook of Research on Creative Problem-Solving Skill Development in Higher Education*, IGI Global, pp. 513–532, 2017.

[8] A. Konar, *Artificial intelligence and soft computing: behavioral and cognitive modeling of the human brain*, CRC press, 2018.

[9] T. Veale, R. Pérez y Pérez, "Leaps and bounds: An introduction to the field of computational creativity," *New Generation Computing*, vol. 38, pp. 551–563, 2020.

[10] D. Mateja, A. Heinzl, "Towards machine learning as an enabler of computational creativity," *IEEE Transactions on Artificial Intelligence*, vol. 2, no. 6, pp.460–475, 2021.

[11] G. Franceschelli, M. Musolesi, "Creativity and machine learning: A survey," *ACM Computing Surveys*, vol. 56, no. 11, pp.1–41, 2024.

[12] G. Raayoni, S. Gottlieb, Y. Manor, G. Pisha, Y. Harris, U. Mendlovic, D. Haviv, Y. Hadad, I. Kaminer, "Generating conjectures on fundamental constants with the ramanujan machine," *Nature*, vol. 590, no. 7844, pp.67–73, 2021.

[13] T. Wu, M. Tegmark, "Toward an artificial intelligence physicist for unsupervised learning," *Physical Review E*, vol. 100, no. 3, pp.033311, 2019.

[14] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, et al., "Accurate

structure prediction of biomolecular interactions with alphafold 3," *Nature*, vol. 630, no. 8016, pp. 493–500, 2024.

[15] D. Dudka, R. B. Akins, M. A. Lampson, "Freeda: an automated computational pipeline guides experimental testing of protein innovation," *Journal of Cell Biology*, vol. 222. no. 9, 2023.

[16] V. Zambaldi, D. La, A. E. Chu, H. Patani, A. E. Danson, T. O. Kwan, T. Frerix, R. G. Schneider, D. Saxton, A. Thillaisundaram, et al., "De novo design of high-affinity protein binders with alphaproteo," *arXiv preprint arXiv:2409.08022*, 2024.

[17] Y. Hakuk, Y. Reich, "Automated discovery of scientific concepts: Replicating three recent discoveries in mechanics," *Advanced Engineering Informatics*, vol. 44, pp.101080, 2020.

[18] R. Iten, T. Metger, H. Wilming, L. Del Rio, R. Renner, "Discovering physical concepts with neural networks," *Physical review letters*, vol. 124, no. 1, pp.010508, 2020.

[19] T. H. Trinh, Y. Wu, Q. V. Le, H. He, T. Luong, "Solving olympiad geometry without human demonstrations," *Nature*, vol. 625, no. 7995, pp.476–482, 2024.

[20] B. Chen, K. Huang, S. Raghupathi, I. Chandratreya, Q. Du, H. Lipson, "Automated discovery of fundamental variables hidden in experimental data," *Nature Computational Science*, vol. 2, no. 7, pp.433–442, 2022.

[21] A. Garcon, J. Vexler, D. Budker, S. Kramer, "Deep neural networks to recover unknown physical parameters from oscillating time series," *Plos one*, vol. 17, no. 5, pp.e0268439, 2022.

[22] M. Ismayilzada, D. Paul, A. Bosselut, L. van der Plas, "Creativity in ai: Progresses and challenges," *arXiv preprint arXiv:2410.17218*, 2024.

[23] M. B. Fazi, "Can a machine think (anything new)? automation beyond simulation," *Ai & society*, vol. 34, no. 4, pp.813–824, 2019.

[24] G. Malik, D. K. Tayal, S. Vij, "An analysis of the role of artificial intelligence in education and teaching," in: *Recent Findings in Intelligent Computing Techniques: Proceedings of the 5th ICACNI 2017*, Vol. 1, Springer, pp. 407–417, 2019.

[25] M. Krenn, R. Pollice, S. Y. Guo, M. Aldeghi, A. Cervera-Lierta, P. Friederich, G. dos Passos Gomes, F. Häse, A. Jinich, A. Nigam, et al., "On scientific

understanding with artificial intelligence," *Nature Reviews Physics*, vol. 4, no. 12, pp.761–769, 2022.

[26] Z. Liu, Y. Li, Z. Liu, L. Li, Z. Li, "Learning to prove trigonometric identities," *arXiv preprint arXiv:2207.06679*, 2022.

[27] M. Pearce, J. McKinney, C. Alvin, "Query-based generation of trigonometric identity problems and solutions.," in: *FLAIRS*, pp. 445–450, 2020.

