Quality Assessment of Tea Employing Electronic Tongue (ET): Exploration of Feature Extraction Techniques and Prediction Algorithms

Thesis submitted

By

Srikanta Acharya

Doctor of Philosophy (Engineering)

Department of Instrumentation and Electronics Engineering

Faculty Council of Engineering & Technology

Jadavpur University, Kolkata, India

2024

JADAVPUR UNIVERSITY

KOLKATA- 700032

INDIA

INDEX NO. 168/19/E

Title of the Thesis **Quality Assessment of Tea Employing Electronic Tongue (ET): Exploration of Feature Extraction Techniques and Prediction Algorithms**

Professor
Dept. of Instrumentation & Electronics Engg
Jadavpur University
Saltlake, 2nd Campus
Kolkata-700 106

Name, Designation& Institution of theSupervisor

Dr. Runu Banerjee Roy
Professor
Dept. of Instrumentation and Electronics Engineering Jadavpur University, Salt Lake Campus, Sector-III, Block-LB, Plot-8, Kolkata- 700106, India. Phone: 08240235742, Fax: 03323357254

<u>List of Publications</u>

Journals

[1] **S. Acharya**, D. Das, T. N. Chatterjee, S. Mukherjee, R. B. Roy, B. Tudu, *et al.*, "Voltammetric electrode array optimization for black tea discrimination using computational intelligence approach," *IEEE Sensors Journal*, vol. 21, pp. 20589-20595, 2021.

[2] **S. Acharya,** D. Das, S. Nag, S. Mukherjee, A. K. Hazarika, S. Sabhapondit, *et al.*, "Optimization techniques for a voltammetric signal to predict green tea quality parameters using MIP electrode," *IEEE Sensors Journal*, vol. 23, no. 17, pp. 19842-19847,2023.

[3] **S. Acharya,** S. Nag, D. Bandyopadhyay, D. Das, A. Mandal, and R. B. Roy, "A Molecular Imprinted Polymer Tethered Capacitive Sensor for Epicatechin Detection in Green Tea," *IEEE Sensors Journal*, vol. 24, no. 4, pp. 4213-4220, 2024.

Conference Proceedings

[1] **S. Acharya,** T. N. Chatterjee, S. Mukherjee, D. Das, B. Tudu, R. Bandyopadhyay, *et al.*, "Optimization of electrode array in electronic tongue for classification of black tea," in *2018 IEEE Applied Signal Processing Conference (ASPCON)*, 2018, pp. 134-137.

[2] **S. Acharya,** T. N. Chatterjee, S. Mukherjee, D. Das, R. B. Roy, B. Tudu, *et al.*, "Selection of optimum number of sensors of an electronic tongue for efficient classification of black tea: A combinatorial approach based on discrete cosine transform and artificial neural network," in *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, 2018, pp. 108-111.

[3] **S. Acharya,** S. Nag, D. Bandyopadhyay, S. Mukherjee, D. Das, and R. B. Roy, "Unveiling the capacitive sensor performance developed for Epicatechin detection in Green tea: A Clustering Approach," in *2024 IEEE 3rd International Conference on Control, Instrumentation, Energy & Communication (CIEC)*, 2024, pp. 175-179.

[4] **S. Acharya,** D. Das,S. Nag, D.Bandyopadhyay,A. Mondal, B. Tudu, P. Pramanik, R. Bandyopadhyay, R. B. Roy, "Gallic Acid Detection Using A PMMA/PDA Coated Capacitive Sensor", Nano-architectures for Chemical Biological and Therapeutic Applications (NACBTA)2021,November 12-14th, 2021, Department of Chemistry GLA University, Mathura, UP and Jadavpur University**,** Kolkata .

Statement of Originality

*I, **Srikanta Acharya**, registered on **20th June, 2019**, do hereby declare that this thesis entitled "**Quality Assessment of Tea Employing Electronic Tongue (ET): Exploration of Feature Extraction Techniques and Prediction Algorithms**" contains literature survey and original research work done by the undersigned candidate as part of Doctoral studies.*

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

I also declare that I 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 **7%.**

Srikanta Acharya

Professor
Dept. of Instrumentation & Electronics Engg
Jadavpur University
Saltlake, 2nd Campus
Kolkata-700 106

Srikanta Acharya
File No. 168/19/E
Date: 23.10.2024

Dr.Runu Banerje Roy
Professor
Dept. of Instrumentation and Electronics Engineering
Jadavpur University, Salt Lake Campus, Kolkata

Certificate from the Supervisor

Date: 23.10.2024

*This is to certify that the thesis entitled "**Quality Assessment of Tea Employing Electronic Tongue (ET): Exploration of Feature Extraction Techniques and Prediction Algorithms**", submitted by Srikanta Acharya, who got his name registered on 20th June, 2019, for the award of Ph.D. (Engineering) degree of Jadavpur University, is absolutely based upon his own work under supervision of Prof. Runu Banerjee Roy and that neither his thesis nor any part of the thesis has been submitted for any degree/diploma or any other academic award anywhere before.*

Professor
Dept. of Instrumentation & Electronics Engg
Jadavpur University
Saltlake, 2nd Campus
Kolkata-700 106

Dr. Runu Banerjee Roy
Professor
Dept. of Instrumentation and Electronics Engineering
Jadavpur University, Salt Lake Campus,

This thesis is dedicated to My Parents,

My Wife and My Daughter

Acknowledgement

Any success requires the effort of many, and this is no exception. First and foremost, I would like to extend my sincere gratitude to my supervisor, Dr. Runu Banerjee Roy, Professor at Jadavpur University. Without her able guidance, support, and constant encouragement, this thesis would not have materialized. Her technical expertise and motivation have helped me overcome the difficulties during this course of this work. Her kind assistance and her patience with my knowledge gaps in this area have been unparalleled. It was a great privilege and honor to work and study under her guidance.

I would also like to extend my profound gratitude and heartfelt thanks to Prof. Rajib Bandyopadhyay, Dept. of IEE, Jadavpur University, for his invaluable advice at critical junctures of this course. His lectures have left a strong impression on me, and I have always cherished the positive memories of our interactions. I am grateful to him for being the foundation upon which this laboratory, and all the work and emotions within it, have been built.

Furthermore, I would like to extend my gratitude to Prof. Bipan Tudu, Dept. of IEE, Jadavpur University. His innovative teaching methods and problem solving capability have consistently inspired me through challenging times during this work.

I would also thank Prof. Amit Mandal, Department of Chemistry, Behala College, Kolkata, India for helping me with technical details of material characterizations and selection. He has been a constant motivator and advisor throughout my research duration. I thank my colleagues; Mr. Dipan Bandyopadhyay, Dr. Sreya Nag, Dr. Trisita Nandi Chatterjee, Dr. Debangana Das and Dr. Soumen Mukherjee for making a cheerful ambience and support.

Finally, I would like to extend my love and gratitude to my parents, wife and my daughter whose unwavering support, vigilance, and constant encouragement enabled this dissertation to blossom from a mere bud into a beautiful flower. Their understanding of the stress I experienced during my research and their tactful handling of my mood fluctuations were immensely gratifying. Every time I was ready to quit, they would not let me, and for that, I will remain forever grateful. This thesis stands as a testament to their genuine love, compassion, and support throughout the course of this work. Lastly, I express my gratitude to the Almighty God for His abundant blessings.

Abstract

Tea is very popular and well consumed beverage due to its immense contributions to human health. It is a rich source of antioxidants, which protect us from various diseases when consumed in the recommended quality and quantity. The quality of tea is typically assessed based on the amount of phenolic compounds it contains, such as caffeine, gallic acid, epicatechin, catechin, epicatechin gallate, epigallocatechin gallate, and theaflavin. These compounds are primarily responsible for the variations in the taste and flavor of tea. Therefore, determination of the taste-affecting compounds and sensor data analysis are of utmost necessity. Electronic tongue (ET) emulates the human taster, it can proficiently discriminate the different taste modalities. Different sensors forming the arrays in an ET system emulate the human receptors which are responsible for sensing taste. Advanced pattern recognition techniques allow for establishing correlations for both qualitative and quantitative analysis of test samples. This research endeavor seeks to improve the efficiency of the ET system through the utilization of machine learning methodologies. The investigation focuses on three specific categories of study aimed at tackling issues related to pattern identification and the development of sensor.

First, a total of nine polymer graphite composite electrodes (PGE) were fabricated by employing different ratio of three monomers namely acrylamide, aniline, and pyrrole, in conjunction with graphite, for the purpose of evaluating the overall characteristics of black tea. Subsequently, optimization of the electrode array was carried out through the utilization of four distinct techniques for feature extraction Principal Component Analysis (PCA), Discrete Cosine Transform (DCT), Singular Value Decomposition (SVD), and Independent Component Analysis (ICA) along with the application of five classification algorithms: Support Vector Machine (SVM), k-Nearest Neighbor (KNN), ensemble, decision tree, and discriminant analysis, followed by a polling process.

Second, a machine learning model that has been developed to improve the performance of predicting green tea quality in electrochemical systems by employing sensor data obtained from molecularly imprinted polymers (MIPs) electrodes. The data from two MIP electrodes, specifically MIP-GAL and Q-IPG, are subjected to a transformation process using the Discrete Cosine Transform (DCT) technique. Three bio-inspired metaheuristic algorithms namely the genetic algorithm (GA), bat algorithm (BA), and whale optimization algorithm

(WOA) are used to optimize the features. Optimized features are then utilized in conjunction with partial least squares regression (PLSR) and principal component regression (PCR) to improve the predictions accuracy of green tea sample.

Third, a robust, reusable, cost-effective MIP modified capacitive sensor was developed for epicatechin (EC) detection in green tea. Two statistical regression approaches, namely PLSR and PCR, were employed for the estimation of this non-volatile compound in tea. Subsequently, the performance of the sensor is evaluated through the application of two commonly utilized clustering techniques: K-Means and Agglomerative Clustering.

All the proposed methods show excellent performance in pattern recognition of sensor datasets and sensor fabrication. This research will significantly impact the tea and pharmaceutical industries.

Table of Contents

List of Figure

List of Tables

List of Abbreviation

Ag/AgCl	Silver/ Silver Chloride
ANN	Artificial neural networks
ACO	Ant colony optimization
AI	Artificial intelligence
BA	Bat Algorithm
BPMLP	Back propagation multilayer perceptron
BPNN	Back-propagation neural networks
CV	Cyclic voltammetry
CNT	Carbon nanotubes
CCD	Charge-couple device
CA	Clustering analysis
CNN	Convolutional neural network
DPV	Differential pulse voltammetry
DT	Decision trees
DCT	Discrete cosine transform
DWT	Discrete wavelet transform
ET	Electronic tongue
EC	Epicatechin
EIS	Electrochemical impedance spectroscopy
ELM	Extreme learning machine
GAL	Gallic acid
GA	Genetic algorithm
ISE	Ion selective electrodes
ICA	Independent Component Analysis
KNN	K-nearest neighbors
LDA	Linear discriminant analysis
MIP	Molecularly imprinted polymer
MVDA	Multivariate data analysis
MLP	Multilayer Perceptron
ML	Machine learning
NLP	Natural language processing
NIP	Non-imprinted polymer

PSO	Particle swarm optimization
PLSR	Partial least squares regression
PCR	Principal component regression
PCA	Principal component analysis
Q-IPG	Quercetin imprinted electrode
QDA	Quadratic Discriminant Analysis
RF	Random forest
RMSE	Root mean square error
RMSEC	Root mean square error of calibration
RMSEP	Root mean square error of prediction
RMSEV	Root mean square error of validation
SWV	Square wave voltammetry
SAW	Surface acoustic wave
SOM	Self-organizing map
SVD	Singular value decomposition
SVM	Support vector machine
SEM	Scanning electron microscope
WOA	Whale Optimization Algorithm
WE	Working Electrode
XRD	X-ray diffraction

CHAPTER 1

INTRODUCTION AND SCOPE OF THESIS

This chapter presents the purpose and objectives of this study. It begins with an introduction to the thesis, followed by an overview of the taste-sensing mechanism and tea chemistry. In the tea chemistry section, the focus is on tea processing and a brief discussion of the chemical compounds responsible for its color, aroma, and taste. The chapter also covers conventional methods of tea quality assessment, such as human tea tasting and chemical analysis techniques. Additionally, the role of pattern recognition systems, which are integral part to this research, is explained, detailing the various multivariate systems used. Finally, a review of existing literature and the motivations driving this research are discussed.

LIST OF SECTION

Chapter 1

Introduction and Scope of thesis

1.1 Introduction

A sensory experience encompasses the reception of information through the five senses : sight, hearing, touch, taste, and smell. These experiences occur when sensory organs detect stimuli from the environment, sending signals to the brain for interpretation. For instance, witnessing a beautiful sunset, feeling the warmth of sunlight on skin, savoring a favorite food, hearing music, and inhaling the fragrance of a flower are all examples of different types of sensory experiences that shape our understanding and interactions with the world around us.Our senses play a crucial role in memory formation, both in recalling past experiences and creating new ones. For example, a particular scent can trigger memories even before our conscious mind identifies the smell. Utilizing a combination of visual, auditory, and other sensory stimuli in learning experiences takes advantage of the brain's natural connectivity.

The taste sensing mechanism in humans involve specialized receptors on the tongue and oral cavity that detect different chemical compounds in food, while the pattern recognition process in the brain integrates these sensory inputs to identify and distinguish between various tastes. Taste information is transmitted to specific regions of the brain responsible for processing the taste. Here, neural circuits analyze and interpret the patterns of neural activities corresponding to different taste attributes. The brain employs pattern recognition algorithms to identify and categorize the complex patterns of neural activity associated with different tastes.

Human perception and recognition system is very complex and subjective which may vary from person to person or even for a same person depends on the physical status. In order to mimic human perception, in a more objective, repetitive and instrumental way, researchers have attempted to develop electronic nose, tongue, eye etc. in conjunction with advanced sensing technology and application of artificial intelligence (AI). In this work, an electronic taste sensing system has been considered for quality assessment of green and black tea samples. Employing various data transformation, classification and prediction algorithms, quality analysis of tea samples can be achieved with high degree of accuracy. The highly beneficial, globally cherished beverage, tea has been selected as the test sample of the whole research work. As the appreciation for tea continues to grow, the demand for high-quality tea products becomes increasingly imperative within the tea industry. Quality assessment and commitment of tea quality in the tea industry are pivotal not only for consumer satisfaction but also for ensuring the economic viability of tea cultivation.

The primary objective of this thesis is to explore the tea quality evaluation employing multivariate data analysis techniques on the sensor response used for taste attributes measurement. Specifically, this research aims to investigate various feature transformation methods, classification and clustering techniques, and prediction algorithms to extract meaningful insights from the sensor response and accurately predict tea quality. It proposes that by leveraging the capabilities of electronic tongue (ET) technology in conjunction with advanced machine learning methodologies, the accuracy and efficiency of tea quality assessment can be significantly enhanced.

The introductory chapter offers an insight into the importance of tea quality evaluation, the limitations of traditional approaches, and the potential of ET technology coupled with machine learning techniques as a precious tool in the subsequent sections. Sensing methods and sensor types have been discussed in section1.4. Section 1.7 delves into an extensive exploration of relevant literature, followed by the motivation and objectives of the research in section 1.8.

