Some Studies on Breast Cancer Detection Using Deep Learning Approaches

Thesis submitted by

Payel Pramanik

DOCTOR OF PHILOSOPHY (Engineering)

Department of Computer Science and Engineering
Faculty Council of Engineering & Technology
Jadavpur University
Kolkata-700032, India
2025

JADAVPUR UNIVERSITY
KOLKATA-700032, INDIA

Index No.: **128/22/E**

1. Title of The Thesis:
   Some Studies on Breast Cancer Detection Using Deep Learning Approaches

2. Name, Designation & Institution of The Supervisor:
   Prof. Ram Sarkar
   Professor,
   Department of Computer Science and Engineering,
   Jadavpur University, Kolkata-700032

3. List of Publications:

   (a) **Journal:**

   (i) Ayush Roy, **Payel Pramanik** and Ram Sarkar. "EU2-Net: A Parameter Efficient Ensemble Model With Attention-Aided Triple Feature Fusion for Tumor Segmentation in Breast Ultrasound Images", ***IEEE Transactions on Instrumentation and Measurement (2024) (IF: 5.6)***.

   (ii) **Payel Pramanik**, Ayush Roy, Erik Cuevas, Marco Perez-Cisneros and Ram Sarkar. "DAU-Net: Dual attention-aided U-Net for segmenting tumor in breast ultrasound images", ***Plos one (2024)***.

   (iii) **Payel Pramanik**, Rishav Pramanik, Friedhelm Schwenker, and Ram Sarkar. "DBU-Net: Dual branch U-Net for tumor segmentation in breast ultrasound images", ***Plos one (2023)***.

   (iv) Rishav Pramanik, **Payel Pramanik**, and Ram Sarkar. "Breast cancer detection in thermograms using a hybrid of GA and GWO based deep feature selection method", ***Expert Systems with Applications (2023) (IF: 7.5)***.

   (v) **Payel Pramanik**, Souradeep Mukhopadhyay, Seyedali Mirjalili, and Ram Sarkar. "Deep feature selection using local search embedded social ski-driver optimization algorithm for breast cancer detection in mammograms", ***Neural Computing and Applications (2023) (IF: 4.8)***.

(vi) Samridhha Majumdar, **Payel Pramanik**, and Ram Sarkar. "Gamma function based ensemble of CNN models for breast cancer detection in histopathology images", ***Expert Systems with Applications (2023) (IF: 7.5)***.

(vii) Arnab Bagchi, **Payel Pramanik**, and Ram Sarkar. "A multi-stage approach to breast cancer classification using histopathology images", ***Diagnostics (2022) (IF: 3.0)***.

(b) **Book Chapter:**

(i) **Payel Pramanik**, Rishav Pramanik, Anurup Naskar, Seyedali Mirjalili, and Ram Sarkar. "U-WOA: an unsupervised whale optimization algorithm based deep feature selection method for cancer detection in breast ultrasound images", Handbook of Whale Optimization Algorithm, pp. 179-191, 2024.

(c) **Conference:**

(i) Annada Dash, **Payel Pramanik** and Ram Sarkar. "Attention-Enhanced Residual Network for Breast Cancer Detection in histopathological Images". In ***International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC 2025). IEEE, pp. 697-703,*** https://doi.org/10.1109/ISACC65211.2025.10969381.

(ii) Ayush Roy, **Payel Pramanik**, Sohom Ghosal, Daria Valenkova, Dimitrii Kaplun, and Ram Sarkar. "GRU-Net: Gaussian Attention Aided Dense Skip Connection Based MultiResUNet for Breast Histopathology Image Segmentation". In ***Medical Image Understanding and Analysis (MIUA 2024). Lecture Notes in Computer Science, vol 14859. Springer, Cham.*** https://doi.org/10.1007/978-3-031-66955-2_21.

(iii) Ayush Roy, **Payel Pramanik**, Dmitrii Kaplun, Sergei Antonov and Ram Sarkar. "Awgunet: Attention-Aided Wavelet Guided U-Net for Nuclei Segmentation in Histopathology Images". ***IEEE International Symposium on Biomedical Imaging (ISBI 2024), pp. 1-4,*** https://doi.org/10.1109/ISBI56570.2024.10635449.

(iv) **Payel Pramanik**, Soradeep Mukhopadhyay, Dmitrii Kaplun, and Ram Sarkar, "A deep feature selection method for tumor classification in breast ultrasound images", ***In International conference on mathematics and its applications in new computer systems (MANCS 2021), vol 424. Springer, Cham.*** https://doi.org/10.1007/978-3-030-97020-8_22.

4. List of Patents: None

5. List of Presentations in National/International/Conference/Workshops:

   (i) **Payel Pramanik**, Souradeep Mukhopadhyay, Dmitrii Kaplun, and Ram Sarkar. "A deep feature selection method for tumor classification in breast ultrasound images". In ***International conference on mathematics and its applications in new computer systems (MANCS 2021). Lecture Notes in Networks and Systems, vol 424. Springer, Cham.*** https://doi.org/10.1007/978-3-030-97020-8_22

PROFORMA-1

STATEMENT OF ORIGINALITY

I, **Payel Pramanik** registered on **20/05/2022** do hereby declare that this thesis entitled **"Some Studies on Breast Cancer Detection Using Deep Learning Approaches"** 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 to all materials and results that are not original to this work.

I also declare that I have checked this thesis as per the "Policy on Anti Plagiarism, Jadavpur University, 2019", and the level of similarity as checked by iThenticate software is 8%.

Signature of Candidate:
Date :

Certified by Supervisor:
(Signature with date, seal)

PROFORMA-2

CERTIFICATE FROM THE SUPERVISOR

Index No.: **128/22/E**

This is to certify that the thesis entitled **"Some Studies on Breast Cancer Detection Using Deep Learning Approaches"** submitted by Smt. Payel Pramanik, who got her name registered on 20/05/2022 for the award of Ph.D. (Engg.) degree of Jadavpur University is absolutely based upon her own work under the supervision of Prof. Ram Sarkar, Professor, Department of Computer Science and Engineering, Jadavpur University, Kolkata, India, 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.

_____

Signature of the Supervisor
with date and Office Seal

Dedicated to my Family

Acknowledgement

The successful completion of this research would not have been possible without the guidance, support, and encouragement of several individuals. I take this opportunity to express my deepest gratitude to all those who have contributed to this work, both directly and indirectly.

First and foremost, I would like to express my earnest gratitude and deep respect to my supervisor **Prof. Ram Sarkar** for his relentless guidance, tremendous support, endless care, invaluable suggestion, and constant monitoring throughout this journey. His expertise and mentorship have been instrumental in shaping the direction of this research, and I am truly grateful for his patience and belief in my work.

I would also like to express my sincere respect and gratitude to Prof. Nirmalya Chowdhury, Head, Department of Computer Science and Engineering, Jadavpur University as well as to Prof. Mita Nasipuri, Former Coordinator of Centre for Microprocessor Application for Training, Education and Research (CMATER) Laboratory, Jadavpur University (JU), Prof. Mahantapas Kundu, Coordinator CMATER Laboratory, JU, Prof. Debotosh Bhattacharjee, Prof. Subhadip Basu, and Prof. Nandini Mukherjee from the same department for their continuous encouragement, invaluable support and insightful suggestions throughout the course of this thesis. I am also thankful to the faculty and staff members of the Department of Computer Science and Engineering, Jadavpur University, for their assistance and support during the course of my research work.

I am deeply grateful to Dr. Neelotpal Chakraborty and Dr. Samir Malakar for their support and encouragement in my research. I also extend my heartfelt thanks to my co-researchers Mr. Rishav Pramanik, Mr. Ayush Roy, Mr. Souradeep Mukhopadhyay, Mr. Samridhha Majumdar, Mr. Anurup Naskar, and Mr. Arnab Bagchi for their unwavering support during my research journey. Additionally, I sincerely appreciate the encouragement and solidarity of my fellow researchers, including Mr. Sujan Sarkar, Sk Mohiuddin, Mr. Asfak Ali and Ms. Nandita Gautam from the CMATER Laboratory, Jadavpur University, Kolkata, India.

I owe my encompassing debt to my respected father (Mr. Ganesh Pramanik) and mother (Mrs. Chandana Pramanik) for their unconditional love, encouragement and support throughout my academic journey. I sincerely thank my parents-in-law for their encouragement throughout the course of my research work. I am also thankful to my little brother (Mr. Pritam Pramanik) and my husband (Mr. Anirban Dutta) whose love, support, encouragement, good wishes and inspiration enabled me to complete this thesis. Their strength and kindness have been a source of inspiration throughout this journey.

Finally, I offer my sincere gratitude to the Almighty for the countless blessings, which have guided me in making this journey a success.

Date:
Place: Kolkata

(PAYEL PRAMANIK)
Index No. 128/22/E
Department of Computer Science & Engineering
Jadavpur University, Kolkata, India.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Breast Cancer and its Impact

Breast cancer is a major global health concern and one of the most common malignancies affecting women worldwide. According to the World Health Organization (WHO), cancer is the leading cause of non-accidental deaths globally, with breast cancer ranking as the third most fatal cancer after lung and liver cancers. In 2020, an estimated 2.3 million women were diagnosed with breast cancer globally, resulting in approximately 685,000 deaths [1]. The burden is disproportionately higher in low- and middle-income countries, where limited access to early detection and treatment contributes to elevated mortality rates. In the United States alone, 2022 saw around 287,850 cases of invasive breast cancer and 51,400 cases of non-invasive breast cancer, with 43,250 related deaths reported [2]. In India, the scenario is equally alarming. Data from the National Institute of Cancer Prevention and Research (NICPR) indicate that for every two women diagnosed with breast cancer, one succumbs to the disease [1]. As per GLOBOCAN 2020 data, breast cancer accounted for 13.5% (178,361) of all cancer cases and 10.6% (90,408) of all cancer-related deaths in India [3].

Breast cancer begins when cells in the breast grow uncontrollably due to mutations in the DNA. It typically originates in the milk-producing glands (lobules) or the ducts that carry milk to the nipple. The breast is composed of various connective tissues, blood vessels, lymph nodes, and lymph vessels. When cell division becomes unregulated, tumors may form. These tumors can be classified as benign (non-cancerous, do not invade nearby tissues) or malignant (cancerous, capable of spreading to other parts of the body

via the lymphatic system) [4].

The impact of breast cancer extends far beyond physical health, influencing emotional well-being, family dynamics, and socioeconomic status. For patients, a diagnosis often brings anxiety, fear, and uncertainty about the future. Families and caregivers also bear emotional and financial burdens, as treatment can be long, costly, and physically taxing. On a broader scale, breast cancer places a substantial burden on healthcare systems worldwide, particularly in low- and middle-income countries where access to early detection and advanced treatment is limited. Despite advances in diagnostic imaging, targeted therapies, and awareness campaigns, disparities in survival rates still exist due to factors such as late detection, lack of screening programs, and variations in healthcare access. As a result, breast cancer continues to be not only a medical challenge but also a social and economic one, underscoring the urgent need for improved detection methods, equitable healthcare, and continued research.

1.2 Breast Cancer Imaging Techniques

In breast cancer diagnosis, different imaging techniques play specific roles, each with its own significance. Techniques such as mammography, ultrasonography, histopathology, and thermography are commonly used, offering diverse approaches that enhance the understanding and diagnosis of breast cancer. This section begins by outlining these imaging modalities, followed by a discussion of the advantages and limitations of each technique.

1.2.1 Digital Mammography

Digital mammography is the most widely used and primary imaging modality for early-stage breast cancer screening, as recommended by the American Cancer Society [5, 6]. It is typically performed on asymptomatic women to detect early and clinically undetectable breast abnormalities, and remains the gold standard for breast imaging and diagnosis [7]. In this technique, X-ray imaging captures two standard views of each breast—cranio-caudal (top-down) and mediolateral oblique (angled side view)—by compressing the breast in a near-vertical plane. These images help examine masses and calcifications, which are key indicators of breast cancer. However, the contrast between

these features is often minimal, and their morphology can be highly irregular, making clinical diagnosis challenging, particularly in dense breast tissue. Digital detectors convert X-ray data into high-resolution mammographic images, as illustrated in Fig. 1.1.

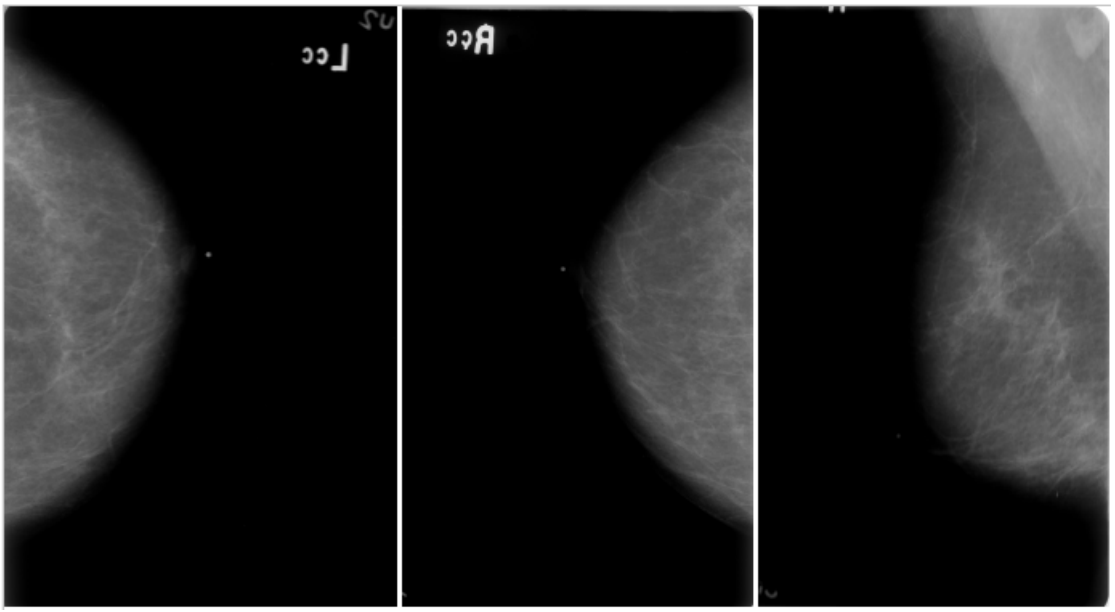


FIGURE 1.1: Sample breast mammography images.

1.2.2 Breast Ultrasound (or Sonography)

Ultrasound is the second most commonly used imaging modality for medical diagnosis after mammography [8]. Although it is operator-dependent, ultrasound has proven effective in the detection and differentiation of breast lesions, offering fast visualization and diagnosis of breast tissue. It is a noninvasive, safe, cost-effective, and widely accessible technique—particularly valuable for women under 35 years of age, as it is more sensitive than mammography for detecting abnormalities in dense breasts. In fact, the diagnostic accuracy for identifying simple cysts with ultrasound can reach 96–100% [9].
This technique works by transmitting high-frequency sound waves through the breast tissue, with returning echoes converted into real-time images on a monitor. A handheld transducer is used to scan the breast, allowing for visualization of the internal structures, including their shape and movement. Fig. 1.2 presents sample ultrasound images.

However, ultrasound images often suffer from reduced visual clarity due to noise and speckling artifacts. Malignant lesions frequently appear with indistinct margins, irregular shapes, and blurred edges, making it difficult to delineate them clearly from

FIGURE 1.2: Sample breast ultrasound images belonging to Benign, Malignant and Normal classes.

surrounding tissues. These vague boundaries increase diagnostic uncertainty, yet they also reflect the complex interactions between tumors and adjacent structures. Despite its limitations in image resolution and contrast compared to other modalities, ultrasound is particularly adept at distinguishing between cystic and solid masses—something mammography often struggles with. A key advantage of ultrasound (or sonography) is that it does not involve ionizing radiation, making it a safer choice, especially for pregnant women and those with dense breast tissue [10].

1.2.3 Breast Histopathology Imaging

Histopathology refers to biopsy-based imaging used for breast tissue analysis, wherein a tissue specimen is extracted from a suspicious area for detailed examination by medical experts [11]. Tissue samples are commonly obtained through procedures such as fine-needle aspiration (FNA), core needle biopsy, vacuum-assisted biopsy, and surgical biopsy (either excisional or incisional). These samples are then processed, stained with Hematoxylin and Eosin (H&E), and mounted on glass slides for microscopic analysis [12]. Fig. 1.3 displays sample histopathological images of breast cancer.

A pathologist examines the stained slides under a microscope, and these are subsequently digitized into high-resolution whole slide images (WSIs). From these WSIs, multiple region of interest (ROI) images can be extracted for further analysis. These ROI patches serve as valuable input for training deep learning models aimed at classifying invasive and non-invasive breast cancer [13].

Histopathological imaging plays a critical role in breast cancer diagnosis due to its ability to capture intricate phenotypic details. One of its key advantages is the potential to identify a variety of cancer subtypes—not just malignant versus benign—thanks to the

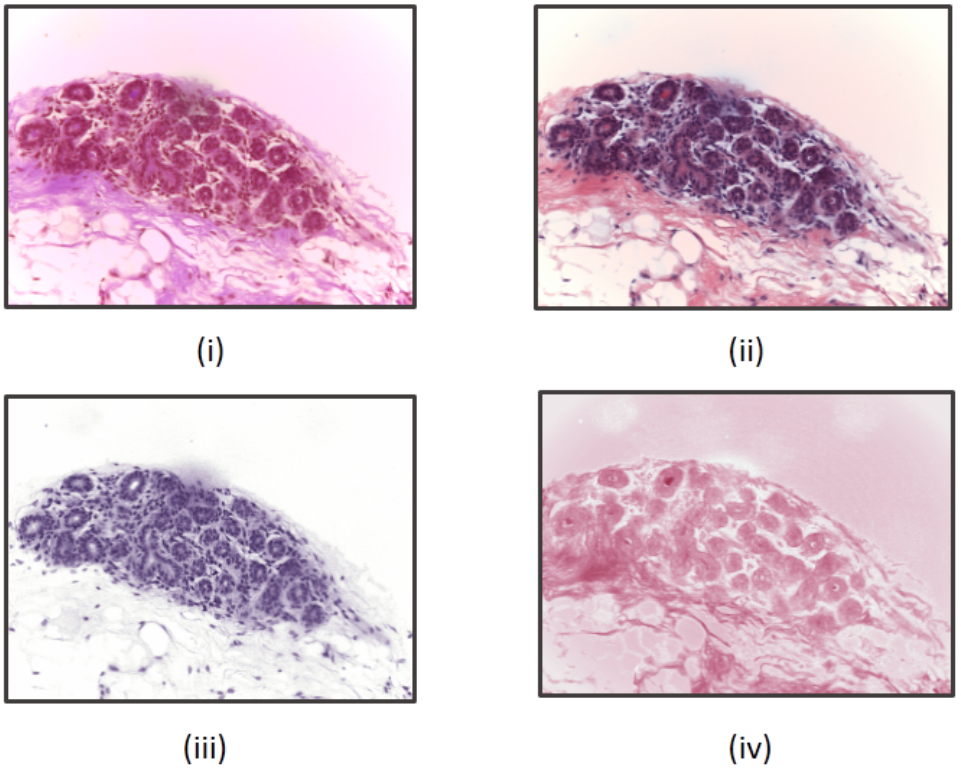


FIGURE 1.3: Sample breast histopathology images: (i) original image, (ii) stain normalized image (iii) hematoxylin stain, and (iv) eosin stain.

use of high-resolution, multi-colored images. Additionally, the generation of multiple ROI patches from a single slide reduces the likelihood of missing early-stage cancer. However, this method also presents notable challenges. Manual slide analysis is labor-intensive, time-consuming, and requires a high level of expertise. Variations in staining techniques and color distribution can introduce inconsistencies, potentially leading to misdiagnosis. Moreover, the extremely large size of gigapixel WSIs—often exceeding 1 GB per image—poses significant challenges for deep learning models, which struggle to process these high-resolution images in their entirety [14].

1.2.4 Breast Thermal Imaging

Breast thermography, also known as thermal imaging, is a non-invasive, radiation-free, and cost-effective technique for the detection and classification of breast cancer. This method leverages temperature variations on the breast surface as indicators of potential abnormalities, based on the fact that malignant cells typically exhibit elevated metabolic

activity and increased blood flow. These physiological changes result in localized heat emissions, which are captured using infrared cameras to produce thermal maps of the breast [15].

In thermographic images, different colors represent temperature intensities—cooler regions (typically healthy tissue) appear blue, while warmer regions indicating potential abnormalities are shown in red, orange, or yellow, depending on the camera's settings [16]. Fig. 1.4 displays sample thermographic images of the breast.

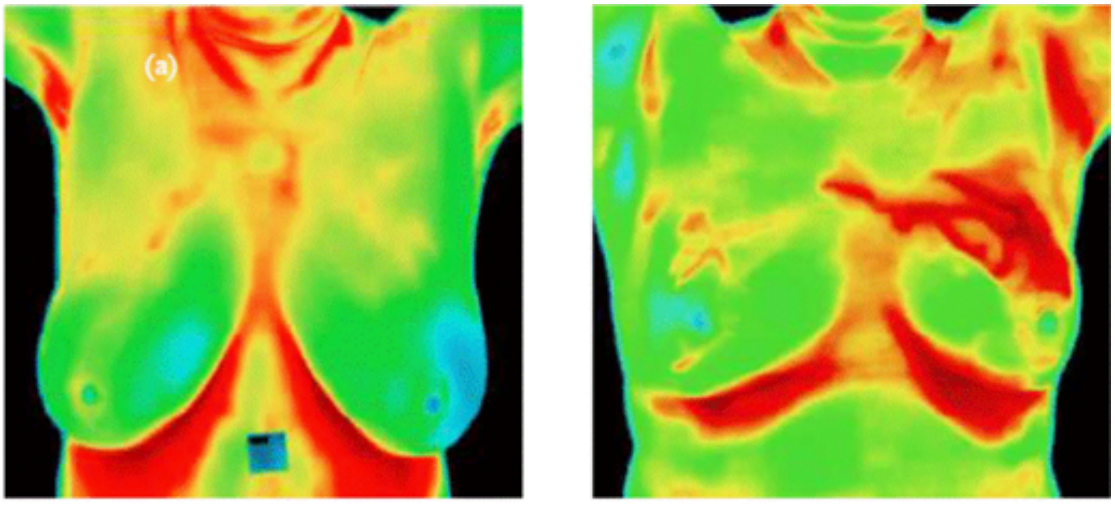


FIGURE 1.4: Sample breast thermal imaging.

One of the key advantages of thermography is its ability to potentially detect cancerous changes at an early stage. Studies suggest that thermographic methods can identify cancerous activity 8–10 years before it becomes visible through mammography [17]. When used as a standalone technique, thermography achieves a sensitivity of approximately 83%, which increases to 95% when combined with mammography [18].

Despite these benefits, thermography has its limitations. Its primary drawback lies in its inability to provide detailed information about deep tissue structures, as it only captures surface-level temperature changes. This restricts its utility as a sole diagnostic tool, though it remains a valuable adjunct in multi-modal breast cancer screening strategies.

1.3 Impact of Deep Learning in Breast Cancer Detection

1.3.1 Computer-Aided Diagnosis Systems

The rising global incidence of cancer underscores the urgent need for accurate and efficient diagnostic tools. In the context of breast cancer, early detection is vital for effective treatment planning and improved survival outcomes. Traditional diagnosis methods often rely heavily on the expertise and interpretation of radiologists, which can introduce variability and potential error. To address this, Computer-Aided Diagnosis (CAD) systems have emerged as a powerful second-opinion tool, enhancing diagnostic accuracy and reducing reliance on manual interpretation. CAD integrates Artificial Intelligence (AI), Computer Vision, and Medical Image Processing (MIP) to improve diagnostic sensitivity and specificity, particularly during the early stages of breast cancer when timely intervention is most critical.

CAD systems support healthcare professionals by automating the interpretation of medical images, such as mammograms, ultrasound scans, and histopathological slides. These systems help minimize false positives (FP) and false negatives (FN), ultimately reducing unnecessary biopsies and lowering treatment costs. A high FN rate is especially concerning, as it may result in delayed or missed treatment. By providing consistent, objective analyses, CAD systems contribute significantly to reducing diagnostic errors and enhancing early detection rates [19].

A typical CAD model operates through four main phases:

- Pre-processing: Medical imaging data often contains unwanted artifacts such as noise, blur, and haze. To enhance the quality and consistency of the data, various image pre-processing techniques are applied in this initial phase.
- Segmentation: In this phase, the relevant regions of interest are segmented from the pre-processed image. However, this step is not mandatory in every CAD model—some systems are designed to analyze the entire image without the need for segmentation.
- Feature Extraction and Selection: After segmentation (if performed), diverse features are extracted from the imaging data. Feature Selection (FS) methods are

then applied to identify the most relevant features, improving the model's performance and efficiency.

- Classification: In the final phase, the selected features are used to classify the imaging data into specific categories, as defined by the clinical or research objectives.

A schematic overview of the automatic cancer detection pipeline for breast cancer medical imaging is illustrated in Fig. 1.5.

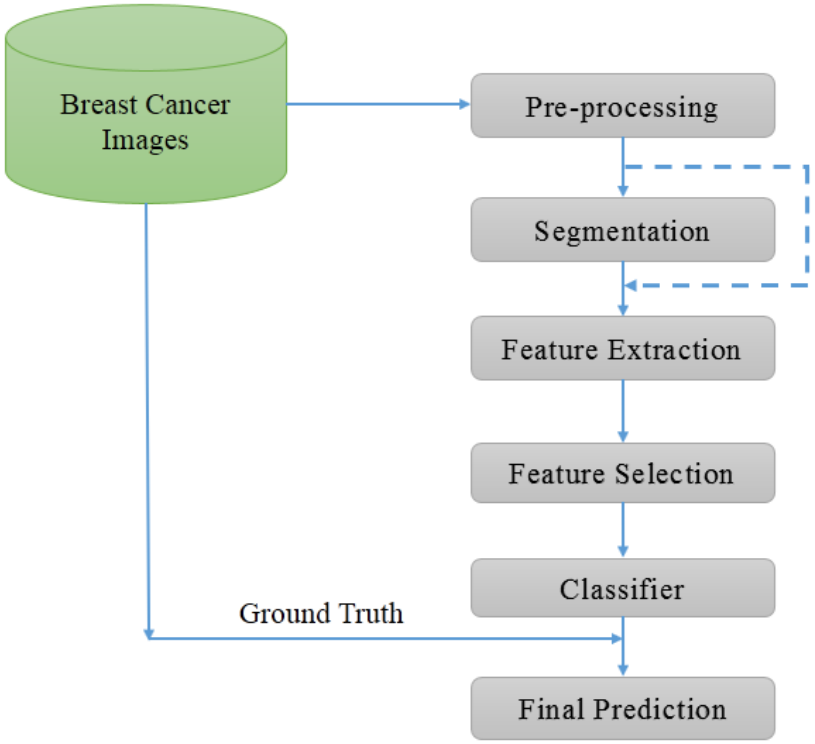


FIGURE 1.5: Flow chart showing basic architecture of CAD system for breast cancer detection from breast cancer imaging data. Dashed arrow indicates that segmentation can be bypassed.

1.3.2 Evolution of DL-based Systems

Within the AI paradigm, both Machine Learning (ML) and Deep Learning (DL) have demonstrated notable success in breast cancer diagnosis. ML algorithms enable predictive analysis by learning from structured, pre-processed data. However, ML depends heavily on manual feature extraction, where domain experts must identify and engineer relevant features for the model. This process is time-consuming and may fail to capture

subtle or complex patterns critical for identifying early malignancies. While ML can handle large datasets, its performance is limited by its reliance on experts input and predefined rules.

DL, particularly with the use of Convolutional Neural Networks (CNNs) [20], has transformed breast cancer diagnostics by addressing many of the challenges faced by traditional ML approaches. Unlike ML, DL models automatically learn hierarchical features directly from raw data, requiring minimal manual intervention. This ability makes DL especially effective for image-based tasks such as detecting minute abnormalities in mammograms, differentiating between benign and malignant tumors, or segmenting regions of interest in histopathological slides. DL networks excel at identifying complex, non-linear patterns and textural nuances that might be overlooked by conventional methods or human observers. With the rise of DL especially CNNs, significant advances have been made in medical image segmentation as well as classification.

One of the main contributing factors to this immense popularity of DL has to do with the good generalization traits DL has showcased in the past. DL leverages the use of derivatives or gradients as we commonly know it to optimize its internal weights. A typical DL method for image processing applications consists of two internal components: a) a Feature extractor and b) a fully connected neural network-based classifier. Generally, both of these components work in synchronization for the model to work as an end-to-end model which is fed with an image and the desired class is obtained as the output. Majorly in the biomedical field, the process of knowledge distillation is used which is known as Transfer Learning (TL) to alleviate the need for datasets of larger size. In this process, the model is typically trained on a standard large-scale dataset and the internal weights are then transferred to a relatively smaller-sized dataset to be fine-tuned on.

The U-Net architecture [21], with its encoder-decoder structure and skip connections, has emerged as a cornerstone DL-based model for biomedical image segmentation due to its ability to capture both global context and local spatial details. The U-Net framework has an encoder-decoder setup, wherein the encoder downsamples the image to extract features while the decoder utilizes these features (from the encoder) to upsample the output segmentation mask through the help of skip connections. This allows the network to obtain features of different granularity, leading to generate improved segmentation masks. The presence of skip connections facilitates the transfer of low-level to high-level

features from the encoder to the decoder, which ultimately leads to an enhanced comprehension of the contextual information. Overall, U-Net's effectiveness in concatenating contextual information through skip connections has made it a popular choice for breast cancer image segmentation [22].

Thus, DL has significantly advanced breast cancer diagnosis by improving accuracy, enhancing early detection, and supporting clinical decision-making. As the volume and complexity of medical imaging data continue to grow, DL-powered CAD systems are poised to play an increasingly central role in modern diagnostic workflows, ultimately contributing to better patient outcomes and more personalized healthcare delivery.

1.4 Problem Definition: Inherent Challenges

Malignant breast tumors pose a significant health risk and must be detected and treated at an early stage to improve patient outcomes. Early detection is crucial, as it enables timely clinical interventions, personalized treatment planning, and effective rehabilitation, thereby reducing breast cancer-related mortality. Breast cancer is commonly detected through physical examinations, imaging techniques such as mammography, thermography, ultrasound as well as histopathological biopsy. Among these, mammography is the most widely used and cost-effective screening tool. However, studies have shown considerable inter-reader variability among radiologists interpreting mammograms, even when reviewing the same image at different times [15, 23]. Moreover, mammograms are less effective for women with dense breast tissue, increasing the risk of false-negative results[1]. Although biopsy remains the gold standard for determining the malignancy of a lesion, it is invasive, painful, and often inconvenient, sometimes requiring multiple attempts. Additionally, access to biopsy services may be limited in remote or low-resource areas, resulting in diagnostic delays. Ultrasound imaging offers a promising alternative due to its affordability, lack of radiation, improved sensitivity in certain breast tissue types, and broader accessibility. Nevertheless, ultrasound suffers from several limitations, such as low contrast, poor spatial resolution, fuzzy lesion boundaries caused by speckle noise, acoustic shadowing, and indistinct tissue interfaces. These factors make ultrasound-based diagnosis subjective, time-consuming, and challenging for radiologists. Traditional diagnosis methods heavily depend on the expertise of radiologists, which

[1] cancer.org

introduces variability and increases the likelihood of human error.

To mitigate these issues, CAD systems—particularly DL-based models—have emerged as valuable second-opinion tools that enhance diagnostic accuracy and reduce dependency on manual interpretation. To construct a reliable DL model, a very large amount of data is required for training. This also leads to a huge training time for those models. A major limitation in developing robust DL models for breast cancer detection is the scarcity of high-quality, labeled, and publicly accessible medical datasets. DL models—particularly CNNs—require large volumes of annotated data for effective training. However, acquiring such datasets is often labor-intensive, time-consuming, and susceptible to human error during annotation. Additionally, ethical, legal, and privacy concerns further hinder the sharing and accessibility of medical imaging data, limiting both the diversity and scale of available datasets. This challenge is exacerbated by the fact that many datasets used in academic and clinical research are not publicly released. For this reason, the concept of TL is widely adopted where the CNN models are trained on larger datasets like ImageNet and further, the weights are transferred and fine-tuned on a smaller (i.e., target) dataset. However, more advanced TL-based approaches are required to improve classification accuracy across different imaging modalities.

Another major limitation in DL based models is their lack of interpretability and robustness, which are critical for clinical adoption. DL models, although powerful, often function as "black boxes", making it difficult to understand their decision-making process. To address these limitations, it is crucial to explore techniques such as attention mechanisms, optimization-driven feature selection, and contextual feature enhancement. These approaches aim to improve the model's focus on clinically meaningful regions, select the most informative features, and preserve structural details, thereby improving both diagnostic accuracy and model explainability.

Accurate segmentation of tumor regions in breast cancer images is a particularly complex and challenging task. While some diagnostic models operate on entire images without explicit segmentation, accurate identification and isolation of ROIs significantly enhance diagnostic precision. After the emergence of U-Net, various improvements have been proposed to address its limitations and enhance its capability to extract spatial and contextual information for generating accurate segmentation mask. These models focus mainly on modifying contextual feature extraction concepts, and passing information between the encoder and the decoder. This approach may cause mis-classification on the pixel of boundary areas, since they do not consider the simultaneous extraction of

spatial and contextual information. Also, there are very less methods that focus on enriching the encoded features through the use of edge detection. This requires to explore further improvements of U-Net based models for accurate segmentation of breast cancer images.

Histopathology and ultrasound images present additional challenges for segmentation and feature extraction. Histopathology images, for instance, suffer from high variability in color, intensity, and brightness, which makes it difficult to standardize preprocessing without losing essential features. Furthermore, segmentation of small lesions or nodules in ultrasound images remains challenging. Developing robust segmentation techniques for these modalities is crucial for improving the reliability of breast cancer detection systems.

The large computational and memory requirements of DL models pose a significant challenge, especially when large labeled datasets are scarce. There is a need for lightweight DL models that can perform well with smaller amounts of data while maintaining accuracy. Developing models that are both efficient and effective is critical to ensure that breast cancer detection systems can be deployed in real-world clinical environments, where computational resources may be limited.

While DL models have shown great promise in breast cancer detection, these inherent challenges need advanced detection techniques to improve the reliability, accuracy, and generalizability of these models. Overcoming these challenges will pave the way for the development of more effective AI-based diagnostic tools in breast cancer detection.

1.5 Objectives of the Thesis

The primary objective of this thesis is to design, develop, and evaluate automated computer-based models for accurate and interpretable breast cancer detection across multiple medical imaging modalities. The thesis focuses on the following specific objectives:

- To perform a comprehensive literature review of breast cancer detection techniques across various imaging modalities, critically analyze their limitations and challenges, and leverage these insights to the development of improved, modality-specific DL approaches for breast cancer detection.

- To investigate and implement advanced DL based architectures capable of extracting complex and clinically relevant features from diverse breast cancer imaging modalities, including mammography, ultrasound, histopathology, and thermography.
- To address the challenges of model interpretability and robustness by incorporating attention mechanisms and optimization-driven feature selection techniques.
- To develop lightweight and computationally efficient models that balance performance with feasibility for real-time or resource-constrained clinical settings.
- To perform a comprehensive evaluation of the proposed models for segmentation and classification tasks, demonstrating improvements in accuracy, interpretability, and overall performance.

1.6 Scope of the Thesis

This thesis aims to address the challenges associated with different breast cancer imaging analysis and proposes effective solutions to enhance diagnostic performance. The primary objective is to design and develop a breast cancer detection framework that utilizes different imaging modalities to: (a) accurately segment the tumor region, verifying the results against the provided ground truth data and (b) classify breast tumors as benign or malignant. To achieve these goals, a series of systematic investigations have been carried out, covering multiple aspects of breast cancer imaging analysis.

Initially the focus is to explore and analyze existing techniques for breast cancer detection across various imaging modalities. To achieve this, an extensive literature review has been conducted, covering methodologies specific to each modality. The review highlights a common trend across the field: a transition from purely handcrafted approaches to DL-based models, increasingly incorporating optimization techniques and attention mechanisms [24]. Recent advancements in DL, particularly CNNs, have significantly improved the performance of medical image segmentation tasks. The present study focuses on to develop two types of frameworks: Segmentation frameworks based on advanced U-Net variants for breast tumor segmentation and DL based FS frameworks for breast cancer image classification.

From a visual perspective, the edges for a particular image can be used for localizing

the ROI. For this, a dual branch U-Net [25] is proposed that fuses information from edges with original images of breast ultrasound. It involves two separate input paths for the encoding process. One of these branches incorporates the original image, and the other path uses the edge information by applying the different edge operators [26] on the original image. The encoded information learned by each of the branches fuses edge information with the encoding from original images to localize the context in order to produce a better segmentation mask.

To accurately segment the ROI from the breast ultrasound images a dual attention-enhanced U-Net model, DAU-Net is introduced. In this, two powerful attention mechanisms: the novel Positional Convolutional Block Attention Module (PCBAM) and Shifted Window Attention (SWA) [27] are integrated into a Residual U-Net model [28]. The PCBAM mechanism improves the encoder features before connecting them with the decoder features via residual connections. PCBAM extends the Convolutional Block Attention Module (CBAM) by integrating the Positional Attention Module (PAM), thereby enriching contextual information and improving the model's ability to capture spatial relationships within local features. Furthermore, the SWA module is incorporated into the bottleneck layer of the Residual U-Net to further boost segmentation performance. These attention mechanisms help capture both spatial and global contextual information, allowing the network to focus on tumor regions while suppressing irrelevant features from the surrounding tissue. This model demonstrates substantial improvements in accuracy, particularly in distinguishing boundaries of tumors that often appear indistinct due to the noisy and heterogeneous nature of ultrasound images.

The complexity of ultrasound images characterized by intricate patterns, closely resembling intensity distributions, varying tumor shapes, and unclear boundaries, presents significant obstacles in achieving accurate segmentation of breast lesions. These challenges motivates to explore for further enhancement of the U-Net framework. To this, $EU^2 - Net$ model is introduced which is a novel ensemble model that extends the capabilities of the $U^2 - Net$ architecture, a standard framework used for image segmentation [29]. $EU^2 - Net$ introduces a lightweight ensemble mechanism by incorporating separable convolutions to reduce the number of trainable parameters, making the model more computationally efficient and suitable for real-time applications. Unlike the conventional ensemble techniques that require multiple models and a parameter explosion, a

weighted averaging ensemble mechanism with learnable weights is introduced that seamlessly integrated within the $U^2 - Net$ architecture. To enhance model performance further, an attention-aided triple feature fusion technique is proposed. This enhances both encoder and decoder features using feature similarity-based attention and Squeeze-and-Excitation (SE) channel attention. Additionally, edge features extracted by the Sobel filter from the encoder features is incorporated, enriching the decoder with boundary-aware information. These combined multi-domain features provide the decoder with a comprehensive set of generic and detailed features, thereby improving segmentation accuracy.

Accurate nuclei segmentation in histopathological images is crucial for cancer diagnosis. Automating this process offers valuable support to clinical experts, as manual annotation is time-consuming and prone to human errors. However, automating nuclei segmentation presents challenges due to uncertain cell boundaries, intricate staining, and diverse structures. To address this challenge, a segmentation approach is introduced that combines the U-Net architecture with a DenseNet-121 backbone, harnessing the strengths of both to capture comprehensive contextual and spatial information. In this, the classical U-Net is extended by incorporating a DenseNet-121 encoder backbone to strengthen feature propagation and reuse. To improve boundary localization, a Wavelet-Guided Channel Attention Module (WGCAM) is proposed that captures boundary-aware spatial features. Furthermore, a learnable weighted global attention pooling (lw-GAP) is employed to model channel-specific dependencies, enabling the network to focus on more relevant feature maps. The decoder, comprising structured upsampling and convolutional blocks, refines the segmentation output by handling small and complex regions influenced by staining artifacts. The method is evaluated on two benchmark datasets—MoNuSeg and TNBC—demonstrating superior performance in accurately segmenting nuclei in breast histopathology images.

At the early stage of this thesis while exploring the breast cancer classification approaches it is seen that the DL based frameworks in particular CNNs has considerably boosted the performance of correctly classifying the breast cancer images. However, researchers overlook one principal concept of eliminating redundant features which ultimately hamper the performance of the overall framework. For this very reason, strategies like feature selection are getting more popular these days. The objective of these methods is to reduce the presence of redundant and noninformative features.

For the above said facts various deep feature selection strategies are explored. In an

initial experiment, a deep feature selection method [30] is proposed for tumor classification in breast ultrasound images. The method begins by extracting features using fine-tuned TL-based models. Then these extracted features are ranked based on the correlation coefficient values that is obtained after applying a standard statistical method and select the top ranked features which are most important and relevant, and classify the unseen breast ultrasound tumors using ML-based classifiers. While further exploring the FS strategies it is seen that meta-heuristic algorithms [31] perform well for FS because they are designed to efficiently explore large, complex, and non-linear search spaces—which is exactly what FS entails. So, another method called U-WOA [32] is proposed for tumor classification in breast ultrasound images that introduced a novel unsupervised optimization technique. By incorporating rank-based statistical measures into a meta-heuristics algorithm called the Whale Optimization Algorithm (WOA), the framework identifies the most discriminative features without requiring class labels. In this, the leader agent is selected in U-WOA using a pseudo-fitness value which is calculated using three rank-based feature filter methods. The overall framework aims to reduce the presence of redundant and noninformative features extracted from the deep learner to build a more robust framework.

Meta-heuristics are designed to explore (search broadly to avoid local optima) and exploit (refine promising solutions). This balance helps find near-optimal feature subsets without getting stuck in sub-optimal solutions. This fact is realized while exploring classification approaches in thermograms. A lightweight model is introduced to classify breast thermograms using a combination of TL-based model and FS approaches [33]. In this, a TL model (pre-trained on the ImageNet dataset) namely SqueezeNet 1.1 is employed for the feature extraction purpose, which is very light in terms of the number of parameters, thus leading to low processing time, storage, and computational efforts. Then these extracted features are fed into the proposed hybrid FS model, which selects distinctive features and discards the redundant features. This hybrid FS model used a chaotic Genetic Algorithm (GA) initialized using a sinusoidal chaotic map integrated with Grey Wolf Optimizer (GWO) to reduce computational time and to select an optimal subset of features using the exploration ability of GA and the exploitation ability of GWO.

However, choosing the optimal feature subset in FS is difficult in wrapper-based techniques where the chosen subsets must be assessed using a learning algorithm at each iteration. Ensemble of local search improves a method's exploitation ability and hence

improves the overall learning model's performance. This fact is realized while exploring classification approaches in mammograms. One of the common approaches to identifying breast cancer is through breast mammograms. However, the identification of malignant breasts from mass lesions is a challenging research problem. To address this issue a method is introduced for the classification of breast mass using mammograms [34]. In this, an attention-enhanced TL model called VGG16 is used for feature extraction. Then to select an optimal subset of features, a local search embedded with the Social Ski-Driver (SSD) algorithm is introduced. The attention mechanism exploits the most important regions of an image by paying more attention to the same. Furthermore, FS approaches reduce the number of features, whereas local search helps to increase the exploitation capability of the FS method and produces the most optimal subset of features.

Further, it has been explored that ensemble learning based methods along with DL approaches in various domains are achieving significantly improved classification performance. Ensemble learning in the decision-making stage is a powerful strategy in which confidence scores of multiple base learners are fused to obtain the final prediction about an input sample. However, two classifiers may contain redundant information and if they work in a group the results of that group are not significant. To resolve this limitation, a rank based ensemble technique [35] is proposed to solve a 2-class classification problem of breast histopathological images using the Gamma function. This rank-based fusion works better than standard ensemble approaches because it uses adaptive priority depending on the confidence scores of the classifiers for each sample to be predicted. It integrates the predictions from three pre-trained CNN models—GoogleNet, VGG11, and MobileNetV3_Small—using a rank-based ensemble method.

1.7 Organization of the Thesis

This thesis provides a comprehensive study of automatic breast cancer detection techniques across various imaging modalities, with a strong emphasis on DL-based approaches. It investigates prior research while presenting the proposed solutions developed in this study. Following this introductory chapter (Chapter 1: *Introduction*), the remaining chapters are organized as follows:

(i) **Chapter 2** presents a comprehensive review of related work in breast cancer classification and segmentation. It discusses existing techniques across multiple imaging modalities and identifies key research gaps addressed in this thesis.

(ii) **Chapter 3** focuses on the proposed models for segmenting tumor regions in breast cancer images. It explores various attention mechanisms, boundary-aware enhancements, and lightweight architectures. For each method, the chapter details the dataset used, model architecture, training strategy, and performance evaluation.

(iii) **Chapter 4** describes the proposed DL and optimization-based FS methods for classifying breast cancer using images from different modalities, including ultrasound, histopathology, mammography, and thermography. It discusses various FS strategies and evaluates their effectiveness across the proposed models. Additionally, the chapter investigates the performance benefits of ensemble approaches compared to individual models.

(iv) **Chapter 5** summarizes the key findings of the thesis and discusses the strengths and limitations of the proposed methods. It emphasizes the role of attention mechanisms, optimization-based FS techniques, and ensemble learning in enhancing breast cancer detection accuracy. The chapter also outlines potential future research directions, including data-efficient learning, temporal modeling, and multimodal approaches for breast cancer detection.

Chapter 2

Literature Review

This chapter comprehensively reviews literature on breast cancer segmentation and classification approaches across various imaging modalities, as introduced in Chapter 1, Section 1.2. It examines publicly available datasets corresponding to these imaging techniques and analyzes a range of methodologies, including both ML and DL approaches, highlighting their respective limitations. Finally, the chapter identifies the existing gaps and challenges in the field and outlines potential directions for required research.

2.1 Cancer Detection using Breast Mammograms

Digital mammography is the most common and primary imaging method for early breast cancer screening, especially in asymptomatic women, and is endorsed by the American Cancer Society [5]. It uses X-rays to capture two standard views of each breast—craniocaudal and mediolateral oblique—by compressing the breast vertically. This technique is effective in detecting masses and calcifications, though dense breast tissue can make interpretation challenging due to low contrast and irregular morphology.

Publicly Available Datasets: The Digital Database for Screening Mammography (DDSM) [36] is the largest publicly available dataset for mammographic image analysis, developed by Massachusetts General Hospital, Sandia National Labs, and the University of South Florida. It includes 2,620 studies featuring normal, benign, and malignant cases, each with verified pathology. Each study provides two images per breast, along with patient data such as age, ACR breast density, and abnormality subtlety ratings. The

Curated Breast Imaging Subset of DDSM (CBIS-DDSM) [37] is an enhanced and more accessible version of the original DDSM, designed to support advanced mammography research and CAD system development. It includes 2,620 scanned mammograms comprising 753 cases with calcifications and 891 with masses. The dataset features updated, accurately re-annotated ROIs by trained mammographers, improved segmentation using a local level set method, and removal of 339 low-quality annotations. Images are provided in modern 16-bit grayscale TIFF and DICOM formats, along with extensive metadata (e.g., BI-RADS scores, pathology, breast density, view type, mass/calcification details). Data is split into training and testing sets for standardized evaluation.
Another public dataset is The Mammographic Image Analysis Society (MIAS) [38] database, a widely used benchmark dataset for breast cancer research and CAD system development. It contains 322 mammogram images from 161 patients (one image per breast), categorized as normal, benign, or malignant. The 8-bit grayscale images, standardized to 1024×1024 pixels at 50-micron resolution, include various abnormalities such as masses, microcalcifications, architectural distortions, and asymmetries. Each image is annotated with the type, location, and size of abnormalities, along with metadata on patient age, breast density, and lesion characteristics, making it a valuable resource for algorithm evaluation and comparison.

Classification Approaches: In recent years, DL approaches, particularly CNNs and TL models, have significantly enhanced the accuracy and reliability of mammographic image classification. For instance, Shen et al.[39] proposed a CNN-based method for classifying mammograms from the CBIS-DDSM dataset, reporting an AUC of 0.88, sensitivity of 86.1%, and specificity of 80.1%. Khamparia et al.[40] used VGG16 with fine-tuning and regularization, reaching an accuracy of 88.3%. L. G. Falconi et al. [41] evaluated several TL models (e.g., NasNet, MobileNet, ResNet, and VGG16) for classifying malignancy in breast abnormalities. They found that fine-tuned VGG16 achieved the best accuracy (84.4%) on the CBIS-DDSM dataset. Another approach by Al-antari et al. [42] proposed a model leveraging You Only Look Once (YOLO) for lesion detection and multiple CNN architectures (e.g., ResNet-50, InceptionResNet-V2) for classification. Their framework showed promising results across the DDSM and INBreast datasets, indicating the generalizability of YOLO-based detectors in diverse mammographic environments.
From these studies, it is evident that deep learning techniques—especially those utilizing

transfer learning and model fine-tuning—have shown significant promise in mammography image classification. Despite the challenges of heterogeneous datasets and image quality variations, these approaches have enhanced diagnostic precision and can reduce both false positives and false negatives in clinical breast cancer screening. However, the effectiveness of these models could be further improved by integrating attention mechanisms, which help to localize and focus on the most informative regions within mammograms.

Beyond DL, a number of studies have investigated the integration of meta-heuristic FS strategies with mammographic image analysis. For example, Oyelade et al. [43] implemented a nature-inspired meta-heuristic optimization strategy to fine-tune the weights and biases of a CNN model designed for breast cancer image classification, including mammograms. The optimization process involved multiple algorithms such as GA, WOA, Multiverse Optimizer (MVO), Satin Bowerbird Optimization (SBO), and Life Choice-Based Optimization (LCBO). This comprehensive optimization scheme aimed to improve the CNN's performance across various imaging types, including mammograms. However, despite its broad applicability, the model faced limitations in accuracy, likely due to the difficulty in capturing the subtle variations in mammographic features with evolutionary tuning alone. Furthermore, Rezaee et al. [44] introduced a hybrid descriptor-based approach tailored to identify multi-mass breast cancer, which is especially relevant to mammography where masses of varying sizes and textures frequently appear. They employed a memetic meta-heuristic learning strategy to enhance the FS process. Although the method offered enhanced adaptability to complex breast lesion patterns, its effectiveness was undermined by high data processing times and relatively lower precision in dense breast scenarios, a common issue in mammographic interpretation.

Limitations of Existing Techniques: In summary, while meta-heuristic FS methods have shown increasing promise in mammogram-based cancer detection, existing approaches often face limitations such as computational inefficiency, limited accuracy gains, and poor generalization. There is a requirement of developing more adaptive and scalable hybrid frameworks that combine DL with swarm intelligence or evolutionary algorithms. Such models could better accommodate the inherent variability of mammographic images, ultimately enhancing the precision and robustness of automated breast cancer screening systems.

2.2 Cancer Detection in Breast Ultrasound Images

Ultrasound is the second most common imaging modality for breast cancer diagnosis after mammography. It is noninvasive, safe, cost-effective, and particularly useful for younger women and those with dense breast tissue, achieving up to 96–100% accuracy in detecting simple cysts. Ultrasound utilizes high-frequency sound waves to create real-time images, distinguishing cysts from solid masses without ionizing radiation. However, it is operator-dependent and often produces lower-resolution images with blurred details and indistinct margins, making accurate diagnosis challenging. While it reduces unnecessary biopsies and is widely accessible, it is prone to false positives and negatives.

Publicly Available Datasets: Publicly accessible breast ultrasound datasets for general researchers are limited, with many being private. The Breast Ultrasound Imaging Dataset or popularly known as BUSI dataset was generated in 2018 and made publicly accessible by Dhabyani et al. [45]. This dataset includes breast ultrasound images from 600 female patients aged 25 to 75 years. It consists of 780 images, each with an average size of 500 $\times$ 500 pixels in PNG format and its own ground truth (mask image). The images are categorized into three classes including 437 benign, 210 malignant and 133 normal images. Another popular publicly accessible dataset is Dataset B, also known as UDIAT [46]. The UDIAT dataset was shared by the UDIAT Diagnostic Centre in Sabadell, Spain, for research purposes. It comprises 109 benign and 58 malignant breast ultrasound images. Experienced radiologists carefully annotated each image, creating binary masks as ground truth for regions of interest. Both the ultrasound images and their corresponding masks have an average resolution of 760 $\times$ 570 pixels, providing high-quality data for breast cancer detection and classification studies.