[28] S. Ghosh and A. Konar, "Computational Creativity by Diversity-Optimized Intelligent Search: An Automatic Approach to Artificial Synthesis of Trigonometric Identities," in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 55, no. 6, pp. 4385-4395, June 2025.

[29] [29] L. Yu, W. Jiang, H. Shi, J. Yu, Z. Liu, Y. Zhang, J. T. Kwok, Z. Li, A. Weller, W. Liu, "Metamath: Bootstrap your own mathematical questions for large language models," *arXiv preprint arXiv:2309.12284*, 2023.

[30] [30] I. Drori, S. Zhang, R. Shuttleworth, L. Tang, A. Lu, E. Ke, K. Liu, L. Chen, S. Tran, N. Cheng, et al., "A neural network solves, explains, and generates university math problems by program synthesis and few-shot learning at human level," *Proceedings of the National Academy of Sciences*, vol. 119, no. 32, pp.e2123433119, 2022.

[31] X. Zhao, W. Wu, J. Guan, L. Kong, "Promptcot: Synthesizing olympiad-level problems for mathematical reasoning in large language models," *arXiv preprint arXiv:2503.02324*, 2025.

[32] S. Ghosh, A. Konar and A. K. Nagar, "Computational Creativity by Generative Adversarial Network with Leaked Information," 2024 *International Joint Conference on Neural Networks (IJCNN)*, Yokohama, Japan, 2024.

[33] H. Ding, N. Huang, X. Cui, "Leveraging gans data augmentation for imbalanced medical image classification," *Applied Soft Computing*, vol. 165, pp.112050, 2024.

[34] J. Guo, S. Lu, H. Cai, W. Zhang, Y. Yu, J. Wang, "Long text generation via adversarial training with leaked information," in: *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, 2018.

[35] C. Közkurt, A. Diker, A. Elen, S. Kılıçarslan, E. Dönmez, F. B. Demir, "Trish: an efficient activation function for cnn models and analysis of its effectiveness with optimizers in diagnosing glaucoma," *The Journal of Supercomputing*, vol. 80, no. 11, pp.15485–15516, 2024.

[36] E. Reiter, "A structured review of the validity of bleu," *Computational Linguistics*, vol. 44, no. 3, pp.393–401, 2018.

[37] I. Brito, J. J. Almeida, G. J. Machado, "Mvgen: Multi version question generation for math courses," *Open Education Studies*, vol. 1, no. 1, pp.146–150, 2019.

[38] J. Zhang, F. Shen, "Sumudu transform for automatic mathematical proof and identity generation," in: *Information Technology-New Generations: 14th International Conference on Information Technology*, Springer, pp. 677–683, 2018.

[39] B. T. Tabuguia, "Hypergeometric-type sequences," *Journal of Symbolic Computation*, vol. 125, pp.102328, 2024.

[40] P. V. L. Pham, A. V. Duc, N. M. Hoang, X. L. Do, A. T. Luu, "Chatgpt as a math questioner? evaluating chatgpt on generating pre-university math questions," in: *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*, pp. 65–73, 2024.

[41] A. Deroy, S. Maity, "Exploring the mathematical reasoning capabilities of gemini," *Authorea Preprints*, 2024.

[42] D. Tito Svenstrup, J. Hansen, O. Winther, "Hash embeddings for efficient word representations," *Advances in neural information processing systems*, vol. 30, 2017.

[43] P. Bachman, D. Precup, "Data generation as sequential decision making," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[44] H. Pratiwi, A. P. Windarto, S. Susliansyah, R. R. Aria, S. Susilowati, L. K. Rahayu, Y. Fitriani, A. Merdekawati, I. R. Rahadjeng, "Sigmoid activation function in selecting the best model of artificial neural networks," in: *Journal of Physics: Conference Series*, vol. 1471, IOP Publishing, 2020.

[45] R. S, A. S. Bharadwaj, D. S K, M. S. Khadabadi and A. Jayaprakash, "Digital Implementation of the Softmax Activation Function and the Inverse Softmax Function," 2022 *4th International Conference on Circuits, Control, Communication and Computing (I4C)*, pp. 64-67, Bangalore, India, 2022.

[46] A. Sherstinsky, "Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network," *Physica D: Nonlinear Phenomena*, vol. 404, pp.132306, 2020.

[47] J. S. Steele, K. J. Grimm, "Using sympy (symbolic python) for understanding structural equation modeling," *Structural Equation Modeling: A Multidisciplinary Journal*, vol. 31, no. 6, pp.1104–1115, 2024.

[48] Z. Wang, J.-C. Liu, "Translating math formula images to latex sequences using deep neural networks with sequence-level training," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 24, no. 1, pp.63–75, 2021.