Chapter 2 discusses various machine learning techniques applied to pattern recognition and data analysis. Section 2.2 explores different supervised and unsupervised methods used for analyzing sensor data. Sections 2.3 and 2.4 delve into various feature transformation and optimization techniques. Different regression model used to address the problem have been detailed in section 2.5. Chapter 3 explored a novel technique for assessing black tea quality employing electrode array optimization method. Nine polymer graphite composite electrodes were synthesized by varying the ratios of three polymers: polyaniline (PANI), polyacrylamide (PAM), and polypyrrole (PPY) with graphite. Electrodes preparation and sensor data analysis have been detailed in sections 3.2 and section 3.3. Chapter 4 presents the enhancement of green tea quality prediction method employing feature optimization technique and voltammetric MIP electrodes. Two voltammetric MIP electrodes have been synthesized for gallic acid and epicatechin detection of green tea samples. The fabrication processes are detailed in Section 4.2. Section 4.3 discusses the feature transformation and various feature optimization methods, while Section 4.4 presents the results and discussion of the work. Chapter 5 explores the development of a low-cost, reusable MIP-tethered capacitive sensor for detecting epicatechin in green tea samples. The performance of the synthesized sensor was evaluated using two widely used clustering algorithms: K-Means and Agglomerative clustering. Chapter 6 presents the conclusion and future directions of the research. Section 6.2 highlights the major findings of the work, while Sections 6.4 and 6.5 outline the future scope and conclusion of the research, respectively.

1.2 Taste sensing system

Taste serves as one of the fundamental senses for all creatures, providing essential chemical data regarding nourishment from the oral cavity to the central nervous system to modulate the process of consumption. Two types of taste sensing mechanism have been described here to address the system.

1.2.1 Human taste sensing system

The sensing organ responsible for taste perception in human is called “taste buds”.Taste buds are multicellular receptor organs, located in the epithelial lining of the mouth and throat. Studies have shown that taste cells are continuously replaced throughout adulthood [1, 2]. Every taste bud consists of a barrel-shaped structure comprising of numerous spindle-shaped taste receptor cells bundled within its epithelial sack [3]. The Fig.1.1 illustrates that taste buds consist of 150 to 300 densely arranged cylindrical cells derived from epithelial tissue [3, 4] .Variability in taste bud numbers among individuals, ranging from 500 to 20,000, explains differences in taste sensitivity. When food chemicals engage with the taste receptor cells, a series of biochemical reactions occur in the saliva, resulting in the generation of electrical potentials within the cells. The generated electrical signal due to transduction transmits to the brain via nerve cell for interpretation. Different pattern for the diverse taste of the food have been stored to the brain cell .Brain employs complex pattern recognition system for identification, discrimination and assessment the food. Input taste signal coming to the brain is always compared with the stored pattern which was learned by the previous experience. Human gustatory system exhibits remarkable sensitivity and accuracy in discerning the basic taste types and assessing the various food items [5].

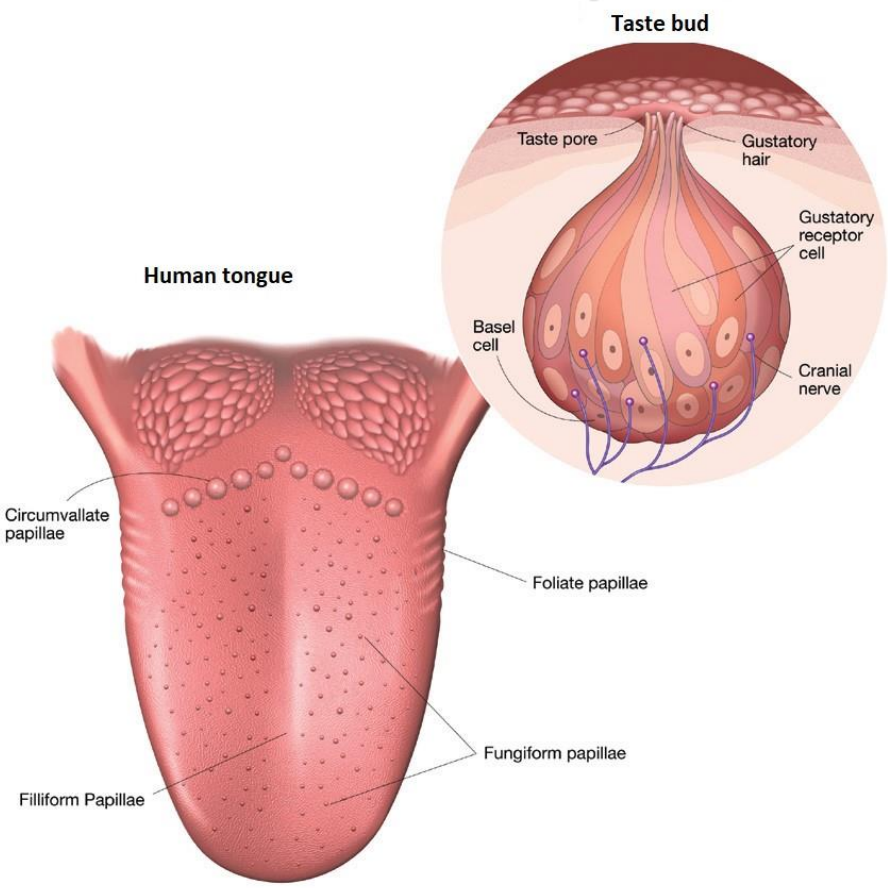


Fig.1.1. Anatomy of papillae and taste buds in the human tongue [3].

1.2.2 Artificial taste sensing system

The schematic representation of the human gustatory system, in conjunction with an electronic tongue (ET), is depicted in Fig. 1.2. Here ET acts as an analytical system that can mimic the human gustatory system. When test sample interacts with the submerged ET sensors, electrical signals are generated due to the different chemical reaction. These transduced electrical signals are then captured using data capturing process. Once response data is captured, it undergoes transformation and analysis to extract meaningful signatures of the samples. Visualization and clustering of captured data can also be possible by using various clustering and visualization algorithm. Advanced multivariate machine learning and data analysis techniques are employed to predict and classify the target analyte of the test sample.

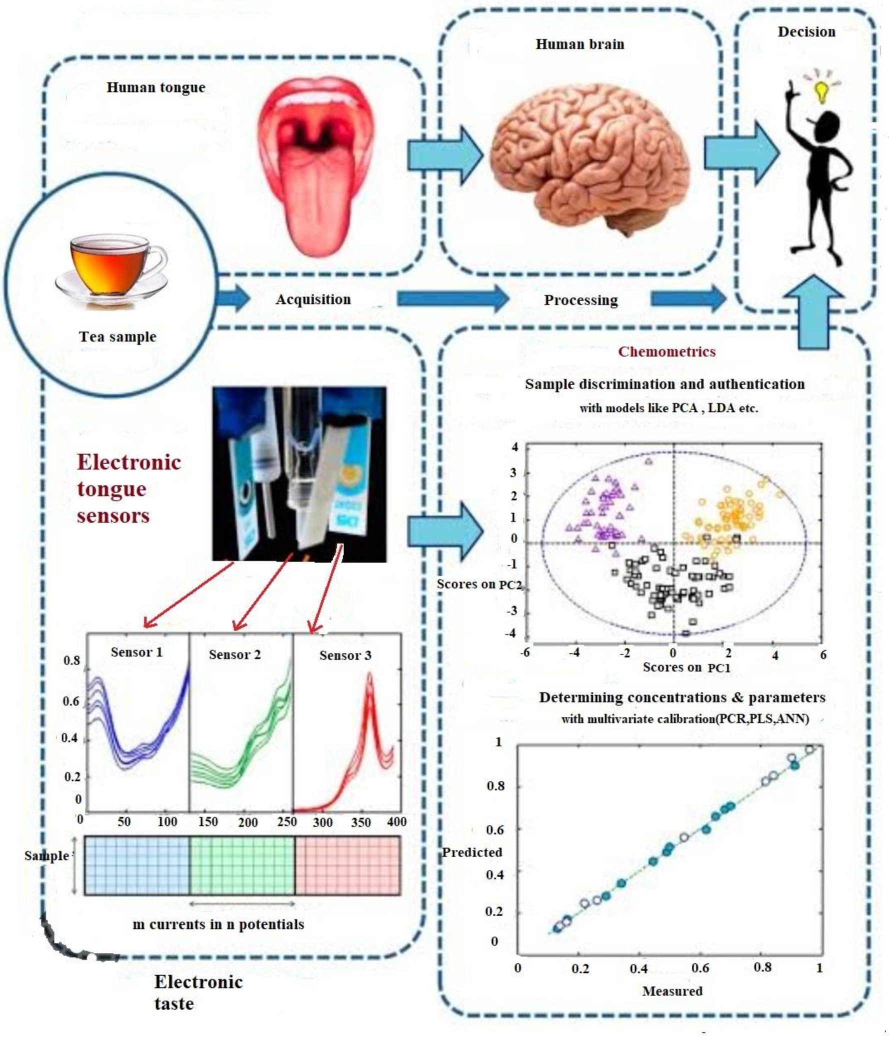


Fig.1.2.Block diagram of human taster taste modality and ET.

1.3 Tea and its chemistry

There exist numerous classifications of tea that are accessible in the market; the two primary classifications of tea enjoy widespread popularity on a global scale.

(i) Non fermented green tea

(ii) Fully fermented black tea

When tea leaves come into contact with oxygen, significant chemical transformations occurs. Black tea undergoes full oxidation during the fermentation process, which is a crucial step in developing its distinctive flavor, aroma, and color profile. Fully oxidized tea, such as black tea, typically boasts higher caffeine levels and exhibits a robust aroma. Conversely, green tea experience minimal oxidation or none at all during their production. It's worth noting that black tea tends to maintain its flavor for extended periods, whereas green tea may lose its flavor in comparison to black tea quickly [6]. The change of tea quality often occurs when tea plants face challenging conditions, such as intense summer sunlight or cool autumn temperatures. Around 90% of black tea is consumed in the Western world. Black tea is available in two primary forms: Orthodox tea, which is in leaf form, and CTC tea, which is granular.

1.3.1 Tea Processing

The details of the manufacturing process for both black and green tea are illustrated in Fig.1.3. The process begins with collection of fresh green leaves from the garden. The collected leaves are then sorted and cleaned. Withering is the common stage of processing for green and black tea manufacturing. Black tea processing encompasses the complete oxidation of tea leaves. Following the plucking of tea leaves, they undergo subsequent withering for about 8 to 24 hours to evaporate the water. Then the leaves are rolled to initiate the reaction of oxygen with enzymes and subsequently the oxidation process begins. After complete oxidation the tea leaves turn into deep black colour.

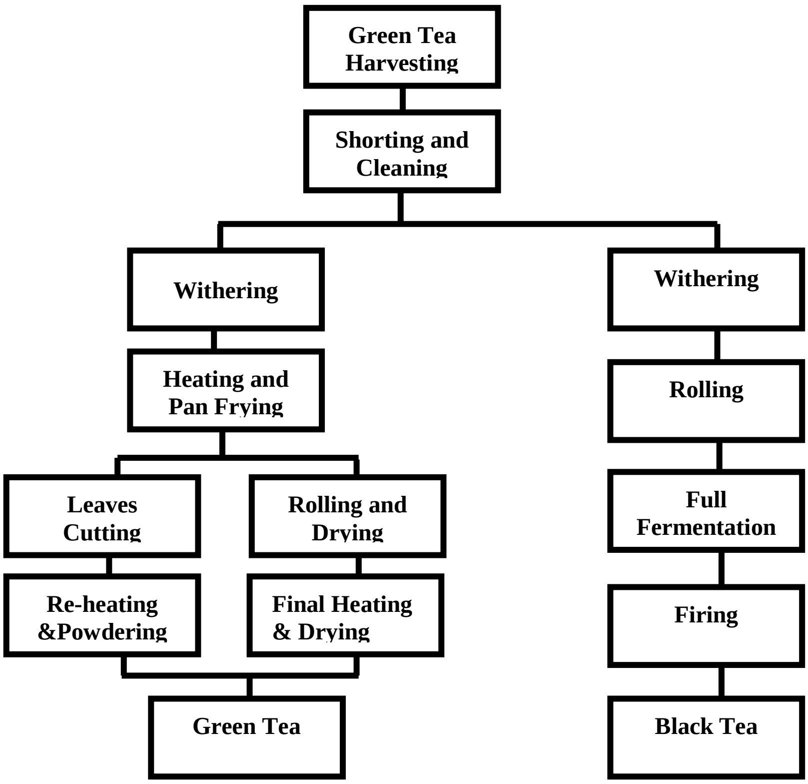


Fig.1.3. Tea Processing and manufacturing.

1.3.2 Tea chemistry

Tea is a complex compounds consists of innumerable chemical constituents that enhance its taste, aroma, and color characteristics. According to the findings presented by Kawakami, there are 239 distinct chemical compounds that contribute to the sensory attributes of taste, fragrance, and color in tea. The color and aroma characteristics of the tea are thoroughly detailed in Table 1.1 and Table 1.2, while Table 1.3 provides an in-depth explanation of the factors affecting the sensory attributes related to taste. The major taste-contributing chemicals in tea includes catechins, amino acids like L-theanine, caffeine, and various phenolic compounds such as epicatechin (EC), epicatechin gallate (ECG), epigallocatechin (EGC), and epigallocatechin gallate (EGCG). Among these, L-theanine is the most abundant, comprising over 60% of the free amino acids, by glutamic acid (9%), arginine (7%), serine (5%), and aspartic acid (4%) [7, 8]. Catechin also a vital constituent of tea, characterized as a colorless, odorless, and soluble compound with a relatively low molecular weight of 290.26 g/mol [9]. These elements constitute around 25% of the total dry matter of tea. During the fermentation process of tea, catechins are oxidized by polyphenol oxidase or plant ferment and known as oxidisable matter of tea [10]. Caffeine is another important constituent of tea, which is 2.5 to 4.5% [11] of the mass of tea leaves, is a colorless and harsh alkaloid that contributes to the stimulating effects of the brewed tea. The vigor and liveliness of the brewed tea are contingent upon the caffeine concentration present in the tea leaves. Polyphenol compounds present in tea reduces the risk of a variety of diseases and responsible for astringent taste [12]. Epigallocatechin-3-gallate and theaflavin-3 are the two essential polyphenols found in green tea and black tea. Another fundamental element of tea is Thearubigins, which represent the predominant category of phenolic pigments found in black tea, constituting nearly 60% of the solid matter in a standard black tea infusion [13]. The presence of phenolic compounds in tea contributes the ashy and astringent taste.

Table 1.1. *Chemical compounds of tea liable for color.*

Chemical compounds	Colour
Theaflavins	Yellowish brown
Thearubigins	Reddish brown
Flavonol glycosides	Light yellow
Carotene	Yellow
Pheophytin	Blackish
Pheophorbide	Brownish

Table 1.2. *Chemical compounds of tea liable for aroma.*

Chemical compounds	Flavor
Trans-2Hexenal, nHexenal, Cis-3-Hexenol, Grassy, b-Ionone	Fresh
Linalool, Linalool oxide	Sweet
Nerolidol, Benzaldehyde, Phenyl ethanol	Fruity
Geraniol , Phenylacetaldehyde	Floral

Table 1.3. *Chemical compounds of tea responsible for Taste.*

Chemical compounds	Taste
Amino acids	Brothy
Caffeine	Bitter
Catechin(CAT)	Strongly bitter
Polyphenol	Astringent
Theaflavins	Astringent
Thearubigin	Ashy and slight astringent
Epigallocatechin gallate(EGCG)	Astringent and less bitter

1.4 Conventional methods for tea quality estimation

There are two traditional methods for liquid tea quality analysis.

i) Human tea tasting method and ii) Chemical analysis method

i) <u>Human tea tasting method</u>

The commonly employed tea evaluation process in industry is performed by human panel tasters. The tasters develop their palate by professional training [14, 15]. The tea tasting process involves evaluation of tea quality by considering the aroma and visual presentation of dry tea leaves, as well as the taste, flavor, and color of the brewed tea. Furthermore, the aroma and color of the infused tea leaves are also taken into account during the assessment .The aroma and color of the infused tea leaves are considered, with phrases such as "Mixed," "Delicate," "Bright greenish," and "Dark" used to represent their appearance. The quality score of aroma of tea samples are assigned in terms like “Fruity”, “Delicate”, “Sour”, “Burnt”, and “Smoky”. Tea sample is infused in the boiling water for the period of five to six minutes. The liquor is then filtered out from the infused leaves and poured into the transparent cup to ensure the proper view of color. The quality scores are assigned on a scale of 1 to 10 for each attribute of the sample by the tasters. This method of quality assessment is highly susceptible to errors and lacks reproducibility. It is influenced by subjective factors such as mental state and environmental conditions, personal feeling of individual taster making it inherently subjective in nature.

ii) Chemical analysis method

Analytical instruments like gas chromatograph mass spectroscopy (GC-MS), gas chromatograph(GC) [16], UV-VIS spectrophotometer and high performance liquid chromatography (HPLC) are used to detect the bio-chemical compound present in varieties of tea .When GC combine with mass spectroscopy (MS) become a powerful analytical technique used to separate, identify, and quantify complex mixtures of volatile compounds . Though these sophisticated instruments yields accurate result, the methods are time consuming, expensive and complicated. These instruments are largely time consuming and need human expertise [17]. Since the methods are purely laboratory based, online monitoring of tea chemicals are not feasible with these high end instruments. They are hardly used in industries for routine quality assessment of tea.