Segmentation Approaches: Accurate segmentation of breast tumors in ultrasound imaging is crucial for early diagnosis and clinical decision-making. Traditional image processing techniques often fall short in handling the challenges posed by ultrasound images, such as low contrast, speckle noise, and fuzzy lesion boundaries. With the rise of DL, especially CNNs, significant advances have been made in medical image segmentation, particularly using architectures such as Fully Convolutional Networks (FCNs) [47] and U-Net [21]. These architectures serve as the backbone for many SOTA models used in medical imaging tasks [48–52].

Inspired by the success of U-Net, various improvements have been proposed by the

researchers to address its limitations and enhance its ability to extract spatial and contextual information. Variants like Residual U-Net [28], Attention U-Net [53], U-Net++[54], and RDAU-Net[55] have demonstrated significant promise for breast lesion segmentation in ultrasound images. In particular, attention mechanisms have emerged as powerful tools to improve focus on relevant features and suppress irrelevant regions, thereby enhancing segmentation accuracy. Channel and spatial attention modules, such as those used in CBAM [56], have shown effectiveness in modeling interdependencies between feature maps.

Studies like Vakanski et al. [57] and Lee et al. [58] demonstrated how visual saliency and grid-based channel attention can guide networks to focus on tumor-relevant areas in ultrasound images. Others, such as Zhang et al. [59] and Han et al. [60], integrated CBAM and adversarial learning to develop models capable of joint lesion detection and segmentation, even with limited annotated data.

Transformer-based models have also gained traction. Swin Transformer [27] introduces a sliding window-based attention mechanism, which was later incorporated into hybrid CNN-transformer frameworks such as BTS-ST [61]. These models further improve segmentation in challenging imaging modalities like breast ultrasound by capturing long-range dependencies.

Classification Approaches: Over the years, a range of methods have been proposed leveraging different combinations of feature extraction strategies and classification techniques for breast cancer classification in ultrasound images. Early approaches relied heavily on handcrafted features. For instance, Shi et al. [62] extracted histogram, texture, and fractal features from the BUSI dataset and used an Support Vector Machine (SVM) classifier. However, their method did not account for key diagnostic indicators such as tumor size, limiting its clinical applicability. Yang et al. [63] focused on Gray Level Co-Occurrence Matrix (GLCM)-based texture features derived through multi-resolution ranklet transformations. Although the method provided robustness against scale and orientation variations, it suffered from high computational cost and achieved only modest accuracy. Similarly, Cai et al. [64] integrated phase congruency and Local Binary Pattern (LBP)-based features with SVM for classification. While the approach captured both structural and textural cues, it lacked consideration for region-specific tumor features, which are critical for effective diagnosis.

With the advent of DL, researchers began exploring CNN-based methods. Xiao et al. [65]

implemented TL using pretrained models like InceptionV3, Xception, and ResNet50 on a large ultrasound dataset, achieving an accuracy of 89.44%. While this marked an improvement over classical methods, the performance left room for further refinement, particularly in handling inter-class variability. Hybrid techniques have also gained popularity. Sadad et al. [66] proposed a CAD system that first segmented the lesion region using marker-controlled watershed transformation and then extracted shape and GLCM-based texture features. These were used to train an ensemble classifier, which achieved a commendable accuracy of 96.6%. However, the method did not address class imbalance, which can bias results toward the dominant class. Mishra et al. [67] introduced a more refined handcrafted approach by combining Histogram of Oriented Gradients(HOG), GLCM, and additional texture features. Feature selection was performed using recursive feature elimination (RFE), and data imbalance was mitigated using Synthetic Minority Oversampling Technique (SMOTE). While this enhanced classification accuracy, the method was computationally intensive and relied on tuning the number of selected features, affecting generalizability. A more comprehensive ensemble approach was proposed by Moon et al. [68], who fused multiple CNN architectures with handcrafted descriptors. They selected the best-performing models for each input modality and combined their outputs using a weighted average strategy. This framework improved classification accuracy but remained semi-automated, requiring manual steps in tumor localization and segmentation. To overcome the limitations of static feature extraction, Jabeen et al. [69] integrated DL with meta-heuristic optimization. Features extracted using DarkNet-53 were refined through reformed Binary Grey Wolf Optimization (BGWO) and Differential Evaluation algorithms. This hybrid model significantly enhanced classification accuracy and showed better generalization across BUSI categories. Further advancing this direction, Mishra et al. [70] proposed a fusion technique that combined handcrafted and CNN-extracted features. After normalization and dimensionality reduction, the final feature set was optimized using the WOA. Their method outperformed several SOTA approaches, achieving accuracy beyond 95% and demonstrating the effectiveness of hybrid and optimized feature spaces.

In summary, the trajectory of research in ultrasound image-based breast cancer detection reflects a clear shift from purely handcrafted techniques to hybrid and deep learning-driven approaches. The integration of intelligent feature selection, ensemble

learning, and optimization algorithms—especially when tested on standardized datasets like BUSI—has significantly advanced the accuracy, robustness, and applicability of these models in real-world diagnostic settings.

Limitations of Existing Techniques: Despite these advancements, many models still face limitations such as high computational costs, redundant parameters, and insufficient attention to boundary regions. To overcome these issues, there is a need for end-to-end models that not only improves segmentation and classification accuracy across varied tumor shapes and textures but also ensures computational efficiency suitable for real-time clinical applications.

2.3 Cancer Detection using Breast Thermograms

Breast thermography is a non-invasive, radiation-free, and cost-effective technique for detecting and classifying breast cancer by capturing heat patterns using infrared cameras. Malignant cells generate more heat due to higher metabolic activity and blood flow, making temperature variations on the breast surface potential indicators of tumors. Thermography uses color maps—blue for healthy tissue and warmer colors for abnormalities.

Publicly Available Datasets: The Database for Mastology Research with Infrared Thermal Imaging or popularly known as DMR-IR database[1] provided by Federal Fluminense University, is a widely used open-access dataset for breast cancer research. It includes infrared (IR) images of 640×480 resolution and detailed patient information such as medical history and symptoms. Images were captured using static and dynamic protocols with a FLIR-SC 620 camera. The dataset supports comprehensive diagnostic algorithm development, with some achieving up to 100% accuracy.
Another recent publicly available thermographic dataset is The DBT-TU-JU breast thermogram database [71], developed by the Department of Biotechnology, Tripura University, and Jadavpur University. It provides annotated thermograms with ground-truth images of suspicious regions. It contains 1100 thermograms from 100 subjects, including original and zoomed-in views across six angles (frontal, lateral, oblique, and supine). The dataset is enriched with supporting clinical data, including mammography results,

[1] http://visual.ic.uff.br/dmi/

patient history, and diagnostic reports, to validate infrared breast thermography findings.

Classification Approaches: Over the past decade, researchers have explored various techniques to leverage texture and thermal information embedded in thermograms. Initial efforts such as those by Roslidar et al. [72] focused on leveraging TL with CNN architectures like ResNet-101, DenseNet-201, MobileNet V2, and ShuffleNet V2. Remarkably, these models achieved near-perfect classification accuracy, with MobileNet V2 offering an efficient trade-off between training time and resource usage, making it suitable for real-time applications. However, the models were only partially automated and relied on TL rather than task-specific customization. To enhance feature extraction, Pramanik et al. [73] proposed a novel local image descriptor named the Local Instant-and-Centre-Symmetric Neighbor-based Pattern, designed to capture high-level texture information from extrema-images of thermal scans. This handcrafted feature design improved representation capability but remained computationally demanding. Some studies have applied swarm intelligence for classification of thermograms. Gonçalves et al. [74] applied Particle Swarm Optimization to optimize hyperparameters of a VGG-16-based CNN model, showing performance improvements in infrared images. CNN-based feature learning has also been explored for classification using both raw images and auxiliary data. Sánchez-Morillo et al. [75] used a multi-view approach that incorporated frontal and lateral thermal views along with clinical and personal metadata as input to a CNN. This strategy boosted classification accuracy by integrating contextual information with imaging features. Furthermore, studies like [76] examined the influence of data augmentation techniques and CNN hyperparameter tuning on classification accuracy across different thermal datasets. The study provided important insights into dataset partitioning strategies and their impact on model generalizability.

Limitations of Existing Techniques: Overall, the literature demonstrates increasing interest in developing thermography-based CAD systems using hybrid methods. These include combining handcrafted and learned features, metaheuristic optimization for feature selection or hyperparameter tuning, and context-aware models incorporating both image and non-image data. Despite the progress, challenges remain in reducing computational cost and achieving fully automated, lightweight models suitable for hardware-constrained environments.

2.4 Cancer Detection in Histopathology Images

Histopathology, or biopsy imaging, is a tissue-based diagnostic method for breast cancer screening. It involves collecting tissue samples from suspicious areas using techniques like fine-needle aspiration, core needle biopsy, or surgical biopsy. Samples are stained with Hematoxylin and Eosin (H&E), fixed on slides, and examined under a microscope. These are then digitized into whole slide images (WSIs), enabling extraction of ROIs for training models to classify invasive and non-invasive cancers.

Publicly Available Datasets: The BreakHis dataset, introduced by Spanhol et al. [77], contains 7,909 histopathological images—2,480 benign and 5,429 malignant—collected from 82 patients by P&D Laboratory in Brazil. Images were captured at 40x, 100x, 200x, and 400x magnification using an Olympus BX-50 microscope and a Samsung digital color camera, stored as 700×460 pixel PNG files in 24-bit RGB. Prepared using standard histopathology protocols (formalin fixation, paraffin embedding, H&E staining), the slides support detailed morphological analysis. Pathologists identify regions of interest at 40x and examine them further at higher magnifications. The dataset includes four benign tumor types (adenosis, fibroadenoma, phyllodes tumor, tubular adenoma) and four malignant types (ductal, lobular, mucinous, and papillary carcinoma), making it a valuable resource for developing and evaluating breast cancer classification models. The ICIAR Breast Cancer (BACH) dataset [78], developed for the BACH Grand Challenge at ICIAR 2018, supports automated classification and segmentation of H&E-stained histology images. It includes two components: (i) 500 annotated microscopy images (400 for training, 100 for testing) equally distributed across four classes—normal, benign, in situ carcinoma, and invasive carcinoma—collected from hospitals in Portugal and annotated by medical experts; and (ii) pixel-level labeled WSIs for segmentation tasks. Images were captured using a Leica DM2000 LED microscope and ICC50 HD camera.

Another publicly accessible dataset is the TNBC (Triple-Negative Breast Cancer) dataset [79], which supports research in detecting, grading, and classifying TNBC. It contains 50 high-resolution (512×512) histology images with 4,022 meticulously annotated cells, averaging 80 cells per image. Annotations cover various cell types, including normal epithelial, myoepithelial, invasive cancer cells, fibroblasts, endothelial cells, adipocytes, and immune cells (macrophages, lymphocytes, plasmocytes). Annotations were performed by

a pathologist and two trained fellows, with each image reviewed and finalized through expert consensus, ensuring high accuracy. This dataset is a valuable resource for developing advanced breast cancer analysis algorithms.

Segmentation Approaches: Over the years, a wide range of DL-based segmentation techniques have been proposed to address the complex challenges of histopathological image analysis. These methods aim to handle the irregular shapes, varying sizes, low contrast, and intricate structural details characteristic of ROIs in clinical images. To address the diversity in ROI appearance and size, Kanadath et al. [80] introduced AIR-UNet++, an atrous inception-ResNet-based U-Net architecture augmented with dense skip connections. This model yields segmentation results through enhanced feature reuse and multiscale context extraction. Yang et al. [81] presented ADS-UNet, a stage-wise additive training framework that incorporates deep supervision in shallower layers. It utilizes performance-weighted ensemble learning by combining multiple sub-UNets. Kiran et al. [82] proposed DenseRes-UNet, a modified U-Net framework that integrates dense blocks at the encoder's final stages and introduces residual connections with Atrous convolutions. These modifications help reduce the semantic gap between encoder and decoder paths. Additional post-processing techniques, including distance maps and binary thresholding, are applied to enhance the delineation of nuclei interiors. Singha et al.[83] proposed AlexSegNet, an encoder-decoder model inspired by AlexNet. It combines low- and high-level features via skip connections and channel-wise feature map fusion for effective nucleus segmentation.

Classification Approaches: In this section, studies that have utilized ensemble techniques and deep feature extraction for breast cancer classification in histopathology images are explored. Ensemble learning has become a prominent technique in the classification of breast histopathology images, as it enhances the robustness and accuracy of DL models. It enhances the prediction capability of the overall model which is beyond the research of the individual base models. This approach also makes the model more robust. Ensemble learning helps to resolve the issues when a model makes a biased decision for a particular test sample. This approach is explored in respect with the histopathology images where the dataset contains high imbalance between benign and malignant types. It minimizes the variance of the prediction errors by adding some bias to the competitive base learners. Average probability, majority voting and weighted average probability are the most popular ensemble techniques used in different studies

as found in the literature. Several approaches have employed ensemble methods to integrate the outputs of multiple models, combining the strengths of individual classifiers to improve overall performance. For instance, Bhowal et al. [84] proposed a fuzzy ensemble model using Choquet integrals, integrating fuzzy measures from information theory and coalition game theory. This approach combines the strengths of individual classifiers in a fuzzy framework to improve the classification of breast histopathology images. Wang et al. [85] employed manifold learning to fuse CNN features extracted at different layers of the network. This approach utilized adaptive neighborhood preservation to measure similarity among samples, further enhancing the feature representation and improving the classification performance. Wang et al. [86] integrated CNN and CapsNet features to simultaneously capture semantic and spatial information. This hybrid approach aims to leverage the strength of CNNs in feature extraction and the spatial awareness provided by Capsule Networks, which improves the classification of breast cancer histopathology images. Zerouaoui et al. [87] evaluated seven deep learning architectures and used the Borda count method to rank the best combination of models. This voting-based approach enabled the authors to select the most effective ensemble of CNN models for breast cancer classification. Kassani et al. [88] proposed a three-path ensemble approach, utilizing VGG19, MobileNetV2, and DenseNet201 as base models. The fully connected layers of these models were concatenated to create a final feature representation, which was then classified to improve prediction accuracy.

These studies highlight the effectiveness of combining multiple CNN models and leveraging ensemble learning techniques to enhance the performance of breast histopathology image classification. The integration of TL, feature fusion, and advanced ensemble strategies has led to significant improvements in classification accuracy, providing a strong foundation for developing more reliable and robust automated diagnostic systems for breast cancer classification.

Limitations of Existing Approaches: Despite notable advancements, DL-based cell segmentation methods continue to face significant challenges in clinical applications. A primary obstacle lies in the inherent complexity of histopathological images, which often feature intricate structures and noisy backgrounds. These characteristics make it difficult to accurately segment ROIs, particularly at the cellular level. Moreover, deep neural networks rely heavily on large volumes of high-quality annotated data—resources that are limited in the pathological domain. This data scarcity hampers model generalization

and reduces performance on unseen samples. Additionally, the increasing complexity of SOTA models often comes at the cost of high computational demands, limiting their scalability and practical deployment in resource-constrained clinical environments. Although lightweight models offer a more computationally efficient alternative, achieving high segmentation accuracy with such models remains a persistent challenge. These limitations highlight the pressing need for the development of segmentation approaches that are not only lightweight and accurate but also robust and generalizable for real-world clinical use.

Also, the traditional ensemble methods have some flaws in it. For instance, the majority voting method gives equal weightage to each classifier irrespective of their classification ability. The weighted average technique has the flaw of giving higher weight to classifiers with higher accuracy and ignores a group of classifiers working together and classifiers sharing redundant data. For example, a classifier may not be able to produce good classification results when working alone but can produce substantially good results if it works in a group with other classifiers.

2.5 Discussion

The literature surveyed across multiple imaging modalities—ultrasound, thermography, mammography, and histopathology—reveals a clear evolution from traditional handcrafted methods toward DL-based and optimization-enhanced classification frameworks. Each modality introduces unique challenges and opportunities, which have been addressed through increasingly sophisticated computational strategies.

In **ultrasound imaging**, early reliance on handcrafted features such as texture, shape, and fractal descriptors provided initial baselines but suffered from limited generalizability, especially in the presence of variable lesion morphology and low-quality scans. The integration of DL, particularly CNNs and TL, improved classification accuracy and robustness. However, standalone CNN models often lacked precision in feature localization. Recent trends have thus shifted toward hybrid models, combining deep feature extraction with metaheuristic optimization and ensemble classifiers. These methods, especially when applied to datasets like BUSI, have demonstrated improved generalization and interpretability. Nonetheless, computational complexity remain ongoing concerns.

In **thermographic imaging**, the emphasis has been on leveraging temperature patterns and textural cues embedded in infrared scans. While CNN-based TL has yielded promising results with lightweight models like MobileNetV2, handcrafted descriptors and bio-inspired segmentation approaches (e.g., PSO, GWO) have been essential in capturing thermal variance and ROI-specific features. Despite these advances, computational overhead and lack of task-specific model optimization remain barriers to real-time implementation. Future progress in this domain may lie in developing streamlined, end-to-end pipelines that incorporate contextual metadata and dynamic model tuning without increasing system load.

For **mammography**, DL techniques, particularly through fine-tuned CNNs and TL, have significantly enhanced classification accuracy and reduced diagnostic errors. Models such as YOLO for lesion localization, when combined with CNN classifiers, have shown good generalizability across datasets like DDSM and INBreast. Moreover, attention mechanisms and FS algorithms like BMFO and WOA have been explored to improve model focus and reduce dimensionality. Despite the gains, variability in breast density, inter-patient differences, and limited interpretability of some black-box models still limit clinical adoption. Therefore, future research should emphasize attention-based and interpretable hybrid models that can adapt to varying image quality and complexity.

In the domain of **histopathology**, ensemble DL frameworks have become the dominant trend. These models, especially those using multi-CNN backbones, capsule networks, or attention-based fusion, have delivered high accuracy in classification tasks. The use of ensemble voting strategies (e.g., majority voting, weighted averaging, Borda count) has proven effective in mitigating individual model weaknesses. However, the need for large annotated datasets and high processing requirements for patch-based methods remain significant challenges. Addressing these issues may involve weakly-supervised learning, attention-guided sampling, and TL strategies that can scale efficiently while preserving accuracy.

On the **segmentation** front, architectures like U-Net and its variants (U-Net++, Attention U-Net, and CBAM-integrated models) have revolutionized the detection of tumors, particularly in ultrasound and histopathology images. Attention modules and transformer-based designs have further improved feature focus and boundary sensitivity. Nonetheless, the trade-offs between accuracy, computational cost, and model complexity are still unresolved. While attention-enhanced CNNs and Swin Transformers offer

improved spatial awareness, they often come at the cost of inference speed and memory usage.

In summary, the current landscape reflects a common trajectory across modalities: a shift from purely handcrafted methods to advanced DL-based models, with increasing adoption of optimization techniques and attention mechanisms. Despite notable performance gains, several key challenges persist:

- Computational efficiency and scalability remain major challenges for clinical deployment, especially in thermography and histopathology-based approaches, highlighting the need for lightweight, automated architectures suitable for real-time integration.
- Data imbalance and annotation scarcity affect performance in ultrasound and histopathology image classification.
- Model interpretability and adaptability are crucial for adoption in mammography, where dense tissues and subtle lesions challenge even advanced systems.

From the above discussion it is clear that there is a requirement of developing unified frameworks that effectively balance accuracy, efficiency, and interpretability. Attention-driven, context-aware DL systems—potentially optimized through evolutionary algorithms—offer promising directions for next-generation CAD tools. These advanced systems have the potential to improve diagnostic accuracy across various imaging modalities while also supporting clinical decision-making, particularly in diverse and resource-limited settings.

Chapter 3

Advanced U-Net Models for Breast Tumor Segmentation

Breast tumor segmentation plays a crucial role in the CAD pipeline, enabling clinicians to localize, characterize, and assess tumor progression with greater precision. Accurate segmentation in medical imaging not only assists in early detection but also enhances treatment planning and monitoring. This chapter focuses on breast tumor segmentation across two critical imaging modalities—ultrasound and histopathology—both of which provide complementary insights into tumor morphology. Ultrasound imaging offers a real-time, non-invasive view of soft tissue structures, commonly used in screening and diagnostic workflows. Histopathological images, on the other hand, deliver cellular-level detail essential for cancer grading and prognosis.

Despite their diagnostic potential, precise and automated segmentation in both ultrasound and histopathology images remains a major challenge. In ultrasound imaging, factors such as speckle noise, low contrast, blurred lesion boundaries, and tumor shape variability often obscure the precise contours of breast lesions [30]. Similarly, in histopathology, inconsistent staining, overlapping nuclei, variable nuclear morphology, and complex background textures hinder accurate nuclei segmentation, which is vital for detecting malignancy.

Recent advancements in DL, particularly CNNs, have significantly improved the performance of medical image segmentation tasks. The U-Net architecture [21], with its encoder-decoder structure and skip connections, has emerged as a cornerstone model for biomedical image segmentation due to its ability to capture both global context and

local spatial details. However, the vanilla U-Net faces limitations when applied to complex breast imaging data. In ultrasound, U-Net may struggle to delineate lesions from low-contrast backgrounds, while in histopathology, the skip connections can propagate irrelevant features, and the network may overlook fine-grained structures like nuclear boundaries.

To address these modality-specific challenges, this chapter proposes enhanced U-Net-based architectures that leverage attention mechanisms and guided feature enhancement to improve segmentation accuracy in both domains. In this chapter, three models are presented, each advancing the field of breast tumor segmentation.

DAU-Net: Dual Attention-Enhanced U-Net for Tumor Segmentation in Breast Ultarsound: [89] The first method focuses on an attention-enhanced U-Net model, where a hybrid architecture is proposed that integrates PCBAM and SWA. These attention mechanisms help capture both spatial and global contextual information, allowing the network to focus on tumor regions while suppressing irrelevant features from the surrounding tissue. This model demonstrates substantial improvements in accuracy, particularly in distinguishing boundaries of tumors that often appear indistinct due to the noisy and heterogeneous nature of ultrasound images.

$EU^2 - Net$: A Lightweight Ensemble U-Net Architecture for Breast Ultrasound Image Segmentation [90]: The second model involves the development of a novel ensemble architecture, $EU^2 - Net$, which builds upon the $U^2 - Net$ framework [29]. $EU^2 - Net$ introduces a lightweight ensemble mechanism by incorporating separable convolutions to reduce the number of trainable parameters, making the model more computationally efficient and suitable for real-time applications. Additionally, a weighted averaging ensemble technique is incorporated within the network, allowing for enhanced segmentation performance without significantly increasing model complexity. This approach leverages attention-aided triple feature fusion, combining feature similarity-based attention, Squeeze-and-Excitation (SE) channel attention, and Sobel edge features for improved segmentation of breast lesions in ultrasound images.

AWGUNet: Attention-aided Wavelet Guided U-Net for Tumor Segmentation in Breast Histopathology [91] In this approach, the classical U-Net is extended by incorporating a DenseNet-121 encoder backbone to strengthen feature propagation and reuse. To improve boundary localization, WGCAM is introduced that captures boundary-aware spatial features. Furthermore, a learnable weighted global attention

pooling is employed to model channel-specific dependencies, enabling the network to focus on more relevant feature maps. The decoder, comprising structured upsampling and convolutional blocks, refines the segmentation output by handling small and complex regions influenced by staining artifacts. The method is evaluated on two benchmark datasets—MoNuSeg and TNBC—demonstrating superior performance in accurately segmenting nuclei in breast histopathology images.

Objectives of the Chapter

- To explore and integrate advanced DL architectures, particularly U-Net and its variants, for improving segmentation performance in breast ultrasound and histopathology imaging.
- To enhance segmentation accuracy using dual attention mechanisms by proposing an attention-augmented U-Net incorporating PCBAM and SWA.
- To develop a lightweight ensemble segmentation model (EU^2-Net) that reduces computational complexity while maintaining high accuracy through separable convolutions and a weighted triple feature fusion strategy.
- To effectively address segmentation challenges posed by histopathological images, such as overlapping nuclei, staining artifacts, and heterogeneous tissue morphology, by enhancing both spatial precision and contextual understanding.
- To demonstrate how attention mechanisms and edge information can guide the model to focus on tumor-specific regions, thus improving the precision of CAD systems in clinical settings.

3.1 Model 1: Dual Attention-Enhanced U-Net (DAU-Net)

A novel segmentation method for detecting tumors in breast ultrasound images is proposed that incorporates two attention mechanisms—PCBAM and SWA—which are designed to effectively capture context-aware features, spatial dependencies, and global contextual information. The complete architecture of the proposed segmentation model is illustrated in Fig. 3.1.

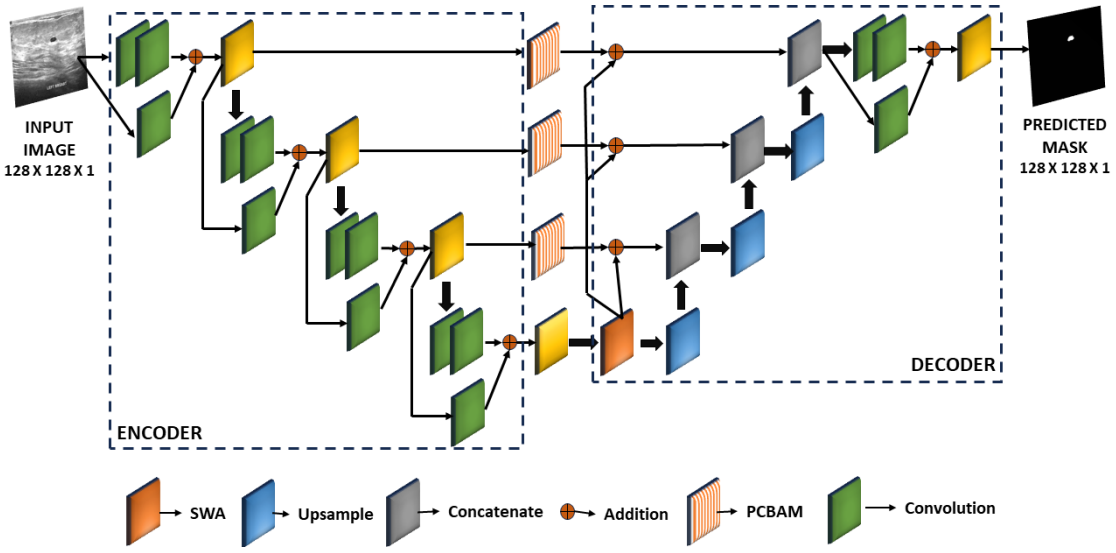


FIGURE 3.1: Block diagram of the proposed DAU-Net model used for segmentation of tumor in breast ultrasound images. An input image of size $128 \times 128 \times 1$ is processed by the encoder for feature extraction, while the decoder upsamples the encoded features to generate a binary segmentation mask of the same dimensions. To enhance performance, PCBAM and SWA attention mechanisms are integrated into the skip connections between the encoder and decoder.

3.1.1 Encoder Module

The encoder extracts hierarchical features from the input data using convolutional layers with 3×3 filters and a stride of 1. Each convolutional layer is followed by batch normalization and Rectified Linear Unit (ReLU) activation to ensure feature stability and promote efficient information flow. To ensure smooth gradient flow during training and to retain essential information, residual connections are utilized. To downsample, 2×2 stride convolutional layers are employed. Additionally, the PCBAM mechanism improves the encoder features before connecting them with the decoder features via residual connections. More information on this mechanism can be found in the subsequent Section.

3.1.2 Decoder Module

In the process of upsampling and reconstructing segmented output, the decoder plays a crucial role. By combining upsampled feature maps with those from the attention-aided encoder features, it gains access to both low-level and high-level features. This happens through strategic fusion, where features are refined using convolutional layers, followed by batch normalization and ReLU activation, leading to a richer, higher-dimensional

representation of spatial relationships. To further enhance the spatial dimensions, the decoder uses residual blocks, which contribute to its exceptional performance. Moreover, the SWA layer is incorporated in the decoder, capturing global dependencies and improving spatial coherence in the segmentation results.

3.1.3 Positional Convolutional Block Attention Module

CBAM attention mechanism [56] is applied to the last feature map of dimension $C \times H \times W$ generated from any CNN architecture. Here, C, H and W represent a feature map's number of channels, height, and width, respectively. The CBAM attention mechanism contains two modules: one dimensional Channel Attention Module (CAM) and two dimensional Spatial Attention Module (SAM). The CAM assigns weights to the feature map channels, emphasizing those that contribute more significantly to improving the model's performance. It is formulated as per Eq. 3.1.

$$F_c = \sigma(mlp(gap(F)) + mlp(gmp(F))) \tag{3.1}$$

In Eq. 3.1, σ denotes the sigmoid activation function, *gap*, *gmp* and *mlp* represents the global average pooling layer, global max pooling layer and the multi-layer perceptron respectively. *mlp* consists of two successive fully connected i.e., dense layers (D_L) with C and $C/8$ units, respectively and F is the feature map. Now, $F'_c = F_c \otimes F$ is fed to the SAM ($\otimes$ denotes the element-wise matrix multiplication). The SAM module processes the feature map F'_c, produced by the CAM, by applying a spatial attention mask to refine and enhance the feature representation. The SAM is formulated according to Eq. 3.2.

$$F^{''} = f^{7\times7}[D_L(gap(F'_c)); D_L(gmp(F'_c))] \tag{3.2}$$

In Eq. 3.2, $f^{7\times7}$ is a convolutional layer of kernel size 7×7 and ';' denotes the concatenation operation. F_{CBAM}, the final output feature map, is obtained by element-wise multiplication between $F^{''}$ and F'_c as shown in Eq. 3.3.

$$F_{CBAM} = F^{''} \otimes F'_c \tag{3.3}$$

The CBAM attention mechanism effectively captures both channel-wise and spatial dependencies, enabling the model to focus on relevant features, and enhance its performance in image segmentation.

Similarly, the Position Attention Module (PAM) is designed to enrich local features by incorporating a broader context, thereby enhancing their representational capacity. To achieve this, starting with a local feature map denoted as $F \in \mathbb{R}^{H\times W\times C}$. The feature map is passed through a convolutional layer, producing two new feature maps, B and Z, both of size $R^{H\times W\times C}$. Afterward, B and Z are reshaped into matrices of size $\mathbb{R}^{N\times C}$, where $N = H \times W$, representing the number of pixels in the feature map. The spatial attention map $S \in R^{N\times N}$ is generated by performing matrix multiplication between the transpose of Z and B, followed by a softmax operation. This attention map captures the spatial relationships between different pixels in the feature map. PAM allows local features to leverage a wider contextual understanding by employing the attention mechanism to emphasize relevant spatial information. This enables the local features to better represent complex patterns and structures in the input data. The formula is shown in Eq. 3.4.

$$s_{ji} = \frac{\exp(B_i \cdot Z_j)}{\sum_{i=1}^{N} \exp(B_i \cdot Z_j)} \tag{3.4}$$

where s_{ji} quantifies the influence of the ith position on the jth position.

Next, the feature map F is fed into a convolutional layer to generate a new feature map $D \in R^{H\times W\times C}$, which is reshaped to $R^{N\times C}$. A matrix multiplication is performed between D and the transpose of S, resulting in a feature map of size $R^{N\times C}$. It is then reshaped back to $R^{H\times W\times C}$. Finally, multiply it by a scale parameter α and an element-wise sum operation with the features F is performed to obtain $F_{PAM} \in R^{H\times W\times C}$. The calculation is done in accordance with Eq. 3.5.

$$F_{PAM_j} = \alpha \sum_{i=1}^{N} (s_{ji} \cdot D_i) + F_j \tag{3.5}$$

where α is initialized as 0. The model learns the parameter α and gradually adjusts it to assign greater emphasis where needed. The resulting feature F_{PAM} at each position represents a weighted combination of features from all positions and the original features,

enabling a global contextual perspective and selective aggregation based on the spatial attention map. This process enhances intra-class compactness and preserves semantic consistency within the feature representations. Leveraging the strengths of both CBAM and PAM, these modules are integrated using Eq. 3.6 to form PCBAM, where the input feature map F is fed into both CBAM and PAM. The block diagram of the PCBAM is shown in Fig 3.2.

$$F_{PCBAM} = F_{PAM} + F_{CBAM} \tag{3.6}$$

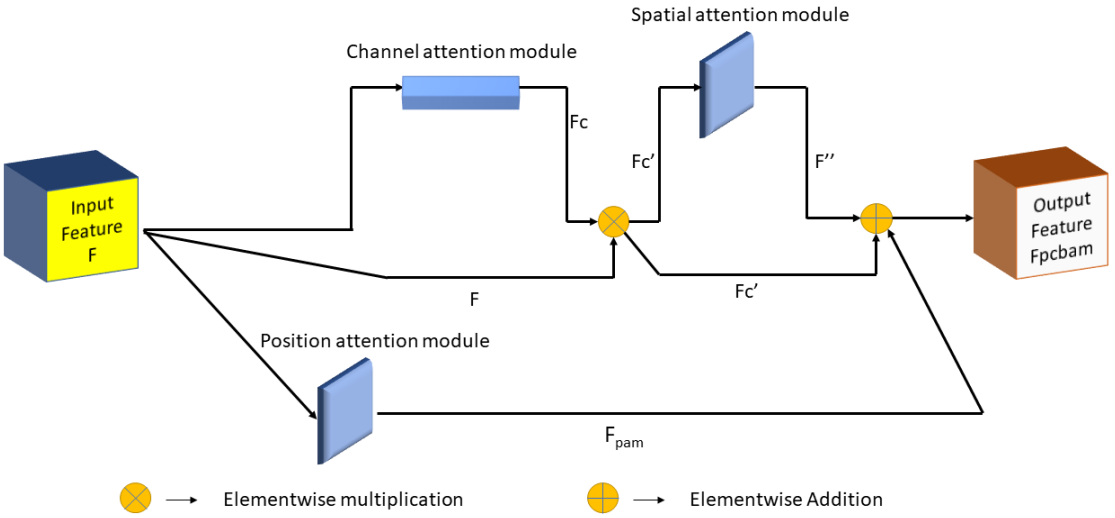


FIGURE 3.2: An illustration of the PCBAM attention block. CBAM and PAM are applied to the input feature F. The addition of the outputs of CBAM and PAM is the output of the PCBAM attention mechanism, F_{PCBAM}.

3.1.4 Shifted Window Attention

The SWA [27] is a powerful attention mechanism used to capture global dependencies and improve spatial coherence in the segmentation results of our proposed model. It enhances the model's ability to focus on relevant regions and strengthens its contextual understanding of the input images. In image segmentation tasks, understanding the contextual relationships among different regions is crucial. However, traditional convolutional operations might not fully capture these long-range dependencies. To this end, SWA module is used in order to address this limitation by introducing a window-based attention mechanism, which allows the model to attend to relevant information from different parts of the image. The SWA mechanism can be mathematically defined as follows. Let F be the input feature map of size $H \times W \times C$, where H, W, and C represent the height, width, and number of channels, respectively. To compute the attention

map, position-aware query matrix q, key matrix k, and value matrix v are calculated as follows:

$$q_{i,j} = F_{i,j} \cdot w_q \tag{3.7}$$

$$k_{i,j} = F_{i,j} \cdot w_k \tag{3.8}$$

$$v_{i,j} = F_{i,j} \cdot w_v \tag{3.9}$$

where w_q, w_k, and w_v are learnable weight matrices for query, key, and value projections, respectively.

Next, a convolution operation $f^{1\times1}$ (1×1 is the kernel dimension) is performed on q, k, and v to compute the attention map A as per Eq. 3.10.

$$A = f^{1\times1}(q, k, v) \tag{3.10}$$

The attention map A is then added element-wise to the original feature map F using a residual connection to obtain the final output of the SWA mechanism X_{out} using the following Eq. 3.11.

$$X_{\text{out}} = F + A \tag{3.11}$$

The SWA mechanism is integrated into the decoder part of the Residual U-Net architecture. By introducing SWA, the model can effectively capture long-range dependencies and achieve better spatial coherence in the segmentation results, leading to improved performance in segmenting breast tumor regions in ultrasound images.

3.1.5 Loss Function

Dice loss [92], Binary Cross-Entropy (BCE) loss [93], and Focal loss [94] are widely used in image segmentation tasks. These loss functions guide the training process by measuring the similarity between the predicted masks and the ground truth, thereby enhancing the model's learning efficiency.

The Dice loss quantifies the overlap between the predicted masks and the ground truth, aiming to maximize their similarity. Eq 3.12 defines the Dice loss, where TP represents the number of true positive pixels, FP signifies the number of false positive pixels, and FN denotes the number of false negative pixels.

$$Dice\ Loss = 1 - \frac{2 \times TP}{2 \times TP + FP + FN} \tag{3.12}$$

Conversely, the BCE loss measures the dissimilarity between predicted masks and their corresponding actual masks. Eq 3.13 outlines the BCE loss, where N represents the total number of pixels, y_i represents the ground truth label (foreground or background) for pixel i, and p_i represents the predicted probability of the foreground class for pixel i.

$$BCE\ Loss = -\frac{1}{N}\sum_{i=1}^{N}(y_i \log(p_i) + (1 - y_i)\log(1 - p_i)) \tag{3.13}$$

The Focal loss is designed to address class imbalance in segmentation tasks and provide more focus on hard-to-classify pixels. It assigns higher weights to misclassified pixels and thus reduces the impact of easy-to-classify pixels during training. The Focal loss is computed using the following Equation 3.14.

$$Focal\ Loss = -\frac{1}{N}\sum_{i=1}^{N}(\alpha(1 - p_i)^{\gamma}\log(p_i)) \tag{3.14}$$

In Equation 3.14, α serves as a balancing parameter that adjusts the contribution of each class, and γ is a focusing parameter that controls how strongly the loss emphasizes hard-to-classify pixels.

During training, the model is optimized using a composite loss function comprising Dice loss, BCE loss, and Focal loss, as defined in Eq. 3.15. Minimizing this combined loss enables the model to effectively learn and accurately segment the target regions of interest.

$$Loss = Dice\ Loss + BCE\ Loss + Focal\ Loss \tag{3.15}$$

3.2 Experimental Results and Analysis

3.2.1 Dataset Description

In the present study, the BUSI database [45] is used. This database contains 780 breast ultrasound images (each 500×500 pixels) including 437 benign cases, 210 malignant cases, and 133 images of normal breast tissue. To train and evaluate the proposed segmentation model, both benign and malignant cases are used, along with their corresponding segmentation masks, which serve as ground truth annotations.

3.2.2 Evaluation Metrics

The performance of the DAU-Net model is assessed using Dice score, Intersection over Union (IoU) score, accuracy, recall, and precision. These metrics offer quantitative insights into the model's effectiveness in accurately delineating regions of interest. The Dice score integrates both precision and recall into a single value for evaluation and is defined in Eq. 3.16.

$$Dice\ score = \frac{2 \times \mathrm{TP}}{2 \times \mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{3.16}$$

IoU quantifies the overlap between the ground truth mask and the predicted binary segmentation mask. It is computed as the ratio of intersection area to the union area of the two masks and is defined in Eq. 3.17.

$$IoU = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{3.17}$$

The accuracy metric assesses the overall correctness of binary segmentation and is calculated as the ratio of correctly classified pixels to the total number of pixels. It is defined in Eq. 3.18.

$$Accuracy = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{3.18}$$

Precision evaluates the fraction of true positive predictions among all positive predictions and is defined in Eq. 3.19.

$$Precision = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{3.19}$$

Recall, commonly referred to as sensitivity or true positive rate, quantifies the proportion of true positive predictions out of all the actual positive instances and is defined in

Eq. 3.20.

$$Recall = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.20}$$

3.2.3 Experimental Setup

The model is trained for 50 epochs, where every epoch represents one complete pass through the entire dataset. To address the issue of non-uniform sizes in the original BUSI images, all images are resized to a consistent resolution of 128×128 pixels before being fed into the model for segmentation. In the architecture's convolutional layers, 'He Normal' weight initialization is used, which has proven to be effective in deep neural network architectures. This initialization strategy contributes to better convergence and performance during training. During training, the model is optimized using the Adam optimizer with a learning rate of 0.0001. This choice of optimizer allows to efficiently update the model's parameters, enhancing convergence during training. To ensure a comprehensive assessment, the data is divided into a 70-10-20% train-test-validation split.

3.2.4 Ablation study

A series of experiments are performed to refine the segmentation model and assess the effects of various modifications. These include:

(i) Base Residual U-Net model, serving as the initial benchmark.

(ii) Residual U-Net model with PAM applied to the skip connections.

(iii) Residual U-Net model with CBAM applied to the skip connections.

(iv) Residual U-Net model with PCBAM, combining the strengths of PAM and CBAM.

(v) Proposed model with PCBAM and SWA, emphasizing global features.

Results in Table 3.1 showcase the efficacy of each modification. Each model is trained using the loss function defined in Eq 3.15. The addition of PAM and CBAM improves performance, while SWA further enhances accuracy and segmentation quality.

Fig 3.3 visually demonstrates the performance improvements achieved through the incorporation of attention mechanisms. The integration of PCBAM and SWA enhances

TABLE 3.1: Performance metrics of the Ablation study. All values are in %. Bold values indicate superior performance. The results are in $x(\pm y)$ format, where x is the mean and y is the standard deviation of the evaluation metric for the five runs of the model.

Model	Dice	Accuracy	Precision	Recall	IoU
(i)	68.27 (±0.60)	92.85 (±0.60)	68.15 (±0.41)	71.71 (±0.77)	55.82 (±0.53)
(ii)	71.72 (±0.32)	94.55 (±0.27)	**78.28** (±0.16)	65.45 (±0.39)	61.08 (±0.49)
(iii)	72.08 (±0.50)	94.38 (±0.25)	74.42 (±0.39)	70.86 (±0.59)	62.35 (±0.24)
(iv)	71.97 (±0.39)	94.26 (±0.45)	73.76 (±0.58)	69.66 (±0.76)	62.99 (±0.43)
(v)	**74.23** (±0.67)	**95.88** (±0.42)	73.81 (±0.43)	**74.59** (±0.65)	**65.32** (±0.56)

segmentation accuracy for both small and large regions of interest by refining feature representations and effectively capturing global and local spatial dependencies.

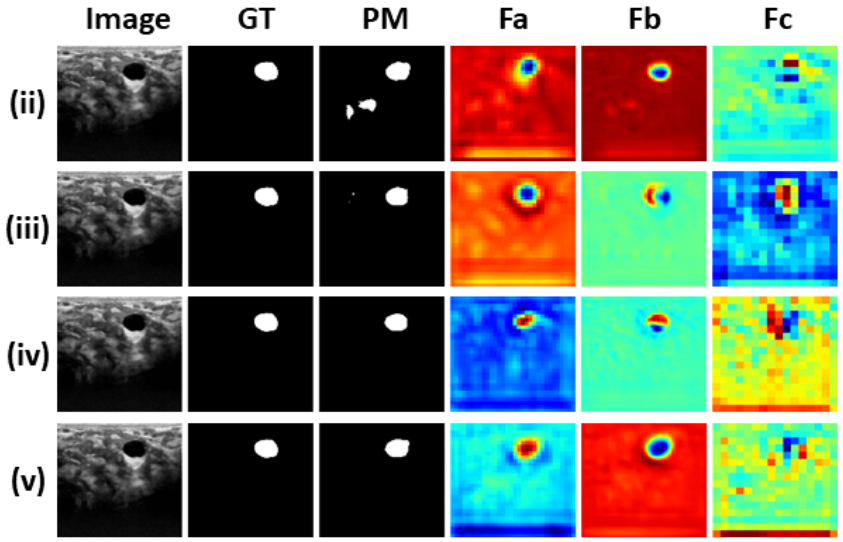
Through these experiments and analyses, the segmentation model shows the improvement iteratively, identifying the most effective modifications and attention mechanisms. These advancements make a significant contribution to enhancing the accuracy and robustness of the proposed DAU-Net model, positioning it as an advanced solution for segmenting breast tumors in ultrasound images.

Additionally, a five-fold cross validation [95] approach for assessing the model's generalizability is conducted, and tabulate the results under Table 3.2.

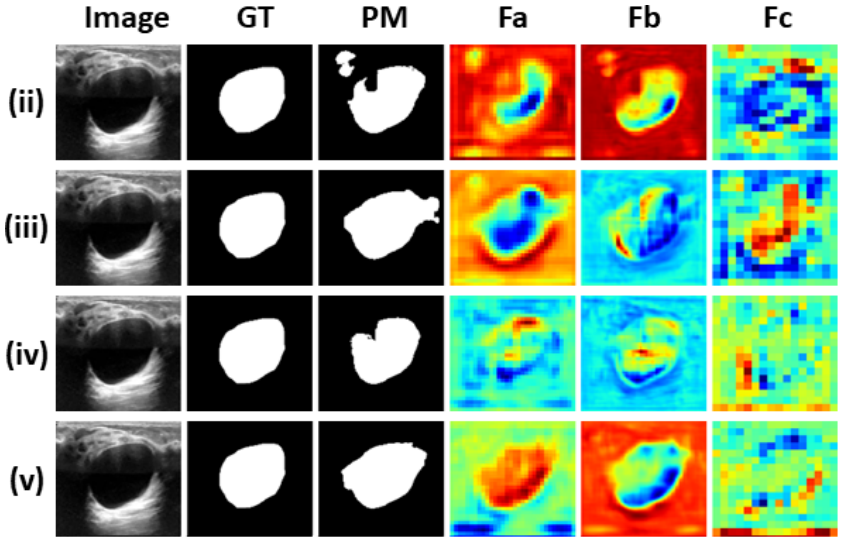
TABLE 3.2: Results of the proposed DAU-Net model with 5-fold cross-validation on the BUSI dataset.

5-Fold CV	IoU(%)	DSC(%)
Fold 1	64.85	73.58
Fold 2	65.43	74.14
Fold 3	65.17	74.29
Fold 4	64.79	73.21
Fold 5	65.91	74.61
Mean	65.23	73.97
Std. Dev.	0.411	0.504

Fig 3.4 showcases the segmentation results of the proposed DAU-Net model, demonstrating its ability to segment breast tumor regions in ultrasound images accurately. The heatmaps showcase the spatial regions where the SWA and PCBAM layers focus. Furthermore, the heatmap visualization of the proposed model as shown in Fig 3.3 highlights the spatial regions where the model places greater emphasis, closely aligning with the ground truth areas in the BUSI dataset. This demonstrates the model's ability to focus on relevant regions, thereby contributing to its accurate segmentation performance.



(A) Heatmap and predicted segmentation mask for a smaller region of interest.



(B) Heatmap and predicted segmentation mask for a larger region of interest.

FIGURE 3.3: Results of the ablation study indicate the improvement in model performance with each experimental modification. GT and PM are Ground Truth and Predicted Mask, respectively. F_c is the heatmap of the bottleneck layer and it demonstrates the improved model's performance in focusing on the ROI after the addition of the SWA in the bottleneck layer. F_a and F_b are heatmaps of the features flowing from the first and second encoder layers to the first and second decoder layers via skip connections. It can be seen that F_a and F_b get more enriched with the use of attentions such as CBAM, PAM, and PCBAM.

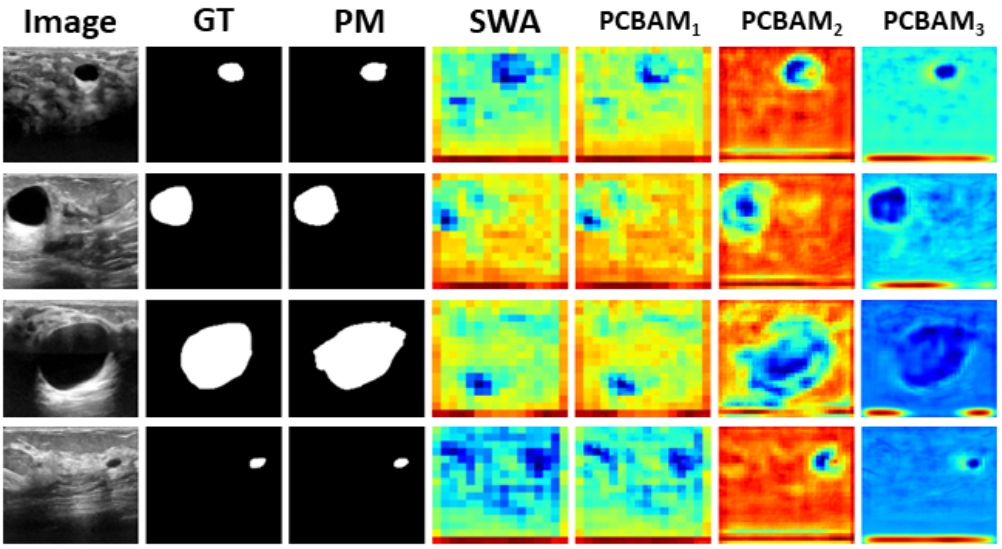


FIGURE 3.4: Results of the proposed segmentation model on images of the BUSI dataset and the heatmaps of SWA and $PCBAM$ layers. $PCBAM_1$ corresponds to the $PCBAM$ layer just above the SWA layer, $PCBAM_2$ corresponds to the PCBAM layer just above $PCBAM_1$ layer, and $PCBAM_3$ corresponds to the $PCBAM$ layer just above $PCBAM_2$ layer.

3.3 Model 2: EU^2-Net: A Lightweight Ensemble U^2-Net Architecture

For further enhancement of performance in breast ultrasound image segmentation EU^2-Net model is proposed. A novel ensemble architecture EU^2-Net is built upon the foundations of U^2-Net. The primary contribution lies in the development of a lightweight ensemble U^2-Net architecture, EU^2-Net, which strategically replaces standard convolutional layers with separable convolutions. This architectural modification brings about a significant reduction in the number of trainable parameters, a key factor in enabling real-time usability, and enhances the model's training efficiency. While ensemble methods are known to enhance the performance, they often come at the cost of a significant increase in the number of model parameters due to the involvement of multiple models, leading to resource-intensive training and deployment. In contrast, EU^2-Net incorporates a weighted averaging ensemble mechanism with learnable weights internally within the U^2-Net architecture. To further improve the performance of the model, an attention-aided triple feature fusion technique is introduced. This enhances both encoder and decoder features by leveraging feature similarity-based attention and Squeeze-and-Excitation (SE) channel attention, respectively. Additionally, edge features extracted using the Sobel filter from the encoder features are incorporated that provide valuable

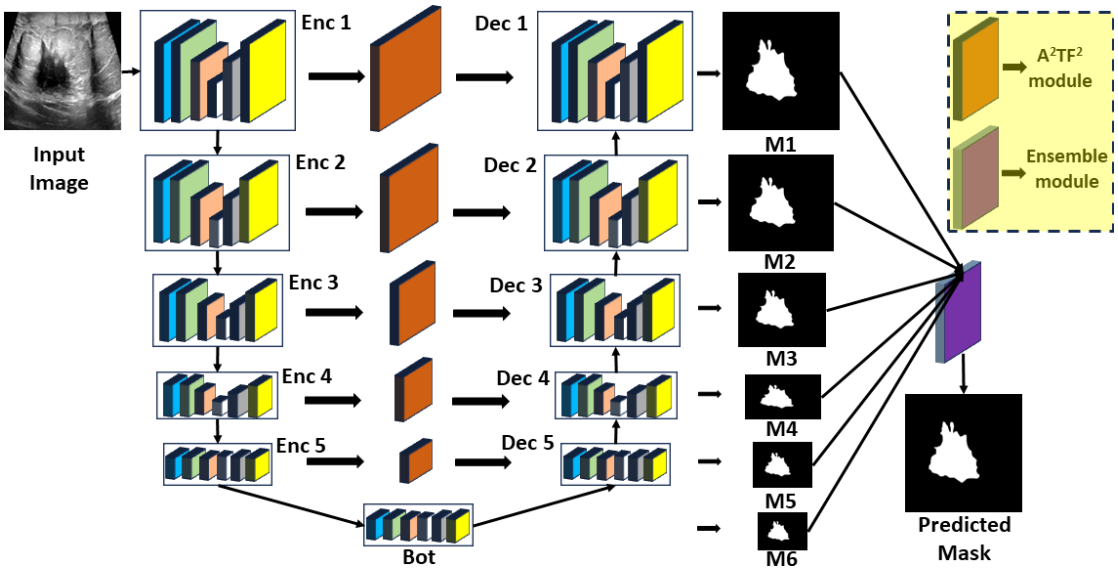


FIGURE 3.5: A block diagram of the $EU^2 - Net$. The standard $U^2 - Net$ is enhanced by the A^2TF^2 module for a multi-feature aware segmentation. The ensemble module internally combines the masks $M1$, $M2$, $M3$, $M4$, $M5$, and $M6$ to leverage the masks generated from various network depths, thus focusing on capturing both detailed and general features. The weights of the weighted averaging used in the ensemble module are made trainable along with the model training procedure.

boundary-aware information. These multi-domain features are skillfully fused together, reinforcing the decoder with a rich set of both general and detailed features. This fusion plays a pivotal role in improving segmentation accuracy for the images, where complex ultrasound patterns, similar intensity distributions, variable tumor morphologies, and blurred boundaries are found.