[49] M. Zhou, M. Cai, G. Li, M. Li, "An end-to-end formula recognition method integrated attention mechanism," *Mathematics*, vol. 11, no. 1, pp.177, 2022.

[50] Y. Deng, A. Kanervisto, J. Ling, A. M. Rush, "Image-to-markup generation with coarse-to-fine attention," in: *International Conference on Machine Learning*, PMLR, pp. 980–989, 2017.

[51] M. Hossam, T. Le, M. Papasimeon, V. Huynh, D. Phung, "Text generation with deep variational gan," *arXiv preprint arXiv:2104.13488*, 2021.

[52] M. Hashemi, "Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation," *Journal of Big Data*, vol. 6, no. 1, pp.1–13, 2019.

[53] Z. Zhao, A. Kleinhans, G. Sandhu, I. Patel, K. Unnikrishnan, "Capsule networks with max-min normalization," *arXiv preprint arXiv:1903.09662*, 2019.

[54] Dataset for trigonometric identity generation is avaliable at: https://drive.google.com/file/d/1tVq_Pu23uQpbyeIfVTf9yyACi6aOFYeP/view?usp=sharing.

[55] L. Yang, A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp.295–316, 2020.

[56] S. Dereich, A. Jentzen, "Convergence rates for the adam optimizer," *arXiv preprint arXiv:2407.21078*, 2024.

[57] C. Kiddon, L. Zettlemoyer, Y. Choi, "Globally coherent text generation with neural checklist models," in: *Proceedings of the 2016 conference on empirical methods in natural language processing*, pp.329–339, 2016.

[58] S. R. Bowman, L. Vilnis, O. Vinyals, A. M. Dai, R. Jozefowicz, S. Bengio, "Generating sentences from a continuous space," *arXiv preprint arXiv:1511.06349*, 2015.

[59] L. Yu, W. Zhang, J. Wang, Y. Yu, "Seqgan: Sequence generative adversarial nets with policy gradient," in: *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, 2017.

[60] K. Lin, D. Li, X. He, Z. Zhang, M.-T. Sun, "Adversarial ranking for language generation," *Advances in neural information processing systems*, vol. 30, 2017.

[61] H. Wang, Z. Qin, T. Wan, "Text generation based on generative adversarial nets with latent variables," in: *Advances in Knowledge Discovery and Data Mining: 22nd Pacific-Asia Conference*, PAKDD 2018, Melbourne, VIC, Australia, Proceedings, Part II 22, Springer, pp. 92–103, 2018.

[62] G. L. Guimaraes, B. Sanchez-Lengeling, C. Outeiral, P. L. C. Farias, A. Aspuru-Guzik, "Objective-reinforced generative adversarial networks (organ) for sequence generation models," *arXiv preprint arXiv:1705.10843*, 2017.

[63] H. Li, X. Yue, L. Meng, "Enhanced mechanisms of pooling and channel attention for deep learning feature maps," *PeerJ Computer Science*, vol. 8, pp.e1161, 2022.

[64] G. Dhara, R. K. Kumar, "Spatial attention guided cgan for improved salient object detection," *Frontiers in Computer Science*, vol. 6, pp.1420965, 2024.

[65] G. Li, J. Lv and C. Wang, "A Modified Generative Adversarial Network Using Spatial and Channel-Wise Attention for CS-MRI Reconstruction," in *IEEE Access*, vol. 9, pp. 83185-83198, 2021.

[66] S. Woo, J. Park, J.-Y. Lee, I. S. Kweon, "Cbam: Convolutional block attention module," in: *Proceedings of the European conference on computer vision (ECCV)*, pp. 3-19, 2018.

[67] E. Lu, X. Hu, "Image super-resolution via channel attention and spatial attention," *Applied Intelligence*, vol. 52, no. 2, pp.2260–2268, 2022.

[68] C.-Y. Lee, P. W. Gallagher, Z. Tu, "Generalizing pooling functions in convolutional neural networks: Mixed, gated, and tree," in: *Artificial intelligence and statistics*, PMLR, pp. 464–472, 2016.

[69] L. Zhao, Z. Zhang, "A improved pooling method for convolutional neural networks," *Scientific Reports*, vol. 14, no. 1, pp.1589, 2024.

[70] Q. Wang, J. Xie, W. Zuo, L. Zhang and P. Li, "Deep CNNs Meet Global Covariance Pooling: Better Representation and Generalization," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 8, pp. 2582-2597, 1 Aug. 2021.

[71] T. K. R. Arvind, M. Brand, C. Heidorn, S. Boppu, F. Hannig and J. Teich, "Hardware Implementation of Hyperbolic Tangent Activation Function for Floating Point Formats," *2020 24th International Symposium on VLSI Design and Test (VDAT)*, Bhubaneswar, India, 2020.