1.4.1 Quality evaluation of tea with electronic tongue

The functional schematic drawing of the ET set-up is depicted in Fig. 1.4. An ET set-up is divided into two primary types: amperometric and potentiometric. Voltammetry, a subset of amperometry, involves applying varying voltage levels to the working electrode and recording the resulting current spectra.

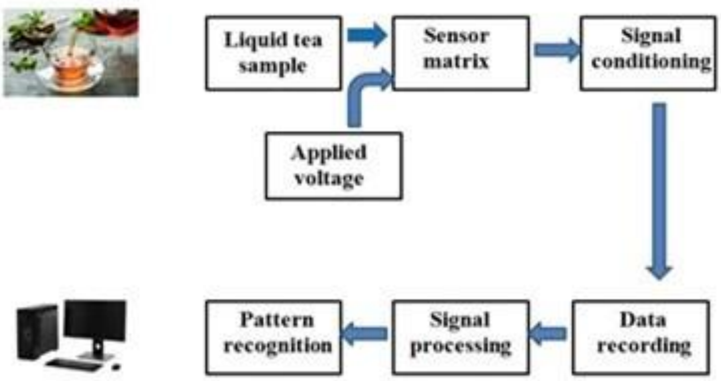


Fig.1.4. Schematic drawing of ET set-up.

In cyclic voltammetry (CV), a triangular voltage signal is applied to the working electrodes, and the current is analyzed relative to the applied voltage. In contrast, pulse voltammetry applies voltage pulses of varying amplitudes over time. The resulting current signal, measured against time, provides valuable insights into the chemodynamic properties of the test electrolyte. In this ET system, both broadly tuned non-selective sensors and target-specific selective sensors are used. Non-selective sensors detect the overall response pattern of the tea sample, while selective sensors focus on identifying specific biochemical components present in the tea. The collected sensor data is then subjected to signal conditioning and pre-processing, and pattern recognition techniques to extract the meaningful insights.

1.4.2 Broadly tuned non selective sensor system

Preliminary an ET had been developed by Jadavpur University comprising of multiple electrodes of noble metals and associated electronic system were used to detect the overall quality of test samples [17]. The ET system comprised an array consisting of five nonselective working electrodes, alongside a counter and a reference electrode. These five working electrodes are constructed from gold, iridium, palladium, rhodium, and platinum respectively. The setup also includes a platinum counter electrode and an Ag/AgCl reference electrode for three electrode voltametric measurement [18]. The response of the tongue sensor varies depending on the composition of the test sample, owing to the presence of diverse compounds and ions. ET configuration employed for discriminating black tea is delineated in Fig.1.5, whereas Fig.1.6 shows the Voltammogram response in a standard solution. The non-selective graphite electrode unveiled the overall quality of tea. However, it's important to note that despite its utility, the electronic tongue is limited in providing detailed chemical information about the test sample.

Fig.1.5. ET setup.

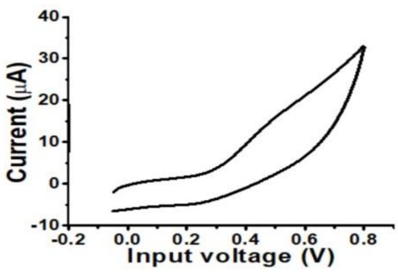


Fig.1.6.Voltammogram in standard solution.

1.4.3 Selective detection

Non-selective measurements yields approximate qualitative signature as an information due to inability to disclose the chemical details about the sample under examination. Therefore, it is crucial for the industry to develop a sensor capable of detecting specific molecules or analytes in the test solution. The advent of molecularly imprinted polymer (MIP) [19] technology addresses this need by making the system selective, low-cost, sensitive, and reproducible.

1.4.3.1 MIP electrode

Molecularly imprinting process forms an imprinted matrix through polymerization of the template and functional monomers, assembled around the template molecule [20]. The MIP synthesis method involves three main steps: association, polymerization, and elution [21]. In the association phase, the target molecule and monomers assemble through covalent or non-covalent bonds. Polymerization involves various polymerization techniques. During elution, template molecules trapped within the polymer network are removed using an appropriate elution agent. This results in the formation of cavities with comparable fits, essential for molecular recognition. Here Fig. 1.7 shows the schematic of MIP development.

MIPs offer numerous advantages, including flexibility, adaptability, and excellent mechanical stability etc. Moreover, due to the abundant availability of practical monomers, MIP electrode can be efficiently designed for various target analytes.

The MIP sensors are low cost, highly stable compared to the sensor based on natural biomolecules [22] . The responses obtained from the sensors are recorded which are then processed for the pattern analysis [23].

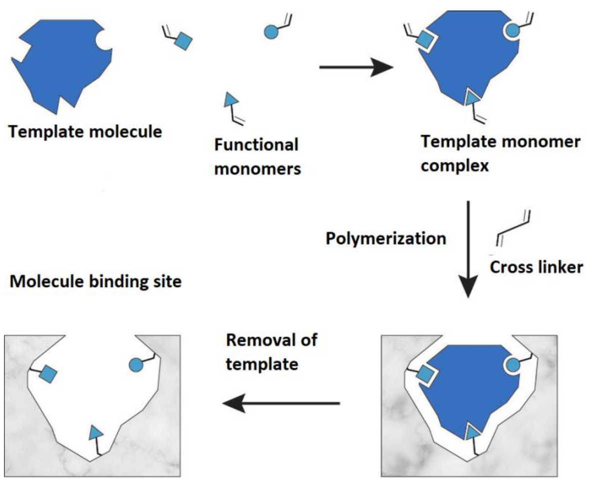


Fig.1.7. Shows schematic of MIP development.

1.5 Electrochemical measurement techniques

Electrochemistry is the branch of chemistry that explores the relationship between electrical and chemical effects [24]. It involves the studying of chemical changes induced by the electric current and the generation of electrical energy through chemical reactions. Electrochemical measurements are typically two-dimensional, involving current (i) versus voltage (V) or time (t). The potential relates to the thermodynamic and kinetic properties of the analyte and can be linked to its qualitative properties, while the current relates to mass transport processes, reaction rates, pH, mass diffusion, and adsorption, which can be connected to the analyte's quantitative properties. Electrochemical measurement can be chosen due to various reasons: to seek the thermodynamic data about a reaction, to generate an unstable intermediate like a radical ion to study its decay rate or spectroscopic properties or to analyze solutions for trace amounts of metal ions or organic species. Numerous electrochemical techniques have been evolved, and implementation of these methods needs knowledge of the basic theories of electrode reactions and the electrical characteristics of electrode-solution interfaces [25].

This process facilitates quantitative analysis of the sample and provides the crucial insights. In electrochemical sensors, the analytical data taken from the electrical signal generated when the target analyte interacts with the recognition layer. Electrochemical techniques involve oxidation at the anode and reduction at the cathode within an electrochemical cell. Depending on the nature of analyte's and requirements for sensitivity or selectivity, a variety of electrochemical instruments can be used. These instruments typically categorize into different types such as amperometric and potentiometric electrochemical sensors, based on their operational principles [26]. Amperometric sensors are particularly adept at detecting electro active species involved in chemical or biological identification. Whereas in potentiometric sensors, the interaction between the sensor and analyte, a local Nernstian equilibrium is established at the surface interface when no current flows, providing insight into the analyte's concentration. Amperometric sensors utilize a voltage applied between references and working electrodes to induce electrochemical oxidation or reduction, with the resulting current measured as a quantitative measure of analyte concentration, as described by the Cottrell equation.

$$i = \frac{NFAK_j^0\sqrt{D_j}}{\sqrt{\pi T}} \qquad (1.1)$$

Where i= Current in ampere;

N= Number of electrons;

A=Area of the planer electrode in $cm^{2;}$

K_j^0= Initial concentration of the reducible analyte in mol/cm3;

D_j= Diffusion coefficient for species in cm^2/s;

T= Time in seconds.

Conversely, conductometric sensors, often termed impedimetric sensors, gauge alterations in surface impedance to identify and measure analyte-specific recognition occurrences on the electrode.

Electrochemical techniques are widely employed due to their affordability in fabrication, stability, sensitivity, ease of retrieval, and minimal interference, as indicated in references [27]. Various sensors are utilized depending on the varied operational principles of ET, and the following section provides a concise overview of different electrochemical sensing technologies.

1.5.1 Voltammetric technique

Voltammetric measurement involves measuring current as a function of an applied potential when equilibrium is not reached. The three electrodes voltammetric ET system comprises of a working electrode, a reference electrode and a counter electrode. In this setup, the potential of the reference electrode remains constant while current flows between the working and counter electrodes when they are immersed in the solution. Three electrodes electrochemical cell is illustrated in Fig.1.8. Electrolytic reaction occurs in this system based on the polarity of the used potential .The redox active compounds are either oxidized or reduced which leads to increase or decrease the current. Different techniques like cyclic voltammetry (CV), differential pulse voltammetry (DPV) and square wave voltammetry (SWV) fall under this category. The working electrodes of this measurement system are basically ion selective electrodes (ISE) [28].The potential of ISE is a function of ion presence in a sample solution. According to the Nikolsky equation, the voltammetric response is influenced by the temperature, although the overall impact of temperature is minimal [29]. These type of sensors are targeted for selective compound and have extensive application amidst the area of liquid sample discrimination [9], milk [30, 31], wine, juice and beverages classification.

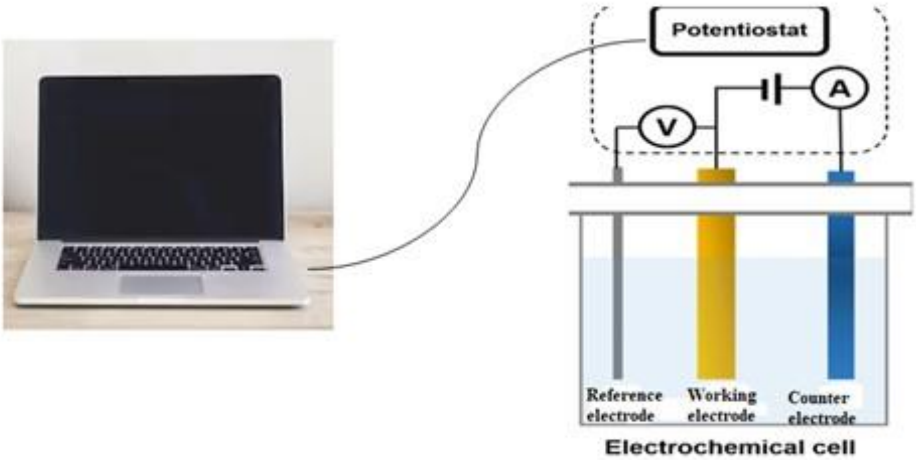


Fig.1.8. Three electrodes electrochemical cell.

1.5.2 Potentiometric technique

In potentiometric sensor, the potential is measured between two electrodes when no current flows between them. The recorded potential serves as the basis for determining the specific analytical parameter of interest, typically the concentration of the target component within the sample solution. There are varieties of potentiometric ET sensors used for measurement of taste. All the sensors work on the common technique that is, all measure the potential over charged membrane. The membranes can be made of different materials like graphene, graphene oxide, and carbon nanotubes (CNTs) depending on the character of the chemical substances. The Fig. 1.9 shows the schematic of potentiometric system. This method require relatively simple measuring system and have wide range of application area like health monitoring, food quality assurance and water and environment analysis etc. using the different sensing elements like different polymeric film, glass and metallic electrodes. The research work on taste was first published on 1920 [32], based on the ion sensitive lipid membrane and immobilized with polymer PVC. In this system eight different membranes are fitted on a multi-channel electrode. This sensor was used to measure the five basic tastes: sour, salt, bitter, sweet, umami, for quality discrimination of mineral water and beverages. Potentiometric sensors encompass various types, including pH, sodium, potassium selective electrodes and specially designed electrodes. An electrode array constitute of gold, platinum, iridium and rhodium discs cast in encore resign into a stainless steel tube can used as potentiometric electrode [33]. This approach has been utilized to detect the different type of beverages like tea, coffee, soft drinks etc. [34]. Potentiometric sensor arrays have also been used in various experiments with milk and urine samples [35].

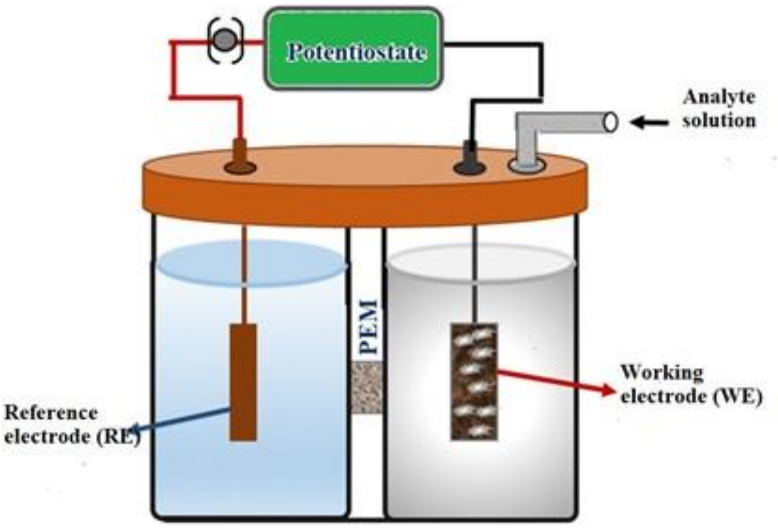


Fig.1.9. Schematic of Potentiometric system.

1.5.3 Amperometric technique

The principle of amperometric measurement technique is based on estimating the current between the working and counter electrodes.Fig.1.10 shows a schematic representation of the amperometric system. The redox reaction takes place on the working electrode and condition applied on the system such that current is directly proportional redox active elements in the test solution. In [36], the amperometric sensor was used to classify and characterize the wine sample of different origin. In another work [37], a disposable amperometric sensor was synthesized by silver nanoparticles (AgNPs) electrode via tea extract .The surface of the sensor is modified by low-cost pencil graphite electrode(PGE) to efficiently detect the nitrite in food. This system can be used in varieties of application such as monitoring the gas [38], determining tannin in tea [39] and catechin in green tea [40].