The $EU^2 - Net$ model, as illustrated in Fig 3.5, represents a significant advancement in image segmentation within the field of architecture. This enhanced framework builds upon the established $U^2 - Net$ model by integrating the Attention Aided Triple Feature Fusion (A^2TF^2) module. This module fosters multi-feature awareness during segmentation tasks by combining the boundary-aware feature information of the encoder with the attention-aided decoder features. The A^2TF^2 module is a novel addition to the proposed model. Moreover, what distinguishes the $EU^2 - Net$ model is its use of the ensemble module, which strategically combines the internally generated masks M1, M2, M3, M4, M5, and M6 as shown in Fig 3.5. This mechanism skillfully harnesses masks generated at different network depths to capture both intricate, detailed features and broader, general features within the input images. It's worth mentioning that the weights governing the averaging process within the Ensemble module are designed to be trainable. This adaptability ensures that the model can fine-tune its feature combination strategy

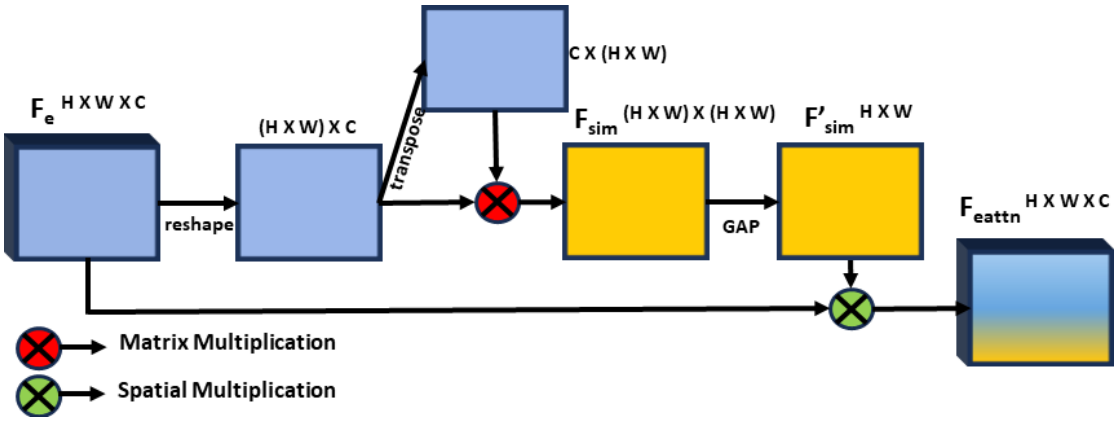


FIGURE 3.6: A block diagram of the Feature Similarity-based Attention module.

during the training procedure, resulting in improved segmentation performance and ensuring the domain-independence. Moreover, unlike traditional ensemble techniques that require multiple models and lead to increased parameters and computational costs, the $EU^2 - Net$ model achieves enhanced performance with the same parameters by using internally generated masks.

3.3.1 Feature Similarity-based Attention Module (FSiAM)

For gray-scale images with low-intensity contrast, convolution operations alone cannot extract diverse features as natural images. This results in redundant features affecting classification performance, especially when distinguishing the background from the foreground. The Feature Similarity-based Attention Module (FSiAM) captures the foreground and background pixel-wise correlation and helps capture intricate patterns. The encoder feature, F_e, is reshaped to a 2D feature matrix of dimensions $(H \times W) \times C$. This reshaped feature is transposed to generate a 2D feature matrix of dimensions $C \times (H \times W)$. These two 2D matrices are then multiplied to generate a similarity-aware feature, F_{sim}, of dimensions $(H \times W) \times (H \times W)$ by Eq. 3.21. F_{sim} is treated by a global average pooling layer (GAP) to generate F'_{sim} of dimensions $(H \times W)$ as shown in Eq 3.22. This matrix is used to generate F_{eattn} to improve the performance as shown in Eq. 3.23, where the sigmoid function is used and $\otimes$ denotes the multiplication across the spatial dimension. Fig. 3.6 presents the block diagram of the proposed FSiAM module. Fig. 3.7 shows the module's ability to highlight the foreground pixel correlation to capture finer details of ultrasound images.

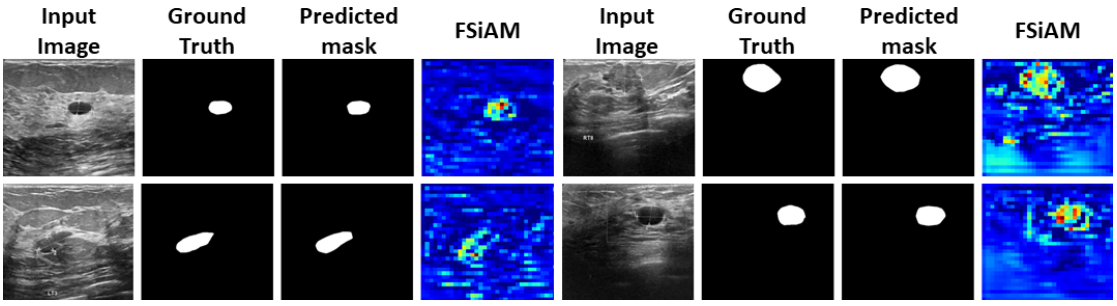


FIGURE 3.7: Heatmap of FSiAM corresponding to the Enc 5 (refer Fig 3.5), which highlights the foreground pixel correlation.

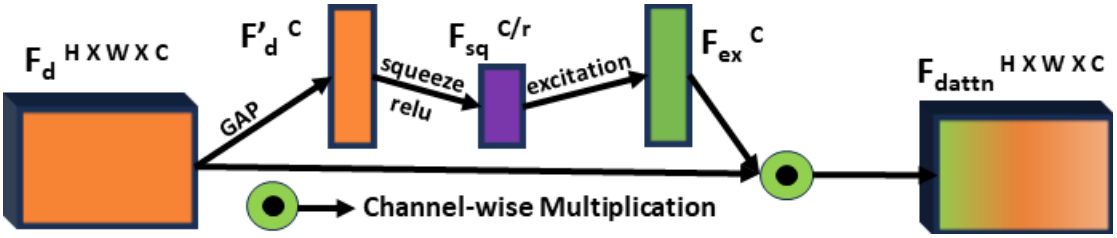


FIGURE 3.8: Squeeze and Excitation Channel Attention module.

$$F_{sim}^{(H\times W)\times(H\times W)} = F_e^{(H\times W)\times C} \times F_e^{C\times(H\times W)} \tag{3.21}$$

$$F_{sim}^{\prime H\times W} = GAP(F_{sim}^{(H\times W)\times(H\times W)}) \tag{3.22}$$

$$F_{eattn} = sigmoid(F_{sim}^{\prime}) \otimes F_e \tag{3.23}$$

3.3.2 Squeeze and Excitation Channel Attention Module (SE-CAM)

This module utilizes two main components to gather global information and adaptively weigh the importance of different channels in feature maps. Firstly, the feature map is squeezed to reduce its spatial dimension and obtain a channel-wise summary of global information. Secondly, the squeezed features are reshaped back to their original form. This adaptive channel-wise scaling helps the network to focus on relevant features and ignore less useful information. The decoder feature map, F_d, of dimensions $H \times W \times C$, is flattened using global average pooling (GAP) to generate a 1D feature map of dimension C as shown in Eq 3.24. This 1D feature map is further reduced to dimension C/r and

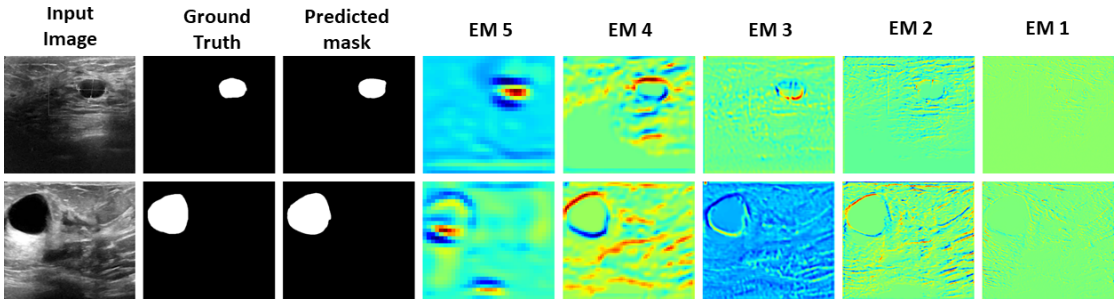
then regained to dimension C using dense layers and ReLU function as per Eq 3.25. The squeezed and excited feature map, F_{ex}, is used to generate F_{dattn}.

$$F_d'^C = GAP(F_d^{H \times W \times C}) \tag{3.24}$$

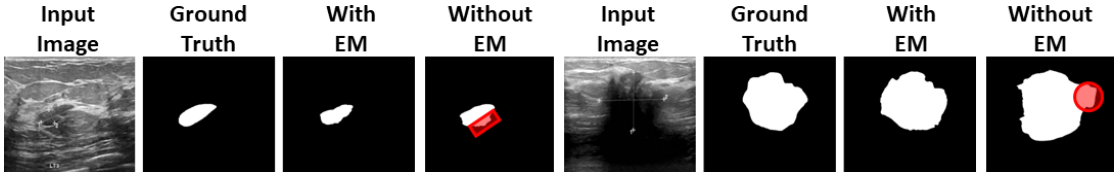
$$F_{ex}^C = dense(relu(dense(F_d'^C))) \tag{3.25}$$

3.3.3 Edge Module (EM)

When it comes to ultrasound breast image segmentation, accurately delineating the boundaries of lesions or tumors becomes crucial. To this end, the Sobel filter is used to extract information across H and W, resulting in more precise edge-aware segmentation maps. The extracted features of dimensions $H \times W \times C$ from the Sobel filter for H and W directions are then concatenated to form a feature map of dimensions $H \times W \times 2C$. This feature map contains necessary edge information of F_e. Finally, F_{edge} of dimensions $H \times W \times C$ is generated by passing on the concatenated feature to two consecutive convolution layers with 2C (kernel size of 3×3) and C (kernel size of 1×1) number of channels, sequentially. F_{edge} complements the spatial features extracted by the encoder, providing a comprehensive view of the ultrasound images. This additional information improves the model's understanding of underlying structures and textures, thereby leading to better segmentation maps. Fig. 3.9a displays the edge information EMs generate. This edge guidance assists the A^2TF^2 module to focus on the areas of interest by emphasizing the significance of the boundary of the regions. From Fig. 3.9a it is evident that the shallower layers, such as EM 5 (corresponding to Dec 5), EM 4 (corresponding to Dec 4), etc., capture more useful boundary information than the deeper layers, like EM 1 (corresponding to Dec 1), which focus on finer details. This indicates that combining the shallower and deeper features results in a better segmentation output by focusing on the relevant shape features and finer details. Furthermore, Fig. 3.9b demonstrates the boundary adherence introduced by the use of EMs. This ensures that the foreground pixels adhere to the boundaries of the tumorous region and prevent false positives.



(A) Visual representation of images after applying EM to capture edge details. EM1–EM5 are the Edge Modules corresponding to the Decoder layers Dec1–Dec5, respectively.



(B) Qualitative analysis of the boundary adherence achieved with the use of EM.

FIGURE 3.9: Illustration of the heatmaps and the enhancement in segmented mask due to the spatial boundary information enrichment by EM.

3.3.4 Learnable parameter-efficient weighted averaging ensemble

Researchers often use ensemble methods to improve classification and segmentation performance, which can be inefficient as they require multiple models to generate and combine outputs for better results. This can lead to an increased number of parameters and computational costs. To address this challenge, a more efficient ensemble strategy has been devised by leveraging masks generated by individual decoders of $U^2 - Net$. In the conventional $U^2 - Net$, decoder features are concatenated and passed through a final convolutional layer with sigmoid activation to generate the final output. However, this approach has limitations as it combines all concatenated features using a single layer with a sole filter, potentially missing intricate information.

The innovation lies in fusing masks $M1$ through $M6$ from different depths of the model, enhancing the model's ability to capture both intricate and generalized features. The weighted averaging ensemble technique as depicted in Eq. 3.26 employs dynamically adjustable learnable weights, w_i during the model's training, optimizing the ensemble process. This dynamic weight adjustment ensures that the ensemble technique evolves and improves over time by providing performance-specific weightage to each mask. The computational complexity of multi-scale feature-level fusion techniques, such as Atrous Spatial Pyramid Pooling (ASPP) and Feature Pyramid Networks (FPNs), is significantly high. This is primarily due to the concatenation of feature maps, which increases the

number of channels C by a factor of N, resulting in a complexity of $O(HWNC)$. Additionally, ASPP introduces further complexity through multiple dilated convolutions, culminating in a total complexity of $O(HWNC^2)$. In contrast, the mask-level weighted fusion method maintains the original feature map dimensions and avoids the need for channel concatenation, thereby achieving a linear complexity of $O(HWn)$ where n is the number of masks to be ensembles (n is much lesser than C and NC). This method is not only computationally efficient but also enhances performance in ensemble learning.

$$M_{\text{pred}} = \frac{\sum_{i=1}^{6} w_i \cdot M_i}{\sum_{i=1}^{6} w_i} \tag{3.26}$$

3.3.5 Loss Function

The proposed $EU^2 - Net$ model employs a composite loss function that combines the Dice loss and BCE loss, as depicted in Eq 3.27. This fusion of loss functions enhances the model's training ability in various image segmentation tasks. Dice loss and BCE loss are defined earlier in Section 3.1.5.

$$L_{seg} = L_{Dice} + L_{BCE} \tag{3.27}$$

3.4 Experimental Results and Analysis

3.4.1 Experimental Setup

The model is evaluated on BUSI and UDIAT dataset with 70-10-20% train-test-validation split. The segmentation model is developed using Python and have leveraged the TensorFlow and Keras libraries for implementation. For data processing, numpy, OpenCV, and scikit-learn libraries are utilized, which have facilitated the efficient handling of data. To speed up training and utilize hardware acceleration, the high-performance NVIDIA TESLA P100 GPU is used. Model is trained for 50 epochs and callbacks are used to save the model with the best validation performance. To address the issue of non-uniform sizes in the original images, all images are resized to a uniform size of 512×512 pixels, which are input into the model for segmentation. All quantitative values in subsequent tables are in %, with bold values indicating superior performance.

3.4.2 Ablation Study

A series of experiments have been conducted on the BUSI dataset to refine the proposed segmentation model and evaluate the impact of various modifications. These experiments include:

(i) Base $U^2 - Net$ model, serving as the initial benchmark.

(ii) Lightweight $U^2 - Net$ model, replacing traditional convolution layer with separable convolution layer.

(iii) $U^2 - Net$ model with FSiAM attention mechanism.

(iv) $U^2 - Net$ model with SE-CAM attention mechanism.

(v) $U^2 - Net$ model with SE-CAM + FSiAM attention mechanism.

(vi) $U^2 - Net$ model with the A^2TF^2 attention mechanism (SE-CAM + FSiAM + EM).

(vii) The proposed model ($EU^2 - Net$).

TABLE 3.3: Performance metrics from the ablation study results of the proposed segmentation model on the BUSI dataset.

Model	Dice	Accuracy	Precision	Recall	IoU
(i)	69.11	95.14	70.64	73.90	56.17
(ii)	70.76	95.57	72.97	73.26	59.05
(iii)	72.25	95.31	74.81	73.65	60.46
(iv)	72.18	95.37	79.12	68.94	61.30
(v)	73.22	95.76	77.98	74.80	61.77
(vi)	73.99	95.81	82.97	74.10	61.93
(vii)	**74.73**	**96.01**	**81.35**	**75.40**	**64.48**

The effectiveness of each modification is displayed in Table 3.3. Convolutional layers are replaced with separable convolution (ii), leading to improved performance and a significant reduction in parameters from 44 M to 5.15 M. FSiAM (iii) identifies foreground and background regions, resulting in a decrease in false negatives, but it fails to keep false positives low, especially in darker regions of most image areas. SE-CAM (iv) recalibrates channel-wise feature responses, effectively reducing false positives, but it is unable to tackle false negatives due to image noise, complex darker patterns, and unrelated features captured by the decoder. The combination of FSiAM and SE-CAM (v) amalgamates the advantages of both modules while compensating for each other's

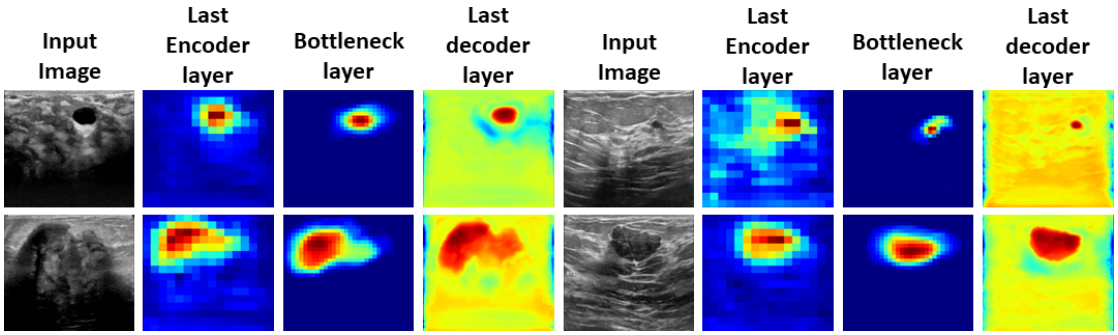


FIGURE 3.10: Visualization of the features at different layers of $EU^2 - Net$.

weaknesses. Furthermore, edge information (vi) brings boundary awareness, further enhancing performance. EM ensures boundary adherence concerning tumorous regions and thus reduces false positives, which, in turn, increases the precision score. The internally occurring ensemble's dynamically updated weights (vii) capture intricate details from each decoder layer, further enhancing the model's performance.

The representation of the features at different layers of the model is shown in Fig. 3.10. Specifically, the visual representations of the last encoder layer (Enc 5 in Fig 3.5), the bottleneck layer (Bot in Fig 3.5), and the last decoder layer (Dec 5 in Fig 3.5). It can be seen from Fig 3.10 that the bottleneck features are more focused on the region of interest compared to the encoder layer, which has a wider focus. The last decoder layer, which comes before the predicted mask generation, accurately pinpoints the regions without getting distracted by other areas. This gradual increase in information demonstrates the importance of having accurate segmentation. Fig 3.11 shows some of the predicted segmentation masks produced by the proposed model. Also, it can be seen that shallower layers, i.e., Dec5 and Dec4 focus more on the contours of the regions, whereas the deeper layers, i.e., Dec1 and Dec2 are more focused on details. The varying information extracted from decoder layers is fused to generate an optimized output mask.

3.5 SOTA comparison

The performance of both the proposed models (DAU-Net and $EU^2 - Net$) are compared against SOTA models and standard segmentation models, and the results are tabulated in Table 3.5 and Table 3.4, respectively. The table reviews different evaluation metrics. Table 3.4 includes well-established image segmentation models, such as FCN [47], U-Net [21], SegNet [51], ResUNet [96], Attention-UNet [53] and ENC-Net [97].

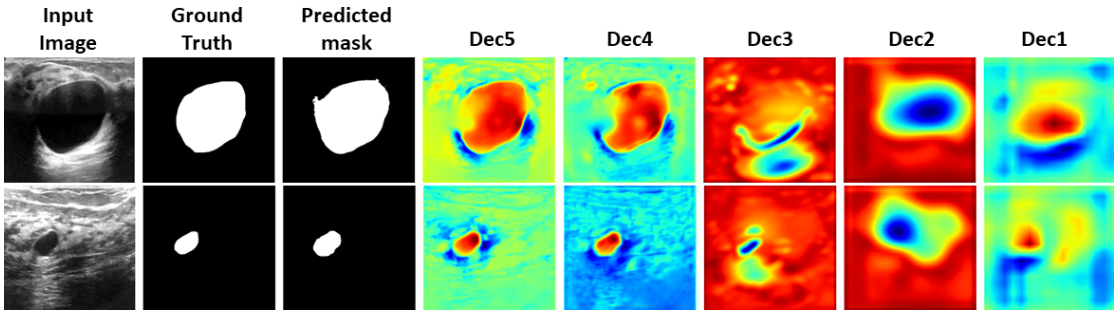


FIGURE 3.11: Some predicted masks of $EU^2 - Net$ on images of the BUSI dataset. Heatmaps show the focus of decoder layers: Dec5, Dec4, Dec3, Dec2, and Dec1. These are visualizations of feature maps before applying the sigmoid activation function to generate masks M1, M2, M3, M4, M5, and M6.

Both of the proposed approaches performs better than the SOTA models, as depicted in Table 3.5. They surpasses these models in Dice score and Precision, underscoring its superior accuracy in segmentation overall with fewer parameters. Furthermore, they attains greater precision than all these models, highlighting its precision in correctly identifying positive predictions. Nevertheless, it's worth mentioning that the proposed methods exhibits a slightly lower recall value than some of these models, indicating a possible trade-off between Precision and Recall.

TABLE 3.4: Performance comparison with standard segmentation models on the BUSI dataset. Here # Params denotes the number of parameters.

Model	# Params	Dice	Precision	Recall	IoU
FCN [47]	134M	71.23	69.07	77.02	56.27
UNet [21]	14.75M	71.32	66.96	78.46	56.13
Attention-UNet [53]	34.88M	69.65	68.30	75.93	60.14
SegNet [51]	33.38M	72.25	68.77	**80.06**	60.01
ResUNet++ [96]	52.61M	70.50	75.80	67.10	54.80
ENC-Net [97]	16.80M	72.66	68.59	79.90	57.70
Proposed DAU-Net	-	**74.23**	**73.81**	74.59	**65.32**
Proposed $EU^2 - Net$	**5.30M**	**74.73**	**81.35**	75.40	**64.48**

To provide specific performance insights, the proposed models delivers a Dice score of 74.23% and 74.73% in DAU-Net and $EU^2 - Net$ respectively, indicating a significant similarity between the ground truth and the predicted segmentation masks. Furthermore, $EU^2 - Net$ yields very good precision, with a score of 81.35%, showing the correct identification of a significant portion of predicted foreground pixels. The recall value of 75.40% underscores the model's capability to precisely identify a significant number of actual foreground pixels. Additionally, the IoU value of 64.48% and 65.32% in

$EU^2 - Net$ and DAU-Net resepectively signifies the model's robust ability to accurately outline regions of interest.

Notably, the proposed $EU^2 - Net$ method achieves the highest Dice score among all models listed in Table 3.5, underscoring its excellence in terms of segmentation accuracy and alignment with the ground truth. It also demonstrates a higher precision score, thereby showing a critical aspect of correct predictions.

TABLE 3.5: Comparison of results with SOTA models on the BUSI dataset.

Model	Dice	Precision	Recall	IoU
DA-Net [98]	67.83	-	80.38	-
ResUNet++ [96]	73.85	**80.10**	71.43	60.02
SK-UNET [99]	70.90	-	80.80	-
SCAN [100]	72.00	73.00	-	-
STAN [101]	72.00	76.00	-	-
ColonSegnet [102]	73.53	76.81	76.43	62.71
MCF-Net [103]	71.06	-	72.23	-
UNext [104]	65.94	-	-	55.22
AE-Unet [105]	73.47	74.44	79.00	64.57
RRC-Net [106]	72.53	71.73	77.72	63.60
MBSNet [107]	72.81	-	-	63.21
U-Net-densenet121 [108]	73.70	-	72.55	62.46
Proposed DAU-Net	**74.23**	73.81	74.59	**65.32**
Proposed $EU^2 - Net$	**74.73**	**81.35**	75.40	64.48

3.6 Additional Experiment on the UDIAT dataset

Additionally, both the proposed models are evaluated on UDIAT dataset. For the evaluation on the UDIAT dataset, the consistency is maintained by using the same set of hyperparameters that are employed in the evaluation of the BUSI dataset. The segmentation results of the proposed DAU-Net model on the UDIAT dataset is shown in Fig. 3.12. Segmentation results of the proposed $EU^2 - Net$ model are displayed in Fig. 3.13. Table 3.6 presents a comprehensive overview of the quantitative performance achieved by both the proposed models when compared to previous notable research efforts conducted on this dataset. It is to be noted that $EU^2 - Net$ model remarkably improved the segmentation performance as compared to the other SOTA models on the UDIAT datasets with Dice score of 83.47%.

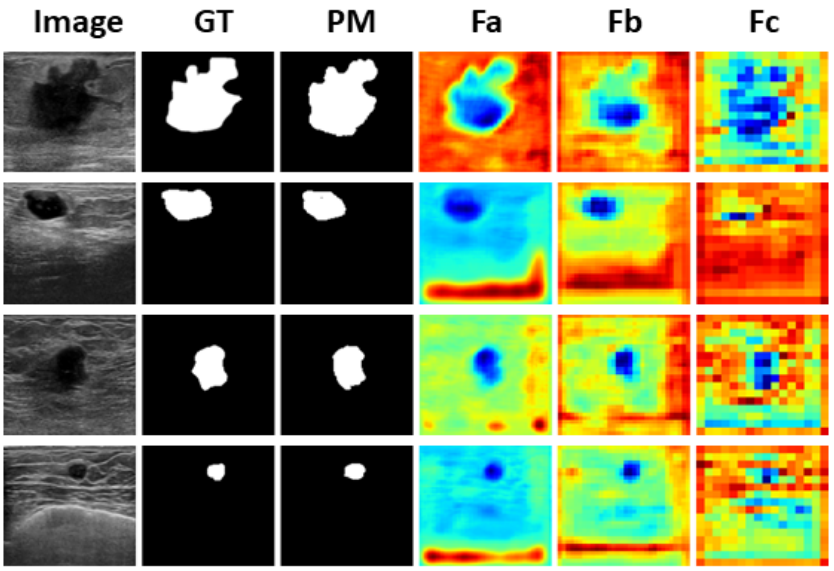


FIGURE 3.12: Predicted mask and heatmap visualization of the proposed DAU-Net model on the UDIAT dataset. GT and PM represent the Ground Truth and Predicted Mask, respectively. F_a, F_b, and F_c are the heatmaps of the features flowing from the first and second encoder layers to the first and second decoder layers via skip connections and the bottleneck layer, respectively.

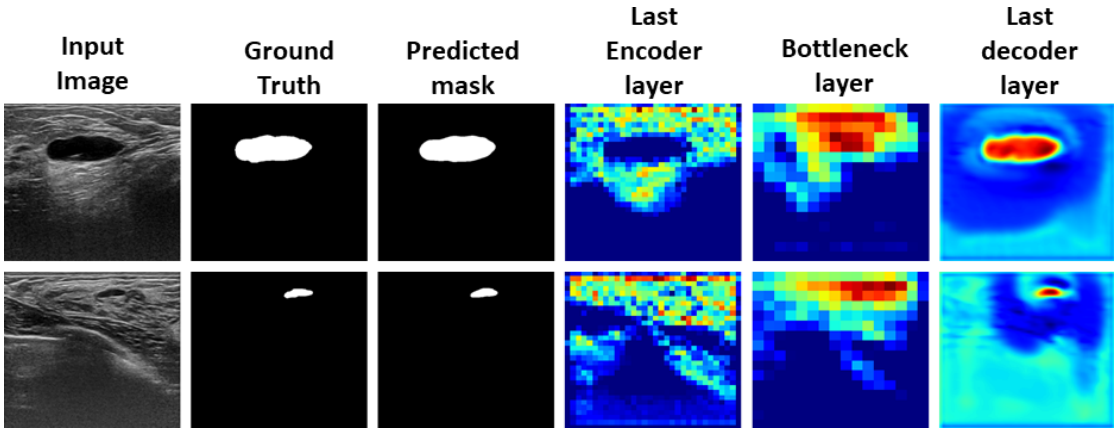


FIGURE 3.13: Some predicted masks of $EU^2 - Net$ on images of the UDIAT dataset. Heatmaps of the last encoder, bottleneck, and last decoder layers.

TABLE 3.6: Comparison of results with SOTA models on the UDIAT dataset.

Model	Dice	Precision	Recall	IoU
UNet [21]	75.00	78.00	-	65.00
SegNet [51]	70.80	85.00	60.00	
CE-Net [109]	72.00	74.00	61.00	
MultiResUNet [110]	75.00	79.00	66.00	
SCAN [100]	74.00	-	75.00	65.00
STAN [101]	78.20	-	80.10	69.50
Proposed DAU-Net	**78.58**	**85.85**	82.25	64.71
Proposed $EU^2 - Net$	**83.47**	**89.28**	**86.30**	**72.11**

3.7 Error Analysis

The proposed DAU-Net model has demonstrated excellent performance across various image segmentation tasks, outperforming SOTA models, as depicted in Table 3.4 and Table 3.5. It is essential to highlight that the precision and recall are relatively lower, indicating instances where non-tumorous regions are misclassified as tumorous and vice-versa. It is important to acknowledge the complexity of the dataset used for evaluation, which presents challenges in achieving perfect segmentation results. Fig. 3.14 illustrates specific cases where this model encounters difficulties, resulting in deviations from the ground truth segmentation. These challenges may arise from dataset complexity, variations in image quality, or the presence of ambiguous features that are hard to accurately delineate.

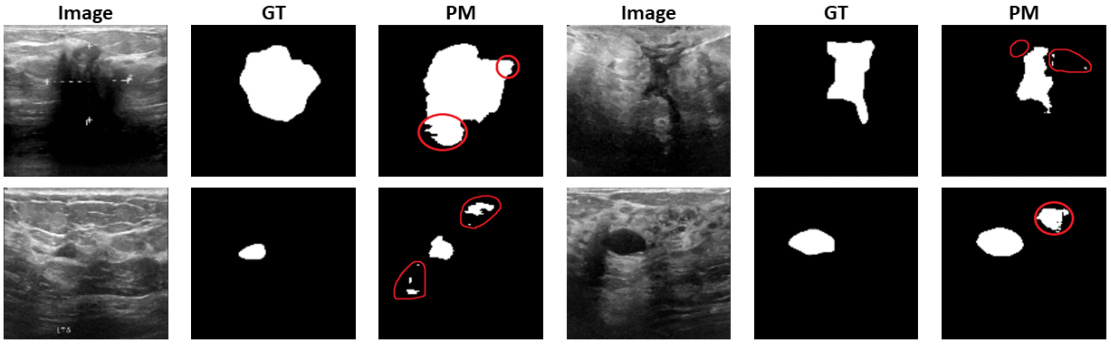


FIGURE 3.14: Illustration of some of the failed cases of DAU-Net model. The encircled regions are the misclassified segmented masks. GT and PM represent the Ground Truth and Predicted Mask, respectively.

From Fig 3.15, it can be seen that $EU^2 - Net$ fails to accurately resemble the ground truth mask in some cases despite achieving SOTA results. Lighter shades and darker complex background regions are possible reasons behind the model's wrong separation of background and foreground pixels.

3.8 Model 3: Attention-aided Wavelet Guided U-Net (AW-GUNet)

The proposed segmentation model combines the structure of U-Net with the feature-rich DenseNet-121 as a backbone network to learn hierarchical features from input data. The WGCAM module captures edge information for spatial guidance, followed by a

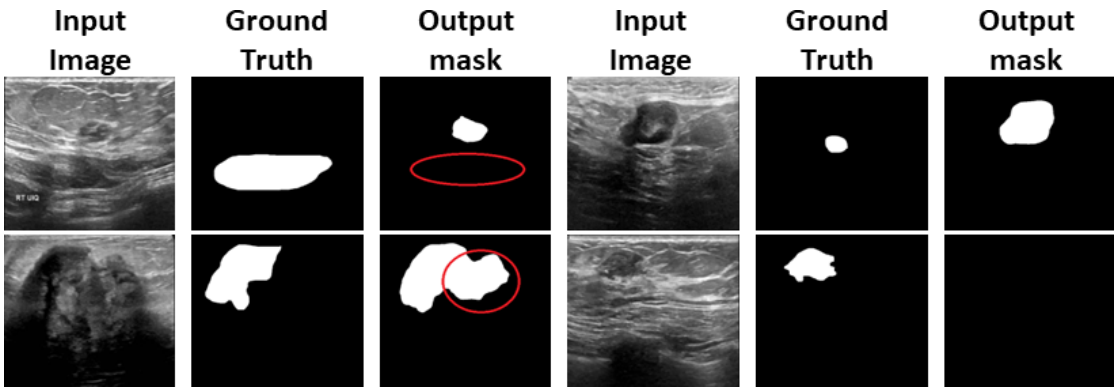


FIGURE 3.15: Error cases of $EU^2 - Net$ on the BUSI dataset. Red-circled regions show where our model fails to segment properly.

lw-GAP module for providing attention to specific channels. The decoder module comprises the upsample block and the convolution block. The upsampling block merges locally extracted features from the transposed convolution layer with upsampled features generated through Gaussian and Lanczos filters, resulting in a noise-suppressed and anti-aliased upsampled feature map. This upsampled feature map, achieved by concatenating the attention-guided features, serves as input for the convolution block. In this block, features from multiple receptive fields are extracted using convolutional layers with kernels of different sizes. The pipeline of the proposed model is shown in Fig. 3.16.

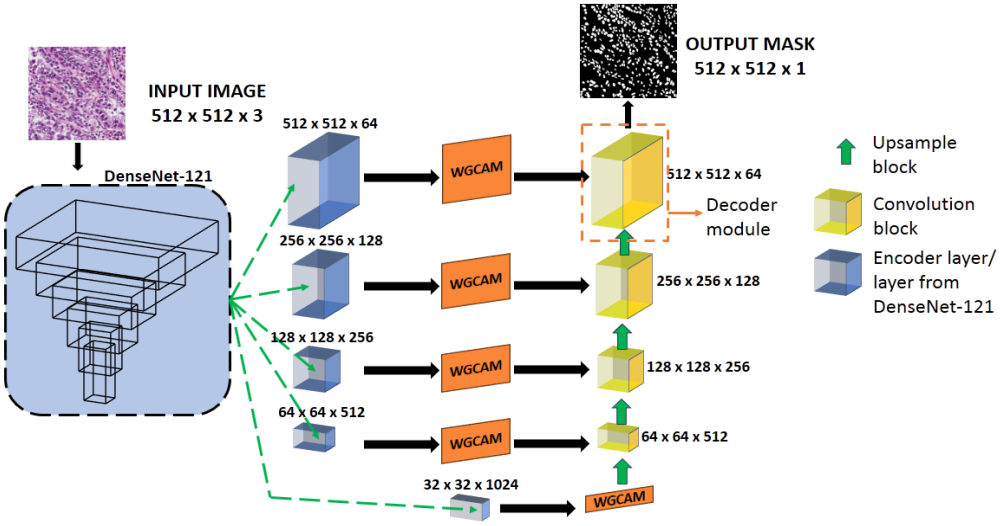


FIGURE 3.16: The proposed segmentation model (AWGUNet) reinforces decoder features with edge information-enhanced features using WGCAM.

3.8.1 Wavelet-guided Channel Attention Module

The WGCAM captures the edge information using wavelet transform [111]. A Haar wavelet is applied to the input feature, F_{inp} of dimensions $H \times W \times C$ to decompose it to form F_{wav} of dimensions $H/2 \times W/2 \times 4C$. F_{wav} is operated with a transposed convolution to modify its dimensions to $H \times W \times 4C$. A separable convolution layer is then utilized to generate the attention weights F'_{wav} of dimensions $H \times W \times C$ as shown in Eq. 3.28, where f_{sc} and f_{ct} are the separable convolution and transposed convolution, respectively.

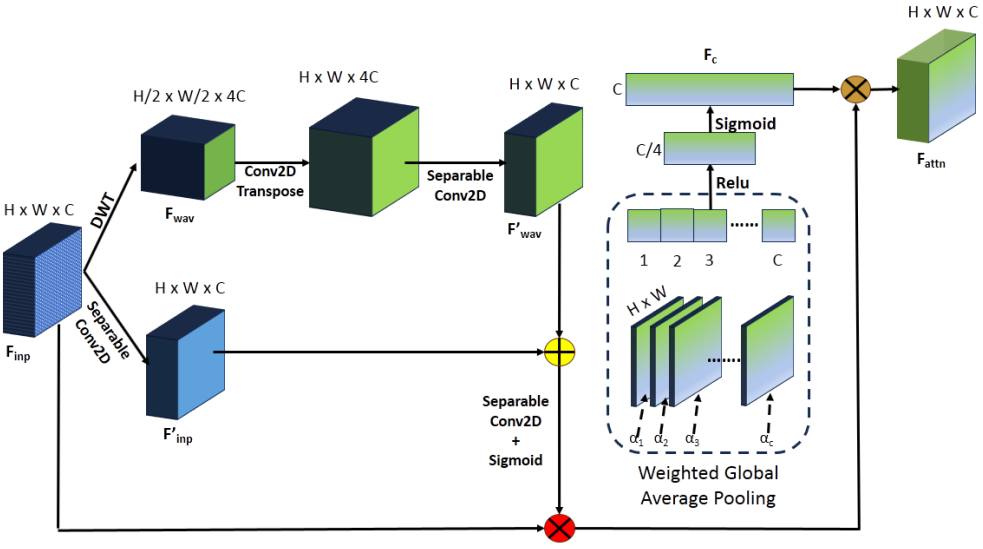


FIGURE 3.17: Wavelet-guided Channel Attention Module

$$F'_{wav} = f_{sc}^{1\times1}(f_{ct}^{1\times1}(DWT_{haar}(F_{inp}))) \tag{3.28}$$

F'_{wav} is then added with F'_{inp} (generated by applying separable convolution on F_{inp}). A separable convolution layer followed by a sigmoid activation function is utilized to generate the attention-aided feature. This attention-aided feature is then element-wise multiplied with F_{inp} followed by a channel-wise multiplication with attention weights F_c across the channel dimension as shown in Eq. 3.29 where $\times$ is element-wise multiplication and $\otimes$ is the multiplication across the channel dimension.

$$F'_{wav} = F_c \otimes (Sigmoid(f_{sc}^{1\times1}(F'_{inp} + F'_{wav})) \times F_{inp}) \tag{3.29}$$

F_c is obtained by a learnable weighted Global Average Pooling (lw-GAP) where learnable weights α_1, α_2,... α_C are assigned to each channel. This flattened 1D tensor of dimension C then passes through two dense layers to generate F_c as shown in Fig. 3.17.

3.8.2 The Decoder Module

In digital images, noise and aliasing can be problematic, especially in histopathological images, given their complex staining patterns and extremely small regions of interest. To address these challenges, the Gaussian filter [112] is applied to suppress noise and the Lanczos filter [113] with a 5×5 kernel size is used for its anti-aliasing properties in the upsampling layers. The upsampled features are combined to create F_{up}, and the feature F_{ct} from the transposed convolutional layer offers detailed local information to compensate for any data loss during interpolation in the upsampling layer. The final feature F_{up-ct} is generated by fusing features from both techniques, as illustrated in Eq. 3.30, where $\oplus$ denotes concatenation.

$$F_{up-ct} = (f_c^{1\times1}(F_{up}) \oplus F_{ct}) \tag{3.30}$$

The convolution block takes the input feature map F_{up-ct} and extracts information from three different receptive fields. It uses convolutional layers with kernel sizes of 5×5 (F_5), 3×3 (F_3), and 1×1 (F_1). These features (F_5, F_3, and F_1) are normalized using Instance Normalization (IN) [114] to account for variations among images in a dataset, followed by a ReLU activation. The normalized features are then concatenated and passed through convolutional layers, instance normalization, and ReLU activation to produce the decoded feature map F_{dec} as shown in Fig. 3.18.

3.9 Results and Analysis

3.9.1 Dataset Description

The proposed AWGUNet model is evaluated on two datasets: MoNuSeg [115] and TNBC [79]. MonNSeg contains 512×512 Hematoxylin and Eosin-stained tissue images, with 30 training images (22,000 annotations) and 14 test images (7000 annotations).

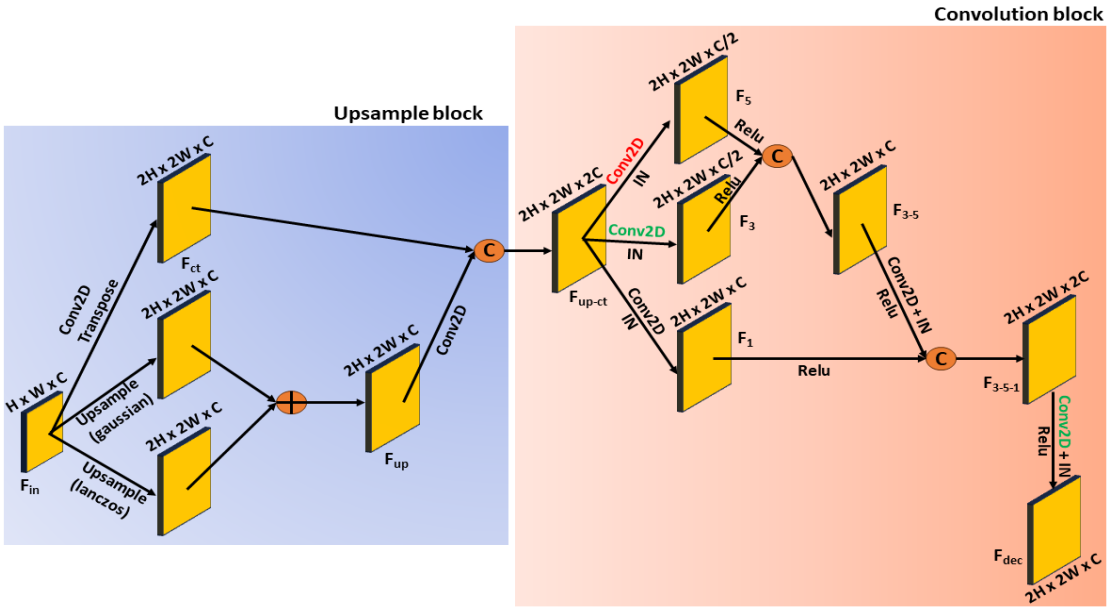


FIGURE 3.18: A block diagram of the Decoder module of the AWGUNet model. The two key components are the upsample and convolution blocks

The TNBC dataset focuses on Triple-Negative Breast Cancer tissues, with 50 images (4022 annotated cells). The original image size (512×512) was used as input for both the datasets.

3.9.2 Experimental Setup

The model is trained using a 70-20-10 % train-validation-test split, using a learning rate of 0.0001, The dice loss and BCE loss are used for training and evaluated using dice score, IoU, precision and recall as standard metrics described in earlier section.

3.9.3 Ablation study

To figure out the optimal setup and parameters for AWGUNet model, an extensive ablation study is performed on the MoNuSeg dataset. The experiments are listed below:

(i) U-Net with DenseNet-121 as the backbone

(ii) (i) + WGCAM with GAP

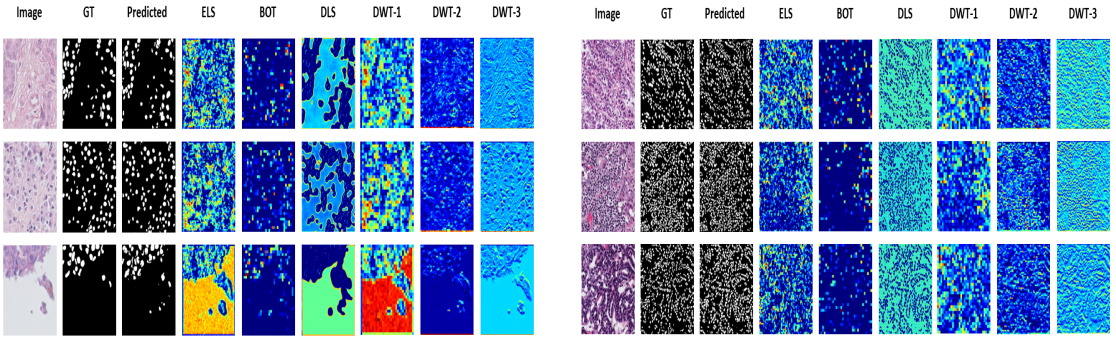
(iii) (i) + lw-WGCAM with GAP

(iv) The proposed model: (iii) + the decoder module

Table 3.7 highlights the substantial impact of WGCAM attention along with the decoder and the upsample module on the performance enhancement of U-Net with DenseNet-121 as the backbone. The denoising and anti-aliasing capacity of the upsample layers proves instrumental in tackling the noisy and complex staining patterns of the images. Also, arbitrary shapes are easily detected with the help of edge information obtained from the wavelet features of the WGCAM. Fig. 3.19 shows the segmentation results and the impact of each module of the proposed model on both datasets.

TABLE 3.7: Performance of the AWGUNet model on MoNuSeg dataset. All values are in %. Bold values indicate superior performance.

Model	Dice	Precision	Recall	IoU
(i)	77.39	72.56	83.17	63.17
(ii)	78.77	73.67	84.89	65.03
(iii)	78.89	75.75	82.62	65.20
(iv)	**79.46**	**76.26**	**84.91**	**66.57**



(A) TNBC (B) MoNuSeg

FIGURE 3.19: Segmentation results of AWGUNet model on MoNuSeg and TNBC datasets. GT represents ground truth, ELS, BOT, and DLS are the Encoder Last Layer, BOTtleneck, and Decoder Last Layer, respectively. DWT-1, DWT-2, and DWT-3 are the wavelet features of WGCAM between the second, third and fourth encoder and decoder layers, respectively.

3.9.4 SOTA Comparison

The proposed model AWGUNet is compared with SOTA methods on both the datasets and tabulate the results in Table 3.8.

TABLE 3.8: Performance comparison of the AWGUNet model with SOTA methods. All values are in %. Bold values indicate superior performance.

Model	**MoNuSeg**		**TNBC**	
	Dice	**IoU**	**Dice**	**IoU**
U-Net[21]	74.67	60.89	68.61	52.92
Attention U-Net[53]	78.67	66.51	71.43	54.21
DIST[79]	77.31	63.77	70.51	56.34
MMPSO-S [116]	72.00	56.00	65.00	49.00
Deep-Fuzz [117]	79.10	66.10	77.80	64.20
AWGUNet	**79.46**	**66.57**	**81.65**	**69.18**

3.10 Discussion

In this chapter, three novel attention-aided U-Net model based segmentation frameworks tailored for breast tumor detection in ultrasound and histopathological imaging have been introduced and evaluated. Each proposed methodology is designed to address key limitations in existing models, such as poor boundary delineation, insufficient contextual awareness, and limited generalizability, while also leveraging advanced DL and attention mechanisms to improve diagnostic performance.

In the first proposed model, PCBAM attention-based Residual U-Net, is developed for the segmentation of breast tumors in ultrasound images. This architecture effectively integrates attention modules to enhance focus on tumor-relevant regions and demonstrates robust performance in delineating complex tumor boundaries with dice score of 74.23% on the BUSI dataset and 78.58% on the UDIAT dataset.

After further exploration, in the second approach, $EU^2 - Net$, is introduced an internally ensembled segmentation model that cleverly leverages the advantages of ensemble learning without increasing the model parameters. By incorporating the A^2TF^2 module and learnable weights for score aggregation, the architecture significantly improves segmentation accuracy, surpassing several SOTA models with dice score of 74.73% on the BUSI dataset and 83.47% on the UDIAT dataset. This architecture highlights the potential of internal ensembling strategies to improve model performance efficiently.

The third contribution of this chapter focuses on histopathological image segmentation, introducing a hybrid architecture composed of three modules: WGCAM, lw-GAP, and a robust decoder. WGCAM effectively captures fine edge details crucial for accurate boundary delineation, while lw-GAP introduces channel-specific attention to refine the

feature representation. The decoder module employs a novel upsampling strategy combining Gaussian and Lanczos filters, helping to mitigate aliasing artifacts and improving the quality of the upsampled segmentation maps. The results of this architecture affirm the value of integrating boundary-aware enhancement and attention mechanisms in achieving high-precision segmentation, particularly in high-resolution histopathology images where cellular structure differentiation is essential.

Thus, the methodologies discussed in this chapter collectively demonstrate that integrating attention mechanisms, boundary-aware enhancement, and ensemble strategies into U-Net-based architectures can significantly improve tumor segmentation accuracy in ultrasound and histopathology domains. Each approach contributes uniquely to solving challenges related to boundary accuracy, model interpretability, and performance efficiency.

Chapter 4

Novel Breast Cancer Classification Techniques

Breast cancer remains one of the most prevalent and life-threatening malignancies among women worldwide, necessitating accurate and timely diagnosis for effective treatment planning. Medical imaging plays a pivotal role in early detection, with modalities such as ultrasound, histopathology, and mammography commonly employed for screening and diagnosis. However, interpreting these images manually is often labor-intensive, subjective, and prone to variability across observers. As a result, the adoption of DL techniques in CAD systems has gained significant momentum in recent years, aiming to enhance diagnostic accuracy and efficiency. This chapter presents a collection of DL-based classification frameworks developed and validated across different imaging modalities for breast cancer classification. The approaches utilize advanced feature extraction and nature-inspired optimization algorithms to boost performance while reducing model complexity.

To be specific, the first method [32] proposes a novel unsupervised optimization technique. By incorporating rank-based statistical measures into the WOA, the framework identifies the most discriminative features without requiring class labels, achieving excellent results on the BUSI dataset.

While the second method [33] presents an efficient hybrid framework that combines TL with a meta-heuristic optimization-based FS approach. Specifically, SqueezeNet 1.1, pre-trained on ImageNet, is fine-tuned on breast thermograms from the DMR-IR dataset for deep feature extraction. To reduce computational overhead, a chaotic GA initialized

using a sinusoidal chaotic map is integrated with GWO to select an optimal subset of features. The proposed method achieved 100% classification accuracy using only 3% of the extracted features, highlighting its effectiveness in delivering high accuracy with reduced computational cost in thermal imaging-based breast cancer diagnosis.

The third approach [34] targets mammographic data using an attention-enhanced VGG16 model for feature extraction. The SSD meta-heuristic algorithm is employed to select an optimal subset of features, with an integrated local search strategy to enhance its exploitation capability. This model achieves high classification accuracy on the DDSM dataset while using only 25% of the extracted features, demonstrating efficiency and robustness.

Finally, the fourth approach is the "Gamma Function based Ensemble of CNN Models for Breast Cancer Detection" [35] which integrates the predictions from three pre-trained CNN models—GoogleNet, VGG11, and MobileNetV3_Small—using a rank-based ensemble method. This method has demonstrated significant improvements in classification accuracy on the BreakHis dataset at various magnification levels, and on the ICIAR-2018 dataset.