[72] C. Banerjee, T. Mukherjee, E. Pasiliao Jr, "An empirical study on generalizations of the relu activation function, " in: *Proceedings of the 2019 ACM Southeast Conference*, pp. 164–167, 2019.

[73] A. K. Dubey, V. Jain, "Comparative study of convolution neural network's relu and leaky-relu activation functions," in: *Applications of Computing, Automation and Wireless Systems in Electrical Engineering: Proceedings of MARC 2018*, Springer, pp. 873–880, 2019.

[74] D.-A. Clevert, T. Unterthiner, S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," *arXiv preprint arXiv:1511.07289*, 2015.

[75] G. Klambauer, T. Unterthiner, A. Mayr, S. Hochreiter, "Self-normalizing neural networks," *Advances in neural information processing systems*, vol. 30, no. 2017.

[76] S. Elfwing, E. Uchibe, K. Doya, "Sigmoid-weighted linear units for neural network function approximation in reinforcement learning," *Neural networks*, vol. 107, pp.3–11, 2018.

[77] D. Misra, "Mish: A self regularized non-monotonic activation function," *arXiv preprint arXiv:1908.08681*, 2019.

[78] H. Zhu, H. Zeng, J. Liu, X. Zhang, "Logish: A new nonlinear nonmonotonic activation function for convolutional neural network," *Neurocomputing*, vol. 458, pp.490–499, 2021.

[79] X. Wang, H. Ren, A. Wang, "Smish: A novel activation function for deep learning methods," *Electronics*, vol. 11, no. 4, pp.540, 2022.

[80] S. Mastromichalakis, "Alrelu: A different approach on leaky relu activation function to improve neural networks performance," *arXiv preprint arXiv:2012.07564*, 2020.

[81] U. Jain, Z. Zhang, A. G. Schwing, "Creativity: Generating diverse questions using variational autoencoders," in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6485–6494, 2017.

CHAPTER 7

[82] Y. Hoshi, "Hemodynamic signals in fNIRS, " *Progress in brain research*, vol. 225, pp.153-179, 2016.

[83] V. Quaresima, and M. Ferrari, "Functional near-infrared spectroscopy (fNIRS) for assessing cerebral cortex function during human behavior in natural/social situations: a concise review," *Organizational Research Methods*, vol. 22, no. 1, pp.46-68, 2019.

[84] S.M. Smith, "Overview of fMRI analysis, " *The British Journal of Radiology*, vol. 77(suppl_2), pp.S167-S175, 2004.

[85] E. Amaro Jr, and G.J. Barker, "Study design in fMRI: basic principles," *Brain and cognition*, vol. 60, no. 3, pp. 220-232, 2006.

[86] J. Goense, Y. Bohraus, and N.K. Logothetis, "fMRI at high spatial resolution: implications for BOLD-models," *Frontiers in computational neuroscience*, vol. 10, p.66, 2016.

[87] C.A. Olman and E. Yacoub, "High-field FMRI for human applications: an overview of spatial resolution and signal specificity," *The open neuroimaging journal*, vol. 5, p.74, 2011.

[88] Z. Liu, J. Shore, M. Wang, F. Yuan, A. Buss, and X. Zhao, "A systematic review on hybrid EEG/fNIRS in brain-computer interface," *Biomedical Signal Processing and Control*, vol. 68, p.102595, 2021.

[89] P. Padmanabhan, A.M. Nedumaran, S. Mishra, G. Pandarinathan, G. Archunan, and B. Gulyás, "The advents of hybrid imaging modalities: a new era in neuroimaging applications," *Advanced Biosystems*, vol. 1, no. 8, p.1700019, 2017.

[90] T. Duan, Z. Wang, S. Liu, Y. Yin, and S.N. Srihari, "UNCER: A framework for uncertainty estimation and reduction in neural decoding of EEG signals," *Neurocomputing*, vol. 538, p.126210, 2023.

[91] P. Liang, S. Wu, and F. Gu, *An introduction to neural information processing*, Berlin: Springer, 2016.

[92] K. Chowdhary, “Natural language processing,” *Fundamentals of artificial intelligence*, pp.603-649, 2020.

[93] S.C. Fanni, M. Febi, G. Aghakhanyan, and E. Neri, Natural language processing, In *Introduction to artificial intelligence*, pp. 87-99, Cham: Springer International Publishing, 2023.

Sayantani Ghosh

16/09/25

PROF. AMIT KONAR
CO-ORDINATOR
INTELLIGENT AUTOMATION
& ROBOTICS
E.T.C.E. DEPTT., J.U., KOL-32