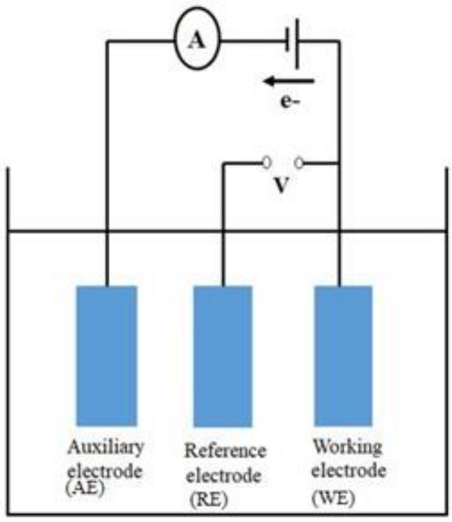


Fig.1.10. Amperometric system.

1.5.4 Impedentiometric technique

This is another type of technique based on the measurement of variation of electrical impedance of the electrode array. Here the contemporary evaluation of three alkaline ions namely potassium, sodium and ammonia, from the single impedance spectrum was investigated [41]. Electrochemical impedance spectroscopy (EIS) and artificial neural networks (ANN) were used to process the data. In [42] , five sensors composed array of three distinct verities based on carbon nanotubes were distributed in polymetric matrices and doped with polythiophenes. The measurements of diverse test samples of five elementary tastes were analyzed using 150Hz frequency. The application areas of impedentiometric sensors vary widely and encompass fields such as: environmental monitoring, biomedical application, industrial process, food and beverage industries [42-44].

In addition to the electrochemical measurement methods mentioned earlier, there exist alternative techniques such as optical sensing, acoustic wave sensing, and enzymatic or bio-sensing, which will be briefly outlined below.

1.5.5 Optical sensing technique

Optical sensing methods have been used to measure the taste of the test sample. Polymeric microspheres with modified surfaces have been utilized as optical sensors in optical ET. The work principle of this tongue is based on the light absorption. The array is formed by the charge-couple device (CCD) combined with chemical indicators in a resin bead [45]. In another study [46], derivatized resin microbeads were placed in micro machined wells fabricated in silicon structures with micro channels, where various samples were injected. These optical sensors have been utilized to investigate pH values in water, food analysis, clinical analysis, and industrial applications [47-53].

1.5.6 Acoustic wave technique

When piezoelectric materials are subjected to pressure, they generate an electric field, and this field is altered when an electric field is applied. When an alternating voltage is applied to the crystal, a stable oscillatory voltage across the crystal is generated to the opposite face of the crystal surface. Quartz crystals are extensively utilized for this application, with chemical sensitivity and selectivity derived from the adsorbent layer present on the crystal material. In quartz crystal microbalance (QCM) system, the analyte adsorption on the layer will leads to the change in frequency of the crystal. Different chemical can be measured depending on the affinity of the adsorption property of the crystal [54]. The surface acoustic wave (SAW) sensor can be used for sensing the gas [55-57] and aqueous phases and shear horizontal mode SAW can measure the liquid characteristics [58-60]. In [61], basic taste identification system was reported by using 36^0 rotated Y-cut, X-propagating $LiTaO_3$ device.

1.5.7 Bio-sensing technique

Biosensors composed ET is an analytical device that combines a biological component (like enzymes, antibodies, cells, tissues, etc.) with a physicochemical detector [62, 63]. These devices are designed to detect and quantify specific biological or biochemical substances by converting the biological response into a measurable signal [64]. The main part of the biosensors are: i) biological elements like tissue, microorganisms, organelles, cell receptors, enzymes, antibodies, nucleic acid etc. ii) transducer or detector elements which operates in a physicochemical way (optical, piezoelectric, electrochemical etc.) to convert the signal arising from influence of analyte with biological element into electrical signal and iii) biosensor reader system accompanied by signal processing unit for displaying the result [65]. Bio-enzymetic ET is an array of three biosensors containing glucose oxidase (GOD) for glucose and ascorbic acid measurement in various fruit juice [66]. The various application area of bio-sensors are: food technology application, medical diagnostics, genetic engineering including cancer, enzyme studies [67-71].

1.6 Pattern recognition methods

Pattern recognition techniques play a crucial role in identifying and declaring the attributes of food products after obtaining the response through the sensors [72-74]. These methods involve detecting, analyzing, and interpreting regularities or patterns within datasets. In the context of analyzing tea and beverages, pattern recognition serves as a robust tool for distinguishing between different types, detecting quality variations, and predicting properties [75-78]. These techniques entail several steps, commencing with data collection, data transformation and optimization, clustering, classification and prediction. A schematic block diagram of the pattern recognition method is illustrated in Fig. 1.11. After data collection, data transformation and data optimization are performed to remove redundant datasets that could affect the overall performance of the system. Classification and clustering techniques have been used to differentiate between various tea samples. Sample prediction and accuracy assessment are achieved through the application of different regression models. The prediction accuracy of the test sample is determined by comparing the predicted values with the actual data values.

Fig.1.11. Basic flow of pattern recognition method for tea analysis.

The ET system utilizes various pattern recognition algorithms and multivariate data analysis techniques, functioning as a model of the human gustatory system. Multivariate data analysis methods are employed for both quantitative and qualitative measurements of food and beverage in conjunction with ET [79, 80]. Broadly, pattern recognition techniques are classified into two categories: model type and learning method type. Model types can be further divided into parametric and non-parametric approaches, while learning methods are categorized as supervised or unsupervised. Parametric techniques rely on statistical methods, assuming that sensor data distribution can be analyzed using probability density functions (PDF). Linear discriminant analysis (LDA) is one such parametric method, used for dimensionality reduction and feature extraction. In contrast, non-parametric methods refrain from explicit assumptions about the probability distributions of the data. These methods learn directly from the dataset and often necessitate sophisticated non-linear tools such as artificial neural networks (ANN) [81, 82] , K-nearest neighbors (KNN) [83-85] , decision trees (DT) [17, 86, 87] , among others.

A supervised pattern recognition system is essentially a type of machine learning system where the model or algorithm is trained on labeled data. In this learning approach, the algorithm learns to recognize patterns and predict the input data corresponding output labels based on the examples provided during the training session [88-90]. Conversely, an unsupervised machine learning system differs from supervised systems in that it does not rely on labeled data. It works to find patterns, structures, and relationships within data without explicit guidance or prior knowledge [91, 92] . Researchers have extensively explored the responses obtained from ET, and the succeeding section offers detailed overviews of reported research works on ET signal processing in various applications.

In this study, pattern recognition methodologies were applied to investigate the overall qualitative discrimination of tea samples, optimize data, measure the key chemical constituents of tea, and optimize the sensor in an array.

1.7 Literature survey

The sensors in the electronic tongue array mimic human receptors responsible for detecting taste and quality in tea or food items. Data acquired from these sensors are complex in nature and characterized by presence of noise. Therefore, it is crucial to extract factual information from the test sample. Depending on the composition of ions and compounds present in the test solution, the combined response of the ET sensor array varies from one solution to another.

Commonly used unsupervised pattern recognition techniques, such as principal component analysis (PCA), clustering analysis (CA), and self-organizing maps (SOM), are valuable tools for visualizing data distribution, often used for data exploration and outlier detection.

For the differentiation of taste substances or biochemical components using electrochemical taste sensor, pattern recognition systems can be categorized into feature transformation and optimization methods [17, 93-95], clustering methods [96-99], classification [100-103] and regression methods [81, 100, 104, 105].

Employing ET combined with machine learning techniques finds practical utility in the beverage industry. Its utilization spans across the qualitative and compositional analysis of a broad spectrum of beverages, encompassing both non-alcoholic and alcoholic variants like milk, tea, fruit juice, honey, water, wine, and beer, among others. Furthermore, it can monitor the fermentation process of food products such as cheese and alcoholic beverages. In tandem with pattern recognition techniques, ET emerges as a crucial asset in the clinical and pharmaceutical domains, facilitating precise chemical investigations. Its versatile and robust data analysis capabilities play a pivotal role in characterizing the analytes under examination. Table 1.4 provides a concise overview of the literature review pertaining to various pattern recognition methods for assessing tea quality using electronic tongue technology.

Table 1.4. *Summary of literature review*

Sl. No.	Measurement Type	Objective	Feature transformation technique	Machine learning technique	Refs
1	Voltammetric ET	Classification of green tea samples	Discrete cosine transform (DCT), discrete wavelet transform (DWT) and singular value decomposition (SVD),	convolutional neural network-based auto features extraction strategy (CNN-AFE) in an ET	[106]
2	Voltammetric ET	Quality estimation of black tea	DWT	One vs one support vector machine (OVO-SVM) &Vector valued regularized kernel function approximation (VVRKFA)	[107]
3	Voltammetric ET	Black tea classification and correlation with tea taster	DWT and PCA	LDA	[18]
4	Potentiometry ET	Green and black tea identification	PCA	PCR and PLS	[108]
5	∞ Astree II ET	Green tea grade discrimination	Canonical discriminant analysis(CDA), PCA	Back-propagation neural networks (BPNN)	[72]
6	Voltammetric ET	Classification of black tea samples	DWT	BPMLP and RBF	[74]
7	Voltammetric ET	Concentration of theaflavins	DWT	Partial least squares(PLS) and	[77]

Sl. No.	Measurement Type	Objective	Feature transformation technique	Machine learning technique	Refs
		(TF) and thearubigins (TR) in black		LDA	
8	Potentiometric ET	Four grades of green tea identification	PCA	KNN and ANN	[90]
9	Voltammetry ET	Black tea bio-chemical content analysis	DCT, Singular Value Decomposition (SVD), Stockwell Transform(ST)	ANN,vector-valued regularized kernel function approximation (VVRKFA), SVR	[93]
10	Pulse voltammetry	Tea sample classification	PCA	SVM	[102]
11	Micro-NIR spectroscopy and chemometric for ET	Characterization of black tea taste quality	Ant colony optimization(ACO), Particle swarm, optimization(PSO), Grey wolf optimization(GWO), and Non-dominated sorting genetic algorithm (NSGA−II), were used both sensors (ET and NIR spectra)	SVM, Extreme learning machine (ELM), KNN	[109]
12	Potentiometry ET	Different grades of black tea identification	PCA	Discriminant analysis(DA), Back-Propagation Neural Network (BPNN)and SVM	[110]
13	Impedance spectroscopy ET	Black tea identification		PCA	[111]
14	Cyclic voltammetry technique	Antioxidant activities evaluation of green tea		ANN	[112]
15	Voltammetric ET	Feature extraction of black tea sample	DWT and NCA	(ELM)-based regression models	[113]
16	Fourier Transform–	Polyphenol and EGCG of tea	Savitzky–Golay smoothing (SG),	Partial least squares regression (PLSR)	[114]

Sl. No.	Measurement Type	Objective	Feature transformation technique	Machine learning technique	Refs
	near-infrared spectroscopy	samples prediction	standard normal variate (SNV), vector normalization (VN), multiplicative scatter correction (MSC) and first derivative (FD)	and least squares support vector regression (LS-SVR)	
17	Voltammetric ET	Discrimination of black tea and green tea samples	PCA	Multivariate data analysis (MVDA)	[115]
18	Cyclic voltammetry ET	Classification different black tea samples	PCA	LDA and BPMLP	[116]
19	Potentiometric all-solid-state ET	Evaluation of black and green tea		PCA	[117]
20	Potentiometry ET	Black tea quality analysis		PLS	[118]
21	Potentiometry ET	Fake green tea identification	PCA,Hierarchical Cluster Analysis (HCA)	DA, ANN	[119]
22	Voltammetric ET	Discriminate the age of Puerh tea	One-dimensional convolutional Neural network (1-D CNN)	Combination of deep learning and transfer learning	[120]
23	Voltammetry ET	Total TF content of black tea		Si-CARS-PLS	[121]
24	Voltammetric ET (LAPV)	Fermentation process investigation of black tea	PCA	PLR	[122]
25	Impedance tongue	Performance analysis of ET by discrimination method	PCA	Genetic algorithm (GA)& particle swarm optimization (PSO)	[123]
26	Voltammetric e-tongue	Different classes of infused tea	PCA	ARMAX modeling	[124]

Sl. No.	Measurement Type	Objective	Feature transformation technique	Machine learning technique	Refs
		evaluation			
27	Electrochemical impedance spectroscopy (EIS)	Flavored green tea classification	PCA	SVM	[125]
28	Impedance-Tongue	Black tea classification	PCA	Social Impact Theory based optimizer (SITO),SVM,GA,PSO	[126]
29	E-Senses systems (Electronic Nose, ET, and Electronic Eyes)	Different type of tea classification	PCA,LDA	SVM, KNN decision trees, naive Bayes, and random forests	[127]
30	Differential Pulse Voltammetry (DPV)	Quality evaluation of green tea samples	Box-Cox Transform, Correlation coefficients	Lasso regression, Ridge regression	[128]
31	Colorimetric artificial tongue	Discrimination of green tea samples	PCA,HCA		[129]
32	Linear sweep voltammetry (LSV)	Different type of tea quality analysis	PCA		[130]
33	Optimal chemometric ET	Quality estimation of black tea	PCA	ACO,SVM,ELM	[131]
34	LCR meter based detection system	Fermentation process of black tea	PCA	Hierarchical clustering analysis (HCA)	[132]
35	Voltammetry ET	Black tea quality analysis		Fuzzy based Response of Signal with Time (FRST)	[133]

1.8 Motivation and Objectives

The literature review reveals that researchers have employed diverse feature transformation techniques to extract meaningful insights from the collected data, including dimension reduction and noise elimination. Additionally, various supervised and unsupervised machine learning techniques have been explored for quality evaluation and tea classification, leveraging biochemical component measurement via sensor integration with machine learning. However, the scope of areas identified in the survey include optimizing electrode arrays using computational methods, developing a cognitive computing framework to enhance prediction accuracy, and creating low-cost, durable, and sensitive sensors for tea biochemical detection.

This study aims to assess tea quality and quantity using ET sensors and a pattern recognition system. The thesis objectives are outlined as follows:

- To investigate more meaningful and appropriate features from the sensor response employing features transformation and compression techniques.
- To minimize the data size in order to decrease computational time and implementing different feature optimization methods to enhance the model efficiency.
- To predict the biochemical content present in black and green tea.
- To optimize the number of sensors in an array of an ET.
- To develop and characterize a low-cost, sensitive and novel sensors for detecting specific molecule present in tea samples.

1.9 Thesis structure

With the aforementioned objectives, the entire research work has been presented into six chapters. An outline of the chapter-wise organization of the thesis is provided below.

Within ***Chapter 1*** of this thesis, the objectives of the study are detailed, focusing on the importance of tea and its chemical composition. It outlines the manufacturing processes for different tea grades, explores tea quality assessment, and discusses ET along with its sensor types. A concise overview of ET and its various sensing technologies has been presented. Literature review on the diverse applications of ET technology and machine learning methodologies for tea quality evaluation has been explored. Objectives and thesis structure presented at the end of the chapter.

Chapter 2 of this thesis delves into various pattern recognition and feature transformation methods. Various feature transformation methods, including Principal Component Analysis (PCA), Discrete Cosine Transform (DCT), Singular Value Decomposition (SVD), and Independent Component Analysis (ICA), have been explored concisely. The chapter also provides a brief overview of unsupervised machine learning techniques such as K-means clustering and agglomerative clustering, as well as supervised learning methods like Support Vector Machines (SVM), k-Nearest Neighbors (KNN), Ensemble Bagging Trees, Decision Trees, and Artificial Neural Networks (ANN). Additionally, it discusses feature optimization methods including Genetic Algorithms (GA), Bat Algorithm (BA), and Whale Optimization Algorithm (WOA), along with regression models such as Partial Least Squares Regression (PLSR) and Principal Component Regression (PCR).