Together, these methodologies illustrate a comprehensive exploration of DL and FS techniques tailored for breast cancer classification across ultrasound, thermography, mammography and histopathology images. The remainder of this chapter discusses the methodology, datasets used, experimental results, and analysis of the results in detail for each proposed approach.

Objectives of the Chapter

- To design and evaluate DL-based classification models for breast cancer detection across multiple imaging modalities, including ultrasound, thermography, mammography and histopathology.
- To explore and implement advanced feature extraction strategies using pre-trained CNN architectures enhanced with TL and attention mechanisms.
- To develop and demonstrate efficient FS techniques using meta-heuristic optimization algorithms (e.g., Whale Optimization Algorithm, Social Ski-Driver, Grey Wolf Optimizer, and Genetic Algorithm) in combination with DL frameworks, aiming

to reduce dimensionality, improve classification accuracy, and minimize computational overhead.

- To explore the effectiveness of ensemble learning in enhancing classification performance, particularly on breast histopathology images.
- To benchmark the proposed models on publicly available breast cancer datasets and compare their performance with SOTA methods.

4.1 Methodology 1: U-WOA for Breast Cancer Classification

In this research work, a method is proposed for categorizing breast cancer from breast ultrasound images that basically involves three steps. At first, the deep feature extraction step that involves extracting deep features from the ultrasound images used as input. Next, the FS step that involves choosing the most important and useful features to create the best feature subset. Lastly, the classification step that involves ML-based classifiers that have been trained on the test dataset to make predictions. Fig. 4.1 shows the entire pipeline of the proposed approach.

4.1.1 Deep Feature Extraction

TL model, pre-trained on the ImageNet database is used as the base model, for feature extraction and fine-tune it using the relevant dataset. To do this, the last layer of the base TL model is removed and replace it with a GAP layer followed by a dense layer with 256 units and a ReLU. Finally, the output layer is modified with a softmax activation function to match the categories of the target dataset, in this case the BUSI database. The fine-tuned TL model structure for deep feature extraction is illustrated in Fig. 4.2.

4.1.2 Feature Selection using U-WOA

The extracted deep features are fed into the U-WOA model which is the unsupervised version of the WOA. WOA by Mirjalili & Lewis [118] is a meta-heuristic optimization algorithm. The WOA meta-heuristic optimization algorithm imitates the humpback

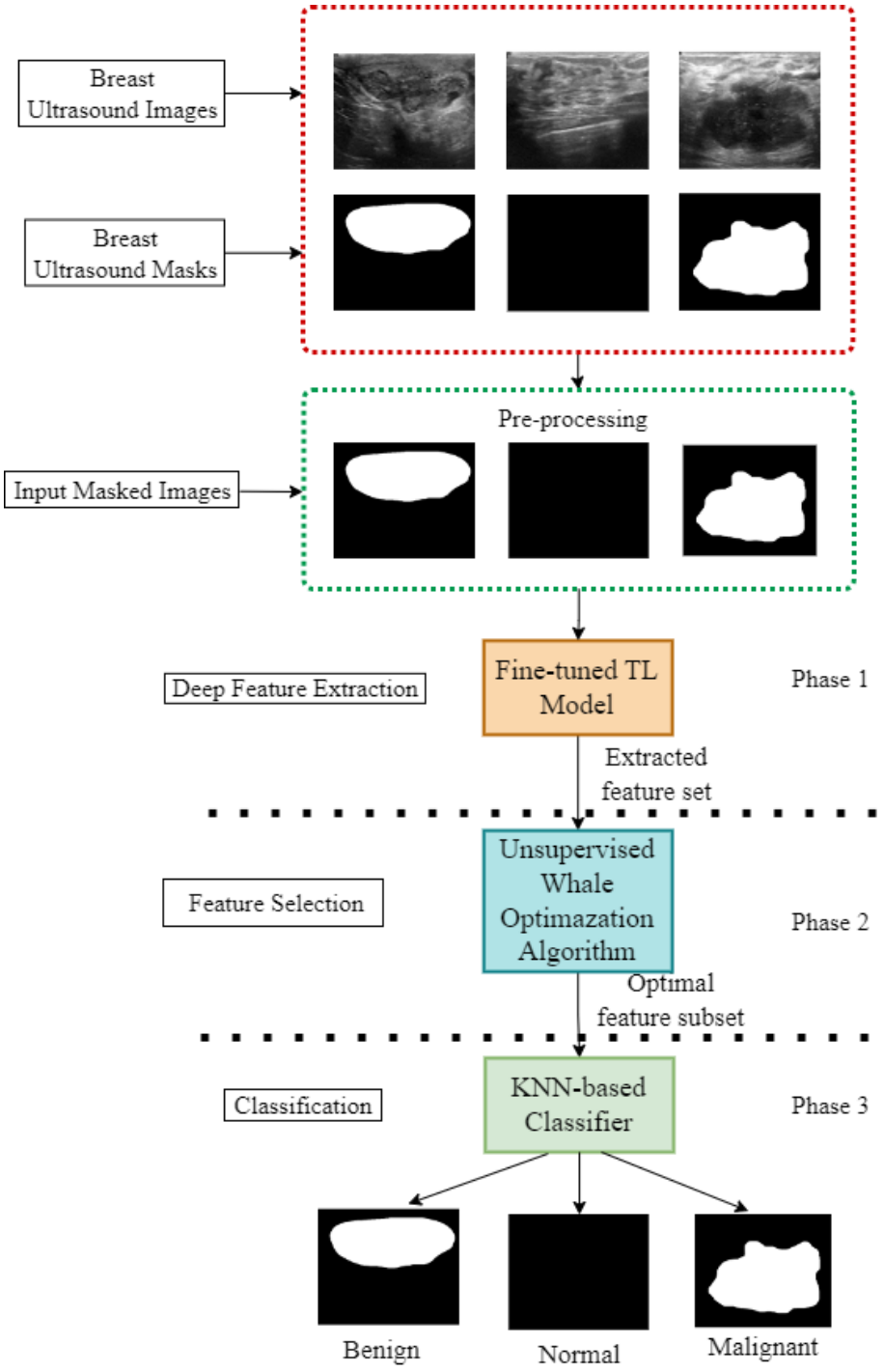


FIGURE 4.1: Pipeline of the proposed U-WOA method for tumor classification in ultrasound images.

whales' bubble-net assault strategy, in which they plunge into the water and wrap a bubble-net around their victim. Three primary processes make up the WOA algorithm: encircling the prey, a bubble-net attacking phase (exploitation), and hunting for prey (exploration).

Encircling prey: In the initial stage, the most suitable search agent is selected with the

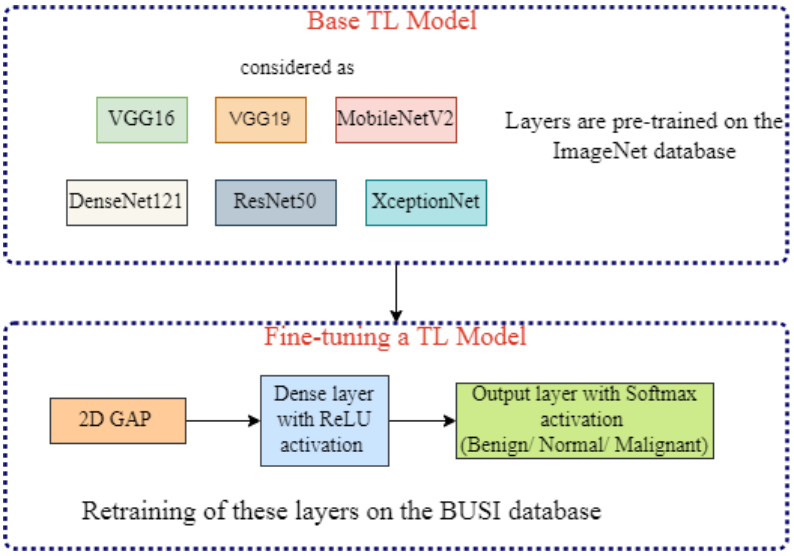


FIGURE 4.2: The TL framework used to extract deep features from the BUSI database consists of six different base TL models: VGG16, VGG19, ResNet50, MobileNetV2, DenseNet121, and XceptionNet. Each of these models is fine-tuned for the purpose of feature extraction.

help of a fitness function and then the distance between the top-performing (best) search agent and the remaining search agents is updated. The current prey is considered to be closer to the global optimum. The methods of spiral updating location and shrinking encircling have a 50% probability of success in the Bubble-net assaulting stage depending on the settings of specific restrictions. In the exploration phase known as "searching for prey", the search agent is free to conduct a haphazard search for prey without using the spiral updating location.

Eq. 4.1& 4.2 are used to update the position of the agent$\vec{P}$ where $\vec{P'}$ is the best agent. The coefficients E & F are evaluated as in Eq. 4.3 & 4.4 respectively.

$$M = |F \cdot \vec{P'}(t) - \vec{P}(t)| \tag{4.1}$$

$$\vec{P}(t+1) = \vec{P'}(t) - \vec{E} * \vec{M} \tag{4.2}$$

$$E = 2 * m * x - m \tag{4.3}$$

$$F = 2 * x \tag{4.4}$$

m drops linearly from 2 to 0 over iterations, where x is an arbitrary variable lying between 0 to 1. The location of the best solution P' is updated after each cycle if there is a better option or the search agent travels beyond the space of all possible physical

solutions. Random variable x helps the search agent to choose the ideal search agent and carry out surrounding the prey.

Bubble-net attacking method (exploitation phase): To model the spiral attacking mechanism, first, the distance is calculated as per Eq. 4.5 and further position is updated as in Eq. 4.6. Here, the form of the logarithmic spiral is determined by the constant b, and the value of *prob* is chosen at random from $(0, 1)$.

$$M' = |\vec{P'}(t) - \vec{P}(t)| \tag{4.5}$$

$$\vec{P}(t+1) = \begin{cases} \vec{P'}(t) - E \cdot M & \text{if } prob \leq 0.5 \\ \vec{M'} * e^{\text{bl}} * \cos(2\pi \text{l}) + \vec{P'} & \text{if } prob > 0.5 \end{cases} \tag{4.6}$$

Search for prey (exploration phase): A randomized search is carried out by the agent in the search space during the algorithm's exploration phase. It serves to draw attention to the random search depending on the proximity of the agents. The strategy for encircling is based on the coefficient vector A's randomized values. To force the search agent to move away from the top ranked whale, values of E in the range[-1,1] is used. For exploration values for which $|E| > 1$ is used. Here, the term "P" denotes a randomly selected agent from the population.

$$\vec{M} = |F \cdot \vec{P''}(t) - \vec{P}(t)| \tag{4.7}$$

$$\vec{P}(t+1) = \vec{P''}(t) - \vec{E} \cdot \vec{M} \tag{4.8}$$

Being a population based algorithm, WOA heavily depends on the best evaluated position ($\vec{P'}$). This position is updated with each iteration and acts as a guiding factor for the other agents, thereby making the entire process expensive to experiment with. Additionally, these approaches focus on a particular agent or a set of features. Notably, this approach also misses out on valuing each of the features. Taking this into account, each of the features are examined and assign importance scores to them using three popular filter-based feature ranking methods. To be specific, the Pearson's correlation coefficient (PCC) [119], Spearman's Correlation Coefficient (SCC) [120] and ReliefF [121] is considered for this. In each iteration, a pseudo-fitness value is assigned to each particle based on these importance scores, while the original fitness function used to rank the agents and evaluate the agents for the target solution remains unchanged. The objective

of this pseudo fitness value is only to replace the leader guiding agent (P').
The pseudo fitness function is constructed using Eq. 4.9.

$$Pseudo\_Fitness \ = \ \sum_{\forall i} S(\sigma(x_i)) \cdot (PCC_i + SCC_i + ReliefF_i) \tag{4.9}$$

In Eq. 4.9, i refers to the i^{th} feature. The feature importance scores for each feature are recorded. By doing this, it is ensured that the proposed solution is guided to select the best set of features obtained so far while at the same time exploring even better solutions through the steps of WOA.

4.1.3 Transfer Function

WOA was originally developed for numerical optimization problems. WOA requires a step to discretize or binarize the continuous data. The optimized values are normalized to $(0, 1)$ using a transfer function. This function is typically referred to as the S−shaped transfer function. In accordance with Eq. 4.11, binarization of continuous variables is performed, where $rand$ is any randomized number within the $\sigma(x)$ range.

$$\sigma(x) = \frac{1}{1 + e^{-\mathrm{x}}} \tag{4.10}$$

$$S(\sigma(x)) = \begin{matrix} 1 \text{ if } \sigma(x) < rand \\ 0 \text{ if } \sigma(x) \geq rand \end{matrix} \tag{4.11}$$

4.1.4 Fitness Function

Using Eq.4.12, the fitness value is calculated to evaluate candidate solutions. Here, α is a hyperparameter. The ratio between the total number of left-out features and the total number of features is represented by f and the classification accuracy is represented by a.

$$Fitness = \alpha * a + (1 - \alpha) * f \tag{4.12}$$

4.1.5 Classification

In this study, a k-Nearest Neighbor (kNN)-based classifier is employed to predict values from the test dataset. The optimal feature set is divided (acquired after feature selection step) into 75 percent training, 10 percent validation, and 15 percent test sets. The training and validation sets are then used as input to the kNN classifier, and then performs prediction on the test set.

4.2 Experimental Results and Analysis

4.2.1 Dataset Description

To train and evaluate the proposed U-WOA model, the BUSI ultrasound dataset is used which is previously described in Chapter 3, Section 3.2.1. All the classes including normal, benign and malignant cases are considered in the present study.

4.2.2 Performance metrics

To evaluate the performance of the U-WOA model the following metrics are used: accuracy, precision, recall and F1-score. Accuracy, precision and recall metrics are defined earlier in Chapter 3, Section 3.2.2. However, in the context of classification task, the proportion of accurately predicted labels for a class to the total samples in the class is measured as precision. Recall is defined as the proportion of samples that are actually positive to all samples that are positively classified in that class. The weighted harmonic mean of the precision and recall values is known as the F1-score. Precision and recall are treated equally. It is determined as:

$$F1 - score = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{4.13}$$

4.2.3 Hyper-parameters for TL models

Initially the grid search method with some common values in practice is used to find the best possible combination of initial learning rate and batch size for training the deep CNN models. In this method, the best possible combination of batch size (BS) and

initial learning rate (L_0) as 64 and 0.001 is considered respectively. The cross entropy loss and the commonly used Adam optimizer is used to optimize internal weights. The learning rate is lowered after the third epoch by a factor of two using a step learning rate scheduler to ensure smooth learning.

4.2.4 Results

To extract deep features, an experiment with different TL models is done and the obtained results are recorded in Fig. 4.3.

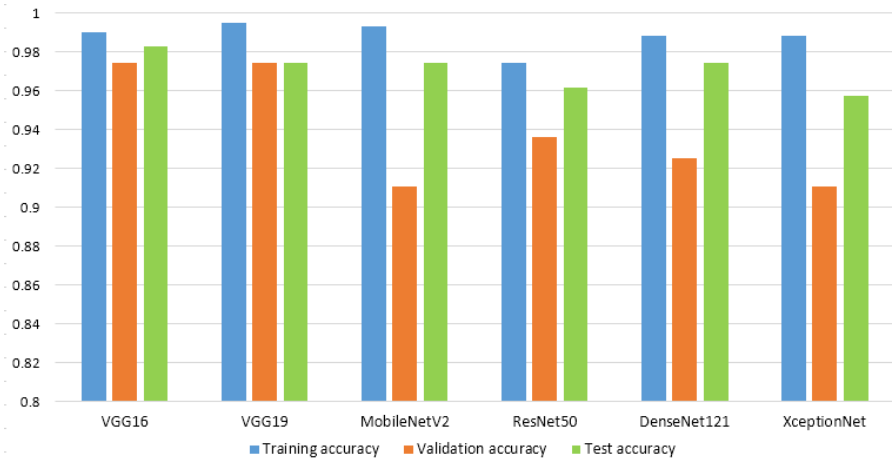


FIGURE 4.3: Performance of different TL models on the BUSI database. All the values are in %.

To train the TL model, the entire dataset is split into 75% training, 10% validation and 15% testing sets. Fig. 4.3 shows that the VGG16 model performs better than other TL models in accuracy over test datasets. So, the deep features extracted by the VGG16 model is considered for the next step. While training, the accuracy and loss curves of the VGG16 model for BUSI dataset are noted and shown in Fig. 4.4 and Fig. 4.5 respectively.

Fig. 4.4 shows the training and validation accuracies for every epoch and it can be seen that the model does not overfit while Fig. 4.5 ensures smooth learning of the model during training. In the loss learning curve the popular categorical cross entropy loss function is used. The confusion matrix of the model is shown in Fig. 4.6. It is clear from this figure that the number of misclassified samples is very less and the model accurately predicts all three classes.

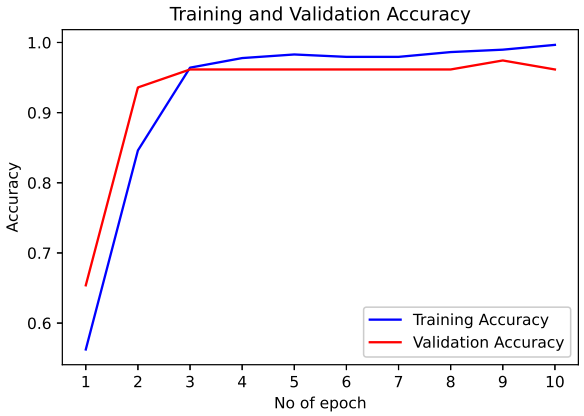


FIGURE 4.4: Learning curve (Accuracy) of the feature extraction model during training of the VGG16 model.

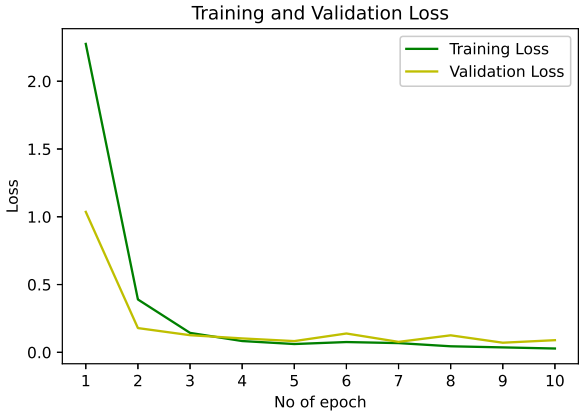


FIGURE 4.5: Learning curve (Loss) of the feature extraction model during training of the VGG16 model.

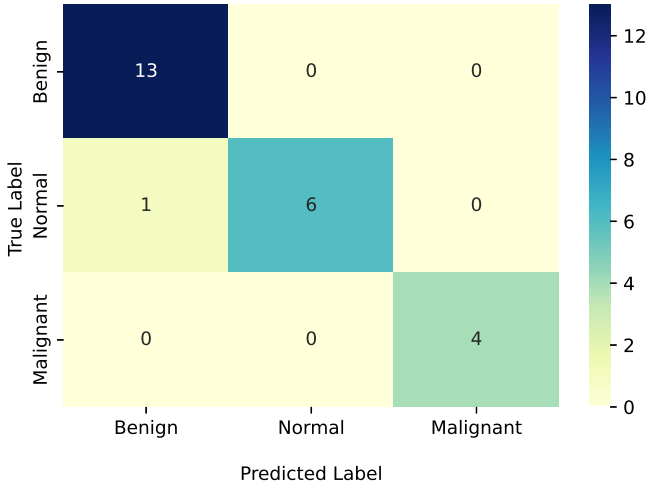


FIGURE 4.6: Confusion matrix of the proposed U-WOA method.

Lastly, the proposed method is compared with previously established methods and present the comparison in Table 4.1. The comparison shows that the proposed method surpasses all the methods considered in this study.

TABLE 4.1: The performance of the proposed U-WOA model is compared with previous methods on the BUSI database, with the highest values indicated in bold font.

Method	**Accuracy**	**Recall**	**Precision**	**F1-score**
Sadad et al. [122]	0.966	0.943	-	-
Moon et al. [123]	0.946	0.923	0.900	0.911
Mishra et al. [124]	0.974	0.96	0.958	0.959
Pramanik et al. [30]	0.987	0.987	0.987	0.987
Proposed method (U-WOA)	**0.991**	**0.991**	**0.991**	**0.991**

One strong possible reason for the proposed method outperforming the existing methods can be linked to the fact that the optimization algorithm selects the most optimal set of features on the basis of different types of statistical importance of features which was ignored in [30]. Specifically, the works reported in [122–124] consider several shape and texture based feature extractors like GLCM to extract features and further use an ML based classification algorithm to classify the same. The fundamental disadvantage of utilizing such feature extractors is that they are not invariant to rotations and scaling-based transformations.

4.3 Methodology 2: GA and GWO based Hybrid Framework for Breast Cancer Classification in Thermograms

In this method, a deep feature selection method is proposed using a hybrid of GA and GWO for identifying breast cancer in thermograms, which does not sacrifice accuracy and is light in terms of storage requirement and computation time. A TL model (pre-trained on the ImageNet dataset) namely SqueezeNet 1.1 is used for feature extraction which is very light in terms of the number of parameters, thus leading to low processing time, storage, and computational efforts. A 1000 dimensional feature vector is extracted from the mentioned model with fine-tuning on the target thermogram dataset. Further, these features are fed into the proposed hybrid FS model, which selects distinctive features and discards the redundant features. A reduced feature set is used as input

to the kNN classifier to get the final classification results. The overall pipeline of the proposed model is presented in Fig. 4.7.

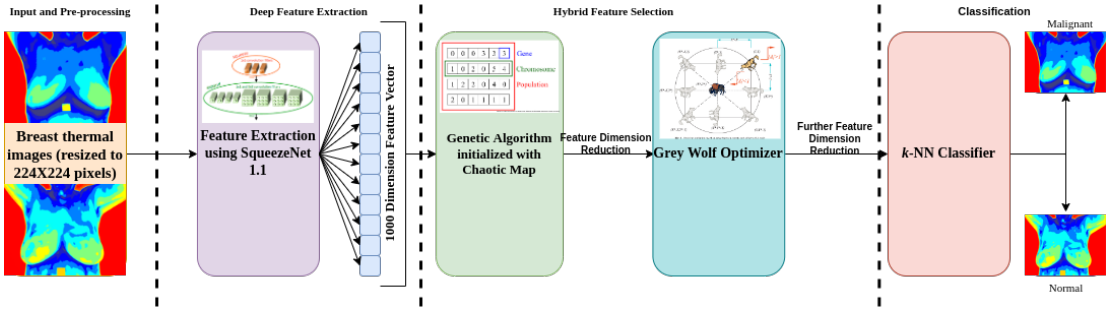


FIGURE 4.7: Pipeline diagram of the proposed Methodology 2 for breast cancer classification in thermograms.

4.3.1 Deep Feature Extraction

A deep CNN architecture proposed by [125], called the SqueezeNet, trains 50X lower parameters than AlexNet and also maintains the accuracy of AlexNet on the ImageNet database. SqueezeNet can be compressed to less than 0.5MB space, which in this case is 512X smaller than the AlexNet. SqueezeNet model contains a convolutional layer, followed by 8 fire modules, a final convolutional layer, and an adaptive average pool layer as shown in Fig. 4.8.

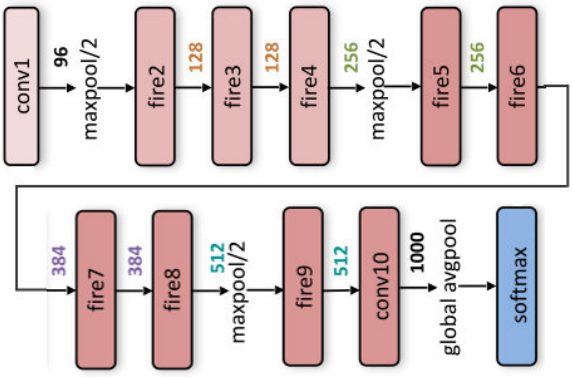


FIGURE 4.8: Diagram of the SqueezeNet architecture.

A fire module consists of a squeeze convolutional layer (with only 1×1 filters) that feeds and expands the layer which is a mix of 3×3 and 1×1 convolutional filters.

In this work, SqueezeNet 1.1 is used for feature extraction which has 2.4X less computation than the original SqueezeNet, and also it has slightly fewer parameters to train.

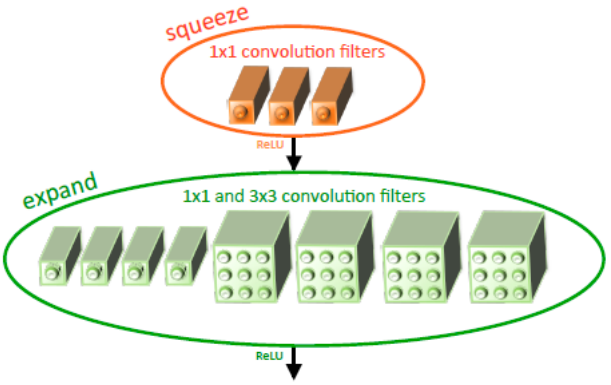


FIGURE 4.9: Fire module of the SqueezeNet model with parameters s_1×1=3, e_1×1=4 and e_3 × 3=4.

4.3.2 Feature Selection using Chaotic GA and GWO Algorithm

Suggested way back in 1975 [126], GA employs the concept of reproduction of the fittest set of features. In GA, candidate solutions—called chromosomes—are binary strings where each bit represents a feature: '1' indicates selection and '0' indicates exclusion. The population comprises multiple such chromosomes, each evaluated using a predefined fitness function. Based on fitness scores, chromosomes are ranked and selected using the Roulette wheel method, which favors higher-fitness individuals. Selected pairs undergo crossover and mutation to explore new feature subsets, reducing the risk of local optima. This evolutionary process iteratively continues until a termination condition is met. As GA has a tendency to get stuck in local optima, and the diversity of the search space is random, which may ultimately result in some unexplored regions. In the present work, a search space is considered as determined by a chaotic map.

Chaos is a special characteristic for non-linear dynamic systems that manifests pseudo random, bounded dynamic, unstable, control parameters and non-period behavior depending on the initial value as described by [127]. Sir Robert May in 1974 [128] had proposed Eq. 4.14 which is known as logistic maps.

$$X_{n+1} = \mu X_n(1 - X_n) \tag{4.14}$$

To be more specific, chaos initialized GA is used, where the initial values are generated using a logistic map as shown in Eq. 4.14. A n-dimensional feature space represented by X_i (where i is the representation of i^th agent) which can be further extended as

$X_i = (x_i^1, x_i^2 ... x_i^j ... x_i^n)$. Here x_i^j is the representation of a feature for the i^{th} particle in j^{th} dimension. The initialization of features is done in different dimensions under Eq. 4.15. The pseudo-code for the chaotic map initialized GA is given in Algorithm 1.

$$x_i^{j+1} = \mu x_i^j (1 - x_i^j) \tag{4.15}$$

Algorithm 1 GA initialized with a logistic map

1: Chaotically initialize the population of the chromosomes X_i (i=1,2...n) in accordance with Eq. 4.14
2: Evaluate the fitness for each chromosome
3: **while** Termination condition is not satisfied **do**
4: Select chromosomes using Roulette Wheel to undergo genetic operations
5: Perform crossover followed by mutation and generate new set of chromosomes
6: Evaluate fitness of the new set of chromosomes
7: **end while**
8: Return best solution

GWO is a nature inspired meta-heuristic algorithm based on echelon and hunting mechanisms of grey wolves (Canis Lupus), proposed by Mirjalili et al. [129]. In a pack of grey wolves, the leadership ranking is as (in decreasing order): (1) Alpha grey wolves (α), (2) Beta grey wolves (β), (3) Omega grey wolves (ω), and (4) The grey wolves which do not fit into any of these categories are called Delta grey wolves (δ). Grey wolves generally encircle their prey before attacking. In this algorithm two coefficients are used and are calculated as $\overrightarrow{C_1}$ and $\overrightarrow{C_2}$ in Eqs. 4.16&4.17.

$$\overrightarrow{C_1} = 2\overrightarrow{a}.\overrightarrow{r_1} - \overrightarrow{a} \tag{4.16}$$

$$\overrightarrow{C_2} = 2\overrightarrow{r_2} \tag{4.17}$$

In these equations, $\overrightarrow{a}$ is linearly decreased from 2 until 0 over iterations, and this represents the wolves approaching the prey. Also $\overrightarrow{r_1}$ and $\overrightarrow{r_2}$ are two distinct random numbers in the interval [0,1]. The best solution in the search space is considered as α followed by second and third bests to be β and δ respectively. The position of each grey wolf is updated as per Eq. 4.18 where $\overrightarrow{X_1}$, $\overrightarrow{X_2}$ and $\overrightarrow{X_3}$ are the three best solutions and *it* is the current iteration.

$$\overrightarrow{X}(it+1) = \frac{\overrightarrow{X_1}(it) + \overrightarrow{X_2}(it) + \overrightarrow{X_3}(it)}{3} \tag{4.18}$$

The value of $\overrightarrow{C_2}$ fluctuates between [-2a,2a], and it is to be noted that $\overrightarrow{C_2}$ is a random vector which is linearly decreased with $\overrightarrow{a}$ when this value of $\overrightarrow{C_2}$ is less than 1 it implies that the grey wolf is in exploitation mode since the grey wolf is positioned between the current position and optimum solution. Conversely, when $\overrightarrow{C_2}$ is greater than 1, it suggests divergence from the prey, leading to the exploration mode. $\overrightarrow{C_1}$ is a random value between [0,2]. The influence of prey is emphasized if $\overrightarrow{C_1} > 1$ and de-emphasized if $\overrightarrow{C_1} < 1$ with reference to Eq. 4.16. Randomization helps proper exploration and avoids the local optimum. The importance and effectiveness of considering three solutions are described statistically by [130]. The pseudo-code for GWO is mentioned in Algorithm 2.

Algorithm 2 GWO

1: Randomly initialize the population of the grey wolves X_i (i=1,2...n)
2: Initialize a, $\overrightarrow{C_1}$ and $\overrightarrow{C_2}$
3: Calculate the fitness value of every grey wolf
4: Rank the grey wolves according to their fitness and name the best, second best and third best grey wolves as X_α, X_β and X_δ respectively
5: **while** termination condition is not satisfied **do**
6: Update each grey wolf position using Eq. 4.18
7: Update value of $\overrightarrow{C_1}$ and $\overrightarrow{C_2}$ using Eqs. 4.16, 4.17 respectively
8: Evaluate fitness of each grey wolf and update X_α, X_β and X_δ.
9: **end while**
10: Return best solution

In this present work, first, the images are resized to 224×224 pixels from the original 640×480 pixels. Then these are used to train the pre-trained SqueezeNet model (pre-trained on ImageNet) from which a 1000 dimensional feature vector is extracted. Out of these 1000 dimensions, a reduced feature set is obtained using a modified version of GA namely chaos initialized GA. Then this set is again reduced to eliminate less-important and less-distinctive features using the GWO algorithm. Finally, kNN classifier is used to classify the images into one of the two classes—cancerous (Sick) or non-cancerous (healthy). The overall experimentation protocol is given in Algorithm 3.

Algorithm 3 Experimentation protocol of the proposed method

1: Construct a heat matrix depicting temperatures.
2: Construct an image from the heat matrix
3: Resize the image to (224, 224) pixels.
4: $Feat \leftarrow$ Feed the image and extract the features using SqueezeNet 1.1
5: $Feat\_reduced \leftarrow$ Feed $Feat$ into chaotic GA to obtain reduced feature set
6: $Feat\_reduced\_Hybrid \leftarrow$ Feed $Feat\_reduced$ into GWO to obtain more reduced feature set.
7: Evaluate $Feat\_reduced\_Hybrid$ using the kNN classifier

4.4 Experimental Results and Analysis

4.4.1 Dataset description

The proposed Methodology 2 is evaluated on popular DMR-IR database[1]. The distribution of the dataset is given under Table 4.2. Each image has a dimension of 640×480 pixels. The software used creates two types of files for each capture:

- Heat Map File
- A matrix comprising of 307200 (640×480) thermal points

TABLE 4.2: Sample distribution of the DMR-IR dataset used for experimentation.

Category	No. of Subjects	No. of Samples
Malignant	19	381
Healthy	37	745

4.4.2 Experimental Setup

All experiments of this method are conducted on a machine equipped with an NVIDIA Tesla T4 GPU. The evaluation metrics used include accuracy, precision, and recall, as defined earlier in Section 4.2.2. The dataset is split into 70% for training, 10% for validation, and 20% for testing to perform and assess the experiments.

[1]http://visual.ic.uff.br/dmi/

4.4.3 Hyperparameters for Transfer Learning

Initially, while trying to come up with a good value of learning rate and batch size for training the deep CNN models, the grid search method is explored with some standard values in practice to determine the best possible combination. The initial learning rate $\in$ {1e-2,1e-3,1e-4,1e-5,1e-6} and batch size $\in$ {8,16,24,32} is considered. The accuracies are reported in Fig. 4.10. The cross entropy loss (see Eq. 4.19) is considered and the widely adopted Adam optimizer is used to optimize internal weights.

$$\mathcal{L}_{CE}(p_i^j, y_i) = -\sum_{j=1}^{c} y_i \log(p_i^j) \tag{4.19}$$

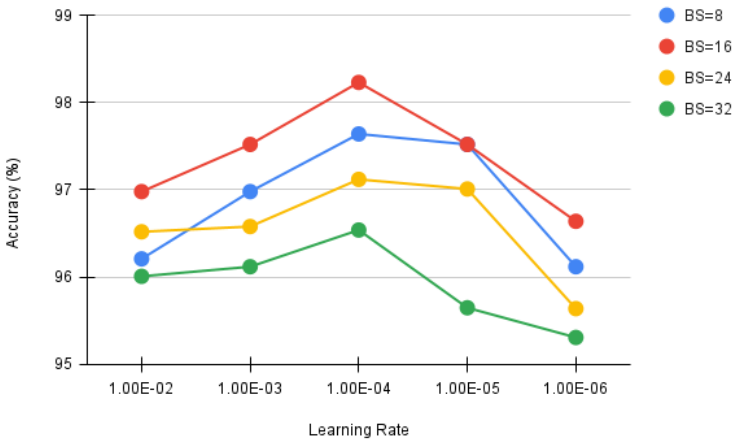


FIGURE 4.10: Accuracies concerning different learning rates and batch sizes for SqueezeNet 1.1 model.

It is visible from Fig. 4.10 that the model does not converge to their optimum maximum during the extremities. From Fig. 4.10 it is clear that the model reaches its optimum for learning rate $1e-4$ and batch size 16. For smooth learning, a step learning rate scheduler is used where the learning rate is reduced by a factor of 10 after 5^{th} epoch.

4.4.4 Analysis of Results

As a lightweight model is used it takes less time to train as well as is low on storage. The SqueezeNet 1.1 TL model is compared with other very popular and widely adopted models (each pre-trained on ImageNet) based on the time taken to train per epoch, size of the model, number of trainable parameters and number of floating-point operations per

second (FLOPs). It is to be noted that the reported sizes are recorded while downloading the uncompressed pre-trained models from PyTorch and the value of FLOPs are recorded for processing a single RGB image of dimension 224×224 pixels. Further details about these TL models can be found here[2].

TABLE 4.3: Comparison of different CNN models in terms of size (in Megabyte), time taken to train for one epoch (in second), number of trainable parameters (in Million) and number of FLOPs (in Gigaflop).

Model	**Size(MB)**	**Time(s)**	**Parameters(M)**	**FLOPs(G)**	**Accuracy (%)**
VGG 16	528	23	138.36	15.52	100
ResNet50	97.8	15	25.56	4.14	100
DenseNet121	30.8	16	7.98	2.90	100
GoogleNet	49.7	9	6.62	1.52	100
MobileNet V2	13.6	9	3.50	0.32	98.23
ResNet18	44.7	8	11.69	1.83	100
EfficientNet_B0	20.5	9	5.29	0.41	97.87
SqueezeNet 1.1	**4.74**	**6**	**1.23**	**0.36**	98.23

From Table 4.3 it is clear that SqueezeNet 1.1 provides a significant edge in terms of computational requirements to train and run the model. However, the accuracy may not be the best but is comparable to the rest of the models. Hence, proceeding further with SqueezeNet 1.1. Also, training a deep CNN may increase the chance of overfitting on the dataset. The model may be biased, particularly towards the samples of the dataset we train it upon, and the model may not recognize the unseen data. While training the SqueezeNet 1.1 model, the training and validation accuracies for every epoch are recorded and given in Fig. 4.11. From these results, it is clear that the model does not suffer from any major over-fitting. Fig. 4.12 suggests smooth learning while training the SqueezeNet 1.1 model. It is to be noted that all the values in Fig. 4.12 are evaluated using widely adopted averaged cross-entropy loss values.

4.4.5 Comparison with Other Feature Selection Algorithms

While applying a FS algorithm, the target is to gain the maximum possible accuracy as well as to reduce the number of features. An experimentation is conducted to compare the performance of the proposed hybrid FS algorithm with some very popular optimization algorithms and their different combinations. Results obtained in such experimentation are in Table 4.4.

[2]pytorch.org/vision/stable/models.html

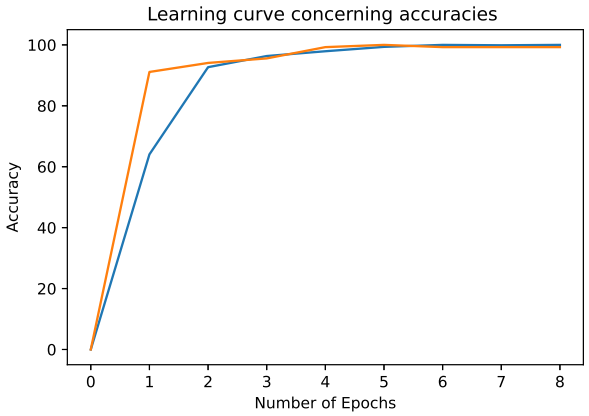


FIGURE 4.11: Learning curves w.r.t. accuracies for the SqueezeNet 1.1 model. The orange colored plot is for the training set, whereas the blue colored plot is for the validation set.

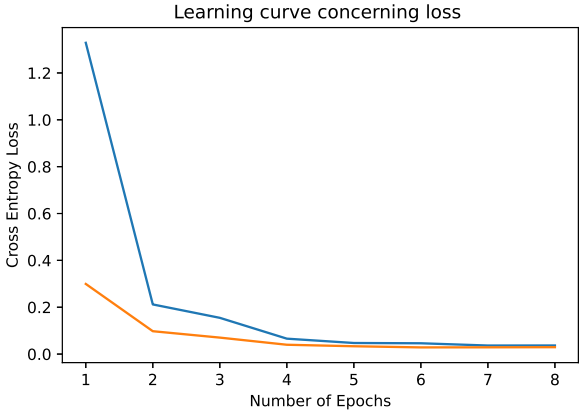


FIGURE 4.12: Learning curves w.r.t. loss values for the SqueezeNet 1.1 model. The loss function used is cross-entropy loss..

Based on experimental outcomes shown in Table 4.4, GA provides the best combination of accuracy and the lowest feature count amongst the well-known optimization algorithms that are used for comparison. The chaotic mapping based initialization to GA is added further. The modification results in a significant reduction of features to **131** features out of 1000 and an accuracy of **100%** is also achieved on the test set. This improvement in performance should be credited to the addition of chaos-based population initialization of GA. This strategy of initialization helps in exploring the search space by GA in a better way. Thus the exploration capability of GA is enhanced by the addition of chaotic mapping based initialization to GA. Since a classifier is used for classification, studies in the past have shown that a growing number of redundant features always lead to a drop in performance of the classifier [131]. To put this in other words, as the classifier maps the similarities between the features, in a higher dimension, optimizing

TABLE 4.4: Deep features of 1000 dimensions extracted from SqueezeNet 1.1 model are fed to different FS algorithms. As GA gives better results, output of the GA is fed to other FS algorithms for further optimization. It is observed that the hybrid of GA and GWO gives the best results among other hybrids. Tabulated results are found on the test set.

(a) Results with naive FS algorithms

FS algorithm	**No. of features**	**Accuracy (%)**
PSO	359	99.2
EO	462	99.2
GSA	463	98.8
GWO	316	98.8
MA	722	98.52
GA	**313**	**99.20**

(b) Results with chaotic GA hybrids

FS algorithm	**No. of features**	**Accuracy (%)**
GA-PSO	46	100
GA-EO	44	100
GA-GSA	36	100
GA-BBA	34	100
GA-MA	62	99.80
GA-GWO	**29**	**100**

Abbreviations used:
PSO: Particle Swarm Optimization, EO: Equilibrium Optimization, GSA: Gravitational Search Algorithm, MA: Mayfly Algorithm, GWO: Grey Wolf Optimization, BBA: Binary Bat Algorithm, GA: Genetic Algorithm

this measure might be more complex, leading to a drop in performance.

More experiments related to FS is conducted, where GA is combined with some other meta-heuristics, and the results obtained are shown in Table 4.4. From this table, further reduction in the number of features is possible without compromising the accuracy. Finally, **29** features out of 1000 features are chosen while maintaining the **100%** classification accuracy. Therefore, only 3% useful deep features are used to evaluate our model. Better results obtained by the combination of the GA-GWO can be attributed to the fact that we use the exploration ability of GA and the exploitation ability of GWO.

Further, the performance of the proposed methodology 2 is compared with some recently published methods (published between 2016 and 2021) and the results is shown in Table 4.5. The gain in performance has to be attributed to both the deep learning and FS approaches. As stated earlier, CNN based models successfully learn distinctive features, however, since here a very deep CNN which can learn much more local features

is not used. Instead, another path is followed to improve the learning capability of the model by removing less informative or redundant features through the use of a hybrid FS approach. Also from this analysis, it may be inferred that a deeper CNN model may have learnt more local features, but would have learnt a lot of redundant features because of the large number of inner layers. This increase in the number of layers may have increased the model performance but would significantly increase the use of computational resources and computational time. However, from Table 4.5, it is seen that the proposed model outperforms the SOTA models with a good margin.

TABLE 4.5: Performance comparison of the proposed Methodology 2 with SOTA methods

Method	**Accuracy(%)**	**Precision (%)**	**Recall (%)**
Pramanik et al. [132]	90	85	95
Gomez et al. [76]	92	94	91
Dey et al. [133]	98.80	99.00	98.00
Pramanik et al. [73]	96.90	90.00	95.70
Sanchez et al. [75]	97	100	83
Silva et al. [134]	98.57	100	97.14
Proposed Methodology 2	**100**	**100**	**100**

4.5 Methodology 3: Local Search Embedded SSD Algorithm for Breast Cancer Classification in Mammograms

In this method, a two-stage breast cancer classification model is designed taking mammograms as input. At first, an attention-aided DL model is used to extract features from the mammograms. Then, a local search embedded SSD algorithm for FS is incorporated to reduce the feature dimension and augment the classification ability. As the base model, a TL model, pre-trained on the ImageNet dataset is used and fine-tuned it on the target mammogram dataset. Furthermore, an attention is introduced by incorporating a global weighted average pooling mechanism on the base model. Features are extracted using this attention-based TL model for the target dataset and lastly process these extracted features with the local search embedded SSD FS method to produce optimal and reduced feature subsets. Finally, these reduced features are used as the inputs to the kNN classifier to produce the final classification results. Fig. 4.13 depicts the whole architecture of the suggested model.

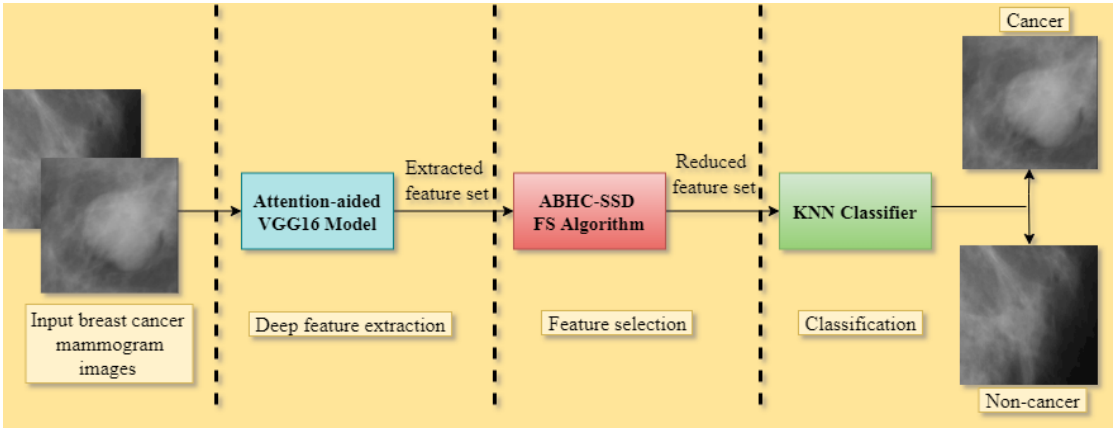


FIGURE 4.13: The pipeline of the proposed Methodology 3 for breast cancer classification.

4.5.1 Feature Extraction from the Attention-aided DL Model

For feature extraction from mammography inputs, a deep attention model is used. Starting with the VGG16 [135] model, which has been pre-trained on the popular ImageNet dataset. The VGG16 model is considered excluding the top layer, which has fully connected layers and freezes the weights of all levels to prevent the layers from learning new information during the model's training. The GAP layer takes the role of the fully connected layer. GAP [136], a pooling operation, is usually applied in place of fully connected layers in classical CNNs. Similar to simple pooling (max-pooling or average pooling) layers, it reduces the spatial dimension of a given tensor. For instance, a three-dimensional tensor having dimensions $h \times w \times d$ gets converted to the dimension of $1 \times 1 \times d$. GAP produces a single value for each feature map of dimension $h \times d$ by taking the average of all hw values. In this work, in the final convolutional layer of the base VGG16 model instead of adding fully connected layers, the generated vector is supplied to the final layer after taking the average of the feature maps. The GAP layer summarizes the spatial records through imposing correspondences among feature maps and categories, consequently making it robust in terms of spatial translation of the input data. However, only GAP is too simplistic as some regions may be more significant than others, thus needing attention. Here, an attention mechanism is introduced to turn on some pixels in the GAP layer. One locally connected convolutional layer of kernel size 1×1 is added here and fan it out to all the layers of the model. Next, a lambda layer is added to account for missing values from the attention model that rescale the results based on the number of pixels. Lastly, the model is fine-tuned by adding dropouts to

prevent over-fitting during training [137]. Thus, a weighted GAP to the existing CNN architecture is designed as depicted in Fig 4.14.

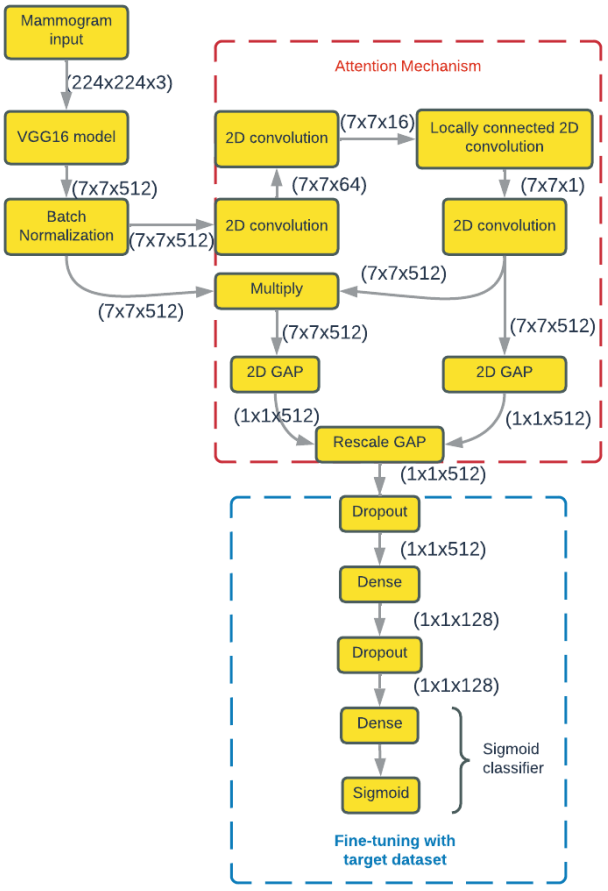


FIGURE 4.14: Architecture of the attention-aided VGG16 model with a weighted average GAP layer (all the layers along with input and output shapes). Area of the attention and the fine-tuning of the model are highlighted by red and blue colored dashed boxes, respectively

4.5.2 FS and classification using local search embedded SSD algorithm

Social Ski-Driver (SSD) Optimization Algorithm: SSD is a recent optimization technique suggested by Tharwat et al. [138]. It is based on the approach taken by the ski-drivers when they go downhill. The pseudo-code for SSD is described in Algorithm 4.

In SSD algorithm, h is a variable that is exploited to maintain the parity to achieve a balance among both two crucial elements of exploitation and exploration and r is exploited to decrease the value of h. The sine and cosine functions guarantees that the directions of movement of the agents are not very straightforward. This is because those

Algorithm 4 Social Ski-Driver (SSD) Optimization Algorithm

Initialize population of agent's location $L_k \in \mathbb{R}^n$, velocity V_k, and parameter h
Evaluate fitness of each agent at position L_k
Set personal best positions $PB_k = L_k$
for each iteration $T = 1$ to MaxIterations **do**
    Evaluate fitness of all agents
    Update personal bests PB_k if current position is better
    Identify top 3 best agents and compute mean global best:

$$MGB = \frac{L_x + L_y + L_z}{3}$$

    for each agent k **do**
        Generate random numbers $r_1, r_2 \sim \text{Uniform}(0, 1)$
        if $r_2 \leq 0.5$ **then**
            $V_k^{T+1} = h \cdot \sin(r_1) \cdot (PB_k - L_k) + \sin(r_1) \cdot (MGB - L_k)$
        else
            $V_k^{T+1} = h \cdot \cos(r_1) \cdot (PB_k - L_k) + \cos(r_1) \cdot (MGB - L_k)$
        end if
        Update position: $L_k^{T+1} = L_k^T + V_k^T$
    end for
    Update parameter: $h^{T+1} = r \cdot h^T$ where $0 < r < 1$
end for
Return the best solution found

functions allow the algorithm to explore and it helps to diversify the searching domain but in a controlled way. Advantage of the SSD algorithm is that it is comparatively more social than other meta-heuristics. The agents in SSD try to get to the mean of the best three options. As a result, if the global best solution is found to be in local minima, the SSD can use the other two best solutions for escaping.

Adaptive Beta Hill Climbing (ABHC) Local Search: Local search is used as a heuristic method for solving Computationally expensive optimization problems. Local search explores from solution to solution in the search space by doing some local changes, until a chosen solution seems to be optimal or the number of iterations (MT) is finished. To overcome the limitation of local optima in traditional Hill Climbing, the present method employs ABHC [139], a popular local search method. The pseudo-code of the ABHC method is discribed in Algorithm 5. This algorithm inputs an agent's location and outputs a modified location of the agent in the search space. This algorithm depends on two operators—one is the Neighborhood operator (N) and another one is the *Beta* operator. N is the greatest probable distance between the present solution and the neighbors.