Chapter 3 of this thesis presents voltammetric electrode array optimization technique using computational approach to distinguish between different varieties of black tea. Polymer graphite composite electrodes (PGE) were synthesized using varying ratios of three monomers: acrylamide, aniline, and pyrrole, combined with graphite. Electrode array optimization was performed using four feature extraction techniques: Principal Component Analysis (PCA), Discrete Cosine Transform (DCT), Singular Value Decomposition (SVD), and Independent Component Analysis (ICA) and five classification algorithms: Support Vector Machine (SVM), k-Nearest Neighbor (KNN), Ensemble, Decision Tree, and Discriminant Analysis, followed by a polling process.

Chapter 4 presents a machine learning model designed to enhance the prediction accuracy of green tea quality in electrochemical systems using sensor data from molecularly imprinted polymers (MIPs) electrodes. Data from two MIP electrodes, MIP-GAL and Q-IPG, are transformed using the Discrete Cosine Transform (DCT) technique. Data optimization was performed using bio-inspired metaheuristic algorithms: Genetic Algorithm (GA), Bat Algorithm (BA), and Whale Optimization Algorithm (WOA). These optimized feature sets were then used to develop prediction models using Partial Least Squares Regression (PLSR) and Principal Component Regression (PCR) to improve prediction accuracy.

Chapter 5 delves into the development of a stable, selective, reusable, and cost-effective capacitive sensor utilizing molecular imprinted polymer (MIP) technology for detecting epicatechin (EC) in green tea. The sensor performance is then assessed using two widely-used clustering algorithms: K-Means and Agglomerative Clustering.

Chapter 6 presents the conclusion and future scope of the research work.

References

[1] L. M. Beidler and R. L. Smallman, "Renewal of cells within taste buds," *The Journal of cell biology,* vol. 27, pp. 263-272, 1965.

[2] A. Farbman, "Renewal of taste bud cells in rat circumvallate papillae," *Cell Proliferation,* vol. 13, pp. 349-357, 1980.

[3] S. A. Gravina, G. L. Yep, and M. Khan, "Human biology of taste," *Annals of Saudi medicine,* vol. 33, pp. 217-222, 2013.

[4] J. Brouwer and A. Wiersma, "Location of taste buds in intact taste papillae by a selective staining method," *Histochemistry,* vol. 58, pp. 145-151, 1978.

[5] S. D. Roper, "Signal transduction and information processing in mammalian taste buds," *Pflügers Archiv-European Journal of Physiology,* vol. 454, pp. 759-776, 2007.

[6] M. A. Bokuchava, N. I. Skobeleva, and G. W. Sanderson, "The biochemistry and technology of tea manufacture," *Critical Reviews in Food Science & Nutrition,* vol. 12, pp. 303-370, 1980.

[7] R. Horanni and U. H. Engelhardt, "Determination of amino acids in white, green, black, oolong, pu-erh teas and tea products," *Journal of Food Composition and Analysis,* vol. 31, pp. 94-100, 2013.

[8] Z. Yu and Z. Yang, "Understanding different regulatory mechanisms of proteinaceous and non-proteinaceous amino acid formation in tea (Camellia sinensis) provides new insights into the safe and effective alteration of tea flavor and function," *Critical Reviews in Food Science and Nutrition,* vol. 60, pp. 844-858, 2020.

[9] A. Arrieta, C. Apetrei, M. Rodríguez-Méndez, and J. De Saja, "Voltammetric sensor array based on conducting polymer-modified electrodes for the discrimination of liquids," *Electrochimica Acta,* vol. 49, pp. 4543-4551, 2004.

[10] K. C. Willson and M. N. Clifford, *Tea: cultivation to consumption*: Springer Science & Business Media, 2012.

[11] T. Atomssa and A. Gholap, "Characterization of caffeine and determination of caffeine in tea leaves using uv-visible spectrometer," *African Journal of Pure and Applied Chemistry,* vol. 5, pp. 1-8, 2011.

[12] N. Khan and H. Mukhtar, "Tea polyphenols for health promotion," *Life sciences,* vol. 81, pp. 519-533, 2007.

[13] N. Kuhnert, "Unraveling the structure of the black tea thearubigins," *Archives of Biochemistry and Biophysics,* vol. 501, pp. 37-51, 2010.

[14] S. Besky, *Tasting qualities: The past and future of tea* vol. 5: University of California Press, 2020.

[15] L. Zhang, Q.-Q. Cao, D. Granato, Y.-Q. Xu, and C.-T. Ho, "Association between chemistry and taste of tea: A review," *Trends in Food Science & Technology,* vol. 101, pp. 139-149, 2020.

[16] N. Togari, A. Kobayashi, and T. Aishima, "Relating sensory properties of tea aroma to gas chromatographic data by chemometric calibration methods," *Food Research International,* vol. 28, pp. 485-493, 1995.

[17] S. Acharya, D. Das, T. N. Chatterjee, S. Mukherjee, R. B. Roy, B. Tudu, *et al.*, "Voltammetric electrode array optimization for black tea discrimination using computational intelligence approach," *IEEE Sensors Journal*, vol. 21, pp. 20589-20595, 2021.

[18] M. Palit, B. Tudu, P. K. Dutta, A. Dutta, A. Jana, J. K. Roy, *et al.*, "Classification of black tea taste and correlation with tea taster's mark using voltammetric electronic tongue," *IEEE Transactions on Instrumentation and Measurement,* vol. 59, pp. 2230-2239, 2009.

[19] T. N. Chatterjee, D. Das, R. B. Roy, B. Tudu, S. Sabhapondit, P. Tamuly, *et al.*, "Molecular imprinted polymer based electrode for sensing catechin (+ C) in green tea," *IEEE Sensors Journal,* vol. 18, pp. 2236-2244, 2018.

[20] F. H. Dickey, "The preparation of specific adsorbents," *Proceedings of the national academy of sciences,* vol. 35, pp. 227-229, 1949.

[21] A. G. Mayes and M. J. Whitcombe, "Synthetic strategies for the generation of molecularly imprinted organic polymers," *Advanced drug delivery reviews,* vol. 57, pp. 1742-1778, 2005.

[22] S. Li, Y. Ge, S. A. Piletsky, and J. Lunec, "Molecularly imprinted sensors: overview and applications," 2012.

[23] S. Acharya, D. Das, S. Nag, S. Mukherjee, A. K. Hazarika, S. Sabhapondit, *et al.*, "Optimization techniques for a voltammetric signal to predict green tea quality parameters using MIP electrode," *IEEE Sensors Journal*, 2023.

[24] R. A. Strobel and G. Czarnopys, "Analysis and detection of explosives and explosives residues," in *16th International Forensic Science Symposium Interpol–Lyon 5 th-8 th October 2010 Review Papers,* 2010, p. 453.

[25] X. Shan, U. Patel, S. Wang, R. Iglesias, and N. Tao, "Imaging local electrochemical current via surface plasmon resonance," *science,* vol. 327, pp. 1363-1366, 2010.

[26] M. D. Meti, J. C. Abbar, J. Lin, Q. Han, Y. Zheng, Y. Wang, *et al.*, "Nanostructured Au-graphene modified electrode for electrosensing of chlorzoxazone and its biomedical applications," *Materials Chemistry and Physics,* vol. 266, p. 124538, 2021.

[27] B. J. Privett, J. H. Shin, and M. H. Schoenfisch, "Electrochemical sensors," *Analytical chemistry,* vol. 80, pp. 4499-4517, 2008.

[28] D. Henn and K. Cammann, "Voltammetric Ion-Selective Electrodes (VISE)," *Electroanalysis: An International Journal Devoted to Fundamental and Practical Aspects of Electroanalysis,* vol. 12, pp. 1263-1271, 2000.

[29] J. Valdes and B. Miller, "Thermal modulation voltammetry response of reversible redox systems: theory and experiment," *The Journal of Physical Chemistry,* vol. 93, pp. 7275-7280, 1989.

[30] P. Ciosek and W. Wróblewski, "Sensor arrays for liquid sensing–electronic tongue systems," *Analyst,* vol. 132, pp. 963-978, 2007.

[31] F. Winquist, P. Wide, and I. Lundström, "An electronic tongue based on voltammetry," *Analytica chimica acta,* vol. 357, pp. 21-31, 1997.

[32] K. Hayashi, M. Yamanaka, K. Toko, and K. Yamafuji, "Multichannel taste sensor using lipid membranes," *Sensors and Actuators B: Chemical,* vol. 2, pp. 205-213, 1990.

[33] F. Winquist, R. Bjorklund, C. Krantz-Rülcker, I. Lundström, K. Östergren, and T. Skoglund, "An electronic tongue in the dairy industry," *Sensors and Actuators B: Chemical,* vol. 111, pp. 299-304, 2005.

[34] K. Toko, "Taste sensor with global selectivity," *Materials Science and Engineering: C,* vol. 4, pp. 69-82, 1996.

[35] C. Di Natale, R. Paolesse, A. Macagnano, A. Mantini, A. D'Amico, A. Legin, *et al.*, "Electronic nose and electronic tongue integration for improved classification of clinical and food samples," *Sensors and Actuators B: Chemical,* vol. 64, pp. 15-21, 2000.

[36] S. Buratti, S. Benedetti, M. Scampicchio, and E. Pangerod, "Characterization and classification of Italian Barbera wines by using an electronic nose and an amperometric electronic tongue," *Analytica Chimica Acta,* vol. 525, pp. 133-139, 2004.

[37] C. Liu, D. Chen, C. Zhu, X. Liu, Y. Wang, Y. Lu, *et al.*, "Fabrication of a Disposable Amperometric Sensor for the Determination of Nitrite in Food," *Micromachines*, vol. 14, p. 687, 2023.

[38] Z. Cao, W. J. Buttner, and J. R. Stetter, "The properties and applications of amperometric gas sensors," *Electroanalysis,* vol. 4, pp. 253-266, 1992.

[39] Y.-T. Hung, P.-C. Chen, R. L. Chen, and T.-J. Cheng, "Determining the levels of tannin in tea by amperometry of ferricyanide pre-reaction with a sample in a flow-injection system," *Sensors and Actuators B: Chemical,* vol. 130, pp. 135-140, 2008.

[40] S. C. Fernandes, R. E.-H. M. Osório, A. d. Anjos, A. Neves, G. A. Micke, and I. C. Vieira, "Determination of catechin in green tea using a catechol oxidase biomimetic sensor," *Journal of the Brazilian Chemical Society,* vol. 19, pp. 1215-1223, 2008.

[41] M. Cortina-Puig, X. Muñoz-Berbel, M. A. Alonso-Lomillo, F. J. Muñoz-Pascual, and M. Del Valle, "EIS multianalyte sensing with an automated SIA system—An electronic tongue employing the impedimetric signal," *Talanta,* vol. 72, pp. 774-779, 2007.

[42] G. Pioggia, F. Di Francesco, A. Marchetti, M. Ferro, and A. Ahluwalia, "A composite sensor array impedentiometric electronic tongue: Part I. Characterization," *Biosensors and Bioelectronics,* vol. 22, pp. 2618-2623, 2007.

[43] T. Nakamoto, "Olfactory display and odor recorder," *Essentials of machine olfaction and taste,* pp. 247-314, 2016.

[44] M. Monique, G. Marco, R. A. Ciro, Z. Renata, C. Gabriella, R. Bruno*, et al.*, "Development of a sensor for the detection of Escherichia coli in brackish waters," *Journal of Coastal Life Medicine,* vol. 4, pp. 200-202, 2016.

[45] E. J. Cho and F. V. Bright, "Optical sensor array and integrated light source," *Analytical chemistry,* vol. 73, pp. 3289-3293, 2001.

[46] Y.-S. Sohn, A. Goodey, E. V. Anslyn, J. T. McDevitt, J. B. Shear, and D. P. Neikirk, "A microbead array chemical sensor using capillary-based sample introduction: toward the development of an “electronic tongue”," *Biosensors and Bioelectronics,* vol. 21, pp. 303-312, 2005.

[47] T. Werner and O. S. Wolfbeis, "Optical sensor for the pH 10–13 range using a new support material," *Fresenius' journal of analytical chemistry,* vol. 346, pp. 564-568, 1993.

[48] K. L. Martin, K. Girma, K. Freeman, R. Teal, B. Tubaṅa, D. Arnall*, et al.*, "Expression of variability in corn as influenced by growth stage using optical sensor measurements," *Agronomy Journal,* vol. 99, pp. 384-389, 2007.

[49] J. Pan, Z. Zhang, C. Jiang, L. Zhang, and L. Tong, "A multifunctional skin-like wearable optical sensor based on an optical micro-/nanofibre," *Nanoscale,* vol. 12, pp. 17538-17544, 2020.

[50] R. Sharma, M. L. Jat, K. L. Martin, P. Chandna, O. P. Choudhary, R. K. Gupta*, et al.*, "Assessment of the nitrogen management strategy using an optical sensor for irrigated wheat," *Agronomy for Sustainable Development,* vol. 31, pp. 589-603, 2011.

[51] J. Solie, W. Raun, R. Whitney, M. Stone, and J. Ringer, "Optical sensor based field element size and sensing strategy for nitrogen application," *Transactions of the ASAE,* vol. 39, pp. 1983-1992, 1996.

[52] M. Li, S. K. Cushing, and N. Wu, "Plasmon-enhanced optical sensors: a review," *Analyst,* vol. 140, pp. 386-406, 2015.

[53] A. Safavi and M. Bagheri, "A novel optical sensor for uranium determination," *Analytica chimica acta,* vol. 530, pp. 55-60, 2005.

[54] J. Devkota, P. R. Ohodnicki, and D. W. Greve, "SAW sensors for chemical vapors and gases," *Sensors,* vol. 17, p. 801, 2017.

[55] Y. Lee, H. Kim, Y. Roh, H. Cho, and S. Baik, "Development of a SAW gas sensor for monitoring SO2 gas," *Sensors and Actuators A: Physical,* vol. 64, pp. 173-178, 1998.

[56] C. Lim, W. Wang, S. Yang, and K. Lee, "Development of SAW-based multi-gas sensor for simultaneous detection of CO2 and NO2," *Sensors and Actuators B: Chemical,* vol. 154, pp. 9-16, 2011.

[57] Z. Yunusa, M. N. Hamidon, A. Ismail, M. M. Isa, M. H. Yaacob, S. Rahmanian*, et al.*, "Development of a hydrogen gas sensor using a double saw resonator system at room temperature," *Sensors,* vol. 15, pp. 4749-4765, 2015.

[58] T. Nomura, A. Saitoh, and Y. Horikoshi, "Measurement of acoustic properties of liquid using liquid flow SH-SAW sensor system," *Sensors and Actuators B: Chemical,* vol. 76, pp. 69-73, 2001.

[59] Z. Li, Y. Jones, J. Hossenlopp, R. Cernosek, and F. Josse, "Analysis of liquid-phase chemical detection using guided shear horizontal-surface acoustic wave sensors," *Analytical Chemistry,* vol. 77, pp. 4595-4603, 2005.

[60] J. Kondoh and S. Shiokawa, "A liquid sensor based on a shear horizontal SAW device," *Electronics and Communications in Japan (Part II: Electronics),* vol. 76, pp. 69-82, 1993.

[61] G. Sehra, M. Cole, and J. W. Gardner, "Miniature taste sensing system based on dual SH-SAW sensor device: an electronic tongue," *Sensors and Actuators B: Chemical,* vol. 103, pp. 233-239, 2004.

[62] R. Arora and R. Saini, "Biosensors: way of diagnosis," *International Journal of Pharmaceutical Sciences and Research*, vol. 4, pp. 2517-2527, 2013.

[63] Q. Liu, C. Wu, H. Cai, N. Hu, J. Zhou, and P. Wang, "Cell-based biosensors and their application in biomedicine," *Chemical reviews*, vol. 114, pp. 6423-6461, 2014.

[64] H. Kaur, A. Bhosale, and S. Shrivastav, "Biosensors: classification, fundamental characterization and new trends: a review," *Int J Health Sci Res*, vol. 8, pp. 315-333, 2018.

[65] R. Karunakaran and M. Keskin, "Biosensors: components, mechanisms, and applications," in *Analytical Techniques in Biosciences*, ed: Elsevier, 2022, pp. 179-190.

[66] A. Gutés, A. Ibanez, M. Del Valle, and F. Céspedes, "Automated SIA e-Tongue Employing a Voltammetric Biosensor Array for the Simultaneous Determination of Glucose and Ascorbic Acid," *Electroanalysis: An International Journal Devoted to Fundamental and Practical Aspects of Electroanalysis*, vol. 18, pp. 82-88, 2006.

[67] L. A. Terry, S. F. White, and L. J. Tigwell, "The application of biosensors to fresh produce and the wider food industry," *Journal of agricultural and food chemistry*, vol. 53, pp. 1309-1316, 2005.

[68] M. Mascini and S. Tombelli, "Biosensors for biomarkers in medical diagnostics," *Biomarkers*, vol. 13, pp. 637-657, 2008.

[69] H. J. Shin, "Genetically engineered microbial biosensors for in situ monitoring of environmental pollution," *Applied Microbiology and Biotechnology*, vol. 89, pp. 867-877, 2011.

[70] I. E. Tothill, "Biosensors for cancer markers diagnosis," in *Seminars in cell & developmental biology*, 2009, pp. 55-62.

[71] A. Amine, H. Mohammadi, I. Bourais, and G. Palleschi, "Enzyme inhibition-based biosensors for food safety and environmental monitoring," *Biosensors and Bioelectronics*, vol. 21, pp. 1405-1423, 2006.

[72] H. Xiao and J. Wang, "Discrimination of Xihulongjing tea grade using an electronic tongue," *African Journal of Biotechnology*, vol. 8, 2009.

[73] C. Söderström, F. Winquist, and C. Krantz-Rülcker, "Recognition of six microbial species with an electronic tongue," *Sensors and Actuators B: Chemical*, vol. 89, pp. 248-255, 2003.

[74] M. Palit, B. Tudu, N. Bhattacharyya, A. Dutta, P. K. Dutta, A. Jana, *et al.*, "Comparison of multivariate preprocessing techniques as applied to electronic tongue based pattern classification for black tea," *Analytica chimica acta*, vol. 675, pp. 8-15, 2010.

[75] T. Artursson and M. Holmberg, "Wavelet transform of electronic tongue data," *Sensors and Actuators B: Chemical*, vol. 87, pp. 379-391, 2002.

[76] W. Collier, D. Baird, Z. Park-Ng, N. More, and A. Hart, "Discrimination among milks and cultured dairy products using screen-printed electrochemical arrays and an electronic nose," *Sensors and Actuators B: Chemical*, vol. 92, pp. 232-239, 2003.

[77] A. Ghosh, B. Tudu, P. Tamuly, N. Bhattacharyya, and R. Bandyopadhyay, "Prediction of theaflavin and thearubigin content in black tea using a voltammetric electronic tongue," *Chemometrics and Intelligent Laboratory Systems*, vol. 116, pp. 57-66, 2012.

[78] A. Rudnitskaya, D. Kirsanov, A. Legin, K. Beullens, J. Lammertyn, B. M. Nicolaï, *et al.*, "Analysis of apples varieties–comparison of electronic tongue with different analytical techniques," *Sensors and Actuators B: chemical,* vol. 116, pp. 23-28, 2006.

[79] L. Lvova, S. S. Kim, A. Legin, Y. Vlasov, J. S. Yang, G. S. Cha, *et al.*, "All-solid-state electronic tongue and its application for beverage analysis," *Analytica Chimica Acta,* vol. 468, pp. 303-314, 2002.

[80] K. Beullens, P. Mészáros, S. Vermeir, D. Kirsanov, A. Legin, S. Buysens, *et al.*, "Analysis of tomato taste using two types of electronic tongues," *Sensors and Actuators B: Chemical,* vol. 131, pp. 10-17, 2008.

[81] R. R. Nidamanuri, "Hyperspectral discrimination of tea plant varieties using machine learning, and spectral matching methods," *Remote Sensing Applications: Society and Environment,* vol. 19, p. 100350, 2020.

[82] R. R. Nidamanuri, "Spectral discrimination of tea plant varieties by statistical, machine learning and spectral similarity methods," in *2014 6th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2014, pp. 1-4.

[83] Q. Chen, J. Zhao, and H. Lin, "Study on discrimination of Roast green tea (Camellia sinensis L.) according to geographical origin by FT-NIR spectroscopy and supervised pattern recognition," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* vol. 72, pp. 845-850, 2009.

[84] S.-F. Zhang, Z. De-Hua, and C. Xiao-Jing, "Analysis of E-tongue data for tea classification based on semi-supervised learning of generative adversarial network," *Chinese Journal of Analytical Chemistry,* vol. 50, pp. 77-85, 2022.

[85] M. B. Banerjee, R. B. Roy, B. Tudu, R. Bandyopadhyay, and N. Bhattacharyya, "Black tea classification employing feature fusion of E-Nose and E-Tongue responses," *Journal of food engineering,* vol. 244, pp. 55-63, 2019.

[86] X. Chen, Y. Xu, L. Meng, X. Chen, L. Yuan, Q. Cai, *et al.*, "Non-parametric partial least squares–discriminant analysis model based on sum of ranking difference algorithm for tea grade identification using electronic tongue data," *Sensors and Actuators B: Chemical,* vol. 311, p. 127924, 2020.

[87] G. Ren, J. Ning, and Z. Zhang, "Intelligent assessment of tea quality employing visible-near infrared spectra combined with a hybrid variable selection strategy," *Microchemical Journal,* vol. 157, p. 105085, 2020.

[88] L. A. Berrueta, R. M. Alonso-Salces, and K. Héberger, "Supervised pattern recognition in food analysis," *Journal of chromatography A,* vol. 1158, pp. 196-214, 2007.

[89] Á. P. Hearty and M. J. Gibney, "Analysis of meal patterns with the use of supervised data mining techniques—artificial neural networks and decision trees," *The American journal of clinical nutrition,* vol. 88, pp. 1632-1642, 2008.

[90] Q. Chen, J. Zhao, and S. Vittayapadung, "Identification of the green tea grade level using electronic tongue and pattern recognition," *Food research international,* vol. 41, pp. 500-504, 2008.

[91] M. Penza, G. Cassano, F. Tortorella, and G. Zaccaria, "Classification of food, beverages and perfumes by WO3 thin-film sensors array and pattern recognition techniques," *Sensors and Actuators B: Chemical,* vol. 73, pp. 76-87, 2001.

[92] P. Papajorgji, R. Chinchuluun, W. S. Lee, J. Bhorania, and P. M. Pardalos, "Clustering and classification algorithms in food and agricultural applications: a survey," *Advances in modeling agricultural systems,* pp. 433-454, 2009.

[93] P. Saha, S. Ghorai, B. Tudu, R. Bandyopadhyay, and N. Bhattacharyya, "Feature fusion for prediction of theaflavin and thearubigin in tea using electronic tongue," *IEEE Transactions on Instrumentation and Measurement,* vol. 66, pp. 1703-1710, 2017.

[94] A. Bakhshipour, A. Sanaeifar, S. H. Payman, and M. de la Guardia, "Evaluation of data mining strategies for classification of black tea based on image-based features," *Food analytical methods,* vol. 11, pp. 1041-1050, 2018.

[95] S. Meng, S. Wang, T. Zhou, and J. Shen, "Identification of tea red leaf spot and tea red scab based on hybrid feature optimization," in *Journal of Physics: Conference Series*, 2020, p. 052023.

[96] A. Moreda-Pineiro, A. Fisher, and S. J. Hill, "The classification of tea according to region of origin using pattern recognition techniques and trace metal data," *Journal of Food Composition and analysis,* vol. 16, pp. 195-211, 2003.

[97] A. Tripathy, A. Mohanty, and M. N. Mohanty, "Electronic nose for black tea quality evaluation using kernel based clustering approach," *International Journal of Image Processing (IJIP),* vol. 6, pp. 1-8, 2012.

[98] F. Song, X. Lu, Y. Lin, Q. Zhou, Z. Li, C. Ling*, et al.*, "Evaluation of black tea appearance quality using a segmentation-based feature extraction method," *Food Bioscience,* vol. 58, p. 103644, 2024.

[99] S. Mukhopadhyay, M. Paul, R. Pal, and D. De, "Tea leaf disease detection using multi-objective image segmentation," *Multimedia Tools and Applications,* vol. 80, pp. 753-771, 2021.

[100] H. Liu, D. Yu, and Y. Gu, "Classification and evaluation of quality grades of organic green teas using an electronic nose based on machine learning algorithms," *IEEE Access,* vol. 7, pp. 172965-172973, 2019.

[101] M. H. Kamrul, M. Rahman, M. R. I. Robin, M. S. Hossain, M. H. Hasan, and P. Paul, "A deep learning based approach on categorization of tea leaf," in *Proceedings of the International Conference on Computing Advancements*, 2020, pp. 1-8.

[102] P. K. Kundu and M. Kundu, "Classification of tea samples using SVM as machine learning component of E-tongue," in *2016 international conference on intelligent control power and instrumentation (ICICPI)*, 2016, pp. 56-60.

[103] D. Yu and Y. Gu, "A machine learning method for the fine-grained classification of green tea with geographical indication using a MOS-based electronic nose," *Foods,* vol. 10, p. 795, 2021.

[104] D. Batool, M. Shahbaz, H. Shahzad Asif, K. Shaukat, T. M. Alam, I. A. Hameed*, et al.*, "A hybrid approach to tea crop yield prediction using simulation models and machine learning," *Plants,* vol. 11, p. 1925, 2022.

[105] S. Chanda, A. K. Hazarika, N. Choudhury, S. A. Islam, R. Manna, S. Sabhapondit, *et al.*, "Support vector machine regression on selected wavelength regions for quantitative analysis of caffeine in tea leaves by near infrared spectroscopy," *Journal of Chemometrics,* vol. 33, p. e3172, 2019.

[106] Y. H. Zhong, S. Zhang, R. He, J. Zhang, Z. Zhou, X. Cheng, *et al.*, "A convolutional neural network based auto features extraction method for tea classification with electronic tongue," *Applied Sciences,* vol. 9, p. 2518, 2019.

[107] P. Saha, S. Ghorai, B. Tudu, R. Bandyopadhyay, and N. Bhattacharyya, "A novel technique of black tea quality prediction using electronic tongue signals," *IEEE Transactions on Instrumentation and Measurement,* vol. 63, pp. 2472-2479, 2014.

[108] L. Lvova, A. Legin, Y. Vlasov, G. S. Cha, and H. Nam, "Multicomponent analysis of Korean green tea by means of disposable all-solid-state potentiometric electronic tongue microsystem," *Sensors and Actuators B: Chemical,* vol. 95, pp. 391-399, 2003.

[109] G. Ren, X. Zhang, R. Wu, L. Yin, W. Hu, and Z. Zhang, "Rapid characterization of black tea taste quality using miniature NIR spectroscopy and electronic tongue sensors," *Biosensors,* vol. 13, p. 92, 2023.

[110] D. Huang, Z. Bian, Q. Qiu, Y. Wang, D. Fan, and X. Wang, "Identification of similar Chinese congou black teas using an electronic tongue combined with pattern recognition," *Molecules,* vol. 24, p. 4549, 2019.

[111] A. P. Bhondekar, M. Dhiman, A. Sharma, A. Bhakta, A. Ganguli, S. Bari, *et al.*, "A novel iTongue for Indian black tea discrimination," *Sensors and Actuators B: Chemical,* vol. 148, pp. 601-609, 2010.

[112] L. Jiang and K. Zheng, "Towards the intelligent antioxidant activity evaluation of green tea products during storage: A joint cyclic voltammetry and machine learning study," *Food Control,* vol. 148, p. 109660, 2023.

[113] S. Kumar and A. Ghosh, "A feature extraction method using linear model identification of voltammetric electronic tongue," *IEEE Transactions on Instrumentation and Measurement,* vol. 69, pp. 9243-9250, 2020.

[114] S. Ye, H. Weng, L. Xiang, L. Jia, and J. Xu, "Synchronously Predicting Tea Polyphenol and Epigallocatechin Gallate in Tea Leaves Using Fourier Transform–Near-Infrared Spectroscopy and Machine Learning," *Molecules,* vol. 28, p. 5379, 2023.

[115] P. Ivarsson, S. Holmin, N.-E. Höjer, C. Krantz-Rülcker, and F. Winquist, "Discrimination of tea by means of a voltammetric electronic tongue and different applied waveforms," *Sensors and Actuators B: Chemical,* vol. 76, pp. 449-454, 2001.

[116] R. Bhattacharyya, B. Tudu, S. C. Das, N. Bhattacharyya, R. Bandyopadhyay, and P. Pramanik, "Classification of black tea liquor using cyclic voltammetry," *Journal of Food Engineering,* vol. 109, pp. 120-126, 2012.

[117] W. He, X. Hu, L. Zhao, X. Liao, Y. Zhang, M. Zhang, *et al.*, "Evaluation of Chinese tea by the electronic tongue: Correlation with sensory properties and classification according to geographical origin and grade level," *Food Research International,* vol. 42, pp. 1462-1467, 2009.

[118] M. Khaydukova, X. Cetó, D. Kirsanov, M. del Valle, and A. Legin, "A tool for general quality assessment of black tea—Retail price prediction by an electronic tongue," *Food Analytical Methods,* vol. 8, pp. 1088-1092, 2015.

[119] Y. Li, J. Lei, and D. Liang, "Identification of fake green tea by sensory assessment and electronic tongue," *Food Science and Technology Research,* vol. 21, pp. 207-212, 2015.

[120] Z. Yang, N. Miao, X. Zhang, Q. Li, Z. Wang, C. Li*, et al.*, "Employment of an electronic tongue combined with deep learning and transfer learning for discriminating the storage time of Pu-erh tea," *Food Control,* vol. 121, p. 107608, 2021.

[121] Q. Ouyang, Y. Yang, J. Wu, Z. Liu, X. Chen, C. Dong*, et al.*, "Rapid sensing of total theaflavins content in black tea using a portable electronic tongue system coupled to efficient variables selection algorithms," *Journal of Food Composition and Analysis,* vol. 75, pp. 43-48, 2019.

[122] A. Ghosh, A. K. Bag, P. Sharma, B. Tudu, S. Sabhapondit, B. D. Baruah*, et al.*, "Monitoring the fermentation process and detection of optimum fermentation time of black tea using an electronic tongue," *IEEE Sensors Journal,* vol. 15, pp. 6255-6262, 2015.

[123] A. P. Bhondekar, R. Vig, A. Gulati, M. L. Singla, and P. Kapur, "Performance evaluation of a novel iTongue for Indian black tea discrimination," *IEEE Sensors Journal,* vol. 11, pp. 3462-3468, 2011.

[124] A. Bhuyan, B. Tudu, R. Bandyopadhyay, A. Ghosh, and S. Kumar, "ARMAX modeling and impedance analysis of voltammetric E-tongue for evaluation of infused tea," *IEEE Sensors Journal,* vol. 19, pp. 4098-4105, 2019.

[125] M. SINGH, "Impedance modeling for classification of flavoured green teas Munendra SINGH, Sunil SEMWAL b, Ashavani KUMAR c, Shailendra SINGHd 2."