Algorithm 5 Function of ABHC for Agent $L_k \in \mathbb{R}^n$

Initialize constants: c, *MaxBeta*, *MinBeta*
Calculate original fitness of agent L using the fitness function
Set $z \leftarrow 1$
while $z \leq$ *MaxIteration* **do**
    Compute $N(z) = 1 - \left(\frac{z^{1/c}}{\text{MaxIteration}^{1/c}}\right)$
    For each dimension $j = 1$ to k, compute $L'_j = L_j \pm \text{rand}(0,1) \times D$
    Compute $\beta(z) = \frac{(\text{MaxBeta} - \text{MinBeta}) \times z}{\text{MaxIteration}} + \text{MinBeta}$
    for $i = 1$ to N **do**
        if $\text{rand}(0,1) < \beta(z)$ **then**
            $L''_j \leftarrow L_j$
        else
            $L''_j \leftarrow L'_j$
        end if
    end for
    if Fitness Function(L'') > Fitness Function(L) **then**
        $L \leftarrow L''$
        Recalculate fitness of L''
    end if
    $z \leftarrow z + 1$
end while

A FS problem seeks to find the best subset of features from the original set to improve the accuracy of a learning model. It can also be considered as a dimension reduction algorithm that removes redundant and/or highly correlated features. Due to the binary nature of the FS problem, most meta-heuristics are not able to solve them. This is due to the assumption of continuous variables in the vanilla version of such algorithms. There are different methods in the literature to convert them to binary algorithms. One of the most computationally cheap ways is to use a transfer function. In [140] a velocity vector is used to update solutions, this transfer function relates velocity to the probability of changing the position in a binary space. The transfer function used in the current work is shown in Eq. 4.20.

$$V(x) = \frac{|x|}{\sqrt{1 + x^2}} \tag{4.20}$$

Exploiting the V-shaped transfer function, the location of an agent is modified as per Eq. 4.21.

$$LB_j^{k+1} = \begin{cases} c(LB_j^k) & if\ V\left(LB_j^{k+1}\right) > rand\,(0,1) \\ LB_j^{k+1} & otherwise \end{cases} \tag{4.21}$$

In Eq. 4.21, LB_j^{k+1} denotes the agent's modified location, LB_j^k signifies the location of the agent at that particular time(here k denotes iteration number and j denotes number of dimention) and $rand(0,1)$ is a function that generates random numbers between 0 and 1. The function $c(x)$ denotes the complement function for all binary x i.e., $c(x) = 1 - x$. Fig. 4.15 shows the graph of the V-shaped transfer function. After altering an agent's location in each iteration, ABHC [139] is used to optimize the position of the agents to get a higher fitness value. The SSD algorithm's exploitation potential is improved by using an ABHC-based local search technique. Pseudo-code of the local search embedded SSD algorithm is shown in Algorithm 6.

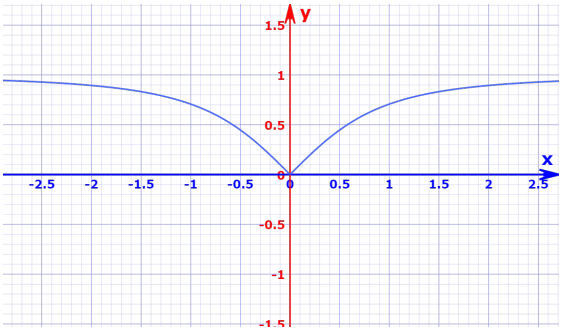


FIGURE 4.15: Graphical representation of the V-shaped transfer function

Algorithm 6 Pseudo-code of the local search embedded SSD algorithm

Initialize the value of Population-size and the Max-iteration
Randomly initialize the value of velocity, position, *pbest*, and *gbest*
for $z = 1$ to $Max - iteration$ do
for $p = 1$ to $Population - size$ do
Set $agent = L$
Using the Fitness Function, determine the agent L's original fitness value.
Change the value of *gbest* by picking one of the top 3 sites at random.
Update the value of *pbest*.
Set the value of $r_1 = rand(0,1), r_2 = rand(0,1)$ and $L_k^{T+1} = V_k^T + L_k^T$
Check if $(r_1 < 0.5)$ then
Set $V_k^{T+1} = h \times \sin(r_2) \times (PB_k^T - L_k^T) + \sin(r_2) \times (MGB_k^T - L_k^T)$
else
Set $V_k^{T+1} = h \cdot \cos(r_2) \cdot (PB_k^T - L_k^T) + \cos(r_2) \cdot (MGB_k^T - L_k^T)$
End if
Calculate $V(x) = \dfrac{|x|}{\sqrt{1+x^2}}$
Update the position of LB_j^{k+1} using Equation 4.21.
Set the new location of LB_j^{k+1}= Local Search (LB_j^{k+1})
End for
Set h^{T+1}=$r \times h^T$
End for

4.5.3 Fitness Function

The fitness function is used to evaluate the quality of a candidate solution. A learning algorithm needs to be exploited for assessing as SSD is a wrapper-based algorithm. Hence, the kNN classifier is exploited for the computation of classification accuracy of a candidate solution. The fitness function mainly contains two components: one is the number of features and another one is classification accuracy. These components are contradictory to each other - while classification accuracy should be maximized, the number of features should be minimized. Due to this the classification error is exploited. As a lesser error value would indicate a better fitness score, so would a lesser number of features. In Eq. 4.22, the fitness function is used to assess a given feature set's strength is defined.

$$Fitness\ Function = w\times \propto + (1-w) \times \frac{|s|}{|d|} \tag{4.22}$$

In Eq. 4.22, $|d|$ is the total count of features in the dataset, $|s|$ denotes the count of features in the chosen feature set (i.e., a candidate solution),$\propto$ is the error in classification using the feature subset, and $w \in [0,1]$ signifies the relative weight value given to the classification error and the number of features.

4.6 Experimental Results and Analysis

4.6.1 Dataset Description

The proposed Methodology 3 is assessed on a publicly accessible breast cancer mammography database [3] which is an open-source and unbalanced binary mammography image dataset that contains normal images i.e., negative samples from the DDSM database [36], and malignant images i.e., positive samples from the CBIS-DDSM database [37]. It has a total of 55,885 training samples out of which 86% are negative samples and 14% are positive samples. Each image in the dataset has a uniform resolution of 299×299 pixels. In this work, only the training samples of mammogram masses are considered which are categorized as benign and malignant masses.

[3] `https://www.kaggle.com/datasets/skooch/ddsm-mammographyg`

4.6.2 Parameter Tuning for Transfer Learning and SSD-based FS

The experiments are conducted on the DDSM database using a data split of 80% for training, 10% for validation, and 10% for testing, which yielded the highest classification accuracy. Various splits are initially tested, but this configuration performed best. To optimize performance, different learning rates 1e-2, 1e-3, 1e-4, 1e-5, 1e-6 and batch sizes 8, 16, 32, 64 are evaluated (refer to Fig. 4.16), with the optimal combination found to be a learning rate of 1e-2 and batch size of 64. The model is trained using the Adam optimizer and a step learning rate scheduler that halves the learning rate after the third epoch. Dropout rates of 0.5 and 0.25 are applied in two respective layers to prevent overfitting.

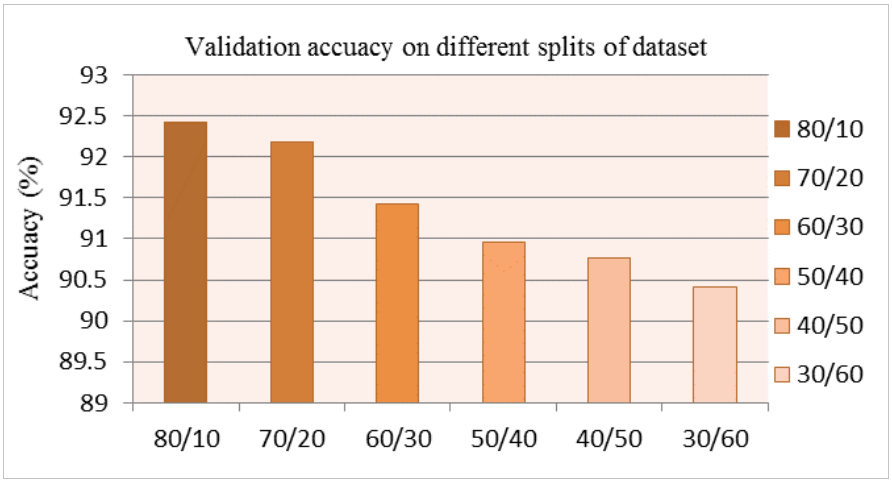


FIGURE 4.16: The attention-aided VGG16 model's validation accuracy for varied splits of training and validation data. On the DDSM database, the model gets the best classification accuracy for 80% training and 10% validation data

The fitness function defined in Eq. 4.22 aims to minimize both the number of selected features and the classification error. The parameter w controls the trade-off between these two objectives: higher values of w emphasize feature reduction, while lower values prioritize minimizing the classification error. Through extensive experimentation, the optimal value was determined to be $w = 0.2$, placing greater focus on improving classification accuracy. Additionally, the algorithm-specific parameters h and r are evaluated across a range of values. Results showed that increasing h consistently improved classification performance. In contrast, lowering r initially enhanced accuracy but led to overfitting when reduced excessively. The highest classification accuracy was obtained when $h = 100$ and $r = 0.9$.

4.6.3 Results

The proposed Methodology 3 is evaluated using accuracy, precision, and recall, as defined in Section 4.2.2. Experimentation is done with some popular pre-trained end-to-end models for deep feature extraction and the result is reported in Table 4.6.

TABLE 4.6: Performance of different pre-trained TL models on the DDSM database

Pre-trained TL Model	**Accuracy (%)**
VGG19	**87.45**
ResNet50	87.23
EfficientNet	87.22
VGG16	87.30

Further, the attention mechanism is incorporated with each of these models. Noteworthy improvement in classification accuracy is achieved due to the effect of weighted average pooling attention. As in the last layer of the CNN model as an alternative to the fully connected layer, the GAP layer is used and it gives a single feature map for the corresponding category. This layer uses spatial information by enforcing correspondences between feature maps and categories. The results obtained on the test dataset are tabulated in Table 4.7. The Mean and the standard deviation (SD) value are shown over five simulations of the obtained results. From Table 4.7, it can be observed that the VGG16 model with attention classifies the mammograms more appropriately. Hence, it is decided to proceed with the attention-based VGG16 model.

TABLE 4.7: Performance of the attention-aided different deep feature extraction models on the DDSM database

Simulation	**VGG16**	**VGG19**	**ResNet50**	**EfficientNet**
1	92.42	91.89	89.96	89.89
2	91.86	91.94	91.14	91.17
3	92.12	91.77	91.32	90.78
4	91.41	90.34	91.23	89.89
5	91.51	90.78	91.86	90.34
Mean ± SD	**91.86 ± 0.42**	91.34 ± 0.73	91.10 ± 0.69	90.41 ± 0.56

A local search-based FS algorithm is applied to gain the maximum possible accuracy as well as to reduce the number of features used for the classification purpose. As finding the best feature subset is a difficult task, especially in a wrapper-based FS model, a local search method is combined with an FS model. The results obtained from the proposed Methodology 3 are shown in Table 4.8. Table 4.8 shows that high accuracy of 96.07%,

the precision of 96.30%, sensitivity (recall) of 99.28% and feature dimension reduction of 75% have been achieved.

TABLE 4.8: Performance of the proposed Methodology 3 on the test set of DDSM database over five simulations

Simulation	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**#FS**
1	96.43	96.97	98.97	32
2	95.98	95.57	100.00	30
3	96.43	96.97	98.97	35
4	95.54	95.54	99.48	34
5	95.98	96.48	98.97	31
Mean ± SD	**96.07 ± 0.37**	**96.30 ± 0.71**	**99.28 ± 0.21**	**32 ± 2**

For experimentation, two well-known local search methods, namely Late Acceptance Hill Climbing (LAHC) [141] and ABHC [139] is used to optimize the feature set. The experimental result on the DDSM database with the said local search-embedded SSD FS method is shown in Table 4.9. From Table 4.9, it is clear that the ABHC embedded SSD technique outperforms the LAHC embedded method in terms of classification accuracy as well as obtaining a smaller optimal feature set.

TABLE 4.9: Performance of the ABHC and LAHC local search embedded SSD algorithm on extracted deep features from the VGG16 model. Classification accuracy is in %.

Simulation	**SSD+LAHC**		**SSD+ABHC**	
	Accuracy	**#Features**	**Accuracy**	**#Features**
1	94.19	52	96.43	32
2	94.64	59	95.98	30
3	94.28	53	96.43	35
4	93.98	56	95.54	34
5	93.86	47	95.98	31
Mean ± SD	94.19 ± 0.30	53 ± 5	**96.07 ± 0.37**	**32 ± 2**

Further experiments are performed using various other meta-heuristic based FS methods namely: Gravitational Search Algorithm (GSA) [142], WOA [118], GWO [129], GA [143], PSO [140], Sine Cosine Algorithm (SCA) [144], Harmony Search (HS) algorithm [145] and Equilibrium Optimizer (EO) [146] for comparison of the proposed method. Meta-heuristic-based FS algorithms need to perform many mathematical operations to identify the best feature subset. For this, the algorithms in the literature use many sets of equations that are aided by different parameters. These parameters are crucial for controlling the optimization process and they have their own significance. The standard

values of these parameters of various algorithms are used in this study. Table 4.10 contains a list of the parameters and their values.

TABLE 4.10: Different sets of hyperparameters and their values for various meta-heuristic based FS algorithms considered for experimentation.

FS Algorithm	**Parameter(s)**	**Value(s)**
Generic parameters	Population size	20
	Number of iterations	100
	Weight for accuracy (α)	$\alpha = 0.98$
GSA	Initial Gravitational constant (G_{init})	$G_{init} = 6$
	Constant (ϵ)	0.00001
WOA	Encircling parameter (a)	$a \in [0, 2]$
	Shape of spiral (b)	$b = 1$
GWO	Convergence operator (a)	$a \in [0, 2]$
GA	Gene Selection	Roulette wheel
	Crossover probability	0.4
	Mutation probability	0.3
PSO	Inertia weight (IW)	$IW \in [0, 1]$
	Coefficients (r_1, r_2)	r_1, $r_2 \in [0, 1]$
SCA	Constant (a)	$a = 3$
	Movement direction (r_1)	$r_1 \in [0, 3]$
HS	Harmony Memory Considering Rate (HMCR)	HMCR = 0.90
EO	Pool size	4
	Constants (a_1, a_2)	a_1=2 and a_2=1
	Generation rate (GP)	GP = 0.5

The simulation results of local search-based different meta-heuristics FS algorithms are shown in Table 4.11.

From Tables 4.8 and Table 4.11, it can be said that ABHC local search-embedded SSD algorithm outperforms others in terms of classification accuracy. Besides, it provides a subset of 32 features which is just 25% features of the given input features obtained from the VGG16 model. The comparative analysis with different combinations of various meta-heuristics and local search is shown in Fig. 4.17 and 4.18. The sine and cosine functions complicate the movement direction of the agents, which is the most essential feature of this SSD-based FS technique. This allows the algorithm to diversify, and the parameter h ensures that the algorithm remains stable between exploration and exploitation, allowing it to converge to better solutions. Furthermore, ABHC aids the algorithm to improve the solutions, thereby overcoming the local optima, thus leading to a better outcome.

TABLE 4.11: Results from five simulations, including average and standard deviation (SD), after deep features from the VGG16 model are extracted and fed to several FS algorithms with embedded local search.

FS Method	LAHC Embedded FS Method		ABHC Embedded FS Method	
	Accuracy (%)	**No. of Features**	**Accuracy (%)**	**No. of Features**
GSA	93.30	65	95.08	70
	94.28	54	94.19	63
	94.19	51	94.28	56
	93.30	67	92.41	57
	93.33	65	93.30	64
	93.68 ± 0.51	**54 ± 7**	**93.85 ± 0.92**	**62 ± 6**
WOA	94.64	48	94.64	55
	93.33	58	93.75	58
	93.75	59	93.33	61
	94.19	68	94.64	69
	94.19	48	94.64	60
	94.02 ± 0.50	**56 ± 8**	**94.19 ± 0.55**	**61 ± 5**
GWO	94.64	76	94.64	69
	93.33	65	93.30	78
	93.75	81	93.33	63
	94.64	74	93.75	79
	93.33	65	94.64	75
	93.94 ± 0.67	**72 ± 7**	**93.93 ± 0.59**	**73 ± 7**
GA	94.64	70	94.64	67
	93.30	59	93.75	62
	93.30	57	93.75	55
	93.33	56	93.33	59
	94.64	70	93.30	59
	93.84 ± 0.73	**62 ± 7**	**93.75 ± 0.48**	**60 ± 4**
PSO	93.30	41	94.64	55
	93.33	42	94.19	52
	94.19	50	93.30	51
	92.86	46	93.33	48
	93.33	52	94.64	57
	93.40 ± 0.48	**46 ± 5**	**94.02 ± 0.59**	**53 ± 4**
SCA	93.30	59	94.19	70
	93.33	71	93.75	67
	94.19	73	93.75	64
	93.33	65	93.33	63
	93.41	57	94.19	70
	93.51 ± 0.38	**65 ± 7**	**93.84 ± 0.32**	**67 ± 3**
HS	93.30	63	92.41	60
	94.28	63	91.16	73
	95.08	56	93.33	53
	93.33	55	91.52	62
	92.42	65	91.96	68
	93.68 ± 1.02	**60 ± 5**	**92.07 ± 0.75**	**63 ± 8**
EO	93.75	53	94.64	55
	94.28	56	94.64	69
	94.64	56	93.30	55
	94.64	50	94.28	63
	93.41	55	93.33	55
	94.14 ± 0.55	**54 ± 3**	**94.04 ± 0.60**	**59 ± 6**

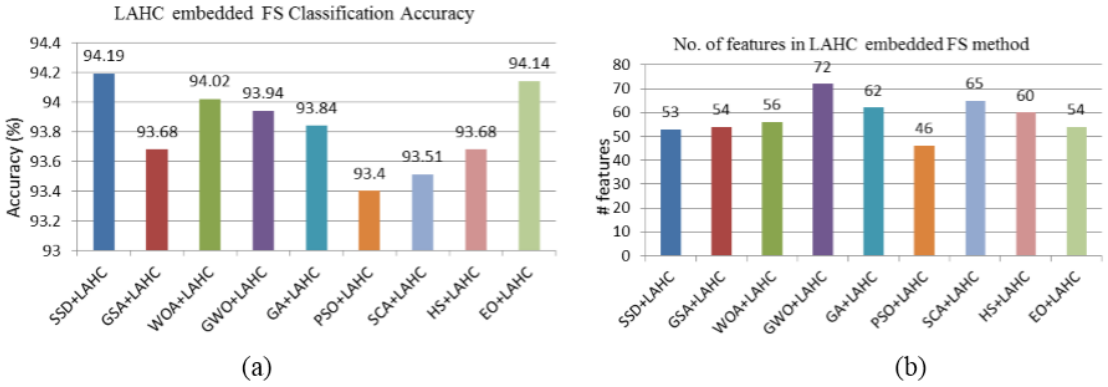


FIGURE 4.17: Results of different FS algorithms embedded with LAHC local search (a) Classification accuracy (b) No. of selected features

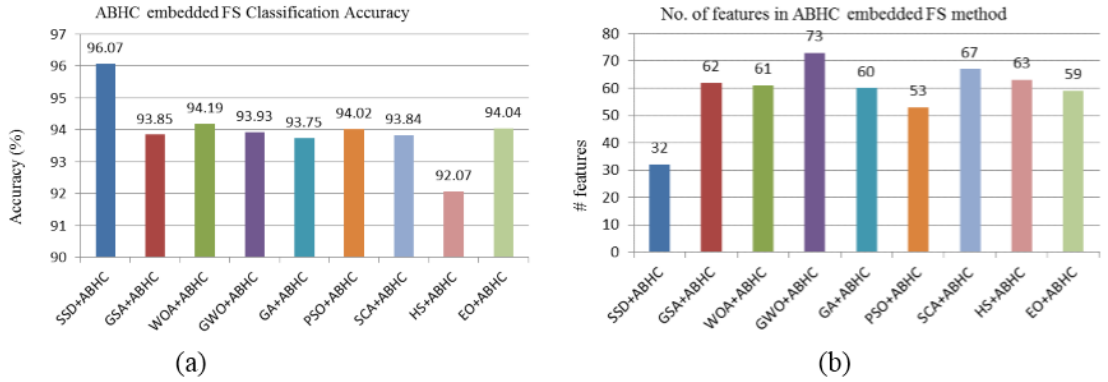


FIGURE 4.18: Results of different FS algorithms embedded with ABHC local search (a) Classification accuracy (b) No. of selected features

Furthermore, experiment is performed with different modules of the proposed method to observe the effectiveness of each module and noted the experimental results in Table 4.12. From the table, it can be observed that an attention-aided VGG16 model yields a classification accuracy of 91.41% on the test set whereas if the kNN classifier is used to it, the classifier produces a classification accuracy of 90.70% which is low compared to end-to-end VGG16 model. Also, FS using the SSD algorithm improved the classification accuracy. Further, if the ABHC local search is incorporated with the SSD-based FS method, improved classification accuracy is acheived. Therefore, from these experimental results, it can be ensured the effectiveness of the individual modules (i.e., deep feature extraction with VGG16, FS with ABHC aided SSD method) of the proposed method. The gain in performance has to be attributed to both the DL and FS approaches.

TABLE 4.12: Comparative results for different combinations of the proposed Methodology 3 in terms of classification accuracy (%) on the test set of the DDSM database

Method	**Classification Accuracy (%)**
End-to-end VGG16	91.41
VGG16 + kNN	90.70
VGG16 + SSD + kNN	92.86
VGG16 + ABHC aided SSD + kNN	95.98

4.6.4 Comparison with Past Methods

Finally, the performance of the proposed Methodology 3 is compared with some recently developed classification methods and tabulate the same in Table 4.13. The performance of the proposed method is superior to that of four DL-based methods used in [147–151]. Hence, from Table 4.13, it is clear that the proposed model outperforms these previous works with a good margin and the experimental results establish the superiority and robustness of the proposed method.

TABLE 4.13: Performance comparison of the proposed Methodology 3 with SOTA models on the DDSM database

Model	**Accuracy (%)**	**Precision (%)**	**Recall (%)**
Levy et al. [147]	92.9	92.4	93.4
Falconi et al. [41]	84.4	-	-
Xiao et al. [148]	-	82.2	94.9
Arias et al. [149]	92.0	-	-
Zhang et al. [150]	94.30	-	89.97
Li et al. [151]	94.7	-	94.1
Proposed Methodology 3	**96.07**	**96.30**	**99.28**

4.7 Methodology 4: Gamma Function based Ensemble Approach

In this method, a rank based ensemble method using the Gamma function is introduced which is used to combine the confidence scores of three TL-based CNN models. The advantage of this technique is that it uses adaptive weights where the confidence scores of each base classifier are emphasized for forming the ensemble to generate the final prediction of each sample. The overall pipeline of the proposed ensemble scheme is presented in Fig. 4.19.

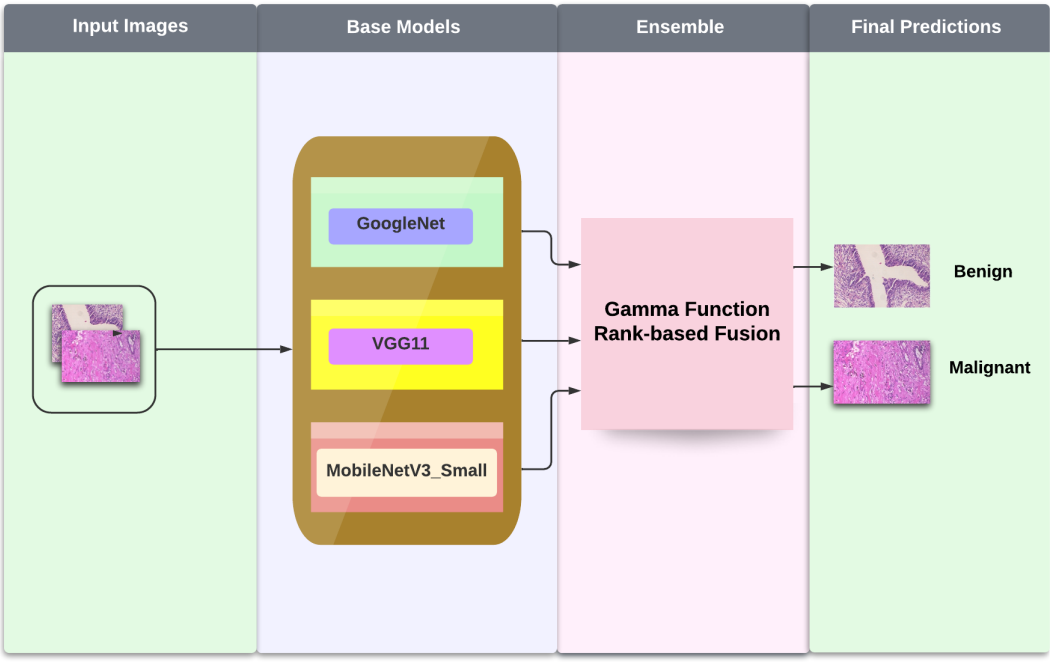


FIGURE 4.19: The overall pipeline of the proposed Gamma function based ensemble of CNN models used for classification of breast histopathology images

4.7.1 Base CNN Models

In this method, three TL-based CNN models (GoogleNet, VGG11, MobileNetV3_Small) pre-trained on the ImageNet dataset is used from which the classification confidence scores are obtained. The BreakHis dataset is split into a train set and a test set for each type of magnification. Data augmentation on the train set is done with random horizontal flipping and resizing of the images. The ImageNet weights are then fine-tuned using the Stochastic Gradient Descent (SGD) optimizer.

GoogleNet:

GoogleNet [152] proposed by the researchers at Google is a 22-layer deep CNN model (refer in Fig. 4.20) containing inception blocks (refer in Fig. 4.21). In an Inception block, several convolutional layers have been working simultaneously, which contains a sparse architecture that helps dimension reduction. This improves the complexity but makes better performance in computer vision tasks.

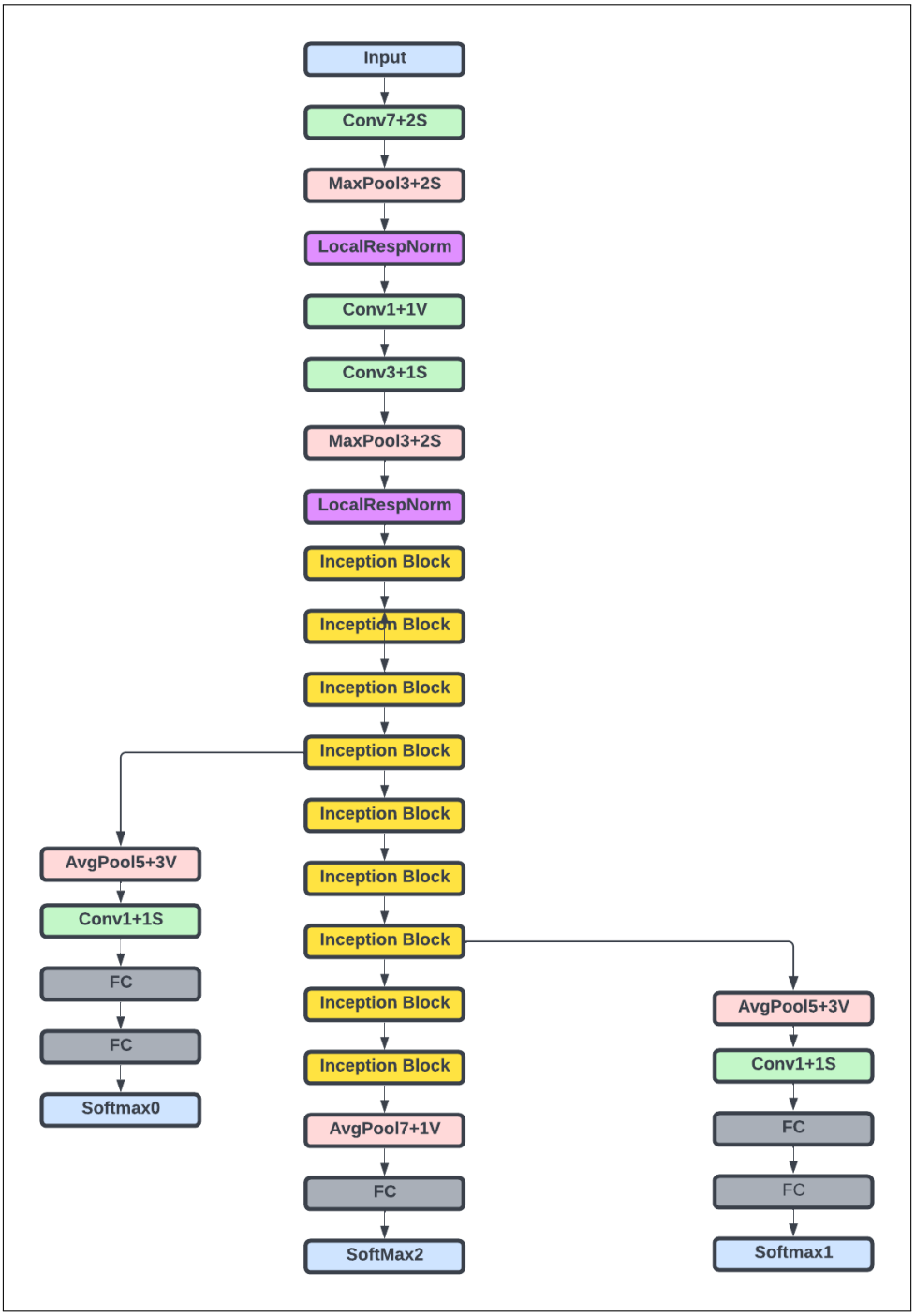


FIGURE 4.20: A pictorial representation of GoogleNet architecture

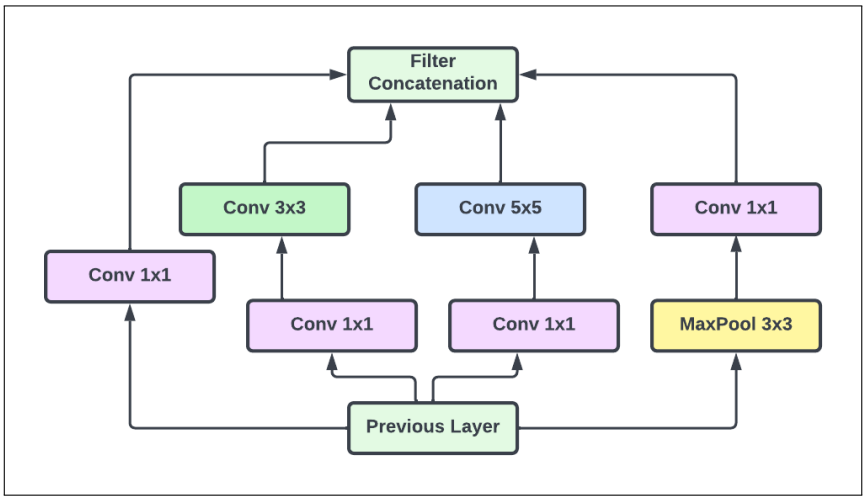


FIGURE 4.21: Inception block used in GoogleNet

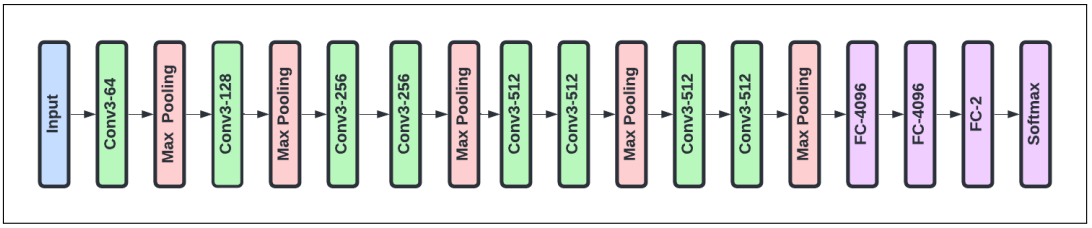


FIGURE 4.22: A pictorial representation of VGG11 architecture

VGG11:

VGG11 is a 11-layer deep CNN [153] which contains 8 convolutional layers and 3 fully connected layers. In the model, RGB images of 224×224 are taken as the input, where it is followed by series of convolutional layers of kernel size as 3×3 and stride of 1, with proper padding. In between some of the convolutional layers, Max Pooling layers of kernel size as 2×2, a padding size as 2 are present. After each of the convolutional layers and some of the fully connected layers Rectified Linear Unit (ReLU) activation function is used. The architecture of the VGG11 model is presented in Fig. 4.22.

MobileNetV3_Small:

MobileNet model has been designed specifically for mobile applications. MobileNet is a small, low-latency, low-power model parameterized to meet the resource constraints of a variety of mobile applications. They can be built upon for classification, detection, embedding and segmentation similar to other large scale models. In the present work, MobileNetV3_Small architecture [154] is used. It takes RGB images of dimension 224×224 as input and the final linear layer contains 1024 units. It comprises of a combination of bottleneck architecture and squeeze and excitation. It uses the swish non-linearity as

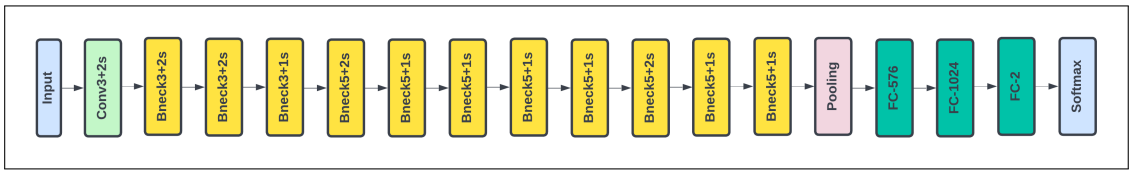


FIGURE 4.23: A pictorial representation of MobileNetV3_Small architecture

activation function. The architecture of MobileNetV3_Small is presented in Fig. 4.23.

4.7.2 Rationality behind choosing the CNN models for the proposed ensembling

The main reason for choosing the **GoogleNet** architecture is because of its inception block. It has nine inception blocks in its overall architecture. Convolutional kernels of different shapes 1×1, 3×3, & 5×5 are used in the inception block as we see in Fig. 4.21. The feature maps obtained from each of the kernels are combined and taken as the output of the inception block. Larger convolutional kernels of size 5×5 cover a large area in an image to get the overall information, whereas the smaller kernels of size 1×1 work on a smaller area in the image to get much finer details of the image. This particular feature of the GoogleNet helps in better feature extraction of the image and also does effective dimensionality reduction. This also makes the architecture quite deep with only 4 million parameters which makes it computationally very effective as compared to other deep CNN architectures.

VGG11 architecture has a low depth of only 11 layers. The less number of layers in the VGG11 architecture results in less number of parameters in the model. As the data volume in the medical domain is less, increasing the number of layers does not help the model to analyze the complicated data patterns. With these points, taken under consideration VGG11 model is selected for experimentation.

In the **MobileNetV3_Small** architecture, there are 11 bottleneck layers. Each bottleneck layer comprises a residual connection along with a squeeze and excitation block which helps in the efficient feature extraction of an image. This results in less depth of the MobileNetV3_Small architecture and the less number of parameters which make the model work computationally faster and help in better feature extraction in low data volume medical domain datasets.

Due to the facts described above and the analysis of results in the later section, the

above-mentioned CNN based models are chosen as the final base models for the Gamma function based ensemble method.

4.7.3 Gamma Function based Ensemble Method

The confidence scores for a histopathological image by all three models are fused using the Gamma function. Gamma function is a mathematical function which is defined for all complex numbers except non-positive integers.
Generally, the Gamma function is defined by Eq. 4.23.

$$\Gamma(\alpha) = \int_0^\infty s^{\alpha-1} e^{-s} ds \tag{4.23}$$

In Eq. 4.23, α is the input of the Gamma function which is defined for all complex numbers except non-positive integers and s is defined as a positive real number over which the following expression is integrated.
For a positive integer n, the Gamma function is defined as Eq. 4.24.

$$\Gamma(n) = (n-1)! \tag{4.24}$$

Graphical representation of the Gamma function for positive real numbers is shown in Fig. 4.24.

The Gamma function is used in this work for finding the rank of samples belonging to class c of the i^{th} classifier's confidence score (CF) as defined in Eq. 4.25.

$$R_c^{(i)} = \Gamma\left(CF_c^{(i)}\right), \forall i, c; i = 1, 2, ..., M; c = 1, 2, ..., C \tag{4.25}$$

In this study the value of $M = 3$ as three CNN models have been used and value of $C = 2$ (Benign and Malignant).
Before, to make the final decision score, some important calculations have been performed. The fuzzy rank sum (FRS) for class c is calculated as per the Eq. 4.26 and the complement of the confidence factor sum (CCFS) for class c is calculated as per the Eq. 4.27.

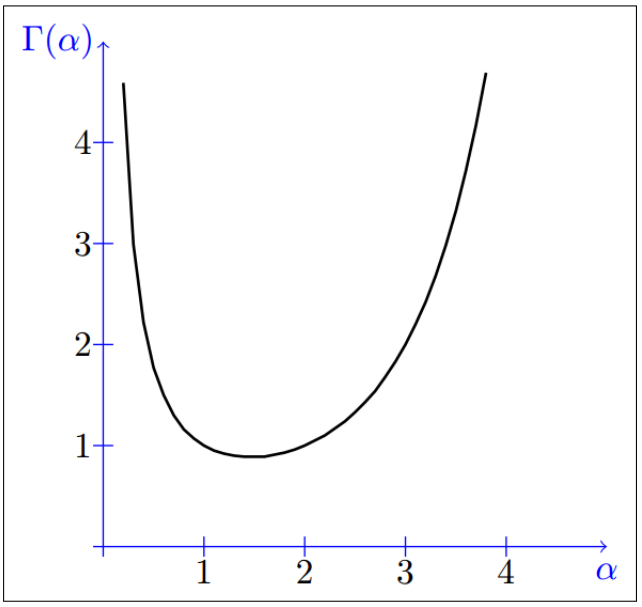


FIGURE 4.24: Graphical representation of the Gamma function for positive real numbers.

$$FRS_c = \sum_{i=1}^{M} R_c^{(i)} \tag{4.26}$$

$$CCFS_c = \frac{1}{M} \sum_{i=1}^{M} (1 - CF_c)^{(i)} \tag{4.27}$$

Now, the final decision score (FDS) for a class can be defined as per the Eq. 4.28.

$$FDS_c = FRS_c \times CCFS_c \tag{4.28}$$

The final predicted class for an image I is calculated by finding the minimum of the FDS values and is given by the following Eq. 4.29.

$$class(\mathbf{I}) = argmin_c(FDS_c) \tag{4.29}$$

The Gamma function is used for rank based ensemble method because on the particular domain (0,1), the function steeply decreases. Due to such nature of the curve, there is a significant difference among the ranks obtained for a sample belonging to a particular class by different base learners, which gives an advantage to design a robust ensemble

model. For better understanding of the proposed ensemble method, this can be verified by the following hypothetical example.

Let us consider, for a particular image I the confidence scores for a particular class (say Benign) by three CNN models are as follows,

$CF_B^1 = 0.81,$
$CF_B^2 = 0.96,$
$CF_B^3 = 0.94$
where the ranks of the confidence scores obtained from each model are as follows,
$R_B^1 = \Gamma\,(0.81) = 1.15$
$R_B^2 = \Gamma\,(0.96) = 1.02$
$R_B^3 = \Gamma\,(0.94) = 1.03$

From the values of R_B we can observe that the minimum difference is of the order of 10^{-2}. Hence, the significant difference among the values of ranks builds the robustness of this ensemble technique.

4.8 Experimental Results & Analysis

4.8.1 Dataset Description

The proposed Ensemble method is evaluated on the BreakHis [77] dataset. It is a histopathological dataset containing 7909 samples which fall under the two main categories: Benign and Malignant. The Benign class contains 2480 samples and the Malignant class contains 5429 samples. The dataset contains images from 82 patients which have a different set of magnification factors: **40X, 100X, 200X and 400X**. The size of each image is 700 X 460 pixels, PNG format, and 8-bit depth in each channel with 3-channel RGB. In Table 4.14, a detailed overview of the BreakHis dataset is provided.

TABLE 4.14: Details of the BreakHis dataset used for experimentation

Magnification	**Benign**	**Malignant**	**Total**
40X	625	1370	1995
100X	644	1437	2081
200X	623	1390	2013
400X	588	1232	1820
Total images	2480	5429	7909

4.8.2 Data Augmentation

Experiments are performed on each magnification factor of the histopathology data independently where each set is split into train and test sets in the ratio of 70%-30% before performing the data augmentation process. In Table 4.15, a detailed overview of the number of benign and malignant images used in the test dataset is provided on each set of magnifications of the BreakHis dataset.

TABLE 4.15: Detailed overview of the test data on each set of magnifications of the BreakHis dataset

Magnification	**Benign**	**Malignant**
40X	188	412
100X	194	432
200X	187	418
400X	177	370

In each set of magnification of the BreakHis dataset, the data augmentation technique is applied on the training data samples to increase the volume of the training dataset and to prevent overfitting of each of the models while training it. The augmentation techniques have been applied as follows: width shift of range 0.125, height shift of range 0.125, horizontal flip and nearest fill mode. The augmented training samples are then divided into training and validation sets in the ratio of 80%-20%. Table 4.16 shows the distribution of the total number of images at each level of magnification in the BreakHis

dataset for the test data and the augmented data which are split into training and validation sets respectively.

TABLE 4.16: Distribution of the total number of images at each level of magnification in the BreakHis dataset for the test data and the augmented data which are split into training and validation sets respectively.

Magnification	**Test set**	**Training set**	**Validation set**
40X	600	3136	784
100X	626	3535	885
200X	605	3149	788
400X	547	2876	720

4.8.3 Results

At first, an experiment is conducted with five different CNN models separately and evaluate the classification accuracy on the test data of each magnification level. In Table 4.17, the values of the hyperparameters utilized for training these CNN models are described.

TABLE 4.17: Hyperparameters along with their values used for training the CNN models

Hyperparamter	**Value**
Optimizer	Stochastic gradient descent
Loss function	Cross entropy
Initial learning rate	0.01
Learning rate scheduler	ReduceLROnPlateau
No. of epochs	30

The proposed model is evaluated using accuracy, precision, and recall, as defined in Section 4.2.2. The classification accuracies of the base CNN models are reported in Table 4.18.

TABLE 4.18: Classification accuracies on the BreakHis test set with different magnification levels obtained by the base CNN models.

Model	**Magnification**	**Accuracy (%)**
GoogleNet	**40X**	99.00
MobileNetV3_Small		97.83
VGG11		97.66
VGG16		97.33
AlexNet		96.50
GoogleNet	**100X**	98.08
MobileNetV3_Small		97.76
VGG11		96.80
VGG16		94.73
AlexNet		95.37
GoogleNet	**200X**	97.19
MobileNetV3_Small		96.52
VGG11		96.85
VGG16		94.73
AlexNet		95.37
GoogleNet	**400X**	95.97
MobileNetV3_Small		96.52
VGG11		92.87
VGG16		93.60
AlexNet		91.22

From Table 4.18 it can be observed that among all the individual base models, GoogleNet yields better classification accuracy for all three magnifications (99% for 40X, 98.08% for 100X and 97.19% for 200X). Whereas, in the case of 400X magnification, the MobileNetV3_Small model produces a better result(96.52%). Now, among the five CNN models, a combination of three models are taken and use their performances to implement the rank based ensemble approach. Rigorous experimentation is perfomred with all possible combinations for each of the magnifications and record the corresponding results. Table 4.19, 4.20, 4.21 and 4.22 show the comparative results we obtain after performing the proposed ensembling over all possible combinations on 40X, 100X, 200X

and 400X magnifications of images.

TABLE 4.19: Classification accuracies of the proposed ensemble method on three base models taken at a time from five CNN models on 40X magnification of the BreakHis test set. The combination of models considered in this work is in bold font. Accuracy values are in % .

Magnification	**Base Model 1**	**Base Model 2**	**Base Model 3**	**Acc**
40X	MobileNetV3_Small	VGG11	VGG16	98.50
	MobileNetV3_Small	VGG11	AlexNet	98.50
	MobileNetv3_Small	VGG16	AlexNet	98.33
	VGG11	VGG16	AlexNet	98.50
	GoogleNet	VGG16	AlexNet	99.12
	GoogleNet	VGG11	AlexNet	99.10
	GoogleNet	VGG11	VGG16	99.10
	GoogleNet	MobileNetV3_Small	AlexNet	99.12
	GoogleNet	MobileNetV3_Small	VGG16	99.13
	GoogleNet	**MobileNetV3_Small**	**VGG11**	**99.16**

TABLE 4.20: Classification accuracies of the proposed ensemble method on three models taken at a time from five CNN models on 100X magnification of the BreakHis test set. The combination of models considered in this work is in bold font. Accuracy values are in % .

Magnification	**Base Model 1**	**Base Model 2**	**Base Model 3**	**Acc**
100X	MobileNetV3_Small	VGG11	VGG16	98.00
	MobileNetV3_Small	VGG11	AlexNet	97.80
	MobileNetv3_Small	VGG16	AlexNet	96.64
	VGG11	VGG16	AlexNet	96.92
	GoogleNet	VGG16	AlexNet	98.15
	GoogleNet	VGG11	AlexNet	98.13
	GoogleNet	VGG11	VGG16	98.19
	GoogleNet	MobileNetV3_Small	AlexNet	98.20
	GoogleNet	MobileNetV3_Small	VGG16	98.12
	GoogleNet	**MobileNetV3_Small**	**VGG11**	**98.24**

TABLE 4.21: Classification accuracies of the proposed ensemble method on three models taken at a time from five CNN models on 200X magnification of the BreakHis test set. The combination of models considered in this work is in bold font. Accuracy values are in % .

Magnification	**Base Model 1**	**Base Model 2**	**Base Model 3**	**Acc**
200X	MobileNetV3_Small	VGG11	VGG16	97.02
	MobileNetV3_Small	VGG11	AlexNet	98.18
	MobileNetv3_Small	VGG16	AlexNet	97.19
	VGG11	VGG16	AlexNet	98.02
	GoogleNet	VGG16	AlexNet	97.02
	GoogleNet	VGG11	AlexNet	98.35
	GoogleNet	VGG11	VGG16	97.85
	GoogleNet	MobileNetV3_Small	AlexNet	97.22
	GoogleNet	MobileNetV3_Small	VGG16	97.52
	GoogleNet	**MobileNetV3_Small**	**VGG11**	**98.67**

TABLE 4.22: Classification accuracies of the proposed ensemble method on three models taken at a time from five CNN models on 400X magnification of the BreakHis test set. The combination of models considered in this work is in bold font. Accuracy values are in % .

Magnification	**Base Model 1**	**Base Model 2**	**Base Model 3**	**Acc**
400X	MobileNetV3_Small	VGG11	VGG16	95.79
	MobileNetV3_Small	VGG11	AlexNet	95.43
	MobileNetv3_Small	VGG16	AlexNet	95.43
	VGG11	VGG16	AlexNet	92.30
	GoogleNet	VGG16	AlexNet	95.79
	GoogleNet	VGG11	AlexNet	95.61
	GoogleNet	VGG11	VGG16	95.61
	GoogleNet	MobileNetV3_Small	AlexNet	96.52
	GoogleNet	MobileNetV3_Small	VGG16	96.52
	GoogleNet	**MobileNetV3_Small**	**VGG11**	**96.16**

It can be observed from Tables 4.19 and 4.20 that rank based ensembling for all the combinations that comprise of GoogleNet produces higher results than the result of the independent GoogleNet model (refer to Table 4.18). Here, the ensembling results is compared only with the GoogleNet as it is the base model which produces the best

results independently shown for the first three levels of magnification (40X, 100X and 200X) as found in Table 4.18. Also, from Table 4.21, it is visible that almost all the rank based ensembling produces higher classification results irrespective of the inclusion of the best base model (i.e., GoogleNet). Whereas from Table 4.22, it is clear that the rank based ensembling produces similar classification result as MobileNetV3_Small in those cases that include this model, which is the best base model that produces higher classification result in 400X magnification.

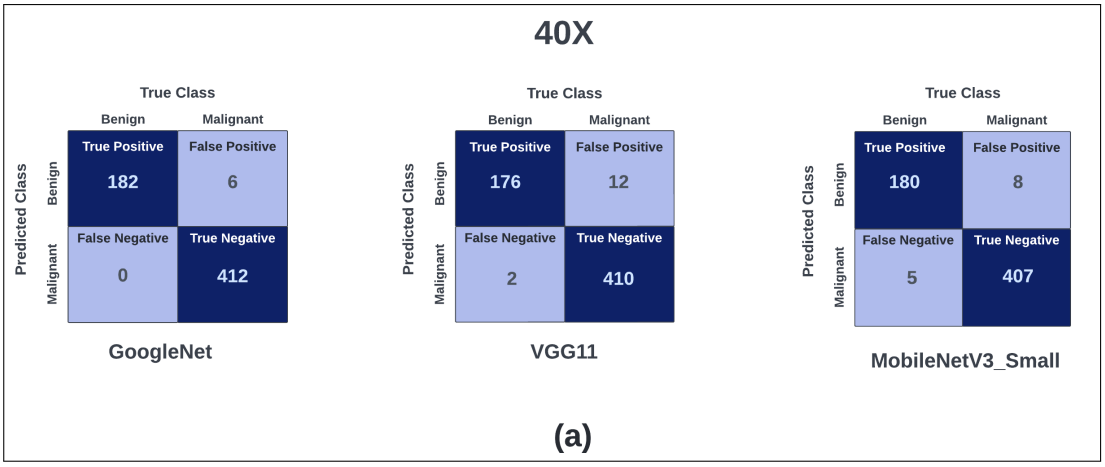
Further, this can be noted that the rank based ensembling of the GoogleNet, VGG11 and MobileNetV3_Small model yields superior classification accuracy for all three levels of magnification i.e., 40X, 100X and 200X. Moreover, it should be remarked that the suggested method surpasses the best individual base model (i.e., GoogleNet) with a classification accuracy improvement of 0.16% for both 40X and 100X magnifications. Furthermore, with a classification accuracy gain of 1.48% at 200X magnification, the performance of the proposed strategy is noticeably more encouraging than that of the GoogleNet model (the best base learner). Whereas, in the case of 400X magnification, the ensemble method yields a bit lower accuracy (96.16%) than the accuracy of the MobileNetV3_Small model (96.52%).

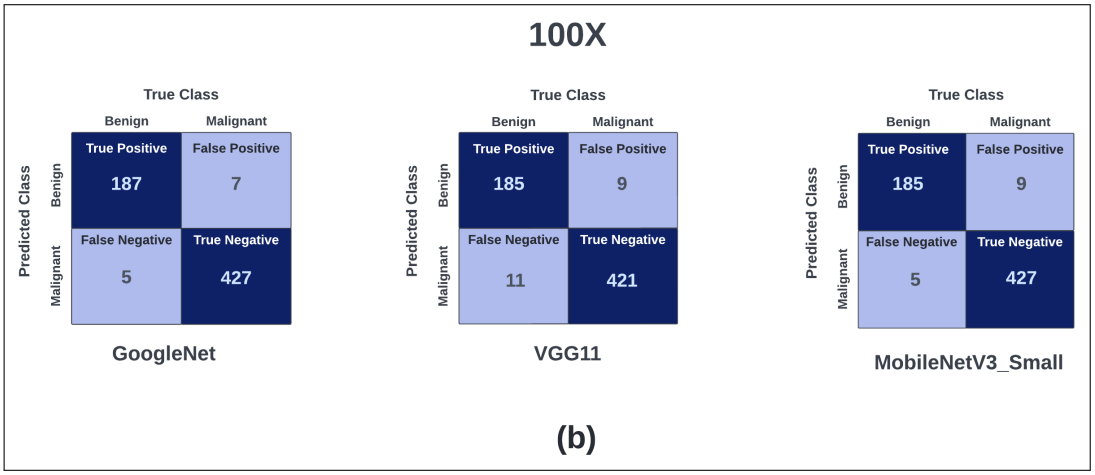
Hence, the above analysis proves that the proposed ensemble approach works efficiently for different combinations of CNN models as per the results in Tables 4.19, 4.20 and 4.21. Also from Tables 4.19, 4.20 and 4.21, it is observed that the combination of models (GoogleNet, VGG11, MobileNetV3_Small), which is considered for the proposed ensemble method, has a higher classification accuracy than the rest of the combinations, and for this reason, it also motivates us to choose this particular combination even for the 400X magnification level.