[126] A. P. Bhondekar, R. Kaur, R. Kumar, R. Vig, and P. Kapur, "A novel approach using Dynamic Social Impact Theory for optimization of impedance-Tongue (iTongue)," *Chemometrics and Intelligent Laboratory Systems,* vol. 109, pp. 65-76, 2011.

[127] J. K. Carrillo, C. M. Durán, J. M. Cáceres, C. A. Cuastumal, J. Ferreira, J. Ramos*, et al.*, "Assessment of E-Senses Performance through Machine Learning Models for Colombian Herbal Teas Classification," *Chemosensors,* vol. 11, p. 354, 2023.

[128] A. Modak, T. N. Chatterjee, S. Nag, R. B. Roy, B. Tudu, and R. Bandyopadhyay, "Linear Regression Modelling on Epigallocatechin-3-gallate Sensor Data for Green Tea," in *2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN)*, 2018, pp. 112-117.

[129] D. Huo, Y. Wu, M. Yang, H. Fa, X. Luo, and C. Hou, "Discrimination of Chinese green tea according to varieties and grade levels using artificial nose and tongue based on colorimetric sensor arrays," *Food Chemistry,* vol. 145, pp. 639-645, 2014.

[130] P. Wei and Z. Wang, "Pattern recognition assisted linear sweep voltammetry sensor for analysis of tea quality," *International Journal of Electrochemical Science,* vol. 18, p. 100275, 2023.

[131] G. Ren, T. Li, Y. Wei, J. Ning, and Z. Zhang, "Estimation of Congou black tea quality by an electronic tongue technology combined with multivariate analysis," *Microchemical Journal*, vol. 163, p. 105899, 2021.

[132] H. Zhu, F. Liu, Y. Ye, L. Chen, J. Liu, A. Gui, *et al.*, "Application of machine learning algorithms in quality assurance of fermentation process of black tea--based on electrical properties," *Journal of food engineering,* vol. 263, pp. 165-172, 2019.

[133] A. Modak, R. B. Roy, B. Tudu, R. Bandyopadhyay, and N. Bhattacharyya, "A novel fuzzy based signal analysis technique in electronic nose and electronic tongue for black tea quality analysis," in *2016 IEEE First International Conference on Control, Measurement and Instrumentation (CMI)*, 2016, pp. 279-283.

CHAPTER 2

MACHINE LEARNING METHODS

This chapter has elaborated on various categories of machine learning methodologies employed in the analysis of sensor data pertinent to this study. The sensor data is processed using different feature transformation methods, including four specific techniques: PCA, DCT, SVD, and ICA. Additionally, three bio-inspired algorithms: Genetic Algorithm, Binary Bat, and Whale Algorithm are employed to optimize the features of the sensor dataset. Five widely used classification algorithms and two clustering techniques are detailed for their roles in data classification and clustering. The chapter also briefly describes two regression models, PLSR and PCR, utilized in this work. Finally, the limitations of these techniques are outlined.

LIST OF SECTION

Chapter 2

Machine Learning Methods

2.1 Introduction

Sensors have been utilized across a diverse fields of disciplines including environmental quality assessment, non-invasive medical diagnostics, food safety evaluation, and industrial procedure examination, due to several inherent advantages: (1) capability to operate effectively in adverse conditions, (2) capacity for continuous and autonomous functionality, and (3) elevated levels of precision and sensitivity. Typically, the advancement of a sensor relies on two primary elements: analytical instruments and data analysis methodologies. Innovative analytical instruments facilitate the generation of substantial quantities of information (data) and also enable the investigation of novel domains. Nevertheless, the sensor data produced may encompass extraneous information, and additionally, the foundational principles of these new domains may be exceedingly intricate and occasionally entirely obscure, thus necessitating that dependable sensor systems increasingly depend on advanced data processing methodologies.

As a formidable instrument for sophisticated data processing, machine learning has emerged as a fundamental technique for the innovative development of sensors, with the objective of unveiling the underlying principles that govern intricate systems. An exhaustive machine learning framework consists of three integral phases: data preprocessing, feature extraction and dimensionality reduction, and system modeling. The data preprocessing phase encompasses noise filtration, data normalization, signal alignment, and various other pertinent data treatments. Given that sensor signals typically comprise a multitude of variables, the subsequent phase employs feature extraction methodologies to transform sensor signals from their original high-dimensional space into a lower-dimensional feature space or to select the relevant variables that effectively characterize the entire system. Upon achieving an optimal feature representation, the final phase of machine learning involves the formulation of system models aimed at addressing either classification challenges or quantitative estimation tasks such as the prediction of chemical concentrations.

The thesis aims to analyze the performance of the sensors developed for tea quality estimation in terms of clustering of samples, classification or by predicting the constituents present in the sample.

Any of these three stages may significantly contribute to the regulation of effects associated with machine learning. This chapter endeavors to present a comprehensive examination of the algorithms presently employed at each stage, juxtapose their distinct characteristics, and explore prospective advancements in machine learning methodologies pertinent to sensor development, encompassing applications related to both classification and quantitative assessment via regression frameworks.

2.2 Machine learning techniques

Machine learning techniques provide an efficient automated way to analyze the various physical and chemical attributes of tea by data driven models based on measureable features. The essential prerequisites for the utilization of machine learning approaches in the assessment of tea quality encompass data collection and preprocessing, feature manipulation, model development, model assessment, as well as execution and deployment. The collected sensor data is subjected to analysis in order to detect patterns within the test samples through the utilization of two main categories of machine learning methodologies: supervised and unsupervised.

In the realm of pattern recognition, initial steps begin with transformation of raw sensor data using various techniques such as Principal component analysis (PCA), Discrete component analysis (DCT), Singular value decomposition (SVD), and Independent component analysis (ICA) to extract relevant and meaningful features. The transformed data are then optimized to reduce the size of data matrix using nature inspired optimization algorithms like Genetic Algorithm (GA), Bat Algorithm (BA), and Whale Optimization Algorithm (WOA).

Here classification model development involves various types of classification algorithms including Support Vector Machines (SVM), Decision Trees (DT), k-Nearest Neighbors (KNN), Ensemble methods, and Multilayer Perceptron (MLP). Additionally, regression models such as Partial Least Squares Regression (PLSR) and Principal Component Regression (PCR) are utilized to predict the biochemical components present in the tea. Performance of capacitive sensor is investigated by two well-known unsupervised clustering techniques: K-Means and Agglomerative.

The following section provides a brief overview of the multivariate machine learning algorithms employed in this study.

2.2.1 Unsupervised Machine Learning Techniques

Unsupervised machine technique pertains to a specific form machine learning wherein the algorithm is not provided with labeled training data. This technique tries to learn the patterns and relationships amidst the data without explicit guidance. The unsupervised techniques are widely used in various applications, including clustering, dimensionality reduction, and generative modeling. The subsequent section outlines the prevalent unsupervised techniques employed within this study

2.2.1.1 K-Means Clustering

The methodology of categorizing specific datasets into K clusters, predicated upon the mean of each cluster, is referred to as K-means clustering. Within the resultant clusters, the degree of intra-cluster similarity is markedly elevated, while the degree of inter-cluster similarity is comparatively diminished. The principal objective of these clusters is to aggregate analogous data points into K clusters. The variable K is a positive integer that delineates the quantity of clusters established. The foundational principle underlying K-means involves the assignment of K centroids, wherein each centroid is associated with a separate cluster. This methodology operates iteratively, allocating each data point to the cluster associated with the nearest centroid until a state of convergence is reached [1-3]. Convergence signifies that there are no further changes in the assignments of data points. The K-means clustering algorithm adheres to the subsequent iterative procedures:

Step1. At initialization stage, K data points are selected randomly from the dataset as initial cluster centroids.

Step2. Based on the Euclidean distance or alternative distance metrics, the assignment of each data point to the cluster linked with the nearest centroid has been accomplished.

Step3. The recalibration of the centroids for each cluster has been accomplished by calculating the arithmetic mean of all data points that have been allocated to that specific cluster.

Step4. Iteration occurs through steps 2 and 3 until convergence achieved. Convergence is reached when the assignments of data points to clusters exhibit minimal change or when a pre-defined number of iterations is achieved.

Step5. At the concluding phase, final cluster centroids represent the centers of the clusters, and each data point is assigned to a specific cluster.

The choice of initial centroids can impact the final clustering result. To diminish this, the algorithm is often run multiple times with different initializations, and the best result is selected based on a criterion such as minimizing the total intra-cluster variance. However, K-means is sensitive to outliers and the determination of the number of clusters (K) in advance is necessary. This may entail leveraging domain knowledge or employing additional techniques. However, this widely used unsupervised machine learning algorithms has good efficiency and scalability for handling extensive datasets.

2.2.1.2 Agglomerative clustering

Agglomerative clustering emerges as a widely used hierarchical clustering method applied across diverse applications, including data science, biology, image segmentation, and social network analysis. This algorithm initiates with each data point forming its own cluster and subsequently merges with the nearest cluster iteratively until the predefined stopping criteria are fulfilled. The key benefits of the agglomerative algorithm include a straightforward computational approach, ease of implementation, flexibility, minimal resource requirements for cluster creation, and the absence of a necessity for periodic re-clustering. The agglomerative hierarchical clustering technique follows the following steps [4, 5].

Step1. The initialization starts by considering each data point as an individual single-point cluster.

Step2. The distance between each pair of clusters or data points is measured using metrics such as Euclidean distance, Manhattan distance, or others depending on the context.

Step3. Two clusters that are closest to each other are identified based on the computed distances.

Step4. The distances between the newly formed cluster and the remaining clusters or data points are recomputed.

Step5. Repeat steps 3 and 4 if more than one cluster or a predefined stopping criterion is met.

Step6. The dendrogram or tree structure generated during the process represents the hierarchical clustering result. The user can cut the dendrogram at a desired level to obtain a specific number of clusters.

The agglomerative hierarchical clustering is flexible and it allows for the creation of a dendrogram which visually represents the merging process. Different linkage methods namely, complete linkage, single linkage, average linkage can be used to determine the distance between clusters. The choice of linkage method can impact the resulting clusters, and users may select the method based on their specific requirements or characteristics of the data.

2.2.2 Supervised Machine Learning Techniques

Supervised machine learning encompasses the process of instructing an algorithm through a labeled dataset, wherein the input variables are associated with their respective output labels. The primary aim of supervised learning is to establish a correspondence between inputs and outputs, enabling the generation of precise predictions or classifications when presented with novel data. Supervised machine learning encompasses the process of instructing an algorithm through a labeled dataset, wherein the input variables are associated with their respective output labels. The primary aim of supervised learning is to establish a correspondence between inputs and outputs, enabling the generation of precise predictions or classifications when presented with novel data. The machines learning algorithms find the patterns and construct the mathematical models. These created models are subsequently assessed by their predictive capability concerning the variations present in the data. Two main supervised models are classification models (classifiers) and regression models. The classification model, map the input data into predefined classes and where in the regression model, map the input space into a real-value domain. There are several options exist for representing classifiers namely, support vector machines (SVM), decision trees (DT), random forest (RF), K-nearest neighbors (KNN) etc.

Whereas linear regression and logistic regression are some regression models used in the supervised ML techniques. In the basic of ML model, the learning process comprises two stages: training and testing. During the training phase, the learning algorithm takes input from samples in the training data, where it learns features and constructs the learning mode. In the testing stages, learned model make the prediction for the test data. The tagged data is the output of learning model which gives the final prediction [6-8]. Among the two supervised models classification model has immense potential benefits for data mining and others application area. It's important to choose the right algorithm for a specific problem based on the nature and size data of the dataset and the desired outcome. Several classification models are employed in our proposed research.

2.2.2.1 Support Vector Machine (SVM)

The Support Vector Machine (SVM) represents a prevalent and potent supervised learning approach utilized for both classification and regression purposes [9]. It is originally designed for binary classification and subsequently by integrating it with binary classification techniques, SVM can effectively extend its application to handle multi-class classification tasks. This method aims to simultaneously reduce the empirical classification error while maximizing the geometric margin, hence often dubbed as maximum margin classifiers. SVM foundation lies in the concept of structural risk minimization (SRM). In a two-class scenario, a hyperplane is constructed in a high-dimensional feature space to facilitate classification. This hyperplane is flanked by two parallel hyperplanes, effectively segregating the data points. The objective is to optimize the distance between these parallel hyperplanes, as a wider margin correlates with improved generalization error of the classifier [10]. Fig.2.1 shows the maximum margin hyperplanes for a SVM, trained with two classes of samples. The SVM is classified into two groups: linear and non-linear SVM.

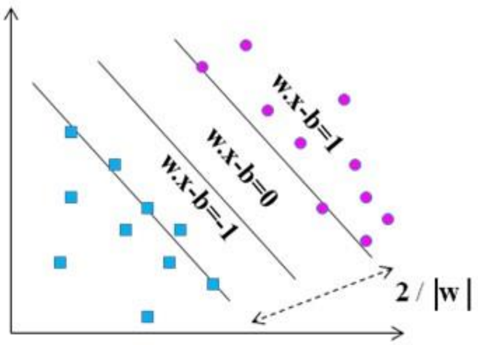


Fig.2.1. Maximum margin hyperplanes for a SVM trained with two classes samples

a) Linear SVM:

It is a simple classifier designed for cases where the training samples can be linearly separated. A linear function is given below

$$f(x) = w^T x + b \tag{2.1}$$

The training samples of two different classes are separated by a hyperplane

$f(x) = w^T x + b = 0$ where, x_i is the data point and **w** is weight vector which is normal to the hyperplane. For a given training set, numerous hyperplanes may maximize the separating margin between the two classes. It relies on the hyperplane that maximizes the separation margin between the two classes. The hyperplane that minimizes the separating margin between the two classes are indicated by data points by red square's and black circles. Elements of the training set that lie on the boundary hyperplane between the two classes are known as support vectors [11]. Fig.2.2 shows the two class linear hyperplane SVM.

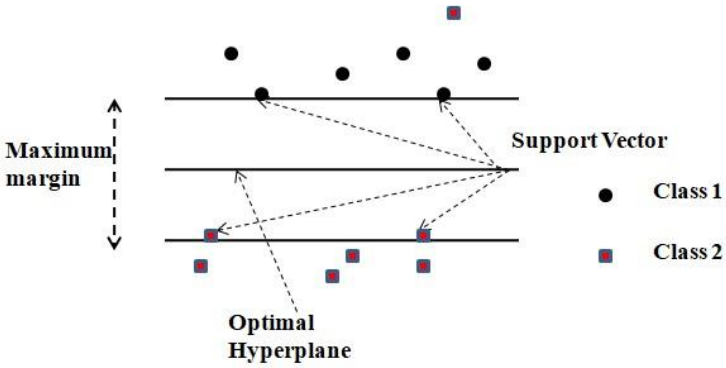


Fig.2.2. Linear SVM classification

The effectiveness of this method is assessed by calculating the classification accuracy rate. The unidentified data is trained, and optimal parameters are subsequently utilized to formulate a model for testing purposes. In this study, 70% of the dataset is allocated for training, 30% for testing, and a 10-fold cross-validation is performed to evaluate the performance.

b) Non-linear SVM: In linear SVM, a straight line or hyperplane is employed to differentiate between two classes. However, it may not always be feasible to separate datasets or data points by simply drawing a straight line between the two classes. But in a nonlinear SVM classifier, an operator is employed to map the input pattern x into a higher-dimensional space H. The non-linear classifier may be expressed as

$$f(x) = W^T \emptyset(X) + b \qquad (2.2)$$

Data exhibiting linear separability can be analyzed using a hyperplane, while linearly non-separable data is examined using various kernel functions such as quadratic and higher order polynomial kernel. The Fig.2.3 shows the non-linear SVM classifier.