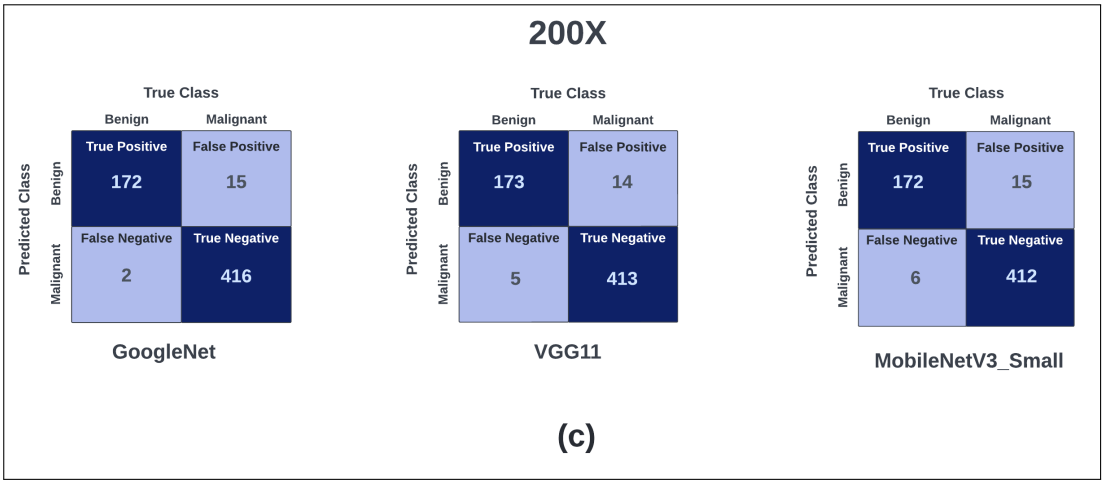
4.8.4 Analysis of Results

The confusion matrix is presented in Fig. 4.25, generated by three standard CNN models i.e., GoogleNet, VGG11 and MobileNetV3_Small on each set of magnifications i.e., 40X, 100X, 200X and 400X respectively. For the first three levels of magnifications i.e., 40X, 100X and 200X, GoogleNet have the highest accuracies of 99%, 98.08% and 97.19% respectively. It can also be interpreted from the confusion matrices that GoogleNet has the highest accuracy as compared to the rest two models. Also, it can be observed that

for the first three levels of magnification the total number of false positives and false negatives for GoogleNet is least compared to the rest two models.







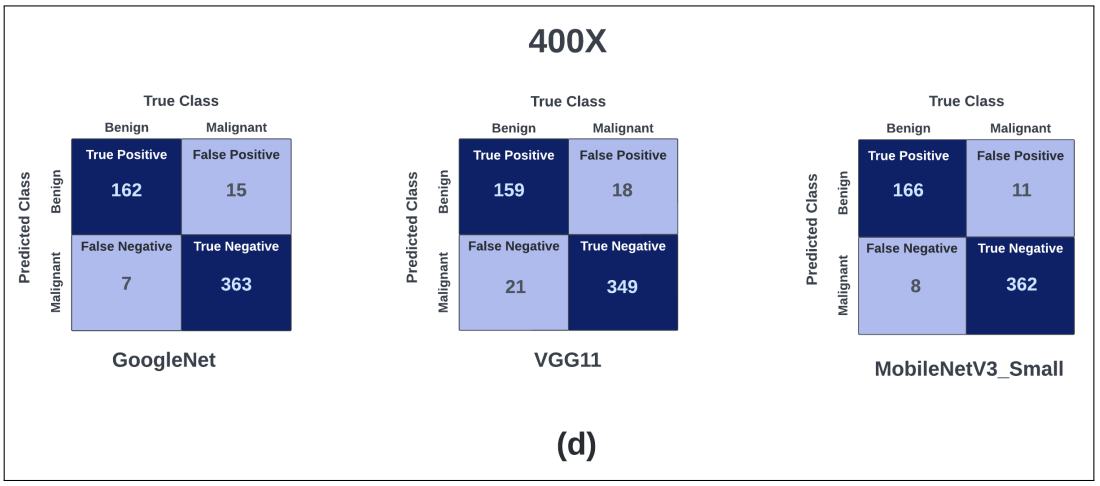


FIGURE 4.25: Confusion matrices generated by the CNN models on each set of a) 40X, b) 100X, c) 200X and d) 400X magnifications respectively

Values of standard performance metrics like Precision, Recall, F1 Score, AUC along with accuracy obtained by the proposed ensemble method on each set of magnifications are presented in Table 4.23.

TABLE 4.23: Detailed results of the proposed ensemble method on each set of magnification on the BreakHis test set. Here Acc, Pr, Re indicate accuracy, precision and recall respectively. All values are in %

Magnification	**Acc**	**Pr**	**Re**	**F1**	**AUC**
40X	99.16	99.40	98.67	99.02	98.67
100X	98.24	97.76	98.16	97.95	98.16
200X	98.67	98.90	98.01	98.44	98.01
400X	96.16	95.81	95.39	95.60	95.39

The confusion matrices of the proposed ensemble model are presented in Fig. 4.26 for each set of magnifications. For the first three levels of magnification (i.e., 40X, 100X and 200X) our proposed ensemble model has higher accuracy than each of the contributing models. It can be seen from the confusion matrices of the proposed ensemble model that the total number of false positives and false negatives is less than the GoogleNet model (in Fig. 4.25) which has the highest accuracy among the contributing models in the first three levels of magnification.

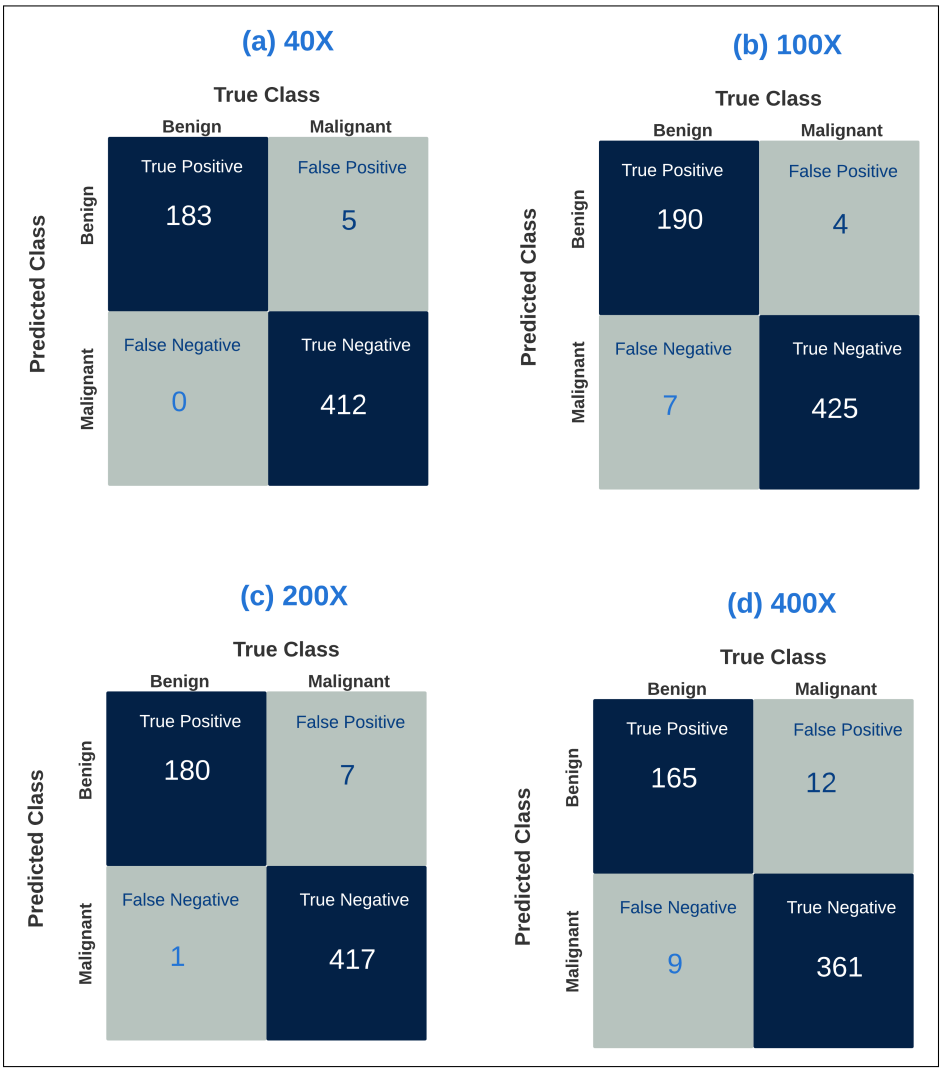


FIGURE 4.26: Confusion matrices generated by the proposed ensemble model obtained from the four levels of magnification

The loss curves of the three standard CNN models on the 40X magnification dataset, whose ensembling provides us with the best accuracy are shown in Fig. 4.27. The TL models are fine-tuned on the BreakHis dataset, where each of them is trained for 30 epochs. From the three curves, it is observed that the performance of the models converges at around 10 epochs. From the curves it can also be noted that among the three CNN models, the VGG11 model suffers from slight overfitting.

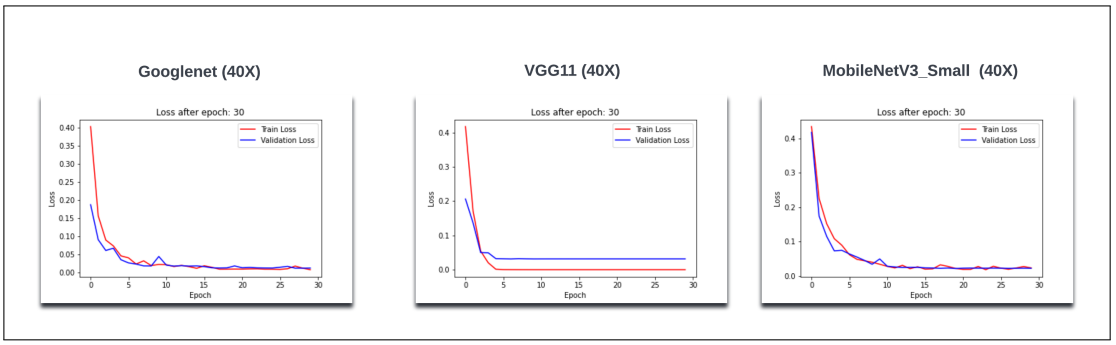


FIGURE 4.27: Loss curves obtained by the three standard CNN models on the 40X magnification

4.8.5 Ablation Study

An ablation analysis is performed to validate the contributions of each of the base classifiers chosen for our proposed ensemble methodology. It is observed that, with the addition of the base learners, there is a gradual increase in overall performance for the 40X, 100X and 200X magnification levels as observed in Tables 4.24, 4.25 and 4.26 respectively. Whereas for 400X magnification, it can be observed from Table 4.27 that combinations of GoogleNet and MobileNetV3_Small yields maximum results. The gain in performance has to be attributed to every model used in the proposed gamma function based ensembling method.

TABLE 4.24: Summary of ablation studies performed on 40X BreakHis test set

Magnification	**Model**	**Accuracy (%)**
40X	VGG11	97.66
	MobileNetV3_Small	97.83
	GoogleNet	99.00
	VGG11 + MobileNetV3_Small	98.5
	VGG11 + GoogleNet	99.0
	GoogleNet + MobileNetV3_Small	99.0
	Proposed Ensemble Method	**99.16**

TABLE 4.25: Summary of ablation studies performed on 100X BreakHis test set

Magnification	**Model**	**Accuracy (%)**
100X	VGG11	96.80
	MobileNetV3_Small	97.76
	GoogleNet	98.08
	VGG11 + MobileNetV3_Small	97.6
	VGG11 + GoogleNet	97.76
	GoogleNet + MobileNetV3_Small	97.72
	Proposed Ensemble Method	**98.24**

TABLE 4.26: Summary of ablation studies performed on 200X BreakHis test set

Magnification	**Model**	**Accuracy (%)**
200X	VGG11	96.85
	MobileNetV3_Small	96.52
	GoogleNet	97.19
	VGG11 + MobileNetV3_Small	98.18
	VGG11 + GoogleNet	98.34
	GoogleNet + MobileNetV3_Small	97.68
	Proposed Ensemble Method	**98.67**

TABLE 4.27: Summary of ablation studies performed on 400X BreakHis test set

Magnification	**Model**	**Accuracy (%)**
400X	VGG11	92.87
	MobileNetV3_Small	96.52
	GoogleNet	95.97
	VGG11 + MobileNetV3_Small	95.43
	VGG11 + GoogleNet	95.61
	GoogleNet + MobileNetV3_Small	**97.44**
	Proposed Ensemble Method	96.16

4.8.6 Comparison with Other Ensemble Approaches

Further experimentation is performed to compare the results of the proposed ensemble approach with some conventional ensemble approaches. Experiment with other ensemble approaches is conducted where the performances of the models which are used: GoogleNet, VGG11, MobileNet_V3_Small for ensembling over the BreakHis dataset. The obtained results are tabulated in Table 4.28. It can be observed from Table 4.28 that our proposed ensemble strategy outperforms all other ensembling methods for the three sets of magnification (40X, 100X, and 200X), while achieving SOTA accuracy for 400X magnification. In the majority voting-based ensemble, the class which obtains the maximum votes from the classifiers is predicted as the class of the sample. In the average probability ensemble, equal weightage is given to each contributing classifier, whereas in the weighted average based ensembling, different weights are assigned to the classifiers according to their performance on the dataset under consideration. In Gompertz function based ensembling [155], the Gompertz function is used for calculating the rank which is an exponential function of decreasing nature. The difference between the values of the ranks obtained from the Gompertz function is low as compared to the Gamma function which gives the Gamma function an advantage to perform well in the rank based ensembling operation.

TABLE 4.28: Performance comparison of the proposed ensemble method with other popular ensemble methods.

Ensemble technique	**Accuracy(%)**			
	40X	**100X**	**200X**	**400X**
Majority Voting	99.00	98.16	97.85	96.16
Average Probability	98.00	97.12	97.52	94.14
Weighted Average	99.00	98.20	98.51	96.00
Gompertz Function	98.06	98.14	98.01	96.16
Proposed Ensemble	**99.16**	**98.24**	**98.67**	**96.16**

4.8.7 Comparison with SOTA methods

Finally, the performance of the proposed method is compared with the results obtained by some of the recent SOTA methods in literature on the BreakHis dataset. The comparative results are tabulated in Table 4.29. It can be seen from Table 4.29 that at all four magnifications, the proposed model performs significantly better than all the preceding methods considered here. It is to be noted that due to the scarcity of publicly accessible medical data, the majority of approaches rely on TL for classification; however, end-to-end classification using TL is insufficient. Ensembling decision scores from different CNN models capture the complementing information offered by the models thereby boosting the overall performance. The findings of the proposed ensemble method is compared against the approaches presented in [156], [157], [158], [159], and [160], all of which extensively utilized TL-based techniques. Also, compared with [85], [86], [87] and [161] where the authors have used the fusion of CNN-based models. It should be noted that to provide a fair comparison, only the current research studies are considered where the performance results are provided for all the magnifications of the BreakHis dataset.

TABLE 4.29: Performance comparison of the proposed ensemble method with SOTA methods.

Method	**Accuracy(%)**			
	40X	**100X**	**200X**	**400X**
Wang et al. [85]	94.91	96.12	95.51	95.42
Wang et al. [156]	96.58	95.65	96.43	95.27
Sharma et al. [157]	96.25	96.25	95.74	94.11
Wang et al. [86]	92.71	94.52	94.03	93.54
Zerouaoui et al. [87]	92.61	92	93.93	91.73
Gour et al. [158]	87.40	87.26	91.15	86.27
Hu et al. [159]	94.03	90.66	90.47	88.79
Senan et al. [160]	95	91.5	91.8	95
Singh et al. [161]	80.8	82.76	86.55	85.8
Proposed Ensemble Method	**99.16**	**98.24**	**98.67**	**96.16**

4.8.8 Additional Experiment on the ICIAR dataset

To ensure the robustness of the proposed ensemble method experiment with another popular histopathology image dataset is conducted. This section analyzes the obtained result on the ICIAR-2018 histopathology dataset [78]. The ICIAR-2018 dataset consists of Hematoxylin and Eosin (H&E) stained breast histopathology microscopy and whole-slide images. It contains a total of 400 microscopy images labelled as normal, benign, in situ carcinoma and invasive carcinoma, where each of the four classes has 100 images. A 2-class classification is performed on this dataset, where the normal and benign labels are taken as in-carcinoma class, and in situ and invasive carcinoma labels are taken as carcinoma class. Therefore, each of the two classes consists of 200 images each. From the dataset, 100 test images are taken randomly and the remaining 300 images are augmented to 4800 images. The augmented dataset is then divided into training and validation sets an 80%-20% ratio. Table 4.30 shows the detailed class-wise distribution of test, validation and training sets of the said dataset.

TABLE 4.30: Class-wise distribution of training, validation and test sets in ICIAR-2018 dataset.

No. of images/class	**Carcinoma**	**In-Carcinoma**
Test set	50	50
Training set	1920	1920
Validation set	480	480

TABLE 4.31: Experimental results of the proposed ensemble method as well as the individual classifiers in ICIAR-2018 dataset. All values are in %.

Model	**Acc**	**Pre**	**Re**	**F1**	**AUC**
GoogleNet	95.95	96.26	95.96	95.95	95.92
MobileNetV3_Small	87.87	88.12	87.88	87.85	87.84
VGG11	90.90	90.92	90.91	90.91	90.90
Proposed Ensemble Method	**96.95**	**96.92**	**96.92**	**96.92**	**96.92**

From the results of Table 4.31, it is observed that the proposed ensemble method outperforms the individual base learners and thus, works well with the ICIAR-2018 dataset as well. Besides, Table 4.32 shows that the proposed ensemble method outperforms many SOTA methods by a good margin. Therefore, the obtained results on the ICIAR-2018

dataset prove the robustness of our proposed method for breast cancer classification using histology images.

TABLE 4.32: Performance comparison of the proposed ensemble method with SOTA methods on ICIAR-2018 dataset.

Method	**Accuracy (%)**
Kassani et al. [88]	95.00
Vizcarra et al. [162]	92.00
Nazeri et al. [163]	95.00
Yang et al. [164]	91.75
Bhowal et al. [84]	96.00
Proposed Ensemble Method	**96.95**

4.9 Discussion

This chapter explores and evaluates four distinct DL and optimization-based methodologies for the classification of breast cancer images derived from different imaging modalities, including ultrasound, thermography, mammography, and histopathology. Each proposed approach addresses unique aspects of breast cancer diagnosis and contribute to a more comprehensive understanding of how different FS techniques combined with DL techniques can aid in early detection and treatment planning.

The first methodology focuses on breast ultrasound imaging, where a novel deep feature selection technique called U-WOA is introduced. This method demonstrate SOTA performance with 99.1% on the BUSI dataset and shows promising generalization capabilities on unseen data.

The second methodology involves a lightweight DL model applied to thermographic images for breast cancer classification. Utilizing SqueezeNet 1.1 for deep feature extraction and a hybrid FS method combining chaotic maps-assisted GA and GWO, the model achieves perfect classification accuracy (100%) on the DMR-IR dataset using only 3% of the original features. This emphasizes the strength of integrating lightweight networks with efficient feature reduction techniques, which is especially beneficial in resource-constrained environments or point-of-care systems.

The third proposed method deals with mammography images and leverages the VGG16 model with a GAP layer to introduce attention mechanisms. This is followed by a FS

strategy involving a ABHC local search embedded SSD algorithm, which significantly reduces the feature dimensionality while maintaining excellent classification performance on the DDSM dataset. With just 25% of the original deep features, the model achieves high accuracy, precision, and recall, demonstrating the effectiveness of combining attention with optimization-based feature selection.

The final methodology introduces an ensemble approach using three established CNN models—GoogleNet, VGG11, and MobileNetV3_Small—for classifying breast histopathology images. A novel Gamma function-based score fusion technique is used to combine model outputs, enabling the ensemble to achieve superior performance across multiple magnification levels of the BreakHis dataset. The robustness of the approach is further validated on the ICIAR-2018 dataset, showcasing its adaptability across different histopathological imaging scenarios. This ensemble strategy underscores the importance of integrating multiple model perspectives for improving diagnostic reliability and accuracy.

Overall, the diverse methodologies presented in this chapter highlight the potential of DL and optimization-based FS algorithms in enhancing the breast cancer classification. Each technique demonstrates unique strengths suited to different imaging modalities and clinical contexts.

Chapter 5

Conclusion and Future Directions

Breast cancer remains a leading cause of cancer-related mortality among women globally. Early and accurate diagnosis is critical for improving survival rates and informing effective treatment strategies. However, diagnostic accuracy is often hindered by the inherent complexity of tumor presentations, overlap with benign conditions, and inconsistencies across imaging modalities.

DL based method has revolutionized medical imaging, offering unprecedented capabilities in automated detection including segmentation and classification. Techniques such as CNNs, attention mechanisms, and advanced segmentation architectures have demonstrated remarkable performance in breast cancer analysis. These models learn directly from raw image data, reducing dependency on handcrafted features and enabling more consistent and scalable diagnostic solutions.

Despite these advances, several challenges remain in deploying DL models in real-world clinical settings. These include limited annotated data, difficulty in interpreting model decisions, and computational constraints. This thesis addressed these challenges through the development of lightweight, interpretable, and robust DL models, integrating domain knowledge, attention modules, and optimization-driven FS to enhance both performance and practical utility. This thesis has explored and developed advanced frameworks for breast cancer detection across multiple imaging modalities, with a focus on image segmentation and classification tasks.

The first aspect of this thesis has been an extensive review of the SOTA techniques in breast cancer detection across different modality of imaging. The literature analysis has revealed a notable shift in the field, moving from traditional handcrafted methods to

more sophisticated DL-based approaches. Notably, the integration of CNNs, attention mechanisms, and optimization techniques has significantly advanced the accuracy and efficiency of breast cancer classification.

In terms of image segmentation, U-Net and its variants played a crucial role. This thesis explored the U-Net architecture in depth, analyzing the benefits of its encoder-decoder structure and skip connections. The combination of these components allows for the extraction of features at varying levels of granularity, leading to more accurate segmentation of breast tumors. Therefore, various enhancements to the U-Net architecture were proposed, such as the DAU-Net (presented in **Chapter 3**), which integrates powerful attention mechanisms to focus on the tumor regions while suppressing irrelevant background features. The dual attention mechanism in DAU-Net combines PCBAM and SWA, effectively captures context-aware features, spatial dependencies, and global contextual information. This resulted in a Dice score of 74.23% on the BUSI dataset and 78.58% on the UDIAT dataset. Another significant contribution of this thesis is the introduction of the $EU^2 - Net$ model (presented in **Chapter 3**), an ensemble approach that enhances the U-Net architecture by reducing computational complexity while maintaining high segmentation accuracy. The incorporation of attention mechanisms, feature fusion strategies, and edge detection further improves the model's ability to segment challenging regions in breast ultrasound images, achieving a Dice score of 74.73% on the BUSI dataset and an impressive 83.47% on the UDIAT dataset.

Additionally, AWGUNet (presented in **Chapter 3**) combining the U-Net model with a DenseNet-121 backbone for nuclei segmentation in histopathological images demonstrated superior performance in capturing both spatial and contextual information, addressing challenges such as uncertain cell boundaries and complex staining patterns. In this model, the wavelet guided attention mechanism captured edge information for spatial guidance, boosting the performance of U-Net. This resulted in a Dice score of 79.46% on the MoNuSeg dataset and 81.65% on the TNBC dataset for the segmentation of neuclei in breast histopathology images.

This thesis has undertaken a comprehensive investigation into the development of advanced DL based frameworks for the classification of breast cancer using various imaging modalities, including ultrasound, thermograms, mammograms, and histopathological images. The goal has been to address the limitations of existing models and introduce more robust, accurate, and computationally efficient solutions for early and reliable

breast cancer classification. One of the earliest contributions of this thesis was the development of a feature selection framework for classifying breast tumors in ultrasound images. This method used TL-based model for feature extraction and employed Spearman Rank Correlation Coefficient-based ranking to select the most relevant features for classification. Recognizing the limitations of deterministic selection approaches, the research further investigated meta-heuristic optimization techniques—specifically, a novel unsupervised FS framework called U-WOA (presented in **Chapter 4**), which integrates statistical ranking filters with the WOA. In this method, feature extraction is done using the VGG16 model and for FS the proposed U-WOA method is used. This unsupervised FS method effectively identified informative features without requiring class labels, enabling more robust classification in challenging ultrasound datasets, achieved a classification accuracy of 99.1% on the widely-used BUSI ultrasound dataset.

The concept of balancing exploration and exploitation in meta-heuristic algorithms was applied to thermogram image based breast cancer classification (discussed in **Chapter 4**). Here, a lightweight and computationally efficient model was proposed using SqueezeNet 1.1 for feature extraction, combined with a hybrid FS mechanism. This hybrid method integrated a sinusoidal chaotic map-based GA with the GWO, successfully identifying optimal feature subsets while reducing memory and processing demands. This GA-GWO hybrid structure achieved the SOTA classification accuracy of 100%, while using only 3% of the superior features extracted from the lightweight SqueezeNet 1.1 model on the DMR-IR thermal breast cancer images.

Further extending this concept to mammographic image classification, an attention-enhanced TL model (VGG16) was used in conjunction with a wrapper-based FS framework utilizing the SSD algorithm (discussed in **Chapter 4**). The integration of ABHC local search with SSD algorithm improved feature exploitation, especially in the challenging task of distinguishing malignant mass lesions. This method achieved an improved classification accuracy of 96.07% on the DDSM mammogram dataset.

Another key outcome of this thesis was the exploration of ensemble learning strategies for breast cancer classification, particularly with histopathological images. Traditional ensemble techniques—such as majority voting and weighted averaging—were found to be limited by their static weighting mechanisms and inability to account for classifier interactions. To overcome this, a novel rank-based ensemble approach was proposed (discussed in **Chapter 4**). This method adaptively fused predictions from multiple pre-trained CNN models (GoogleNet, VGG11, and MobileNetV3_Small) using the Gamma

function to assign dynamic priorities to classifiers based on their confidence scores for each sample. This approach not only improved classification accuracy but also enhanced robustness by minimizing prediction variance and exploiting complementary strengths of individual classifiers. This method has achieved classification accuracies of 99.16%, 98.24%, 98.67%, and 96.16% on the BreakHis dataset at 40×, 100×, 200×, and 400× magnifications, respectively, and an accuracy of 96.95% on the ICIAR dataset.

5.1 Limitations of the Thesis

While the findings and methodologies presented in this thesis demonstrate promising results, some challenges need to be acknowledged.

Segmentation Challenges: The proposed DAU-Net model demonstrates relatively lower precision and recall, which suggests that some non-tumorous regions are misclassified as tumorous and vice versa. The model encounters difficulties, leading to deviations from the ground truth segmentation. These challenges arise from the complexity of ultrasound images, which often include noisy, low-contrast regions or ambiguous features that are difficult to accurately delineate. Similarly, $EU^2 - Net$ fails to consistently match the ground truth mask, particularly in cases involving lighter shades or complex, dark backgrounds. These factors likely contribute to errors in distinguishing foreground from background pixels.

Classification Challenges: In the breast cancer classification techniques employed in this thesis, the FS methods, though effective, are highly sensitive to initialization. A poor initial selection of features can lead to the exclusion of important attributes, which may limit the accuracy of the final classification model. Additionally, meta-heuristic algorithms used in the classification process are highly sensitive to hyperparameters. These algorithms require careful tuning to optimize their performance for specific datasets, which can hinder the model's generalization ability and scalability across different contexts. The reliance on fine-tuned hyperparameters could make the model less adaptable when applied to other datasets or tasks.

5.2 Directions for Future Research

Although the techniques presented in this thesis show promising results compared to current SOTA methods, there is still room for further improvement. Potential directions for future work include the following:

- Investigating data-efficient learning approaches, such as Few-Shot and Self-Supervised learning, to reduce reliance on large annotated datasets, which are often scarce in the medical domain.
- Further investigation into pruning techniques besides feature selection and efficient lightweight architectures may facilitate deployment in low-resource settings without sacrificing accuracy.
- Examining the multimodal data like use of non-imaging data, such as patient history, genetic information, and radiomic features, in conjunction with imaging data for breast cancer diagnosis. DL models that combine both imaging and non-imaging data have the potential to improve the accuracy and robustness of breast cancer detection systems and it could offer a more holistic diagnostic view and improve overall model robustness.
- Contributing to the creation and standardization of large, diverse datasets will enhance reproducibility and model training, paving the way for more generalizable solutions.
- Incorporating temporal data, such as patient history and time-sequenced imaging, using models like LSTMs or Transformers. This could enable better tracking of disease progression and treatment response over time.

In summery, this thesis presents a comprehensive framework for automated breast cancer detection using advanced DL techniques. The proposed models offer a step forward in improving diagnostic accuracy, reducing clinician workload, and advancing the practical deployment of AI-driven medical tools.

References

[1] Barsha Abhisheka, Saroj Kumar Biswas, and Biswajit Purkayastha. A comprehensive review on breast cancer detection, classification and segmentation using deep learning. *Archives of Computational Methods in Engineering*, 30(8):5023–5052, 2023.

[2] Jocelyn Hirschman, Steven Whitman, and David Ansell. The black: white disparity in breast cancer mortality: the example of chicago. *Cancer Causes & Control*, 18:323–333, 2007.

[3] Ravi Mehrotra and Kavita Yadav. Breast cancer in india: Present scenario and the challenges ahead. *World Journal of Clinical Oncology*, 13(3):209, 2022.

[4] Zohre Momenimovahed and Hamid Salehiniya. Epidemiological characteristics of and risk factors for breast cancer in the world. *Breast Cancer: Targets and Therapy*, pages 151–164, 2019.

[5] Bernard Levin and Gerald P Murphy. Revision in american cancer society recommendations for the earlydetection of colorectal cancer. *CA: A Cancer Journal for Clinicians*, 42(5):296–299, 1992.

[6] Fiona J Gilbert, Lorraine Tucker, Maureen GC Gillan, Paula Willsher, Julie Cooke, Karen A Duncan, Michael J Michell, Hilary M Dobson, Yit Yoong Lim, Tamara Suaris, et al. Accuracy of digital breast tomosynthesis for depicting breast cancer subgroups in a uk retrospective reading study (tommy trial). *Radiology*, 277(3):697–706, 2015.

[7] Benjamin Q Huynh, Hui Li, and Maryellen L Giger. Digital mammographic tumor classification using transfer learning from deep convolutional neural networks. *Journal of Medical Imaging*, 3(3):034501–034501, 2016.

[8] Afsaneh Jalalian, Syamsiah BT Mashohor, Hajjah Rozi Mahmud, M Iqbal B Saripan, Abdul Rahman B Ramli, and Babak Karasfi. Computer-aided detection/-diagnosis of breast cancer in mammography and ultrasound: a review. *Clinical imaging*, 37(3):420–426, 2013.

[9] Xiaofeng Qi, Fasheng Yi, Lei Zhang, Yao Chen, Yong Pi, Yuanyuan Chen, Jixiang Guo, Jianyong Wang, Quan Guo, Jilan Li, et al. Computer-aided diagnosis of breast cancer in ultrasonography images by deep learning. *Neurocomputing*, 472:152–165, 2022.

[10] Anton S Becker, Michael Mueller, Elina Stoffel, Magda Marcon, Soleen Ghafoor, and Andreas Boss. Classification of breast cancer in ultrasound imaging using a generic deep learning analysis software: a pilot study. *The British journal of radiology*, 91(1083):20170576, 2018.

[11] Xiaomin Zhou, Chen Li, Md Mamunur Rahaman, Yudong Yao, Shiliang Ai, Changhao Sun, Qian Wang, Yong Zhang, Mo Li, Xiaoyan Li, et al. A comprehensive review for breast histopathology image analysis using classical and deep neural networks. *IEEE Access*, 8:90931–90956, 2020.

[12] Shyam Lal, Devikalyan Das, Kumar Alabhya, Anirudh Kanfade, Aman Kumar, and Jyoti Kini. Nucleisegnet: Robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images. *Computers in Biology and Medicine*, 128:104075, 2021.

[13] Gustavo Carneiro, Jacinto Nascimento, and Andrew P Bradley. Automated analysis of unregistered multi-view mammograms with deep learning. *IEEE transactions on medical imaging*, 36(11):2355–2365, 2017.

[14] Ghulam Murtaza, Liyana Shuib, Ainuddin Wahid Abdul Wahab, Ghulam Mujtaba, Ghulam Mujtaba, Henry Friday Nweke, Mohammed Ali Al-garadi, Fariha Zulfiqar, Ghulam Raza, and Nor Aniza Azmi. Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges. *Artificial Intelligence Review*, 53:1655–1720, 2020.

[15] Deborah A Kennedy, Tanya Lee, and Dugald Seely. A comparative review of thermography as a breast cancer screening technique. *Integrative cancer therapies*, 8(1):9–16, 2009.

[16] Aigerim Mashekova, Yong Zhao, Eddie YK Ng, Vasilios Zarikas, Sai Cheong Fok, and Olzhas Mukhmetov. Early detection of the breast cancer using infrared technology–a comprehensive review. *Thermal science and engineering progress*, 27:101142, 2022.

[17] YK Ng, LN Ung, FC Ng, and E LSJ Sim. Statistical analysis of healthy and malignant breast thermography. *Journal of medical engineering & technology*, 25(6):253–263, 2001.

[18] Rishav Pramanik, Payel Pramanik, and Ram Sarkar. Breast cancer detection in thermograms using a hybrid of ga and gwo based deep feature selection method. *Expert Systems with Applications*, 219:119643, 2023.

[19] Heang-Ping Chan, Ravi K Samala, Lubomir M Hadjiiski, and Chuan Zhou. Deep learning in medical image analysis. *Deep learning in medical image analysis: challenges and applications*, pages 3–21, 2020.

[20] Keiron O'shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.

[21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[22] Ailiang Lin, Bingzhi Chen, Jiayu Xu, Zheng Zhang, Guangming Lu, and David Zhang. Ds-transunet: Dual swin transformer u-net for medical image segmentation. *IEEE Transactions on Instrumentation and Measurement*, 71:1–15, 2022.

[23] EYK Ng and NM Sudharsan. Numerical computation as a tool to aid thermographic interpretation. *Journal of medical engineering & technology*, 25(2):53–60, 2001.

[24] Essam H Houssein, Marwa M Emam, Abdelmgeid A Ali, and Ponnuthurai Nagaratnam Suganthan. Deep and machine learning techniques for medical imaging-based breast cancer: A comprehensive review. *Expert Systems with Applications*, 167:114161, 2021.

[25] Payel Pramanik, Rishav Pramanik, Friedhelm Schwenker, and Ram Sarkar. Dbu-net: Dual branch u-net for tumor segmentation in breast ultrasound images. *Plos one*, 18(11):e0293615, 2023.

[26] Raman Maini and Himanshu Aggarwal. Study and comparison of various image edge detection techniques. *International journal of image processing (IJIP)*, 3(1):1–11, 2009.

[27] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[28] Zhengxin Zhang, Qingjie Liu, and Yunhong Wang. Road extraction by deep residual u-net. *IEEE Geoscience and Remote Sensing Letters*, 15(5):749–753, 2018.

[29] Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R Zaiane, and Martin Jagersand. U2-net: Going deeper with nested u-structure for salient object detection. *Pattern recognition*, 106:107404, 2020.

[30] Payel Pramanik, Souradeep Mukhopadhyay, Dmitrii Kaplun, and Ram Sarkar. A deep feature selection method for tumor classification in breast ultrasound images. In *International conference on mathematics and its applications in new computer systems*, pages 241–252. Springer, 2021.

[31] Tansel Dokeroglu, Ayça Deniz, and Hakan Ezgi Kiziloz. A comprehensive survey on recent metaheuristics for feature selection. *Neurocomputing*, 494:269–296, 2022.

[32] Payel Pramanik, Rishav Pramanik, Anurup Naskar, Seyedali Mirjalili, and Ram Sarkar. U-woa: an unsupervised whale optimization algorithm based deep feature selection method for cancer detection in breast ultrasound images. In *Handbook of Whale Optimization Algorithm*, pages 179–191. Elsevier, 2024.

[33] Rishav Pramanik, Payel Pramanik, and Ram Sarkar. Breast cancer detection in thermograms using a hybrid of ga and gwo based deep feature selection method. *Expert Systems with Applications*, 219:119643, 2023.

[34] Payel Pramanik, Souradeep Mukhopadhyay, Seyedali Mirjalili, and Ram Sarkar. Deep feature selection using local search embedded social ski-driver optimization

algorithm for breast cancer detection in mammograms. *Neural Computing and Applications*, 35(7):5479–5499, 2023.

[35] Samriddha Majumdar, Payel Pramanik, and Ram Sarkar. Gamma function based ensemble of cnn models for breast cancer detection in histopathology images. *Expert Systems with Applications*, 213:119022, 2023.

[36] Michael Heath, Kevin Bowyer, Daniel Kopans, P Kegelmeyer Jr, Richard Moore, Kyong Chang, and S Munishkumaran. Current status of the digital database for screening mammography. In *Digital Mammography: Nijmegen, 1998*, pages 457–460. Springer, 1998.

[37] Rebecca Sawyer Lee, Francisco Gimenez, Assaf Hoogi, Kanae Kawai Miyake, Mia Gorovoy, and Daniel L Rubin. A curated mammography data set for use in computer-aided detection and diagnosis research. *Scientific data*, 4(1):1–9, 2017.

[38] J Suckling, J Parker, D Dance, S Astley, I Hutt, C Boggis, I Ricketts, E Stamatakis, N Cerneaz, S Kok, et al. Mammographic image analysis society (mias) database. In *International Congress Series*, volume 1069, 2015.

[39] Li Shen, Laurie R Margolies, Joseph H Rothstein, Eugene Fluder, Russell McBride, and Weiva Sieh. Deep learning to improve breast cancer detection on screening mammography. *Scientific reports*, 9(1):12495, 2019.

[40] Aditya Khamparia, Subrato Bharati, Prajoy Podder, Deepak Gupta, Ashish Khanna, Thai Kim Phung, and Dang NH Thanh. Diagnosis of breast cancer based on modern mammography using hybrid transfer learning. *Multidimensional systems and signal processing*, 32:747–765, 2021.

[41] Lenin G Falconi, Maria Perez, Wilbert G Aguilar, and Aura Conci. Transfer learning and fine tuning in breast mammogram abnormalities classification on cbis-ddsm database. *Adv. Sci. Technol. Eng. Syst. J*, 5(2):154–165, 2020.

[42] Mugahed A Al-Antari, Mohammed A Al-Masni, Mun-Taek Choi, Seung-Moo Han, and Tae-Seong Kim. A fully integrated computer-aided diagnosis system for digital x-ray mammograms via deep learning detection, segmentation, and classification. *International journal of medical informatics*, 117:44–54, 2018.

[43] Olaide N Oyelade and Absalom E Ezugwu. Characterization of abnormalities in breast cancer images using nature-inspired metaheuristic optimized convolutional neural networks model. *Concurrency and Computation: Practice and Experience*, 34(4):e6629, 2022.

[44] Khosro Rezaee, Adele Rezaee, Negar Shaikhi, and Javad Haddadnia. Multi-mass breast cancer classification based on hybrid descriptors and memetic meta-heuristic learning. *SN Applied Sciences*, 2(7):1297, 2020.

[45] Walid Al-Dhabyani, Mohammed Gomaa, Hussien Khaled, and Aly Fahmy. Dataset of breast ultrasound images. *Data in brief*, 28:104863, 2020.

[46] Moi Hoon Yap, Gerard Pons, Joan Marti, Sergi Ganau, Melcior Sentis, Reyer Zwiggelaar, Adrian K Davison, and Robert Marti. Automated breast ultrasound lesions detection using convolutional neural networks. *IEEE journal of biomedical and health informatics*, 22(4):1218–1226, 2017.

[47] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[48] Ruobing Huang, Mingrong Lin, Haoran Dou, Zehui Lin, Qilong Ying, Xiaohong Jia, Wenwen Xu, Zihan Mei, Xin Yang, Yijie Dong, et al. Boundary-rendering network for breast lesion segmentation in ultrasound images. *Medical image analysis*, 80:102478, 2022.

[49] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*, 39(6):1856–1867, 2019.

[50] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, and Jian Wu. Unet 3+: A full-scale connected unet for medical image segmentation. In *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, pages 1055–1059. IEEE, 2020.

[51] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

[52] Yi Wang, Haoran Dou, Xiaowei Hu, Lei Zhu, Xin Yang, Ming Xu, Jing Qin, Pheng-Ann Heng, Tianfu Wang, and Dong Ni. Deep attentive features for prostate segmentation in 3d transrectal ultrasound. *IEEE transactions on medical imaging*, 38(12):2768–2778, 2019.

[53] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[54] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pages 3–11. Springer, 2018.

[55] Zhemin Zhuang, Nan Li, Alex Noel Joseph Raj, Vijayalakshmi GV Mahesh, and Shunmin Qiu. An rdau-net model for lesion segmentation in breast ultrasound images. *PloS one*, 14(8):e0221535, 2019.

[56] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

[57] Aleksandar Vakanski, Min Xian, and Phoebe E Freer. Attention-enriched deep learning model for breast tumor segmentation in ultrasound images. *Ultrasound in medicine & biology*, 46(10):2819–2833, 2020.

[58] Haeyun Lee, Jinhyoung Park, and Jae Youn Hwang. Channel attention module with multiscale grid average pooling for breast cancer segmentation in an ultrasound image. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, 67(7):1344–1353, 2020.

[59] Xianli Zhang, Yinbin Zhang, Buyue Qian, Xiaotong Liu, Xiaoyu Li, Xudong Wang, Changchang Yin, Xin Lv, Lingyun Song, and Liang Wang. Classifying breast cancer histopathological images using a robust artificial neural network architecture. In *Bioinformatics and Biomedical Engineering: 7th International Work-Conference, IWBBIO 2019, Granada, Spain, May 8-10, 2019, Proceedings, Part I 7*, pages 204–215. Springer, 2019.

[60] Luyi Han, Yunzhi Huang, Haoran Dou, Shuai Wang, Sahar Ahamad, Honghao Luo, Qi Liu, Jingfan Fan, and Jiang Zhang. Semi-supervised segmentation of lesion from breast ultrasound images with attentional generative adversarial network. *Computer methods and programs in biomedicine*, 189:105275, 2020.

[61] Ahmed Iqbal and Muhammad Sharif. Bts-st: Swin transformer network for segmentation and classification of multimodality breast cancer images. *Knowledge-Based Systems*, 267:110393, 2023.

[62] Xiangjun Shi, H.D. Cheng, Liming Hu, Wen Ju, and Jiawei Tian. Detection and classification of masses in breast ultrasound images. *Digital Signal Processing*, 20:824–836, 05 2010.

[63] Min-Chun Yang, Woo Kyung Moon, Yu-Chiang Frank Wang, Min Bae, C. F. Huang, Jeon Chen, and Ruey-Feng Chang. Robust texture analysis using multi-resolution gray-scale invariant features for breast sonographic tumor diagnosis. *IEEE transactions on medical imaging*, 32, 08 2013.

[64] Lingyun Cai, Xin Wang, Yuanyuan Wang, Yi Guo, Jinhua Yu, and Yi Wang. Robust phase-based texture descriptor for classification of breast ultrasound images. *BioMedical Engineering OnLine*, 14, 12 2015.

[65] Ting Xiao, Lei Liu, Kai Li, Wenjian Qin, Shaode Yu, and Zhicheng Li. Comparison of transferred deep neural networks in ultrasonic breast masses discrimination. *BioMed Research International*, 2018, 2018.

[66] Tariq Sadad, Ayyaz Hussain, Asim Munir, Muhammad Habib, Sajid Ali Khan, Shariq Hussain, Shunkun Yang, and Mohammed Alawairdhi. Identification of breast malignancy by marker-controlled watershed transformation and hybrid feature set for healthcare. *Applied Sciences*, 10(6):1900, 2020.

[67] Arnab K Mishra, Pinki Roy, Sivaji Bandyopadhyay, and Sujit K Das. Breast ultrasound tumour classification: A machine learning—radiomics based approach. *Expert Systems*, 38(7):e12713, 2021.

[68] Woo Kyung Moon, Yan-Wei Lee, Hao-Hsiang Ke, Su Hyun Lee, Chiun-Sheng Huang, and Ruey-Feng Chang. Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. *Computer methods and programs in biomedicine*, 190:105361, 2020.

[69] Kiran Jabeen, Muhammad Attique Khan, Majed Alhaisoni, Usman Tariq, Yu-Dong Zhang, Ameer Hamza, Artūras Mickus, and Robertas Damaševičius. Breast cancer classification from ultrasound images using probability-based optimal deep learning feature fusion. *Sensors*, 22(3):807, 2022.

[70] Arnab Kumar Mishra, Pinki Roy, Sivaji Bandyopadhyay, and Sujit Kumar Das. Feature fusion based machine learning pipeline to improve breast cancer prediction. *Multimedia Tools and Applications*, pages 1–29, 2022.

[71] Mrinal Kanti Bhowmik, Usha Rani Gogoi, Gautam Majumdar, Debotosh Bhattacharjee, Dhritiman Datta, and Anjan Kumar Ghosh. Designing of ground-truth-annotated dbt-tu-ju breast thermogram database toward early abnormality prediction. *IEEE journal of biomedical and health informatics*, 22(4):1238–1249, 2017.

[72] Roslidar Roslidar, Khairun Saddami, Fitri Arnia, Maimun Syukri, and Khairul Munadi. A study of fine-tuning cnn models based on thermal imaging for breast cancer classification. In *2019 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom)*, pages 77–81. IEEE, 2019.

[73] Sourav Pramanik, Debotosh Bhattacharjee, Mita Nasipuri, and Ondrej Krejcar. Linpe-bl: A local descriptor and broad learning for identification of abnormal breast thermograms. *IEEE Transactions on Medical Imaging*, 40(12):3919–3931, 2021.

[74] Caroline Barcelos Gonçalves, Jefferson R Souza, and Henrique Fernandes. Cnn architecture optimization using bio-inspired algorithms for breast cancer detection in infrared images. *Computers in Biology and Medicine*, 142:105205, 2022.

[75] Raquel Sánchez-Cauce, Jorge Pérez-Martín, and Manuel Luque. Multi-input convolutional neural network for breast cancer detection using thermal images and clinical data. *Computer Methods and Programs in Biomedicine*, 204:106045, 2021.

[76] J. Zuluaga-Gomez, Z. Al Masry, K. Benaggoune, S. Meraghni, and N. Zerhouni. A cnn-based methodology for breast cancer diagnosis using thermal images. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 9(2):131–145, 2021.

[77] Fabio A Spanhol, Luiz S Oliveira, Caroline Petitjean, and Laurent Heutte. A dataset for breast cancer histopathological image classification. *Ieee transactions on biomedical engineering*, 63(7):1455–1462, 2015.

[78] Guilherme Aresta, Teresa Araújo, Scotty Kwok, Sai Saketh Chennamsetty, Mohammed Safwan, Varghese Alex, Bahram Marami, Marcel Prastawa, Monica Chan, Michael Donovan, et al. Bach: Grand challenge on breast cancer histology images. *Medical image analysis*, 56:122–139, 2019.

[79] Peter Naylor, Marick Laé, Fabien Reyal, and Thomas Walter. Segmentation of nuclei in histopathology images by deep regression of the distance map. *IEEE transactions on medical imaging*, 38(2):448–459, 2018.

[80] Anusree Kanadath, J Angel Arul Jothi, and Siddhaling Urolagin. Air-unet++: a deep learning framework for histopathology image segmentation and detection. *Multimedia Tools and Applications*, 83(19):57449–57475, 2024.

[81] Yilong Yang, Srinandan Dasmahapatra, and Sasan Mahmoodi. Ads_unet: A nested unet for histopathology image segmentation. *Expert Systems with Applications*, 226:120128, 2023.

[82] Iqra Kiran, Basit Raza, Areesha Ijaz, and Muazzam A Khan. Denseres-unet: Segmentation of overlapped/clustered nuclei from multi organ histopathology images. *Computers in biology and medicine*, 143:105267, 2022.

[83] Anu Singha and Mrinal Kanti Bhowmik. Alexsegnet: an accurate nuclei segmentation deep learning model in microscopic images for diagnosis of cancer. *Multimedia Tools and Applications*, 82(13):20431–20452, 2023.

[84] Pratik Bhowal, Subhankar Sen, Juan D Velasquez, and Ram Sarkar. Fuzzy ensemble of deep learning models using choquet fuzzy integral, coalition game and information theory for breast cancer histology classification. *Expert Systems with Applications*, 190:116167, 2022.

[85] Pin Wang, Pufei Li, Yongming Li, Jin Xu, Fang Yan, and Mingfeng Jiang. Deep manifold feature fusion for classification of breast histopathology images. *Digital Signal Processing*, 123:103400, 2022.

[86] Pin Wang, Jiaxin Wang, Yongming Li, Pufei Li, Linyu Li, and Mingfeng Jiang. Automatic classification of breast cancer histopathological images based on deep feature fusion and enhanced routing. *Biomedical Signal Processing and Control*, 65:102341, 2021.

[87] Hasnae Zerouaoui and Ali Idri. Deep hybrid architectures for binary classification of medical breast cancer images. *Biomedical Signal Processing and Control*, 71:103226, 2022.

[88] Sara Hosseinzadeh Kassani, Peyman Hosseinzadeh Kassani, Michal J Wesolowski, Kevin A Schneider, and Ralph Deters. Classification of histopathological biopsy images using ensemble of deep learning networks. *arXiv preprint arXiv:1909.11870*, 2019.

[89] Payel Pramanik, Ayush Roy, Erik Cuevas, Marco Perez-Cisneros, and Ram Sarkar. Dau-net: Dual attention-aided u-net for segmenting tumor in breast ultrasound images. *Plos one*, 19(5):e0303670, 2024.

[90] Ayush Roy, Payel Pramanik, and Ram Sarkar. Eu 2–net: A parameter efficient ensemble model with attention-aided triple feature fusion for tumor segmentation in breast ultrasound images. *IEEE Transactions on Instrumentation and Measurement*, 2024.

[91] Ayush Roy, Payel Pramanik, Dmitrii Kaplun, Sergei Antonov, and Ram Sarkar. Awgunet: Attention-aided wavelet guided u-net for nuclei segmentation in histopathology images. In *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 1–4. IEEE, 2024.

[92] Toufique A Soomro, Ahmed J Afifi, Junbin Gao, Olaf Hellwich, Manoranjan Paul, and Lihong Zheng. Strided u-net model: Retinal vessels segmentation using dice

loss. In *2018 Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–8. IEEE, 2018.

[93] Shruti Jadon. A survey of loss functions for semantic segmentation. In *2020 IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB)*, pages 1–7. IEEE, 2020.

[94] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[95] Payam Refaeilzadeh, Lei Tang, and Huan Liu. Cross-validation. *Encyclopedia of database systems*, pages 532–538, 2009.

[96] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Dag Johansen, Thomas De Lange, Pål Halvorsen, and Håvard D Johansen. Resunet++: An advanced architecture for medical image segmentation. In *2019 IEEE international symposium on multimedia (ISM)*, pages 225–2255. IEEE, 2019.

[97] Hang Zhang, Kristin Dana, Jianping Shi, Zhongyue Zhang, Xiaogang Wang, Ambrish Tyagi, and Amit Agrawal. Context encoding for semantic segmentation. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 7151–7160, 2018.

[98] Ke Sun, Bin Xiao, Dong Liu, and Jingdong Wang. Deep high-resolution representation learning for human pose estimation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5693–5703, 2019.

[99] Michal Byra, Piotr Jarosik, Aleksandra Szubert, Michael Galperin, Haydee Ojeda-Fournier, Linda Olson, Mary O'Boyle, Christopher Comstock, and Michael Andre. Breast mass segmentation in ultrasound with selective kernel u-net convolutional neural network. *Biomedical Signal Processing and Control*, 61:102027, 2020.

[100] Bofei Zhang, Le Lu, Jianhua Yao, Xiaoguang Wang, and Ronald M Summers. Attention-based cnn for kl grade classification: Data from the osteoarthritis initiative. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1006–1009. IEEE, 2020.

[101] Bryar Shareef, Min Xian, and Aleksandar Vakanski. Stan: Small tumor-aware network for breast ultrasound image segmentation. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pages 1073–1076. IEEE, 2020.

[102] Debesh Jha, Aayush Jha, Ashwin Thangali, Hrithik Jha, Deepak Saini, Pratiksha Jha, and Mohsen Guizani. Real-time polyp detection, localization and segmentation in colonoscopy using deep learning. *IEEE Access*, 9:40496–40510, 2021.

[103] Lizhu Liu, Junfeng Liu, Jie Zheng, Shaoquan Chen, Han Li, Xianzhi Wang, Jun Wu, Jia Yang, and Pan Wang. A novel mcf-net: Multi-level context fusion network for 2d medical image segmentation. *Computer Methods and Programs in Biomedicine*, 226:107160, 2022.

[104] Jeya Maria Jose Valanarasu and Vishal M Patel. Unext: Mlp-based rapid medical image segmentation network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 23–33. Springer, 2022.

[105] Yu Yan, Yangyang Liu, Yiyun Wu, Hong Zhang, Yameng Zhang, and Lin Meng. Accurate segmentation of breast tumors using ae u-net with hdc model in ultrasound images. *Biomedical Signal Processing and Control*, 72:103299, 2022.

[106] Gongping Chen, Yu Dai, and Jianxun Zhang. Rrcnet: Refinement residual convolutional network for breast ultrasound images segmentation. *Engineering Applications of Artificial Intelligence*, 117:105601, 2023.

[107] Shangzhu Jin, Sheng Yu, Jun Peng, Hongyi Wang, and Yan Zhao. A novel medical image segmentation approach by using multi-branch segmentation network based on local and global information synchronous learning. *Scientific Reports*, 13(1):6762, 2023.

[108] Mohamed Bal-Ghaoui, My Hachem El Yousfi Alaoui, Abdelilah Jilbab, and Abdennacer Bourouhou. U-net transfer learning backbones for lesions segmentation in breast ultrasound images. *International Journal of Electrical and Computer Engineering (IJECE)*, 13(5):5747–5754, 2023.

[109] Zaiwang Gu, Jun Cheng, Huazhu Fu, Kang Zhou, Huaying Hao, Yitian Zhao, Tianyang Zhang, Shenghua Gao, and Jiang Liu. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE transactions on medical imaging*, 38(10):2281–2292, 2019.

[110] Nabil Ibtehaz and M Sohel Rahman. Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural networks*, 121:74–87, 2020.

[111] Kamarul Hawari Ghazali, Mohd Fais Mansor, Mohd. Marzuki Mustafa, and Aini Hussain. Feature extraction technique using discrete wavelet transform for image classification. In *2007 5th Student Conference on Research and Development*, pages 1–4, 2007.

[112] Yeon Ju Lee and Jungho Yoon. Nonlinear image upsampling method based on radial basis function interpolation. *IEEE Transactions on Image Processing*, 19(10):2682–2692, 2010.

[113] Jagyanseni Panda and Sukadev Meher. A new residual image sharpening scheme for image up-sampling. In *2022 8th International Conference on Signal Processing and Communication (ICSC)*, pages 244–249. IEEE, 2022.

[114] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.

[115] Neeraj Kumar, Ruchika Verma, Deepak Anand, Yanning Zhou, Omer Fahri Onder, Efstratios Tsougenis, Hao Chen, Pheng-Ann Heng, Jiahui Li, Zhiqiang Hu, et al. A multi-organ nucleus segmentation challenge. *IEEE transactions on medical imaging*, 39(5):1380–1391, 2019.

[116] Anusree Kanadath, J Angel Arul Jothi, and Siddhaling Urolagin. Multilevel multi-objective particle swarm optimization guided superpixel algorithm for histopathology image detection and segmentation. *Journal of Imaging*, 9(4):78, 2023.

[117] Nirmal Das, Satadal Saha, Mita Nasipuri, Subhadip Basu, and Tapabrata Chakraborti. Deep-fuzz: A synergistic integration of deep learning and fuzzy water flows for fine-grained nuclei segmentation in digital pathology. *Plos one*, 18(6):e0286862, 2023.

[118] Seyedali Mirjalili and Andrew Lewis. The whale optimization algorithm. *Advances in engineering software*, 95:51–67, 2016.

[119] K Pearson. Notes on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58:240–242, 1895.

[120] Charles Spearman. The proof and measurement of association between two things. *The American journal of psychology*, 100(3/4):441–471, 1987.

[121] Marko Robnik-Šikonja and Igor Kononenko. Theoretical and empirical analysis of relieff and rrelieff. *Machine learning*, 53(1):23–69, 2003.

[122] Tariq Sadad, Ayyaz Hussain, Asim Munir, Muhammad Habib, Sajid Ali Khan, Shariq Hussain, Shunkun Yang, and Mohammed Alawairdhi. Identification of breast malignancy by marker-controlled watershed transformation and hybrid feature set for healthcare. *Applied Sciences*, 10(6):1900, 2020.

[123] Woo Kyung Moon, Yi-Wei Shen, Chiun-Sheng Huang, Li-Ren Chiang, and Ruey-Feng Chang. Computer-aided diagnosis for the classification of breast masses in automated whole breast ultrasound images. *Ultrasound in medicine & biology*, 37(4):539–548, 2011.

[124] Arnab K Mishra, Pinki Roy, Sivaji Bandyopadhyay, and Sujit K Das. Breast ultrasound tumour classification: A machine learning—radiomics based approach. *Expert Systems*, 38(7):e12713, 2021.

[125] Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016.

[126] John Henry Holland et al. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.

[127] Ling Gong and Li Shaoqian. Chaotic spreading sequences with multiple access performance better than random sequences. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 47(3):394–397, 2000.

[128] Robert M May. Simple mathematical models with very complicated dynamics. *The Theory of Chaotic Attractors*, pages 85–93, 2004.

[129] Seyedali Mirjalili, Seyed Mohammad Mirjalili, and Andrew Lewis. Grey wolf optimizer. *Advances in engineering software*, 69:46–61, 2014.

[130] Fehmi Burcin Ozsoydan. Effects of dominant wolves in grey wolf optimization algorithm. *Applied Soft Computing*, 83:105658, 2019.

[131] Avrim L Blum and Pat Langley. Selection of relevant features and examples in machine learning. *Artificial intelligence*, 97(1-2):245–271, 1997.

[132] Sourav Pramanik, Debotosh Bhattacharjee, and Mita Nasipuri. Texture analysis of breast thermogram for differentiation of malignant and benign breast. In *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 8–14, 2016.

[133] Subhrajit Dey, Rajarshi Roychoudhury, Samir Malakar, and Ram Sarkar. Screening of breast cancer from thermogram images by edge detection aided deep transfer learning model. *Multimedia Tools and Applications*, 81(7):9331–9349, 2022.

[134] Lincoln Silva, Flávio Seixas, Cristina Fontes, Débora Muchaluat-Saade, and Aura Conci. A computational method for breast abnormality detection using thermographs. In *2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*, pages 469–474, 2020.

[135] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[136] Min Lin, Qiang Chen, and Shuicheng Yan. Network in network. *arXiv preprint arXiv:1312.4400*, 2013.

[137] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[138] Alaa Tharwat and Thomas Gabel. Parameters optimization of support vector machines for imbalanced data using social ski driver algorithm. *Neural computing and applications*, 32(11):6925–6938, 2020.

[139] Mohammed Azmi Al-Betar, Ibrahim Aljarah, Mohammed A Awadallah, Hossam Faris, and Seyedali Mirjalili. Adaptive β-hill climbing for optimization. *Soft Computing*, 23(24):13489–13512, 2019.

[140] Mojtaba Ahmadieh Khanesar, Mohammad Teshnehlab, and Mahdi Aliyari Shoorehdeli. A novel binary particle swarm optimization. In *2007 Mediterranean conference on control & automation*, pages 1–6. IEEE, 2007.

[141] Edmund K Burke and Yuri Bykov. The late acceptance hill-climbing heuristic. *European Journal of Operational Research*, 258(1):70–78, 2017.

[142] Esmat Rashedi, Hossein Nezamabadi-Pour, and Saeid Saryazdi. Gsa: a gravitational search algorithm. *Information sciences*, 179(13):2232–2248, 2009.

[143] John H Holland. Genetic algorithms and adaptation. *Adaptive control of ill-defined systems*, pages 317–333, 1984.

[144] Seyedali Mirjalili. Sca: a sine cosine algorithm for solving optimization problems. *Knowledge-based systems*, 96:120–133, 2016.

[145] Zong Woo Geem, Joong Hoon Kim, and Gobichettipalayam Vasudevan Loganathan. A new heuristic optimization algorithm: harmony search. *simulation*, 76(2):60–68, 2001.

[146] Afshin Faramarzi, Mohammad Heidarinejad, Brent Stephens, and Seyedali Mirjalili. Equilibrium optimizer: A novel optimization algorithm. *Knowledge-based systems*, 191:105190, 2020.

[147] Daniel Lévy and Arzav Jain. Breast mass classification from mammograms using deep convolutional neural networks. *arXiv preprint arXiv:1612.00542*, 2016.

[148] He Xiao, Qingfeng Wang, Zhiqin Liu, Jun Huang, Yuwei Zhou, Ying Zhou, and Weiyun Xu. Csablock-based cascade rcnn for breast mass detection in mammogram. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2120–2124. IEEE, 2020.

[149] Roberto Arias, Fabián Narváez, and Hugo Franco. Evaluation of learning approaches based on convolutional neural networks for mammogram classification. In *Smart Technologies, Systems and Applications: First International Conference, SmartTech-IC 2019, Quito, Ecuador, December 2-4, 2019, Proceedings 1*, pages 273–287. Springer, 2020.

[150] Qian Zhang, Yamei Li, Guohua Zhao, Panpan Man, Yusong Lin, and Meiyun Wang. A novel algorithm for breast mass classification in digital mammography based on feature fusion. *Journal of Healthcare Engineering*, 2020(1):8860011, 2020.

[151] Hua Li, Jing Niu, Dengao Li, and Chen Zhang. Classification of breast mass in two-view mammograms via deep learning. *IET Image Processing*, 15(2):454–467, 2021.

[152] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Computer Vision and Pattern Recognition (CVPR)*, 2015.

[153] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[154] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 1314–1324, 2019.

[155] Rohit Kundu, Hritam Basak, Pawan Kumar Singh, Ali Ahmadian, Massimiliano Ferrara, and Ram Sarkar. Fuzzy rank-based fusion of cnn models using gompertz function for screening covid-19 ct-scans. *Scientific reports*, 11(1):1–12, 2021.

[156] Chaoqing Wang, Weijun Gong, Junlong Cheng, and Yurong Qian. Dblcnn: Dependency-based lightweight convolutional neural network for multi-classification of breast histopathology images. *Biomedical Signal Processing and Control*, 73:103451, 2022.

[157] Shallu Sharma and Sumit Kumar. The xception model: A potential feature extractor in breast cancer histology images classification. *ICT Express*, 8(1):101–108, 2022.

[158] Mahesh Gour, Sweta Jain, and T Sunil Kumar. Residual learning based cnn for breast cancer histopathological image classification. *International Journal of Imaging Systems and Technology*, 30(3):621–635, 2020.

[159] Chuhan Hu, Xiaoyan Sun, Zhenming Yuan, and Yingfei Wu. Classification of breast cancer histopathological image with deep residual learning. *International Journal of Imaging Systems and Technology*, 31(3):1583–1594, 2021.

[160] Ebrahim Mohammed Senan, Fawaz Waselallah Alsaade, Mohammed Ibrahim Ahmed Al-Mashhadani, HH Theyazn, Mosleh Hmoud Al-Adhaileh, et al. Classification of histopathological images for early detection of breast cancer using deep learning. *Journal of Applied Science and Engineering*, 24(3):323–329, 2021.

[161] Shiksha Singh and Rajesh Kumar. Breast cancer detection from histopathology images with deep inception and residual blocks. *Multimedia Tools and Applications*, pages 1–17, 2021.

[162] Juan Vizcarra, Ryan Place, Li Tong, David Gutman, and May Dongmei Wang. Fusion in breast cancer histology classification. In *Proceedings of the 10th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, pages 485–493, 2019.

[163] Kamyar Nazeri, Azad Aminpour, and Mehran Ebrahimi. Two-stage convolutional neural network for breast cancer histology image classification. In *International Conference Image Analysis and Recognition*, pages 717–726. Springer, 2018.

[164] Zhanbo Yang, Lingyan Ran, Shizhou Zhang, Yong Xia, and Yanning Zhang. Ems-net: Ensemble of multiscale convolutional neural networks for classification of breast cancer histology images. *Neurocomputing*, 366:46–53, 2019.

Neural Computing and Applications
https://doi.org/10.1007/s00521-022-07895-x

ORIGINAL ARTICLE

Deep feature selection using local search embedded social ski-driver optimization algorithm for breast cancer detection in mammograms

Payel Pramanik[1] · Souradeep Mukhopadhyay[1] · Seyedali Mirjalili[2,3,4] · Ram Sarkar[1]

Received: 9 February 2022 / Accepted: 25 September 2022

Abstract
Breast cancer has become a common malignancy in women. However, early detection and identification of this disease can save many lives. As computer-aided detection helps radiologists in detecting abnormalities efficiently, researchers across the world are striving to develop reliable models to deal with. One of the common approaches to identifying breast cancer is through breast mammograms. However, the identification of malignant breasts from mass lesions is a challenging research problem. In the current work, we propose a method for the classification of breast mass using mammograms which consists of two main stages. At first, we extract deep features from the input mammograms using the well-known VGG16 model while incorporating an attention mechanism into this model. Next, we apply a meta-heuristic called Social Ski-Driver (SSD) algorithm embedded with Adaptive Beta Hill Climbing based local search to obtain an optimal features subset. The optimal features subset is fed to the K-nearest neighbors (KNN) classifier for the classification. The proposed model is demonstrated to be very useful for identifying and differentiating malignant and healthy breasts successfully. For experimentation, we evaluate our model on the digital database for screening mammography (DDSM) database and achieve 96.07% accuracy using only 25% of features extracted by the attention-aided VGG16 model. The Python code of our research work is publicly available at: https://github.com/Ppayel/BreastLocalSearchSSD.

Keywords Mammogram images · Breast cancer · Deep learning · Social ski-driver · Optimization · Local search · Algorithm

1 Introduction

Breast cancer is a disorder where the cells of the breast tissue alter and divide uncontrollably, resulting in a lump or mass in the mammary glands or between the mammary glands and the nipple in the majority of instances. Unfortunately, breast cancer is one of the most frequent malignancies among women, and it has a high fatality rate. Early identification of breast cancer can significantly improve women's survival rates, which is critical because breast cancer can be cured in 95 percent of cases if caught early [1]. Reviewing prior diagnostic data and gathering relevant information from past data are key to identifying this disease at an early stage.

Medical images are one of the most important sources of information for the identification and diagnosis of various illnesses and anomalies, allowing radiologists to examine the interior structure of human bodies. It is critical in the diagnosis of clinical diseases, the evaluation of treatment,

✉ Seyedali Mirjalili
ali.mirjalili@torrens.edu.au

Payel Pramanik
ppramanik07@gmail.com

Souradeep Mukhopadhyay
souradeepmukhopadhyay99@gmail.com

Ram Sarkar
ram.sarkar@jadavpuruniversity.in

[1] Department of Computer Science and Engineering, Jadavpur University, Kolkata, India

[2] Centre for Artificial Intelligence Research and Optimisation, Torrens University Australia, Brisbane, QLD 4006, Australia

[3] Yonsei Frontier Lab, Yonsei University, Seoul, South Korea

[4] University Research and Innovation Center, Óbuda University, Budapest 1034, Hungary

Published online: 05 November 2022

and the detection of anomalies in various bodily organs such as the eyes [2], lungs [2], brain, breast [3, 4], and stomach [5]. One of the most efficient ways to diagnose breast cancer is through medical imaging. The research in this field has grown significantly over the last three decades. Breast mammography is the most economical among various other detection methods, namely Thermal imaging, Magnetic Resonance Imaging (MRI), Ultrasound imaging, Computerized Tomography (CT), and Histology imaging [6].

A breast mammogram is an X-ray image of the woman's breast which is a common diagnostic measure for screening for breast cancer. It is useful for the detection of breast swellings, masses, calcifications, and dimpling of breast tissue. All of these are indicating an early stage of breast cancer. However, it is not an easy task to identify these symptoms from the breast images. Moreover, incorrect assessment of these images leads to an incorrect diagnosis with dangerous consequences. Consider the circumstance of a false negative diagnosis, in which an early stage of breast cancer is misdiagnosed as a normal case. As a result, the individual's chances of surviving five years are reduced [3].

Over the last decades, researchers have leveraged various machine learning (ML) techniques in the medical image analysis domain to help with decision-making processes. Data analysis, data cleaning, and meaningful feature extraction or feature representations are the reasons for ML's success to accomplish several tasks. Medical experts are capable of using their knowledge to relate features of a dataset to real-world phenomena or a fact, which is a challenging task for ML techniques. Deep learning (DL) alleviates this drawback as future engineering and processing is a part of the learning process as opposed to traditional methods with manual processes [7].

Many researchers have exploited DL methods in various domain applications such as in image classification [8], image segmentation [9], security [10–12], reinforcement learning [13], letter Recognition [14], partial differential equation solving [15]. In recent times researchers have successfully explored various DL-based methods in the domain of medical imaging, particularly in disease detection like Alzheimer's detection [16], fracture detection [17], COVID-19 detection [18], and many more. Generally, DL models require an ample amount of data for proper training purposes and the accessibility of such enormous volumes of data in the medical realm is quite unusual. As a consequence, experts are increasingly embracing the approach of transfer learning, in which Convolutional Neural Network (CNN) models are trained on bigger datasets such as the ImageNet dataset and then the weights are transferred and fine-tuned on a smaller (i.e., target) dataset [19]. Although transfer learning addresses a lot of challenges faced earlier by the researchers, to improve the performance of the models, several standard machine learning methodologies such as feature selection (FS) are now combined with the DL model.

FS is one of the conventional ways to reduce computational efforts that remove redundant features and selects a subset of distinct features. Also due to the presence of redundant features, distinctive features may not be given the importance they should be for classification purposes [20]. In the past, several optimization algorithms based FS methods have been exhaustively exploited in various domains like image enhancement [21], traveling Salesman problem solving [22], security [23], classification [24], SVM parameter optimization [25], and solving class imbalance problem [26]. Moreover, researchers have effectively been able to solve many image classification problems in the medical image analysis domain using this method such as in prostate cancer detection [27], Alzheimer's disease detection [28], and many more. This encourages us to employ an FS approach that is based on an optimization algorithm in our work.

The concept of embedding a local search method with FS is not very old. Researchers in the recent past have successfully developed and deployed such models and in turn, also proved that these models perform better in terms of classification results [29]. Choosing the optimal feature subset in FS is usually difficult, especially in wrapper-based techniques where the chosen subsets must be assessed using a learning algorithm at each iteration. Ensemble of local search improves an FS method's exploitation ability and hence improves the overall learning model's performance.

Because of the above-mentioned facts, in the current work, we design a two-stage breast cancer classification model taking mammograms as input. At first, we use an attention-aided DL model to extract features from the mammograms. Then, we apply a local search embedded FS approach to reduce the feature dimension and augment the classification ability. As the basis model, we used a transfer-learning model that was pre-trained on the ImageNet dataset and fine-tuned it on the target dataset. Furthermore, we introduce attention by incorporating a global weighted average pooling mechanism on the base model. We extract features using this attention-based transfer learning model for the target dataset and lastly process these extracted features with the local search embedded FS method to produce optimal and reduced feature subsets. Finally, these reduced features are used as the inputs to the KNN classifier to produce the final classification results. After experimenting with different transfer learning models and different FS algorithms, we find that the attention-aided VGG16 model and Adaptive Beta Hill Climbing (ABHC) embedded SSD based FS algorithm on the mentioned dataset outperforms other contemporary methods.

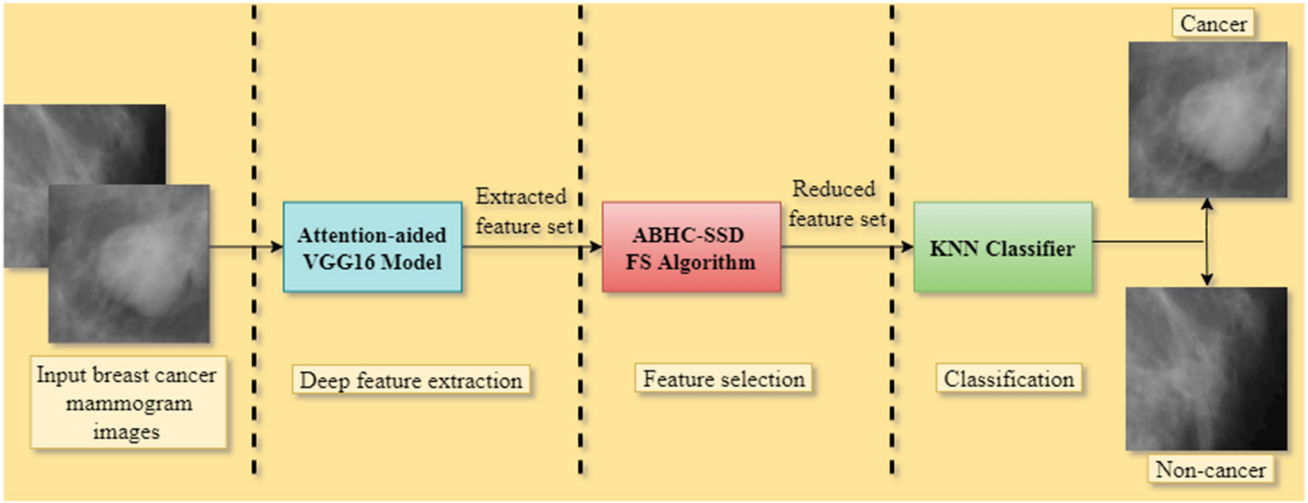


Fig. 1 The pipeline of our suggested breast cancer classification model

Figure 1 depicts the whole architecture of the suggested model.

1.1 Motivation and contributions

In the domains of computer vision and image processing for health and medical assessment, integrating DL approaches with FS algorithms has yielded significant results [29–32]. Researchers have shown the immense potential of DL-based applications for mammogram image processing in terms of providing reliable breast cancer predictions [4, 33–35]. Moreover, the attention mechanism exploits the most important regions of an image by paying more attention to the same [36, 37]. Furthermore, FS approaches reduce the number of features, whereas local search helps to increase the exploitation capability of the FS method and produces the most optimal subset of features [29]. These efforts prompted us to investigate how DL methods perform when paired with FS approaches, as well as to confirm the usefulness of the attention mechanism and the capabilities of local search strategies for breast cancer analysis and assessment from mammography imaging. Therefore, in this paper, we have proposed a model in which an intelligent amalgamation of a DL model with an optimization algorithm-based FS approach has been made. The main contributions of the present research work are listed below.

1. We create a model for breast cancer classification from mammograms that combines the principles of deep learning and optimization algorithms.
2. We introduce an attention mechanism on a deep CNN-based transfer learning model, called VGG16, and fine-tune it for the extraction of deep features from the input images.
3. We ensemble a local search, namely ABHC with SSD based FS algorithm to produce an optimal feature subset from the features produced by the CNN model.
4. We achieve state-of-the-art classification accuracy with just 25% of features of the original feature set obtained by the CNN model when evaluated on the DDSM database.

The rest of this research study is broken down into categories. Section 2 is a literature review in which we look at several techniques of breast cancer diagnosis, starting with mammography and then moving on to local search-based FS algorithms. Section 3 lays out the requirements for this research project, followed by a detailed description of our proposed study in Sect. 4. The metrics we employ to evaluate the proposed model and analyze the experimental results are then discussed in Sect. 5. Finally, in Sect. 6, we make some concluding remarks and discuss some potential future directions.

2 Literature survey

This section contains two subsections, wherein in the first subsection we discuss some recent DL-based methods for the detection and classification of breast cancer, and in the subsequent subsection, we discuss the application of FS algorithms in the medical domain, specifically, in breast cancer detection.

2.1 DL-based methods for breast cancer detection

In health care systems, there is various machine and DL approaches developed by researchers. Machine learning is

widely utilized in several domains like health care, early disease detection, biomedical, etc. [38]. In recent times, the advancement in machine learning, especially in DL has created a significant impact on the medical imaging field. It enhances the precision to identify, classify, and quantify patterns in medical images. To be precise, exploiting complex representations of the features which are understood or realized from the input data is the prime reason for this tremendous progress. Thus, DL models are achieving extraordinary results in different medical applications [7].

For instance, A. Saber et al. have shown in [33] a DL model based on the transfer learning technique for the detection of breast cancer from the mammographic image analysis society (MIAS) database. In this, the authors showed some pre-processing techniques and discussed the evaluation metric results of different pre-trained CNNs for the MIAS database. The experimental results showed that out of different pre-trained CNNs, the VGG16 model yields the best result, which is seen in our research work as well. Similar work can be found in [39]. In 2019, Shen et al. [8] proposed a CNN model to classify mammogram images of the CBIS DDSM dataset. This method yields an AUC score of 0.88, sensitivity of 86.1%, and specificity of 80.1% which is quite low. Furthermore, the INBreast dataset was used to increase the performance of the model yielding a 0.95% AUC score, 86.7% sensitivity, and 96.1% specificity. In the next year, Khatami et al. [40] introduced a regularization scheme for the detection of cancer from mammograms by changing the weights of the convolutional layers using some additive noise and achieved an accuracy of 83%. The limitation of this model is the lower detection accuracy which may not be useful in practical scenarios. A paper published in 2015 by Ertosun and Rubin [34] used a DL-based method which achieves 85% accuracy for identifying breast images with a mass from breast mammograms and an accuracy of 85% for mass localization in mammograms with an average false positive rate of 0.9 per sample image. However, their main focus is to find out whether the breast mass is benign or malignant. In another work, Levy et al. [4] have come up with a DL model with pre-processing and data augmentation to classify pre-detected breast masses from DDSM mammograms and achieved 92.9% accuracy, whereas Khamparia et al. [3] have proposed a method using the fine-tuned VGG16 transfer learning model to diagnose breast cancer from DDSM mammograms, and achieved an accuracy of 88.3%. The authors have used data augmentation and regularization to enhance the performance of the model. A study has been done by L. G. Falconi et al. [41] of different transfer learning models like NasNet, MobileNet, VGG16, ResNet, Xception, and Resnext to train a breast abnormality malignancy classifier. In this study, the authors have concluded that training of DL models tends to overfit and fine-tuning of the models achieves a better classification performance in the case of the VGG16 model which gives an accuracy of 84.4% in the CBIS-DDSM dataset. Al-antari et al. [35] have proposed a CAD system based on You Only Look Once (YOLO) to detect and classify breast lesions. In this work, the authors used the YOLO detector to detect breast lesions from the DDSM and the InBreast mammograms, and the classification was done using three DL classifiers, namely regular feed-forward CNN, ResNet-50, andInceptionResNet-V2. From these research works, it can be said that automatic DL models can achieve better results even on heterogeneous mammography platforms. Also, it holds a strong promise for improving the performance of the clinical tools for reducing false positive and false negative screening mammography results.

Researchers have successfully explored the utilization of ML and DL models not only on mammograms but also on different breast cancer image modalities available. For instance, In Vahadane et al. [42], authors have introduced a structure-preserved stain-normalization technique to deal with histopathological images and achieved 87.50% classification accuracy. In another work, Sarmiento et al. [43] proposed a machine learning-based technique for automatic breast cancer grading of histological images in which the extracted feature from various characteristics of the image such as texture, color, and shape was fed to the Support vector machine (SVM) classifier as the input and with tenfold cross-validation, this method achieved an accuracy of 79.2%. In another research paper by Nawaz et al. [44], the authors used a fine-tuned AlexNet for breast cancer classification in histology images and achieved an image-wise accuracy of 75.73% and patch-wise accuracy of 81.25%. In [45], Silva et al. have suggested a method for abnormality detection in breast thermal images. The authors used Auto-WEKA with some defined settings for the selection of best features and used a K-star classifier with a tenfold cross-validation method for the classification of images. Also, in [46], the authors have introduced a deep CNN method for the automatic cancer tissue nuclei detection, segmentation, and classification of breast cancer cells from whole slide images of hematoxylin and eosin stains. In this work, a multilevel saliency nuclei detection model is used for the detection of nuclei, and the same is integrated with the deep CNN model for the classification of benign and malignant cells. Rakhlin et al. [47] have designed a DL model to classify the images of breast tissues. In this work, pre-trained models of VGG-16, InceptionV3, and ResNet-50 are used for feature extraction, whereas for the classification purpose a tenfold cross-validation scheme with Light Gradient Boosting (GBM) classifier has been used. This approach achieves an accuracy of 87.2% for breast cancer image classification.

However, it can be seen that researchers have extensively used DL models as well as transfer learning models for breast cancer detection from different medical imaging modalities available. But considering breast cancer detection, it is important to focus on the region of interest (ROI) as some of the regions may be more relevant than others, thus justifying the need of adding an attention mechanism to a DL model. In this work, we try to explore the attention mechanism of transfer learning models for breast cancer classification from mammography.

2.2 Nature-inspired meta-heuristic FS algorithms for breast cancer detection

Nature-inspired meta-heuristic techniques and their variants are widely used in solving FS problems [48]. The field of meta-heuristics is vast and it has made significant advancements toward solving complex optimization problems. Since the first meta-heuristic was presented, a considerable amount of progress has been achieved, and countless new algorithms are offered regularly. There is no dispute that research in this area will advance in the coming future. There are two primary groups of meta-heuristic algorithms. One is meta-heuristic algorithms based on a single solution, where optimization begins with a single solution and gets updated as the algorithms move through iterations. Another group is population-based meta-heuristic algorithms, where optimizations start with a population of solutions and update it over the iterations. However, the first group of algorithms can get trapped within local optima and they only partially explore the search space, whereas the latter group helps to prevent local optima since they have a great search space exploration opportunity and various solutions that work together to help one another. Meta-heuristic algorithms can be classified based on their behavior into four categories: algorithms based on physics, swarm intelligence, evolution, and humans [90–93].

Meta-heuristic techniques yield an optimal solution by iteratively exploring as well as exploiting the search space. It assists to select an optimal set of features so that a better classification performance will be ensured with that set of features. Every meta-heuristic method tries to maintain a good balance between exploration and exploitation of the search space to improve the results [49]. Researchers have effectively been able to solve many image classification problems in the cancer detection domain as well as other medical domains [49–51] using several meta-heuristic-based FS algorithms.

For instance, in 2010, Gandhi et al. [52] suggested a cancer detection method based on Pittsburgh Learnt Fuzzy Rule and Particle Swarm Optimization. In 2014, Ahmad et al. [53] designed a breast cancer diagnosis model by using the Genetic Algorithm (GA) for both FS and parameter optimization of an artificial neural network (ANN). The higher computational cost was the main drawback of both these methods. In 2019, Huang et al. [54] proposed a new breast cancer diagnosing technique based on the fruit fly optimization algorithm embedded with a Levy flight strategy. It was mainly used to optimize two key parameters of the SVM classifier. They have exploited two datasets—Wisconsin Prognostic Breast Cancer dataset and Wisconsin Diagnostic Breast Cancer dataset for result computation. In the same year, Sayed et al. [55] suggested a hybrid model on the same dataset that exploited cluster analysis algorithm and binary version of Moth-flame optimization and Whale optimization algorithm (WOA) for FS. In 2020, Fang et al. [56] proposed an intelligent amalgamation of multi-layer perceptrons with WOA for breast cancer detection. Lower accuracy was the main limitation of these techniques. In 2021, Oyelade et al. [57] suggested a nature-inspired meta-heuristic optimized convolutional neural networks model detect abnormalities in breast cancer images. It involved training a CNN network using GA, WOA, multiverse optimizer (MVO), satin bower optimization (SBO), and life choice-based optimization (LCBO) algorithms to optimize only the weights and bias of the model. The main drawback of this model was lower accuracy. In the same year, Tavasli et al. [58] proposed an ensemble with a soft-weighted gene selection-based model for the classification of cancer using an improved version of the Water Cycle Algorithm. This model lacked generalizability and accuracy. Also, Rezaee et al. [59] have suggested a model for identifying multi-mass breast cancer following hybrid descriptors and memetic meta-heuristic learning. Drawbacks were large data processing time and lower precision.

To the best of our knowledge, the SSD optimization algorithm has not been used yet in the domain of breast cancer detection. In this work, we explore this novel optimization algorithm for FS and achieve promising results. However, research works with SSD in the medical domain can be found in [60–62].

3 Preliminaries and essential definitions

In this section, we mention some prerequisites which are needed to describe and understand our proposed model. We briefly discuss the VGG16 model, used for feature extraction from the input images, and SSD and ABHC algorithms that are collectively used to eliminate the irrelevant features obtained in the previous stage.

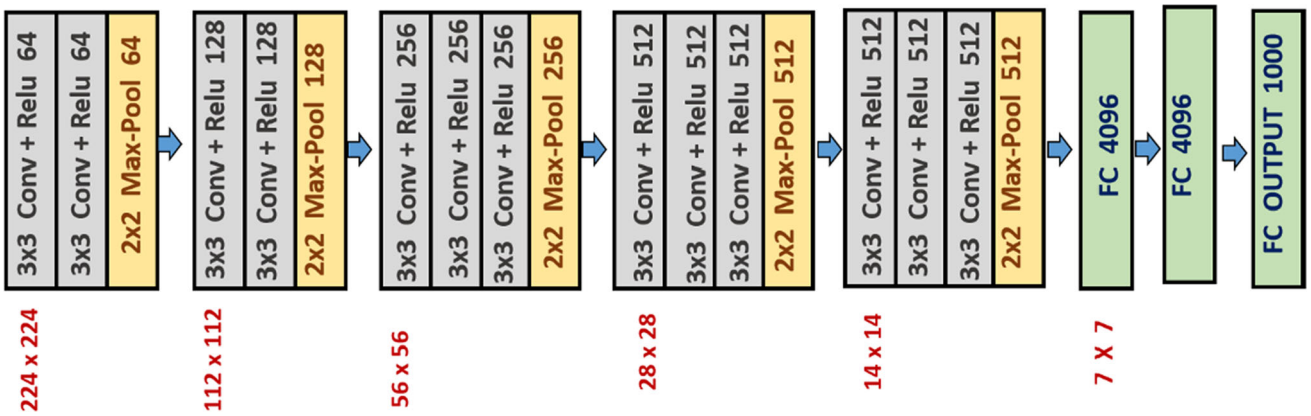


Fig. 2 Detailed architecture of the VGG16 model including fully connected layer and 1000-dimensional output layer for the ImageNet database

3.1 VGG16

VGG16 network is proposed by K. Simonyan and A. Zisserman [63]. It is a very basic CNN model having 13 convolutional layers of 3 × 3 filters with a stride 1, 5 max-pooling layers of 2 × 2 filters with stride 2, and 2 fully connected layers (FC) followed by a softmax for the output layer (Fig. 2). The model obtained a test accuracy of 92.7% (top-5) on the popular large-scale ImageNet dataset [63]. This network has approximately 138 million parameters and 16 in VGG16 refer to that it has 16 weighted layers.

3.2 Social ski-driver optimization algorithm

SSD is a recent optimization technique suggested by Tharwat et al. [25]. It is based on the approach taken by ski drivers when they go downhill. Various modules of SSD are discussed as follows:

3.2.1 Location of the agents

The location of the agents ($L_k^{R^n}$) is exploited to compute the fitness function at the particular position of a multi-dimensional search space.

3.2.2 Best personal location

At each iteration, the fitness measure (described in Sect. 4.3) for every agent gets matched with the personal best fitness measure (previously obtained) and that best location is saved as the personal best location (PB_k) for that agent.

3.2.3 Best mean global location

The agents proceed toward the best global location according to the algorithm. The best global location is computed as the average of the locations of the top 3 solutions as calculated in Eq. 1 and denoted as MGB:

$$MGB = \frac{L_x + L_y + L_z}{3} \tag{1}$$

where L_x , L_y and L_z are the locations of top 3 solutions respectively.

3.2.4 Velocity and location updating

The location and velocity of the agents are modified according to Eqs. (2) and (3), respectively:

$$L_k^{T+1} = V_k^T + L_k^T \tag{2}$$

$$V_k^{T+1} = \begin{cases} h * \sin(rand(0,1))\left(PB_k^T - L_k^T\right) \\ \quad + \sin(rand(0,1))\left(MGB_k^T - L_k^T\right) & if\ rand(0,1) \le 0.5 \\ h * \cos(rand(0,1))\left(PB_k^T - L_k^T\right) \\ \quad + \cos(rand(0,1))\left(MGB_k^T - L_k^T\right) & else \end{cases} \tag{3}$$

In Eqs. (2) and (3),V_k^T, MGB_k^T, L_k^T represent the velocity, mean global best position, and current position of the particle at kth dimension and Tth iteration, respectively. PB_k represent the personal best position of the particle at kth dimension. $Sin(x) and Cos(x)$ are traditional sine and cosine functions. $rand(0,1)$ function chooses any real number between 0 and 1. h is a variable that is exploited to maintain the parity to achieve a balance among both two crucial elements of exploitation and exploration, and it is computed according to Eq. 4:

$$h^{T+1} = r \times h^T \tag{4}$$

In Eq. (4), T denotes the present iteration and r is exploited to decrease the value of h. In Eq. (3), deriving V_k^{T+1}, the sine and cosine functions guarantees that the directions of movement of the agents are not very straightforward. This is because those functions allow the algorithm

to explore and it helps to diversify the searching domain but in a controlled way. Another advantage of the SSD algorithm is that it is comparatively more social than other meta-heuristics. The agents in SSD try to get to the mean of the best three options. As a result, if the global best solution is found to be in local minima, the SSD can use the other two best solutions for escaping [25]. SSD is quicker at discovering optimal solutions than the PSO algorithm.

3.3 Local search (ABHC)

Local search is used as a heuristic method for solving computationally expensive optimization problems. Local search is applicable for the problems that can be framed as searching for a solution maximizing or minimizing a criterion among a huge domain. Local search explores from solution to solution in the search space by doing some local changes, until a chosen solution seems to be optimal or the number of iterations (MT) is finished. ABHC [64] is one of the popular local search methods we have used in this work. Hill climbing sometimes faces problems in local optima. To get rid of this problem, ABHC [64] is proposed. This algorithm inputs an agent L location and outputs a modified location of the agent in the search space. This algorithm depends on two operators—one is the Neighborhood operator (N) and another one is the *Beta* operator. Here, N operator randomly chooses a neighbor $L'\left(L_1', L_2', L_3' \ldots L_k'\right)$ from a solution $L(L_1, L_2, L_3 \ldots L_k)$ as follows:

$$L_j' = L_j \pm rand\ (0,1) \times N \quad where\ j = 1, 2, \ldots k \tag{5}$$

In Eq. 5, N is the greatest probable distance between the present solution and the neighbors, $rand(0,1)$ is a function to generate random numbers between 0 and 1. *Beta* operator gets motivated by the mutation operator used in GA. We assign values to new solutions either arbitrarily from the comparable domain with a probability $Beta = rand(0,1)$ or the present solution as follows:

$$L_j'' = \begin{cases} L_j & if\ Beta > rand\ (0,1) \\ L_j' & otherwise \end{cases} \tag{6}$$

In Eq. (6), $L_j'', L_j\ and\ L_j'$ denote the jth dimension of the updated location of the solution, previous solution, and neighborhood of the previous solution, respectively.

Now, the outcome of this version of hill climbing is mostly dependable on the values of *Beta* and N. Determining the values of these two parameters needs comprehensive experiments. To bypass this shortcoming, ABHC came into existence. In ABHC, *Beta* and N are the functions of the number of iterations.

$N(z)$ is the functional measure of N in the z^{th} iteration. $N(z)$ can be determined according to Eq. (7).

$$N(z) = 1 - \frac{z^{\frac{1}{c}}}{MT^{\frac{1}{c}}} \quad where\ c = cons\tan t \tag{7}$$

Here MT denotes the maximum number of iterations and z is the current iteration number.

The value of *Beta* in zth iteration is denoted as $Beta(z)$ as follows:

$$Beta(z) = \frac{(Ma - Mi) \times z}{MT} + Mi \tag{8}$$

In Eq. (8), Ma, Mi, and z are the maximum and minimum values of *Beta* and the current number of iterations, respectively. Now, if the newly generated neighbor L'' is better than L, then L is replaced with L''.

The function of ABHC (Agent L($L_k^{R^n}$)):

1. Initialize the value of c, Ma, and Mi.
2. Estimate the original fitness value of the agent L using the Fitness Function.
3. Set $z = 1$
4. While ($MT > z$) do
5. Calculate the value of $N(z)$ using Equation 7
6. Compute L_j' using Equation 5
7. Calculate the value of $Beta$ (z) using Equation 8
8. for i =1 to N do
9. Check if($rand(0,1) < Beta$ (z)) then
10. Set $L_j'' = L_j$;
11. End if
12. End for
13. Check if(Fitness Function(L'')> Fitness Function(L)) then
14. Set $L = L''$;
15. Compute the original fitness value of the agent L'' using the Fitness Function
16. End if
17. Increment the value of z by 1
18. End while

4 Proposed model

As previously stated, we create a two-stage breast cancer classification model in this study. To extract features from the mammograms, we first utilize an attention-aided DL model. Then, to minimize the feature dimension and improve classification ability, we use a local search integrated FS technique. We detail the complete process of our work in this part.

4.1 Feature extraction from the attention-aided DL model

For feature extraction from mammography inputs, we employ a deep attention model. We start with the VGG16 model, which has been pre-trained on the popular ImageNet dataset, and add an attention mechanism. We investigate the VGG16 model minus the top layer, which has fully linked layers and freezes the weights of all levels to prevent the layers from learning new information during the model's training. The Global Average Pooling (GAP) layer takes the role of the fully linked layer.

GAP [65], a pooling operation, is usually applied in place of fully connected layers in classical CNNs. Similar to simple pooling (max-pooling or average pooling) layers, it reduces the spatial dimension of a given tensor. For instance, a three-dimensional tensor having dimensions h × w × d gets converted to the dimension of $1 \times 1 \times d$. GAP produces a single value for each feature map of dimension $h \times d$ by taking the average of all hw values. In this work, in the final convolutional layer of the base VGG16 model instead of adding fully connected layers, the generated vector is supplied to the final layer after we take the average of the feature maps. The GAP layer summarizes the spatial records through imposing correspondences among feature maps and categories, consequently making it robust in terms of spatial translation of the input data. However, only GAP is too simplistic as some regions may be more significant than others, thus needing attention. Here, we introduce an attention method to turn on some pixels in the GAP layer. We add one locally connected convolutional layer of kernel size 1×1 and fan it out to all the layers of the model. Next, we add a lambda layer [66] to account for missing values from the attention model that rescale the results based on the number of pixels. Lastly, we fine-tune the model by adding dropouts to prevent overfitting the model during training [67]. Thus, we propose a weighted GAP to the existing CNN architecture as depicted in Fig. 3.

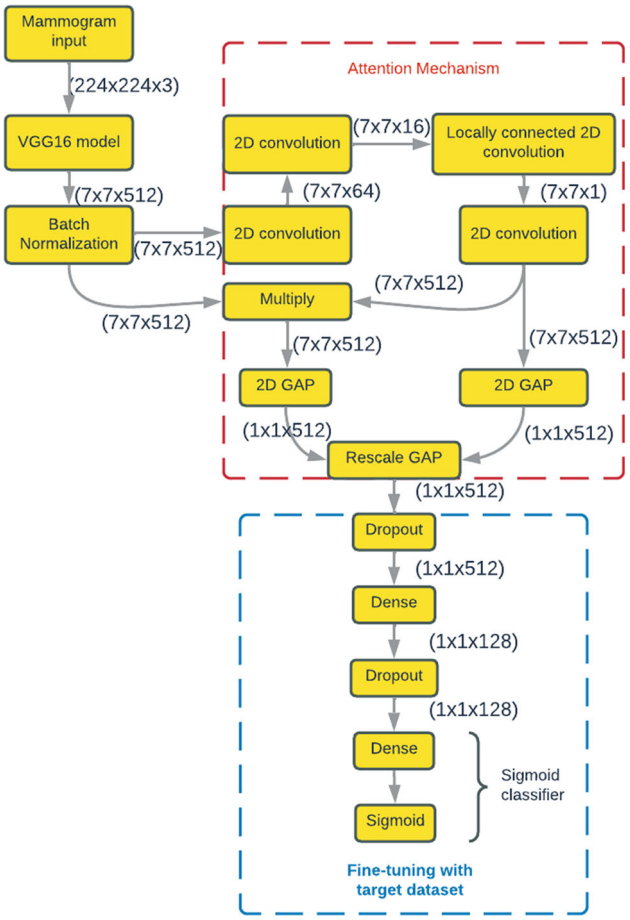


Fig. 3 Architecture of the attention-aided VGG16 model with a weighted average GAP layer (all the layers along with input and output shapes). Area of the attention and the fine-tuning of the model are highlighted by red and blue colored dashed boxes, respectively

4.2 FS and classification using local search embedded SSD algorithm

An FS problem seeks to find the best subset of features from the main set to augment the accuracy of a learning model. It can also be considered as a dimension reduction algorithm that removes redundant and/or highly correlated features. Due to the binary nature of this problem, most meta-heuristics are not able to solve them. This is due to the assumption of continuous variables in the vanilla version of such algorithms. There are different methods in the literature to convert them to binary algorithms. One of the most computationally cheap ways is to use a transfer function. In Particle Swarm Optimization (PSO) algorithm where a velocity vector is used to update solutions, this transfer function relates velocity to the probability of changing the position in a binary space. The transfer function [29] used in the current work is shown in Eq. (9).

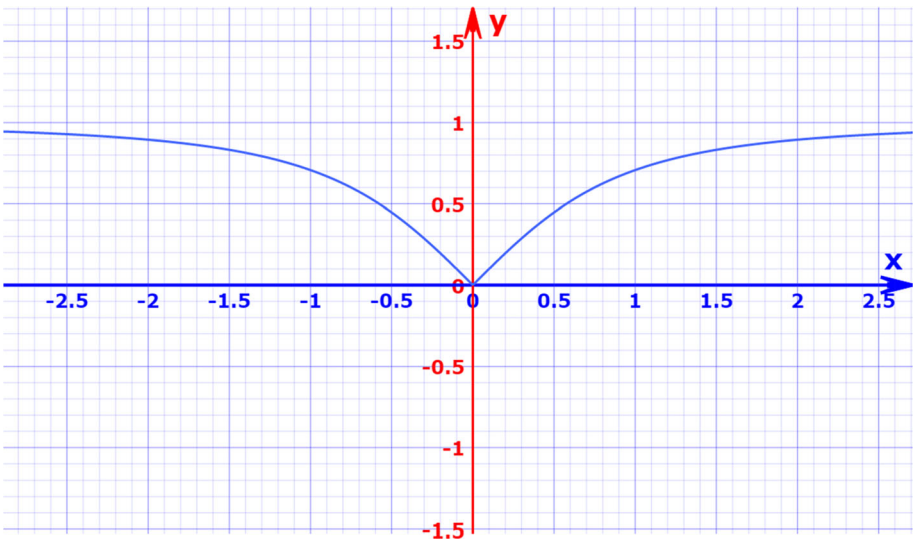


Fig. 4 Graphical representation of the V-shaped transfer function

$$V(x) = \frac{|x|}{\sqrt{1+x^2}} \tag{9}$$

Exploiting the V-shaped transformation function the location of an agent is modified as per the below Eq. (10).

$$LB_j^{k+1} = \begin{cases} c(LB_j^k) & if\ V\left(LB_j^{k+1}\right) > rand(0,1) \\ LB_j^{k+1} & otherwise \end{cases} \tag{10}$$

In Eq. (10), LB_j^{k+1} denotes the agent's modified location, LB_j^k signifies the location of the agent at that particular time(here k denotes iteration number and j denotes number of dimention) and $rand(0,1)$ is a function that generates random numbers between 0 and 1. The function $c(x)$ denotes the complement function for all binary x i.e., $c(x) = 1 - x$. Figure 4 shows the graph of the V-shaped transfer function. After altering an agent's location in each iteration, ABHC is used to optimize the position of the agents to get a higher fitness value. The SSD algorithm's exploitation potential is improved by using an ABHC-based local search technique.

Pseudo-code of the local search embedded SSD algorithm

1. Initialize the value of Population-size and the Max-iteration.
2. Randomly initialize the value of velocity, position, pbest, and gbest.
3. for $z = 1$ to Max-iteration do
4. for $p = 1$ to Population-size do
5. Set agent = L
6. Using the Fitness Function, determine the agent L's original fitness value.
7. Change the value of gbest by picking one of the top 3 sites at random.
8. Update the value of pbest.
9. Set the value of $r1 = rand(0,1)$, $r2 = rand(0,1)$ and $L_k^{T+1} = V_k^T + L_k^T$
10. Check if $(r1 < 0.5)$ then
11. $Set\ V_k^{T+1} = h * \sin(r2)\,(PB_k^T - L_k^T) + \sin(r2)\,(MGB_k^T - L_k^T)$
12. else
13. $Set\ V_k^{T+1} = h * \cos(r2)\,(PB_k^T - L_k^T) + \cos(r2)\,(MGB_k^T - L_k^T)$
14. End if
15. Calculate $V(x) = \frac{|x|}{\sqrt{1+x^2}}$
16. Update the position of LB_j^{k+1}using Equation 10.
17. Set the new location of LB_j^{k+1}= Local Search (LB_j^{k+1})
18. End for
19. Set $h^{T+1} = r \times h^T$
20. End for

4.3 Fitness function

The motive of this particular section is to describe how to evaluate the quality of a candidate solution. A learning algorithm needs to be exploited for assessing as SSD is a wrapper-based algorithm. Hence, we have exploited the KNN [68] classifier for the computation of classification accuracy of a candidate solution, succeeding the works of [69–71]. The fitness function mainly contains two components: one is the number of features and another one is classification accuracy. These components are contradictory to each other. We have to increase classification accuracy but at the same time, we need to decrease the number of features selected. So we have determined to exploit the classification error. As a lesser error value would indicate a better fitness score, so would a lesser number of features. In Eq. (11), the fitness function used to assess a given feature set's strength is defined.

$$Fitness\ Function = w\times \propto +(1-w)\times \frac{|s|}{|d|} \tag{11}$$

In Eq. (11), $|d|$ is the total count of features in the dataset, $|s|$ denotes the count of features in the chosen feature set (i.e., a candidate solution),$\propto$ is the error in classification using the feature subset, and $w \in [0, 1]$ signifies the relative weight value given to the classification error and the number of features.

5 Experimental results and discussion

In this section, we describe the dataset used in the current study and report the results obtained by applying our proposed method on the dataset. To justify the superiority of the framework, comparisons to other published methods on the same dataset have also been discussed.

5.1 Experimental setup

We perform all the experiments on a machine with 12 GB NVIDIA Tesla T4 GPU and the programming language used is Python 3.6. The deep learning model is implemented with the Tensorflow environment using the Keras library.

5.1.1 Dataset description

We assess our model on a publicly accessible breast cancer mammography database [72] which is an open-source and unbalanced binary mammography image dataset that contains normal images i.e., negative samples from the DDSM database [73], and malignant images i.e., positive samples from the CBIS-DDSM database [74]. It has a total of 55,885 training samples out of which 86% are negative samples and 14% are positive samples. The dimension of each sample is 299 × 299. In this work, we consider only the training samples of mammogram masses which are categorized as benign and malignant masses. Sample images of benign and malignant masses of the DDSM dataset are shown in Figs. 5 and 6, respectively.

5.1.2 Performance evaluation metrics

We assess our model with the following evaluation metrics:

The ratio of accurately predicted labels to the overall size of the dataset is referred to as accuracy. [75]. Accuracy is calculated as follows:

$$Accuracy = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{12}$$

The percentage of samples identified as positive that are actually positive is known as Precision. It is the ratio of successfully predicted positive class labels to the total number of positive class samples predicted [75]. It is calculated as follows:

$$Precision = \frac{TP}{(TP + FP)} \tag{13}$$

Recall is calculated by dividing the number of true positive samples by the total number of positive samples in that class. [75].

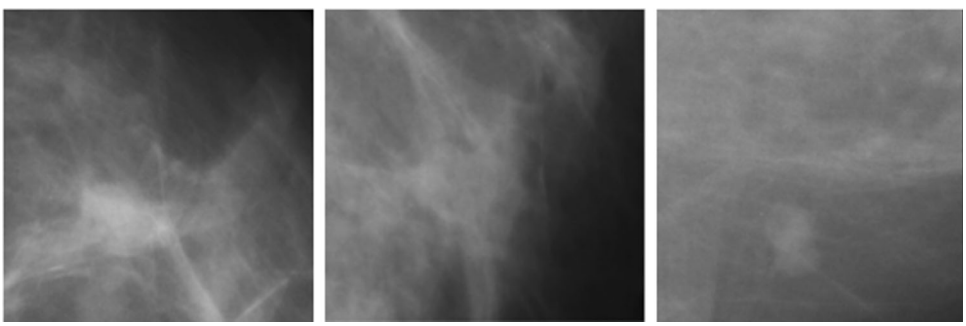


Fig. 5 Sample images of Benign breast mass from the DDSM dataset

Fig. 6 Sample images of Malignant breast mass from the DDSM dataset

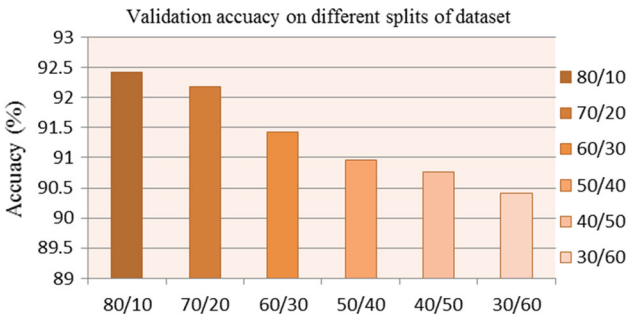


Fig. 7 The attention-aided VGG16 model's validation accuracy for varied splits of training and validation data. On the DDSM database, the model gets the best classification accuracy for 80% training and 10% validation data

$$Recall = \frac{TP}{(TP + FN)} \tag{14}$$

where True Positive (TP) indicates the positive class samples that the classifier correctly labels, True Negative (TN) indicates the negative class samples that the classifier correctly labels, False Positive (FP) indicates the negative class samples that were incorrectly identified as positive class samples, and positive class samples that have been mislabeled as negative class samples are known as False Negatives (FN).

5.1.3 Parameter tuning for transfer learning

The experiments are carried out and assessed on the aforementioned database, which contains 80% training data, 10% testing data, and 10% validation data. We started by experimenting with alternative training and validation data splits. Figure 7 shows the experimental results from the DDSM database. For the above-mentioned splitting, we find that the model has the maximum classification accuracy. In practice, we experiment with various typical parameters for learning rate and batch size to find the best possible combination. We take into account the initial learning rate $\in$\{1e-2, 1e-3, 1e-4, 1e-5, 1e-6\} and batch size $\in$\{8, 16, 32, 64\} and come up with a decent combination of 1e-2 and 64 as the initial learning rate and batch size. In addition, for training purposes, we employ the popularly used Adam optimizer. We employ a step learning rate scheduler for smooth learning, where the learning rate is lowered by a factor of two after the third epoch. The dropout values for the two dropout layers are 0.5 and 0.25, respectively. Table 1 shows the exact parameter values for training the model necessary to perform our strategy.

Table 1 Parameter details used in this work for the training of the DL based transfer learning models

Parameter	Value
Batch size	64
Optimizer	Adam
Initial learning rate	0.001
Loss function	BCE

5.1.4 Parameter tuning for SSD-based FS

In this subsection, we discuss the range of values of various parameters and the fitness function used in the SSD algorithm. We want to reduce our fitness function specified in Eq. (11), i.e., the number of features as well as the classification error, as a critical element of this task. If we increase the value of w in the fitness function then we give more importance to deducting the number of features, whereas if we assign a minimum value to w then more emphasis is given to reduce the classification error. A series of experiments with different values of w leads to the optimal value of w being 0.2. As a result, the algorithm under consideration prioritizes minimizing the classification error, i.e., increasing the classification accuracy. We have tested our strategy using a variety of h and r values in

Table 2 Performance of different pre-trained transfer learning (TL) models on the DDSM database

Pre-trained TL model	Accuracy (%)
VGG19	87.45
ResNet50	87.23
EfficientNet	87.22
VGG16	87.30

the search space throughout the experimentation. Our investigations also show that as the value of h is raised, the accuracy of the classification gets improved. When the value of r is lowered, the classification performance improves but eventually declines after reaching a peak. When the value of r is too low, this results in overfitting. When the values of h and r are fixed at 100 and 0.9, respectively, the maximum classification accuracy is reached.

5.2 Performance of attention-aided deep feature extraction model

As mentioned earlier, in this work, at first we consider a deep attention model for the extraction of deep features from the mammogram inputs. We use the VGG16 model as our base model and incorporate an attention method to this. We experiment with some popular pre-trained end-to-end models for deep feature extraction and report the result in Table 2. Furthermore, we incorporate the attention mechanism with each of these models. Noteworthy improvement in classification accuracy is achieved due to the effect of weighted average pooling attention. As in the last layer of the CNN.

model as an alternative to the fully connected layer, we consider using the GAP layer and it gives a single feature map for the corresponding category. This layer uses spatial information by enforcing correspondences between feature maps and categories. The results obtained on the test dataset are tabulated in Table 3. The Mean and the standard deviation (SD) value are shown over five simulations of the obtained results. From Table 3, it can be observed that the VGG16 model with attention classifies the mammograms more appropriately. Hence, we decide to proceed with the attention-based VGG16 model.

During the training of the attention-aided VGG16 model, training and validation accuracies for every epoch of the first simulation are recorded and shown in Fig. 8. It can be seen from Fig. 8 that initially the model does not suffer from any major over-fitting and later on the validation accuracy does not improve much and the accuracy oscillates between the values 91 to 92, whereas Fig. 9 displays smooth learning of the model during training. It must be noted that all the values in Fig. 9 are evaluated using the widely adopted binary cross-entropy (BCE) loss function. Deep features extracted from the VGG16 model are fed to different local-search embedded FS algorithms and the results are discussed in the subsequent sections.

5.3 Performance of the local search embedded SSD algorithm

We apply a local search-based FS algorithm to gain the maximum possible accuracy as well as to reduce the number of features used for the classification purpose. As finding the best feature subset is a difficult task, especially in a wrapper-based FS model, we combine a local search method with an FS model. The results obtained from the proposed method are shown in Table 4. Table 4 shows that high accuracy of 96.07%, the precision of 96.30%, sensitivity (recall) of 99.28% and feature dimension reduction of 75% have been achieved.

For experimentation, we use two well-known local search methods, namely LAHC [76] and ABHC [64] to optimize the feature set. The experimental result on the DDSM database with the said local search-embedded SSD FS method is shown in Table 5. From Table 5, it is clear that the ABHC embedded SSD technique outperforms the LAHC embedded method in terms of classification accuracy as well as obtaining a smaller optimal feature set.

5.4 Comparison with various meta-heuristic based FS methods

We perform experiments using various other meta-heuristic based FS methods namely: Gravitational Search Algorithm (GSA) [77], WOA [78], Gray-wolf Optimization (GWO) [79], GA [80], PSO [81], Sine Cosine Algorithm (SCA) [82], Harmony Search (HS) algorithm [83] and Equilibrium Optimizer (EO) [84] for comparison of the proposed method. Meta-heuristic-based FS algorithms need to perform many mathematical operations to identify the best feature subset. For this, the algorithms in the literature use many sets of equations that are aided by different

Table 3 Performance of the attention-aided different deep feature extraction models on the DDSM database

Attention-aided TL model's accuracy (%)				
Simulation	VGG16	VGG19	ResNet50	EfficientNet
1	92.42	91.89	89.96	89.89
2	91.86	91.94	91.14	91.17
3	92.12	91.77	91.32	90.78
4	91.41	90.34	91.23	89.89
5	91.51	90.78	91.86	90.34
Mean ± SD	91.86 ± 0.42	91.34 ± 0.73	91.10 ± 0.69	90.41 ± 0.56

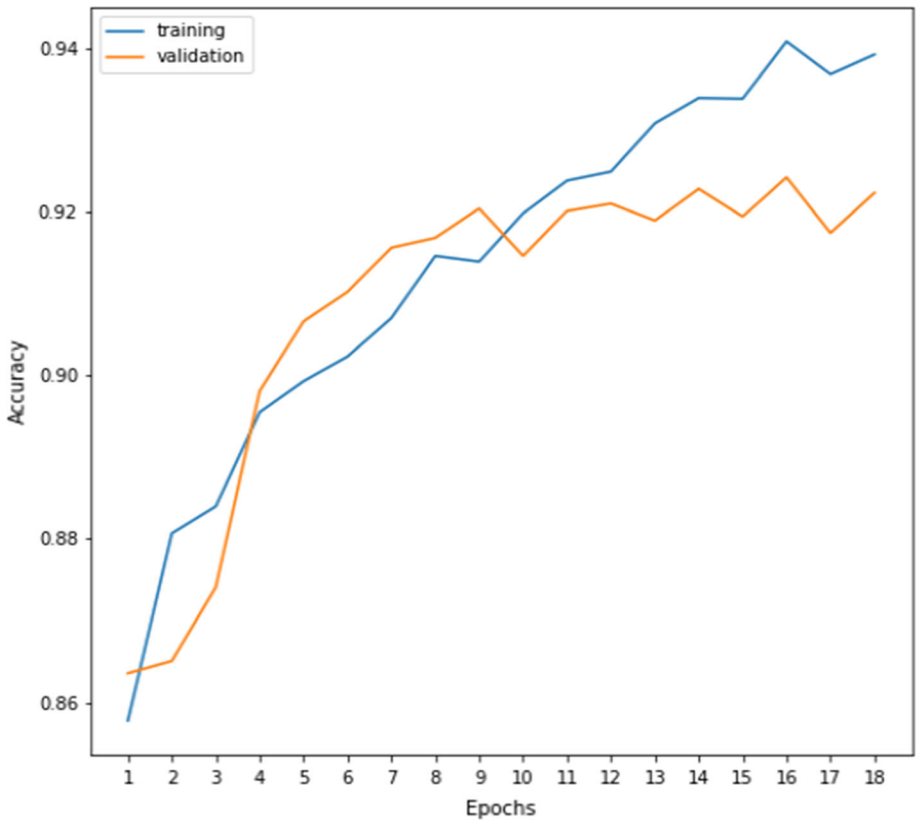


Fig. 8 Training and validation accuracies (%) over the number of epochs of the attention-aided VGG16 model on the DDSM database

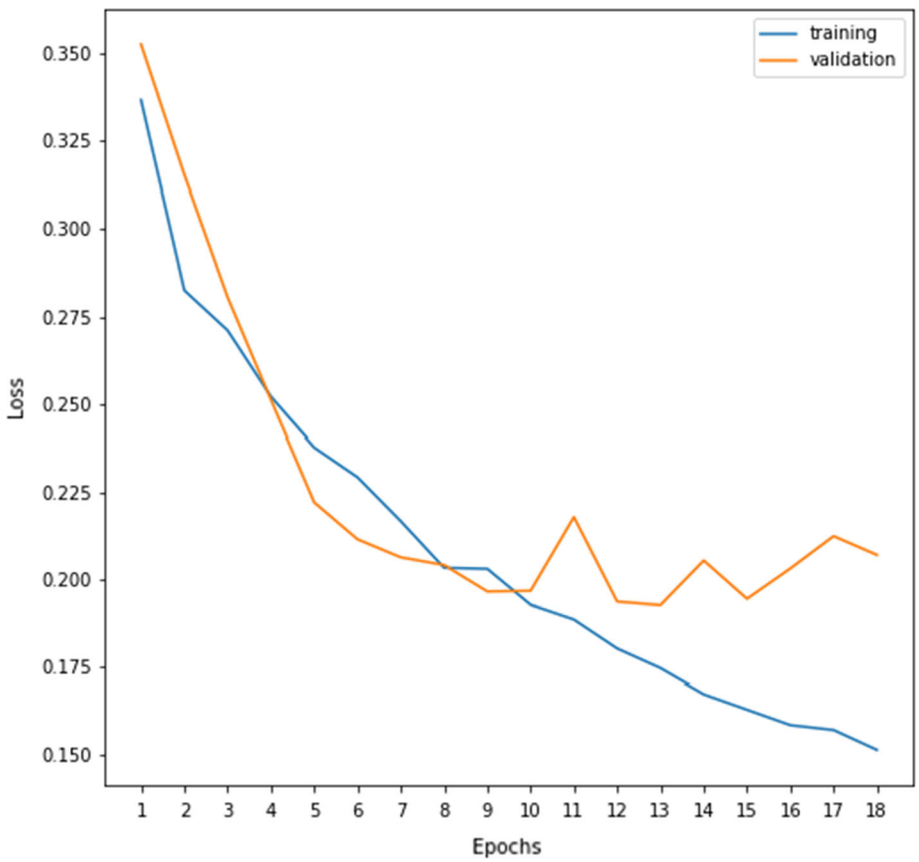


Fig. 9 Training and validation loss values (%) over the number of epochs of the attention-aided VGG16 model on the DDSM database

Table 4 Performance of the proposed method on the test set of DDSM database over five simulations

Simulation	Accuracy (%)	Precision (%)	Recall (%)	#FS
1	96.43	96.97	98.97	32
2	95.98	95.57	100.0	30
3	96.43	96.97	98.97	35
4	95.54	95.54	99.48	34
5	95.98	96.48	98.97	31
Mean ± SD	96.07 ± 0.37	96.30 ± 0.71	99.28 ± 0.21	32 ± 2

Table 5 Performance of the ABHC and LAHC local search embedded SSD algorithm on extracted deep features from the VGG16 model. Classification accuracy is in %

Simulation	Local search embedded SSD			
	SSD + LAHC		SSD + ABHC	
	Classification accuracy	No. of selected features	Classification accuracy	No. of selected features
1	94.19	52	96.43	32
2	94.64	59	95.98	30
3	94.28	53	96.43	35
4	93.98	56	95.54	34
5	93.86	47	95.98	31
Mean ± SD	94.19 ± 0.30	53 ± 5	96.07 ± 0.37	32 ± 2

Table 6 Different sets of hyperparameters and their values for various meta-heuristic based FS algorithms considered for the experimentation purposes

FS algorithm	Parameter(s)	Value(s)
Generic parameters	Population size	20
	Number of iterations	100
	Weight for accuracy (α	$\alpha = 0.98$
GSA	Initial gravitational constant (G_{init})	$G_{init} = 6$
	Constant (ε)	0.00001
WOA	Encircling parameter (a)	a lies in [0 2]
	Shape of spiral (b)	b = 1
GWO	Convergence operator (a)	a lies in [0 2]
GA	Gene selection	Roulette wheel
	Crossover probability	0.4
	Mutation probability	0.3
PSO	Inertia weight (I_W)	I_W lies in [0 1]
	Coefficients (r_1, r_2)	r_1 and r_2 lies in [0 1]
SCA	Constant (a)	a = 3
	Movement direction (r_1)	r_1 lies in [0 3]
HS	Harmony memory Considering rate (HMCR)	HMCR = 0.90
EO	Pool size	4
	Constants (a_1, a_2)	$a_1 = 2$ and $a_2 = 1$
	Generation rate (GP)	GP = 0.5

parameters. These parameters are crucial for controlling the optimization process and they have their own significance. The standard values of these parameters of various algorithms are used in this study. Table 6 contains a list of the parameters and their values. The simulation results of local search-based different meta-heuristics FS algorithms are shown in Table 7.

Table 7 Results from five simulations, including average and standard deviation (SD), after deep features from the VGG16 model are extracted and fed to several FS algorithms with embedded local search

FS method	LAHC embedded FS method		ABHC embedded FS method	
	Classification accuracy (%)	No. of features selected	Classification accuracy (%)	No. of features selected
GSA	93.30	65	95.08	70
	94.28	54	94.19	63
	94.19	51	94.28	56
	93.30	67	92.41	57
	93.33	65	93.30	64
Mean ± SD	93.68 ± 0.51	54 ± 7	93.85 ± 0.92	62 ± 6
WOA	94.64	48	94.64	55
	93.33	58	93.75	58
	93.75	59	93.33	61
	94.19	68	94.64	69
	94.19	48	94.64	60
Mean ± SD	94.02 ± 0.50	56 ± 8	94.19 ± 0.55	61 ± 5
GWO	94.64	76	94.64	69
	93.33	65	93.30	78
	93.75	81	93.33	63
	94.64	74	93.75	79
	93.33	65	94.64	75
Mean ± SD	93.94 ± 0.67	72 ± 7	93.93 ± 0.59	73 ± 7
GA	94.64	70	94.64	67
	93.30	59	93.75	62
	93.30	57	93.75	55
	93.33	56	93.33	59
	94.64	70	93.30	59
Mean ± SD	93.84 ± 0.73	62 ± 7	93.75 ± 0.48	60 ± 4
PSO	93.30	41	94.64	55
	93.33	42	94.19	52
	94.19	50	93.30	51
	92.86	46	93.33	48
	93.33	52	94.64	57
Mean ± SD	93.40 ± 0.48	46 ± 5	94.02 ± 0.59	53 ± 4
SCA	93.30	59	94.19	70
	93.33	71	93.75	67
	94.19	73	93.75	64
	93.33	65	93.33	63
	93.41	57	94.19	70
Mean ± SD	93.51 ± 0.38	65 ± 7	93.84 ± 0.32	67 ± 3
HS	93.30	63	92.41	60
	94.28	63	91.16	73
	95.08	56	93.33	53
	93.33	55	91.52	62
	92.42	65	91.96	68
Mean ± SD	93.68 ± 1.02	60 ± 5	92.07 ± 0.75	63 ± 8
EO	93.75	53	94.64	55
	94.28	56	94.64	69
	94.64	56	93.30	55
	94.64	50	94.28	63
	93.41	55	93.33	55
Mean ± SD	94.14 ± 0.55	54 ± 3	94.04 ± 0.60	59 ± 6

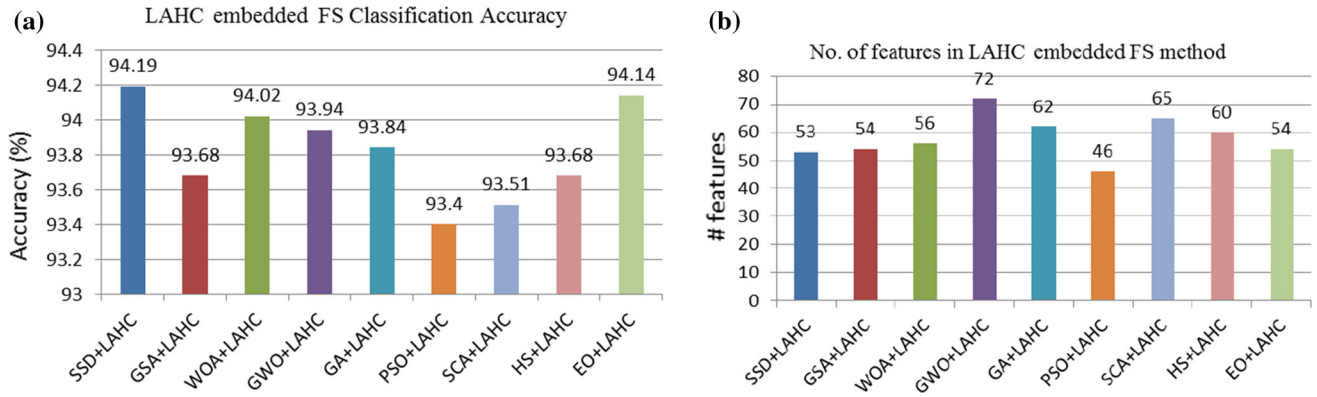


Fig. 10 Results of different FS algorithms embedded with LAHC local search

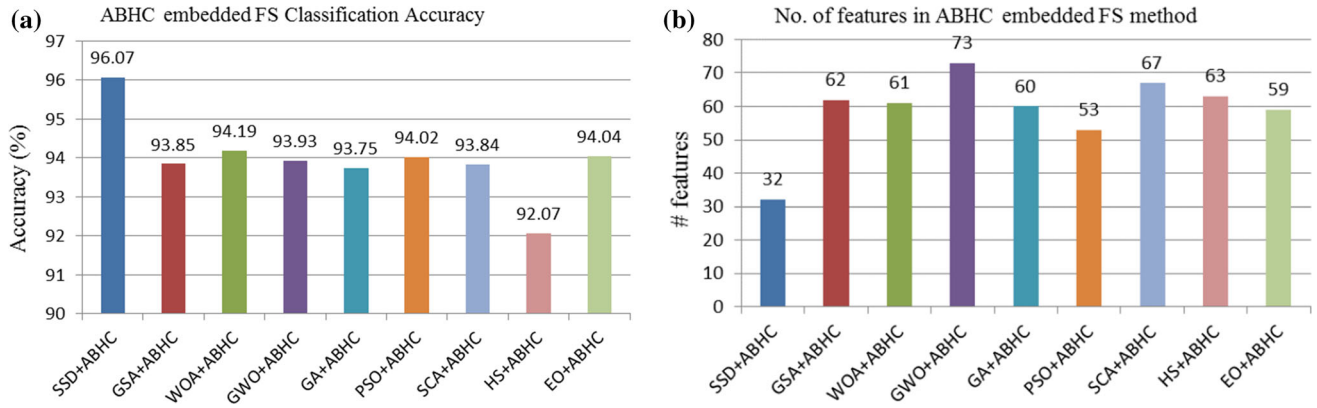


Fig. 11 Results of different FS algorithms embedded with ABHC local search (a) Classification accuracy (b) No. of selected features

From Tables 5 and 7, we can say that ABHC local search-embedded SSD algorithm outperforms others in terms of classification accuracy. Besides, it provides a subset of 32 features which is just 25% features of the given input features obtained from the VGG16 model. The comparative analysis with different combinations of various meta-heuristics and local search is shown in Figs. 10–11. The sine and cosine functions complicate the movement direction of the agents, which is the most essential feature of this SSD-based FS technique. This allows the algorithm to diversify, and the parameter h in Eq. (3) ensures that the algorithm remains stable between exploration and exploitation, allowing it to converge to better solutions. Furthermore, ABHC aids the algorithm to improve the solutions, thereby overcoming the local optima, thus leading to a better outcome.

Besides, ABHC based SSD method achieves not only better classification accuracy but also yields high precision and recall values which is crucial for medical image analysis. However, LAHC and ABHC, both the local search methods are based on the hill-climbing optimization technique. The methods differ in the way of finding the better agent having better fitness value toward the final reduced solution.

Figure 12 depicts the suggested method's Receiver Operating Characteristic (ROC) curve, which has an AUC value of 0.881. The ROC curves depict the trade-off between a classifier's true positive rate (TPR) and false positive rate (FPR). Classifiers with curves that are closer to the top-left corner perform better. If the curve approaches the ROC space's 45-degree diagonal, the classification result becomes less accurate.

Besides, from Fig. 11, it can be said that our proposed model yields a promising result not only in terms of the classification accuracy but also the same is achieved with a very less number of features that we get from the ABHC embedded SSD FS algorithm as compared to other FS algorithms.

Here the FS algorithm produces the optimal subset of features by discarding redundant or less informative features from the deep features obtained from the CNN model and local search increases the exploitation capability of the FS algorithm by finding the location of an agent in the search space having better fitness value.

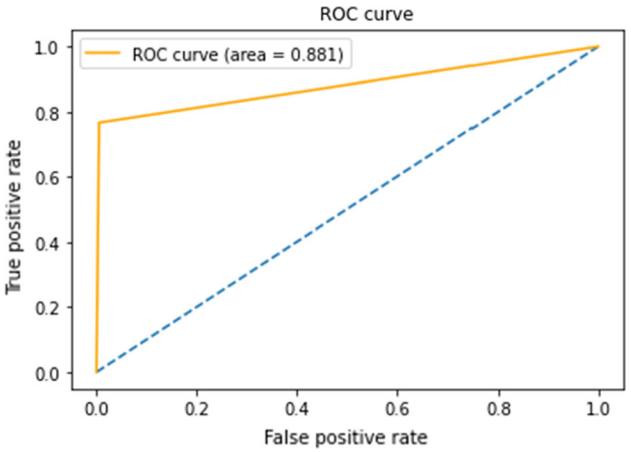


Fig. 12 The ROC curve with the AUC value of the DDSM database using our proposed method

Table 8 Comparative results for different combinations of the proposed model in terms of classification accuracy (%) on the test set of the DDSM database

Method	Classification accuracy (%)
End-to-end VGG16	91.41
VGG16 + KNN	90.70
VGG16 + SSD + KNN	92.86
VGG16 + ABHC aided SSD + KNN	95.98

Furthermore, we experiment with different modules of our proposed method to observe the effectiveness of each module and noted the experimental results in Table 8. From the table, it can be observed that an attention-aided VGG16 model yields a classification accuracy of 91.41% on the test set, whereas if we use the KNN classifier to it, the classifier produces a classification accuracy of 90.70% which is low compared to end-to-end VGG16 model. Also, if we apply FS using the SSD algorithm, improved classification accuracy is observed. Further, if the ABHC local search is incorporated with the SSD-based FS method, we achieve improved classification accuracy. Therefore, from these experimental results, it can be ensured the effectiveness of the individual modules (i.e., deep feature extraction with VGG16, FS with ABHC aided SSD method) of our proposed method. The gain in performance has to be attributed to both the DL and FS approaches.

5.5 Computational complexity analysis of the proposed method

The asymptotic analysis is performed on the proposed ABHC embedded SSD method. The space complexity of the proposed method is $Big - O(PS * Dim)$, where PS is the population size and Dim is the feature dimension. The time complexity of the ABHC-embedded SSD method shows that the worst-case time cost in asymptotic notation is

$$Big - O\big(Max\ \ iteration \times PS \times \big(T_{fit} + T_{knn} + Dim\big)\big).$$

where the Maxiteration is the maximum number of iterations for the local search-based FS method, T_{fit} is the time to calculate fitness function and T_{knn} is the time to run KNN algorithm.

5.6 Statistical analysis of the proposed method

We perform a statistical significance test to assess the proposed algorithm's robustness in comparison to other meta-heuristic algorithms embedded with the ABHC-based local search. The following statement is taken into account as a null hypothesis while we do this test: "The proposed ABHC local search aided SSD method gives similar outcomes when compared to other meta-heuristic techniques embedded with the ABHC based local search." We use the Mann–Whitney U test [91], a widely used nonparametric statistical technique, to reject this null hypothesis. This test is predicated on the notion that two distributions, X and Y, are ranked in ascending order according to their respective values. The majority of the samples in X must be above or below the majority of the samples in Y for a condition to hold [91]. We take into account the classification accuracy of different FS techniques for each of the five simulations to construct the statistical evidence. The results obtained from performing the test are shown in Table 9. If the calculated p-value is higher than 0.05 (5%), we conclude that the null hypothesis has sufficient statistical support to be accepted. If not, we reject this hypothesis. It can be seen from Table 9 that for every case p-value is less than 0.05 which means that the ABHC local search aided SSD

Table 9 Results obtained on statistically analyzing the ABHC aided SSD algorithm with other FS algorithms embedded with the ABHC based local search using Mann–Whitney U test

ABHC embedded FS method	p-value
ABHC + GSA	0.01167
ABHC + WOA	0.01066
ABHC + GWO	0.01141
ABHC + GA	0.01192
ABHC + PSO	0.01141
ABHC + SCA	0.01115
ABHC + HS	0.01218
ABHC + EO	0.01208

method is statistically different from the other methods considered here.

5.7 Comparison with state-of-the-art approaches

Finally, we compare the performance of the proposed method with some recently developed classification methods and tabulate the same in Table 10. The performance of our method is superior to that of four DL-based methods used in [4, 41, 85], and [86]. Moreover, the performances of the method described in [87, 88], based on feature fusion to classify benign and malignant masses, are slightly lower than that of our method. Hence, from Table 10, it is clear that the proposed model outperforms these previous works with a good margin and the experimental results establish the superiority and robustness of our proposed method.

5.8 Advantages and limitations of the proposed method

Although our proposed method poses good results for cancer detection in mamograms, there are some pros and cons to this work. In this section, we discuss the advantages and limitations of our proposed method. The core advantages of this current research work are as follows:

1. We create a model for breast cancer classification from mammograms that combines the principles of deep learning and optimization algorithms.
2. We introduce an attention mechanism on a deep CNN-based transfer learning model, called VGG-16, and fine-tune it for the extraction of deep features from the input images.
3. We embed a local search, namely ABHC with SSD based FS algorithm to produce an optimal feature subset from the features produced by the said CNN model.
4. We achieve state-of-the-art classification accuracy, high precision and recall values with just 25% of features of the original feature set obtained by the CNN model when evaluated on the DDSM database.

The limitations of this research work are described as follows:

1. The initialization in the optimization algorithm is random. So it may sacrifice some results in terms of accuracy and convergence time. Instead of random initialization, techniques like a chaotic map can be explored for better results.
2. Time complexity is a factor on which we need to put more emphasis in the future.
3. Another issue is it may result in early convergence for some inputs.

6 Conclusions and future works

Breast cancer is a significant problem that affects women all over the world, therefore it is critical to recognize any early signs of the disease and treat it with the help of medical specialists. We present a model for breast cancer categorization based on mammography masses in this paper. First, we extract features from the VGG16 model with care. We use the GAP layer instead of the fully connected layer to implement an attention mechanism over the original VGG16 model. Because it enforces the relationship between a feature map and the category, it is more organic to the convolution construction. We then fed the retrieved features into a local search embedded FS model, and it was discovered that the ABHC embedded SSD technique outperforms the local search embedded FS model with a smaller number of features. The FS algorithm minimizes the quantity of features, but local search improves the FS method's exploitation potential and yields the best subset of features. With just 25% of features extracted by the DL model, our proposed model achieves state-of-the-art classification accuracy, precision, and recall on the DDSM dataset. Other medical image datasets of other modalities can be examined in the future to verify the resilience of the proposed model, as the notion of the present study is dataset independent. Also, the present work deals with a classification problem, whereas the segmentation of breast lesions is another challenging area in the

Table 10 Performance comparison of the proposed model with state-of-the-art models on the DDSM database

Model	Dataset	Accuracy (%)	Precision (%)	Recall (%)
Levy et al. [4]	DDSM	92.9	92.4	93.4
Falconi et al. [41]		84.4	–	–
Xiao et al. [85]		–	82.2	94.9
Arias et al. [86]		92	–	–
Zhang et al. [87]		94.30	–	89.97
Li et al. [88]		94.7	–	94.1
Proposed		96.07	96.30	99.28

medical image analysis that can be explored in future research attempts. Furthermore, the SSD algorithm's fitness function and position updating approach are essential variables for feature reduction that may be enhanced further. Also, parallel methods can be used to speed up computations on higher-dimensional datasets.

Acknowledgements The authors would like to thank the Centre for Microprocessor Applications for Training, Education and Research (CMATER) research laboratory of the Computer Science and Engineering Department, Jadavpur University, Kolkata, India for providing the infrastructural support.

Funding Open access funding provided by Óbuda University.

Data availability The dataset used and analyzed during the current study is publicly available at: https://www.kaggle.com/datasets/skooch/ddsm-mammography.

Declarations

Conflict of interest The authors declare that there are no conflict of interests.

References

1. Gautherie M (1980) Thermopathology of breast cancer: measurement and analysis of in vivo temperature and blood flow. Ann N Y Acad Sci 335(1):383–415
2. Akbar S, Akram MU, Sharif M, Tariq A, Khan SA (2018) Decision support system for detection of hypertensive retinopathy using arteriovenous ratio. Artif Intell Med 90:15–24
3. Khamparia A, Bharati S, Podder P et al (2021) Diagnosis of breast cancer based on modern mammography using hybrid transfer learning. Multidim Syst Sign Process 32:747–765. https://doi.org/10.1007/s11045-020-00756-7
4. D Lévy, A Jain (2016) Breast mass classification from mammograms using deep convolutional neural networks. ArXiv, abs/1612.00542
5. Khan MA, Sharif M, Akram T, Yasmin M, Nayak RS (2019) Stomachdeformities recognition using rank-based deep features selection. J Med Syst 43(12):329
6. Al Husaini MAS, Habaebi MH, Hameed SA, IslamMR Gunawan TS (2020) A systematic review of breast cancer detection using thermography and neural networks. IEEE Access 8:208922–208937
7. Shen D, Wu G, Suk HI (2017) Deep learning in medical image analysis. Annu Rev Biomed Eng 19:221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442
8. Shen L, Margolies LR, Rothstein JH et al (2019) Deep learning to improve breast cancer detection on screening mammography. Sci Rep 9:12495. https://doi.org/10.1038/s41598-019-48995-4
9. Sultana F, Sufian A, Dutta P (2020) Evolution of image segmentation using deep convolutional neural network: a survey,-knowledge-based systems 201–202: 106062, ISSN 0950–7051. https://doi.org/10.1016/j.knosys.2020.106062. (https://www.sciencedirect.com/science/article/pii/S0950705120303464)
10. Ray B, Mukhopadhyay S, Hossain S et al (2021) Image steganography using deep learning based edge detection. Multimed Tools Appl 80:33475–33503. https://doi.org/10.1007/s11042-021-11177-4
11. Song F, Ai Z, Zhang H, You I, Li S (2020) Smart collaborative balancing for dependable network components in cyber-physical systems. IEEE Trans Industr Inf 17(10):6916–6924
12. Mukhopadhyay S, Hossain S, Ghosal SK et al (2021) Secured image steganography based on catalan transform. Multimed Tools Appl 80:14495–14520. https://doi.org/10.1007/s11042-020-10424-4
13. Cao Z, Wong K, Lin CT (2021) Weak human preference supervision for deep reinforcement learning. IEEE Trans Neural Networks Learn Syst 32(12):5369–5378
14. Y. Fang, Y. Xu, H. Li, X. He and L. Kang (2020) Writing in the air: recognize letters using deep learning through WiFi signals. In: 2020 6th international conference on big data computing and communications (BIGCOM). pp. 8–14, doi: https://doi.org/10.1109/BigCom51056.2020.00008.
15. Han J, Jentzen A, Ee W (2017) Solving high-dimensional partial differential equations using deep learning. Proc Natl Acad Sci. https://doi.org/10.1073/pnas.1718942115
16. Islam J, Zhang Y (2017) A novel deep learning based multi-class classification method for Alzheimer's disease detection using brain MRI data. Inter-national conference on brain informatics. Springer, London, pp 213–222
17. Lindsey R, Daluiski A, Chopra S, LachapelleA MM, Sicular S et al (2018) Deep neural network improves fracture detection by clinicians. Procee Natl Acad Sci 115(45):11591–11596
18. Karbhari Y, Basu A, Geem ZW, Han GT, Sarkar R (2021) Generation of Synthetic Chest X-ray Images andDetection of COVID-19: A Deep Learning Based Approach. Diagnostics. 11(5). Available from: https://www.mdpi.com/2075-4418/11/5/895
19. Kocmi T (2020) Exploring benefits of transfer learning in neural machine translation. arXivpreprintar Xiv:200101622.
20. Kira Kenji, Rendell Larry A (1992) A practical approach to feature selection. Machine learning proceedings 1992. Elsevier, Amsterdam, pp 249–256
21. Guha R, Alam I, Bera SK, Kumar N, Sarkar R (2022) Enhancement of image contrast using selfish herd optimizer. Multimed Tools Appl 81(1):637–657
22. Bi J, Zhou G, Zhou Y, Luo Q, Deng W (2022) Artificial electric field algorithm with greedy state transition strategy for spherical multiple traveling salesmen problem. Int J Computat Intell Syst 15(1):1–24
23. Hossain S, Mukhopadhyay S, Ray B, Ghosal SK, Sarkar R (2022) A secured image steganography method based on ballot transform and genetic algorithm. Multimed Tools Appl 24:1–30
24. Sheikh KH, Ahmed S, Mukhopadhyay K, Singh PK, Yoon JH, Geem ZW, Sarkar R (2020) EHHM: electrical harmony based hybrid meta-heuristic for feature selection. IEEE Access 27(8):158125–158141
25. Tharwat A, Gabel T (2019) Parameters optimization of support vector machines for imbalanced data using social ski driver

algorithm. Neural Comput Appl. https://doi.org/10.1007/s00521-019-04159-z
26. Shaw SS, Ahmed S, Malakar S, Garcia-Hernandez L, Abraham A, Sarkar R (2021) Hybridization of ring theory-based evolutionary algorithm and particle swarm optimization to solve class imbalance problem. Compl Intell Syst 7(4):2069–2091
27. Ghosh M, Sen S, Sarkar R, Maulik U (2021) Quantum squirrel inspired algorithm for gene selection in methylation and expression data of prostate cancer. Applied Soft Computing. 105:107221. Available from:https://www.sciencedirect.com/science/article/pii/S1568494621001447
28. Chyzhyk D, Savio A, Graña M (2014) Evolutionary ELM wrapper feature selection for Alzheimer's disease CAD on anatomical brain MRI. Neurocomputing 128:73–80
29. Chatterjee B, Bhattacharyya T, Ghosh KK, Singh PK, Geem ZW, Sarkar R (2020) Late acceptance hill climbing based social ski driver algorithm for feature selection. IEEE Access 8:75393–75408. https://doi.org/10.1109/ACCESS.2020.2988157
30. Pramanik P, Mukhopadhyay S, Kaplun D, Sarkar R (2022) A deep feature selection method for tumor classification in breast ultrasound images. In: International Conference on Mathematics and its Applications in new Computer Systems. Springer, Cham. pp. 241–252
31. Ghosh S, Pramanik P (2019) A combined framework for dimensionality reduction of hyperspectral images using feature selection and feature extraction. In: 2019 IEEE recent advances in geoscience and remote sensing: technologies, standards and applications (TENGARSS). pp. 39–44. IEEE.
32. Basu A, Sheikh KH, Cuevas E, Sarkar R (2022) COVID-19 detection from CT scans using a two-stage framework. Expert Syst Appl 193(0957–4174):116377. https://doi.org/10.1016/j.eswa.2021.116377
33. Saber A, Sakr M, Abo-Seida OM, Keshk A, Chen H (2021) A novel deep-learning model for automatic detection and classification of breast cancer using the transfer-learning technique. IEEE Access 9:71194–71209. https://doi.org/10.1109/ACCESS.2021.3079204
34. Ertosun MG, Rubin DL (2015) Probabilistic visual search for masses within mammography images using deep learning. IEEE Int Conf Bioinformat Biomed (BIBM) 2015:1310–1315. https://doi.org/10.1109/BIBM.2015.7359868
35. Al-antari MA, Al-masni MA, Choi M-T, Han S-M, Kim T-S (2018) A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. Int J Med Informatics 117:44–54
36. Chattopadhyay S, Dey A, Singh PK, Sarkar R (2022) DRDA-Net: dense residual dual-shuffle attention network for breast cancer classification using histopathological images. Comput Biol Med 145(0010–4825):105437. https://doi.org/10.1016/j.compbiomed.2022.105437
37. Sanyal R, Jethanandani M, Sarkar R (2021) DAN: breast cancer classification from high resolution histology images using deep attention network. In: Sharma MK, Dhaka VS, Perumal T, Dey N, Tavare JMRS (eds) innovations in computational intelligence and computer vision advances in intelligent systems and computing. Springer, Singapore, pp 319–326
38. Tiwari P, Melucci M (2019) Towards a quantum-inspired binary classifier. IEEE Access 7:42354–42372. https://doi.org/10.1109/ACCESS.2019.2904624
39. Al Mamun R, Rafin GA, Sefat MA. Application of deep convolution neural network in breast cancer prediction using digital mammograms.
40. Khatami A, Nazari A, Khosravi A, Lim CP, Nahavandi S (2020) A weight perturbation-based regularisation technique for convolutional neural networks and the application in medical imaging. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2020.113196
41. Falconi LG, Perez M, Aguilar WG, Conci A (2020) transfer learning and fine tuning in breast mammogram abnormalities classification on CBIS-DDSM database. Adv Sci, Technol Eng Syst J 5(2):154–165
42. Vahadane A et al. (2015) Structure-preserved color normalization for histological images. In: 2015 IEEE 12th international symposium on biomedical imaging (ISBI). pp. 1012–1015. doi: https://doi.org/10.1109/ISBI.2015.7164042
43. Sarmiento A, Fondón I (2018) Automatic breast cancer grading of histological images based on colour and texture descriptors. In: Campilho A, Karray F, terHaarRomeny B (eds) image analysis and recognition ICIAR 2018 lecture notes in computer science. Springer, Cham, pp 887–894
44. Nawaz W, Ahmed S, Tahir A, Khan HA (2018) Classification of breast cancer histology images using ALEXNET. In: Campilho A, Karray F, terHaarRomeny B (eds) image analysis and recognition ICIAR 2018 lecture notes in computer science. Springer, Cham, pp 869–876
45. Silva L, Seixas F, Fontes C, Muchaluat-SaadeD,Conci A (2020) A computational method for breast Ab-normality detection using thermographs. In: 2020IEEE 33rd international symposium on computer-based medical systems (CBMS). p. 469–474.
46. Krithiga R, Geetha P (2020) Deep learning based breast cancer detection and classification using fuzzy merging techniques. Mach Vis Appl 31:63
47. Rakhlin A, Shvets A, Iglovikov V, Kalinin AA (2018) Deep convolutional neural networks for breast cancer histology image analysis. In: Campilho A, Karray F, terHaarRomeny B (eds) image analysis and recognition. ICIAR 2018 lecture notes in computer science. Springer, Cham, pp 737–744
48. Sharma M, Kaur P (2020) A comprehensive analysis of nature-inspired meta-heuristic techniques for feature selection problem. Arch Computat Methods Eng. https://doi.org/10.1007/s11831-020-09412-6
49. Zhang J, Huang Y, Wang Y, Ma G (2020) Multi-objective optimization of concrete mixture proportions using machine learning and metaheuristic algorithms. Constr Build Mater 253:119208. https://doi.org/10.1016/j.conbuildmat.2020.119208
50. Kaur P, Sharma M (2019) Diagnosis of human psychological disorders using supervised learning and nature-inspired computing techniques: a meta-analysis. J Med Syst. https://doi.org/10.1007/s10916-019-1341-2
51. Zhou J, Qiu Y, Zhu S, Armaghani DJ, Li C, Nguyen H, Yagiz S (2021) Optimization of support vector machine through the use of metaheuristic algorithms in forecasting TBM advance rate. Eng Appl Artif Intell 97: 104015. ISSN 0952–1976. doi https://doi.org/10.1016/j.engappai.2020.104015. (https://www.sciencedirect.com/science/article/pii/S0952197620303018)
52. Gandhi KR, Karnan M, Kannan S (2010) Classification rule construction using particle swarm optimization algorithm for breast cancer data sets. In: 2010 International Conference on Signal Acquisition and Processing. doi:https://doi.org/10.1109/icsap.2010.58
53. Ahmad F, Mat Isa NA, Hussain Z, Osman MK, Sulaiman SN (2014) A GA-based feature selection and parameter optimization of an ANN in diagnosing breast cancer. Pattern Anal Appl 18(4):861–870. https://doi.org/10.1007/s10044-014-0375-9
54. Huang H, Feng X, Zhou S et al (2019) A new fruit fly optimization algorithm enhanced support vector machine for diagnosis of breast cancer based on high-level features. BMC Bioinformatics 20:290. https://doi.org/10.1186/s12859-019-2771-z
55. Sayed GI, Darwish A, Hassanien AE (2019) Binary whale optimization algorithm and binary moth flame optimization with clustering algorithms for clinical breast cancer diagnoses. J Classif. https://doi.org/10.1007/s00357-018-9297-3

56. Fang H, Fan H, Lin S, Qing Z, Sheykhahmad FR (2020) Automatic breast cancer detection based on optimized neural network using whale optimization algorithm. Int J Imaging Syst Technol. https://doi.org/10.1002/ima.22468
57. Oyelade ON, Ezugwu AE (2021) Characterization of abnormalities in breast cancer images using nature-inspired metaheuristic optimized convolutional neural networks model. Concurr Computat Pract Exper 22:e6629
58. Tavasoli N, Rezaee K, Momenzadeh M, Sehhati M (2021) An ensemble soft weighted gene selection-based approach and cancer classification using modified metaheuristic learning. J Computat Des Eng 8(4):1172–1189. https://doi.org/10.1093/jcde/qwab039
59. Rezaee K, Rezaee A, Shaikhi N et al (2020) Multi-mass breast cancer classification based on hybrid descriptors and memetic meta-heuristic learning. SN Appl Sci 2:1297. https://doi.org/10.1007/s42452-020-3103-7
60. Ezzat D, Hassanien AE, Ella HA (2021) An optimized deep learning architecture for the diagnosis of COVID-19 disease based on gravitational search optimization. Appl Soft Comput 1(98):106742
61. Majji R, Prakash PG, Cristin R, Parthasarathy G (2020) Social bat optimisation dependent deep stacked auto-encoder for skin cancer detection. IET Image Proc 14(16):4122–4131
62. Ezzat D, Ella HA (2020) GSA-DenseNet121-COVID-19: a hybrid deep learning architecture for the diagnosis of COVID-19 disease based on gravitational search optimization algorithm. arXiv preprint arXiv:2004.05084
63. Simonyan K, Zisserman A (2015) Very deep convolutional networksfor large-scale image recognition.
64. Al-Betar MA, Aljarah I, Awadallah MA et al (2019) Adaptive $\beta-\beta-$hill climbing for optimization. Soft Comput 23:13489–13512. https://doi.org/10.1007/s00500-019-03887-7
65. Min Lin, Qiang Chen, and Shuicheng Yan (2014) Network in network
66. Keras documentation url: https://keras.io/api/layers/core_layers/lambda/
67. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(1):1929–1958
68. Altman NS (1992) 'An introduction to kernel and nearest-neighbor nonparametric regression.' Amer Statist 46(3):175–185. https://doi.org/10.1080/00031305.1992.10475879
69. Emary E, Zawbaa HM, Hassanien AE (2016) 'Binary grey wolf optimization approaches for feature selection.' Neurocomputing 172:371–381. https://doi.org/10.1016/j.neucom.2015.06.083
70. Mafarja MM, Mirjalili S (2017) Hybrid whale optimization algorithm with simulated annealing for feature selection. Neurocomputing 260:302–312. https://doi.org/10.1016/j.neucom.2017.04.053
71. Mafarja M, Qasem A, Heidari AA, Aljarah I, Faris H, Mirjalili S (2020) Efficient hybrid nature-inspired binary optimizers for feature selection. Cognit Comput 12(1):150–175. https://doi.org/10.1007/s12559-019-09668-6
72. Eric A (2021) Scuccimarra, DDSM Mammography. Available online: https://www.kaggle.com/skooch/ddsm-mammography (Accessed on 5July 2021).
73. The digital database for screening mammography Michael Heath (2001) K Bowyer, D Kopans, R Moore, Kegelmeyer WP. In: Proceedings of the Fifth International Workshop on Digital Mammography. Yaffe MJ ed, Medical Physics Publishing. ISBN 1–930524–00–5. pp 212–218
74. Lee R, Gimenez F, Hoogi A et al (2017) A curated mammography data set for use in computer-aided detection and diagnosis research. Sci Data 4:170177. https://doi.org/10.1038/sdata.2017.177
75. Han J, Kamber M, Pei J (2012) Data mining: Concepts and techniques 3rd edn). Kaufmann Publishers, Burlington, pp 19–24
76. Burke EK, Bykov Y (2017) 'The late acceptance hill-climbing heuristic.' Eur J Oper Res 258(1):70–78. https://doi.org/10.1016/j.ejor.2016.07.012.]
77. Rashedi E, Nezamabadi-pour H, Saryazdi S (2009) GSA: a gravitational search algorithm. Informat Sci 179(13):2232–2248
78. Mirjalili Seyedali, Lewis Andrew (2016) The whale optimization algorithm. Adv Eng Soft 95:51–67
79. Mirjalili S, Mirjalili SM, Lewis A (2014) G wolf optimizer. Adv Eng Soft 69:46–61
80. Holland JH (1984) Genetic Algorithms and Adaptation. In: Selfridge OG, Rissland EL, Arbib MA (eds) Adaptive control of Ill-defined systems NATO conference series (II systems science). Boston, MA, Springer
81. Khanesar, Mojtaba Ahmadieh, Teshnehlab Mohammad Aliyari Shoorehdeli, Mahdi. (2007). A novel binary particle swarm optimization. In: 15th IEEE Med. Conf. Control Automation, Athens, Greece. pp 1–6. https://doi.org/10.1109/MED.2007.4433821
82. Mirjalili S (2016) SCA: a sine cosine algorithm for solving optimization problems. Knowl Based Syst 96:120–133. https://doi.org/10.1016/j.knosys.2015.12.022
83. Woo Z, Hoon J, Loganathan GV (2001) A new heuristic optimization algorithm: harmony search. SIMULATION 76(2):60–68. https://doi.org/10.1177/003754970107600201
84. Faramarzi A, Heidarinejad M, Stephens B, Mirjalili S (2020) Equilibrium optimizer: Anovel optimization algorithm. Knowl-Based Syst 191:105190
85. H. Xiao et al. (2020) CSABlock-based Cascade RCNN for Breast Mass Detection inMammogram. In: 2020 IEEE International Conference on Bioinformatics and Biomedicine(BIBM). pp. 2120–2124. doi https://doi.org/10.1109/BIBM49941.2020.9313473.79
86. Arias R, Narváez F, Franco H (2019) Evaluation of learning approaches based on convolutional neural networks for mammogram classification. International conference on smart technologies, systems and applications. Springer, Cham, pp 273–287
87. Zhang Q, Li Y, Zhao G, Man P, Lin Y, Wang M (2020) A novel algorithm for breast mass classification in digital mammography based on feature fusion. J Health Eng 22:2020
88. Li H, Niu J, Li D, Zhang C (2021) Classification of breast mass in two-view mammograms via deep learning. IET Image Proc 15(2):454–467
89. Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 1:50–60
90. Akyol S, Alatas B (2017) Plant intelligence based metaheuristic optimization algorithms. Artif Intell Rev 47(4):417–462
91. Alatas B, Bingol H (2020) Comparative assessment of light-based intelligent search and optimization algorithms. Light Eng 28(6):51–59
92. Alatas B, Bingol H (2019) A physics based novel approach for travelling tournament problem: optics inspired optimization. Informat Technol Contl 48(3):373–388
93. Bingol H, Alatas B (2020) Chaos based optics inspired optimization algorithms as global solution search approach. Chaos, Solitons Fractals 1(141):110434