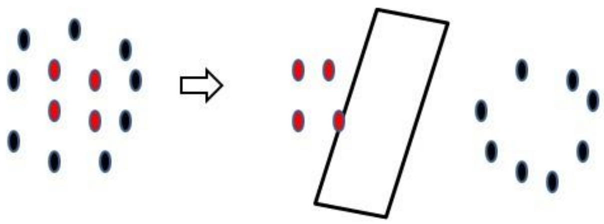


Fig.2.3. Non-linear SVM classification

2.2.2.2 k-Nearest Neighbor (KNN)

The K-Nearest Neighbor algorithm is a straightforward, non-parametric, supervised machine learning method that is grounded in the principles of limit theory. It makes decisions on the classification of an object by considering the majority vote from its nearby data points. The entity is assigned to the category that is most prevalent within its k-nearest neighbors, wherein k is generally a diminutive positive integer [12, 13].

In cases where k is set to 1, the object is directly assigned to the class of its closest neighbor. The emergence of this classifier stemmed from the need to perform discriminant analysis in scenarios where reliable parametric probability density estimates are difficult or impossible to ascertain. The process of classification using KNN entails evaluating the distance between the instance and established instances. The algorithm is proficient in generating multiple class labels for the unknown instance. Presented below is a systematic overview of the KNN classifier algorithm:

Step1. The initialization begins by choosing the value of k for classification.

Step2. Compute distance between input sample and training samples.

Step3. Short the distance by Euclidean or Manhattan distance method to measure the similarity between data points.

Step4. Define the training dataset and process the data.

Step5. Calculate the distance to all data points in the training set using the chosen distance metric and sort the neighbors.

Step6. Apply majority voting to determine the class of the unknown data point by taking a majority vote from the class labels of its k-nearest neighbors.

Step7. Assign the class label to the unknown data point based on the majority vote.

The disadvange of the KNN classifier is the significant time needed to find the closest neighborhood in a large training set [14, 15]. To mitigate this issue, a dimensionality reduction step is often undertaken.

2.2.2.3 Ensemble (Bagging tree)

The ensemble learning model consists of multiple fundamental classifiers that collectively categorize samples based on their collective votes. This strategy aids in alleviating the partiality of a single classifier [16-18]. As various classification techniques can capture unique discriminatory details and categorize samples using diverse criteria, merging a variety of classification methods enhances overall effectiveness. To enhance the diversity and effectiveness of the ensembles fundamental classifiers, we utilize three distinct classification methods to develop a two-layer voting ensemble model. Ensembles are frequently utilized to reduce over fitting, enhance stability, and boost the overall predictive capability of a model. There exist numerous types of ensemble classifiers; with bagging and boosting being two commonly used ones:

a) Bagging (Bootstrap Aggregating): In bagging tree, multiple instances of the same learning algorithm are trained on different subsets of the training data, which are randomly sampled with replacement (bootstrap samples). Each model in the ensemble is trained independently, and their predictions are combined through averaging (for regression) or voting (for classification). Random forest and bagged decision trees are the example of bagging.

b) Boosting: Boosting creates a sequence of weak learners, which are models that perform slightly better than random chance. These individual models are then combined by assigning weights to their predictions to generate a stronger and more precise ensemble prediction. Subsequent models focus on examples that previous models struggled with, assigning higher weights to misclassified instances. This iterative process aids the ensemble in adapting and enhancing its performance over time. Adaboost, gradient boosting machines (GBM), and others are examples of boosting trees. Ensemble algorithms can be applied to various types of base learners, including decision trees, linear models, or neural networks. The key is to incorporate diverse models that make different types of errors, thus enabling their combination to produce a more accurate and resilient prediction. Ensemble classifiers are widely utilized in practical applications and frequently surpass individual models, establishing them as a potent tool in machine learning.

2.2.2.4 Decision Tree

The decision tree algorithm is a supervised machine learning algorithm used for both classification and regression tasks [19]. It works by recursively partitioning the dataset into subsets based on the most significant attributes, creating a tree-like structure of decisions. The decision tree is generated through the training of data, allowing for the classification of multiple classes. This top-down approach initiates from the root node in the process. Three categories of decision trees exist: deep, medium, and shallow, determined by the quantity of splits within the tree. A decision tree is termed "deep" when it has a high number of splits, "shallow" when it has a low number of splits, and "medium" when it has a moderate number of splits. Herein, the medium decision tree has been used to address our research work. Here's a basic overview of how the decision tree algorithm works:

Step1. This algorithm evaluates different attributes and selects the one that provides the best split based on certain criteria.

Step2. Then the dataset is divided into subsets based on the chosen attribute. Each subset represents a branch in the decision tree.

Step3. The algorithm repeats the process recursively for each subset, further splitting the data based on the selected attributes until a stopping condition is met.

Step4. The process continues until the algorithm reaches leaf nodes, where the final predictions or decisions are made. In a classification tree, each leaf node corresponds to a specific class, while in a regression tree, it represents a numerical output.

Different types of decision trees are available depends on the specific situation and the desired outcome like classification tree, regression tree, decision tree forest, classification and regression tree and k-mean clustering [20]. Decision tree algorithms, a significant component of machine learning [21], find diverse applications across various fields. They can be employed for statistical data comparison, text classification, and extraction. Additionally, decision trees prove useful in educational institutions, businesses, and healthcare settings for managing student and staff records, as well as enhancing the diagnosis of various diseases. In the stock market, a decision tree can be developed to serve its purpose, offering an alternative to traditional statistical approaches. Some popular decision tree algorithms are ID3 (iterative dichotomiser 3), C4.5 and CART (classification and regression trees).

2.2.2.5 Discriminant Analysis

Discriminant analysis is a suitable approach for classifying data characterized by diverse feature dimensions. Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA) are commonly employed algorithms due to their prompt responsiveness and user-friendly characteristics [1, 22-24]. The decision surface in the LDA classifier is linear, whereas the decision boundary in QDA is non-linear. Here's an overview of how a discriminant analysis classifier, particularly LDA, works:

LDA: It is very popular and simple discriminant analysis technique which is employed in classification scenarios with the objective of assigning an observation to one of two or more classes. In the normal assumption, features follow a normal distribution within each class where as in homoscedasticity assumption, the variance-covariance matrices of the different classes are equal and in independence assumption, features are independent within each class. Some of the advantages of LDA are: when the assumptions are met, it performs well and LDA offers a reduction in dimensionality while maintaining the ability to discriminate between different classes. This classification method possesses some disadvantage**s** like: if the numbers of features are greater than the number of observations, the covariance matrices are not invertible. In that case another technique may be used. This technique is very much close to the PCA, but differ it as the axes that maximize class separability. In compare to the LDA, QDA formed different covariance matrices for each class. This technique very useful when the relationship between features and classes is nonlinear and for different covariance matrices QDA provide more accurate results compared to LDA. QDA is more flexible and capture complex relationships in the data. However, it comes with certain limitations, such as a susceptibility to over fitting with small datasets and the need to estimate a higher number of parameters. In this current study, the data are subjected to analysis using QDA.

2.2.2.6 Multilayer Perceptron (MLP)

MLP is a kind of artificial neural network (ANN) architecture that comprises of multiple layers of nodes (neurons) arranged in a hierarchical structure. It is a feed forward neural network that means information travels in one direction, from the input layer through the hidden layers to the output layer. The input layer or first layer of the MLP architecture receives the input features of the data. Where each node in the input layer represents a feature and the values are fed into the network. Between the input and output layers, there can be one or more hidden layers. Each node in a hidden layer is connected to every node in the previous and following layers. The inclusion of multiple hidden layers allows MLPs to learn complex patterns. Every link connecting nodes is assigned a weight. These weights are adjusted during the training process to optimize the network's performance and each node has a bias term. The hidden layers and sometimes in the output layer nodes apply an activation function to the weighted sum of their inputs. There are different types of activation functions like: sigmoid, hyperbolic tangent (tanh) has been applied to the network. The step function has been substituted with the sigmoidal function. Specifically, the logical sigmoidal function (logsig) is applied to all layers of the network system [25].The ultimate layer in the network is the output layer, and the number of nodes within this layer is determined by the nature of the task, such as binary classification, multi-class classification, or regression [26]. The network model is trained by supervised learning methods called back propagation and structure of multilayer perceptron is shown in Fig.2.4. In the training phase, the network's output is compared to the actual target values, and the error is propagated backward through the network. The biases and weights are so adjusted that it reduces the error. The loss function of the system is determined by the difference between the predicted value and the true target values. MLPs can be applied to a variety of tasks, like classification, regression, and pattern recognition. The architecture's flexibility and ability to learn complex relationships in data make MLPs a popular choice in machine learning applications. The primary advantages of neural networks include: 1) generalization, enabling them to classify unknown inputs even in noisy situations.2) fault tolerance capability, demonstrating graceful degradation [27].

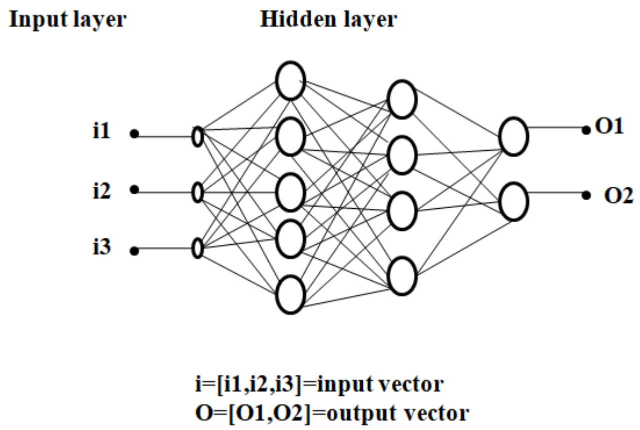


Fig.2.4. A multilayer perceptron with two hidden layer

2.3 Feature transformation Technique

With the continuous advancement of computer science and database technologies, humans increasingly depend on computers for gathering, analyzing, and leveraging the data. Machine learning (ML) and data mining represent a few of the artificial intelligence (AI) tools that help humanity in achieving these objectives. Pre-processing is the pivotal phase before learning, discovering and visualization. In various data analysis techniques, data subset construction and feature selection are to be worthy of focused analysis [28-30]. Feature transformation and extraction are crucial techniques in ML for improving model performance, reducing dimensionality, and capturing relevant information from raw data by eliminating the redundant features. Herein some feature transformation and dimension reduction techniques used in our research work.

2.3.1 Principal Component Analysis (PCA)

PCA [31-34] is an important statistical tool for examining and understanding the data. This particular mathematical methodology utilizes orthogonal transformations to convert a set of observations, which may be linked to correlated variables, into a series of values that represent linearly uncorrelated variables. The quantity of principal components is typically lower than the original variables. This process is designed in a way that ensures the first principal component captures the maximum variability, with each successive component striving to maximize the highest variance possible while also maintaining orthogonality with the previous components and maximizing the remaining variance. PCA offers several advantages, including data compression through dimensionality reduction with minimal loss of information. An essential outcome of PCA is the generation of a 'score plot,' which aids in visualizing differences between various experiments. The process of PCA involves a series of steps.

Step1. First steps involve normalizing or standardizing the data.

Step2. The computation of the covariance matrix is carried out in order to gain insights into the interrelationships among various attributes present in the dataset.

Step3. The derivation of eigenvalues and eigenvectors from the covariance matrix is crucial. Eigenvectors serve as indicators of the directions with the highest variance, while eigenvalues quantify the extent of variance along these specific directions.

Step4. The arrangement of principal components is based on their corresponding eigenvalues. Principal components associated with larger eigenvalues capture a more substantial amount of variance. Typically, a certain percentage of the total variance, such as 95%, is chosen, and the principal components that align with this specified variance are preserved.

Step5. The original dataset is transformed into a new coordinate system defined by the selected principal components. This transformation results in a representation of the data in a lower-dimensional space.

This statistical methodology has widespread applications in various fields including image processing, signal processing, biology, finance, and others. Its utility extends to tasks such as reducing dimensionality, visualizing data, extracting features, and mitigating noise in datasets.

2.3.2 Discrete Cosine Transformation (DCT)

The methodology proposed in this study is based on utilizing DCT for frequency-domain data representation. The conversion of time domain signals into frequency domain signals emphasizes the periodic characteristics of the input signal. In this technique, feature selection is conducted by selecting a suitable subset of the DCT components, facilitating a clearer distinction between groups. The selection of the DCT arises from the following two factors: 1) exploiting the energy compression properties of the DCT during feature selection; and 2) availability of time-efficient implementations of the DCT. This orthonormal time series transformation usefourier-related transformation technique exclusively employs cosine functions instead of both sine and cosine functions. This transformation preserves the length of the original vector and maintains the Euclidean distance between pairs of series [35, 36]. Similar to the fast fourier transform (FFT), it's also feasible to compute the coefficients of the DCT using $O(nlog(n))$ operations. The DCT offers several advantages over the discrete fourier transform (DFT), including: 1) It has good signal handling capabilities compared to the DFT. 2) DCT coefficients are consistently real. 3) For highly correlated input data, the energy concentration of the DCT significantly good that of the DFT. The Discrete Cosine Transform (DCT) finds application in various fields such as image processing [37] and numerous other domains like pattern recognition, data compression, and more. Its utility lies in offering a concise representation of highly correlated data.

The DCT A(k) ,in the time series signal s(x), x=0 ,1 ,2 ,…,n-1 and having length n, can be defined as:

$$A(k) = D(k)\sum_{x=0}^{n-1} s(x)Cos\{(2x+1)k\pi|2n\} \qquad (2.3)$$

For $\forall k = 0, 1,2 , \dots , n-1$, where $D(k) = \sqrt{\frac{1}{n}}$ for k=0 and $D(k) = \sqrt{\frac{2}{n}}$ for $k \neq 0$

DCT becomes orthonormal when it is multiplied by D(k) . Since the energy compactness of DCT coefficients is very high, only a few coefficients are sufficient to retain the characteristics of the whole signal.

2.3.3 Singular Value Decomposition (SVD)

In this research work SVD has been proposed as a way to extract relevant features from the sensor obtained dataset. This is a very important matrix decomposition method in linear algebra. Every matrix A can be divided into three matrixes as shown in equation 2.4.

$$A_{(m\times n)} = U_{(m\times n)} \sum V^{T}_{(m\times n)} \qquad (2.4)$$

The SVD of a given matrix can be simply expressed as $A=USV^T$, where U is the left singular vectors orthogonal matrix, S is the non-negative real numbers on the diagonal, known as the singular values of A and V is the right singular vectors, orthogonal matrix [38]. When decomposing a spectrogram of A using SVD, the singular values in S act solely as scaling factors and do not convey information about the spectrogram. Conversely, both the left and right singular vectors in U and V encapsulate information regarding the time and doppler dimensions of the spectrogram. SVD is widely used in various fields such as data analysis, signal processing, machine learning, and more, due to its ability to capture important patterns and reduce the dimensionality of data while retaining essential information [39, 40].

2.3.4 Independent Component Analysis (ICA)

Independent component analysis (ICA) is a modern statistical and computational technique employed for unleashing the latent factors (sources or features) within a dataset, aiming to ensure these features exhibit maximal independence.

It reduces the linear dependency between two components amidst noisy environments. ICA is a very effective feature extraction method, enhancing the classification performance of both supervised [41, 42] and unsupervised learning technique [43, 44]. Amidst the different execution model of ICA, it has restriction to sort their components. Principal components (PC's) are sorted according to the eigenvalues, but independent component (IC) has no certain rules to order these components. Mathematically, the observed variables x(t) = x1(t), x2(t), ..., xn(t) are composed of linear combination of original and mutually independent source b(t) = b1(t), b2(t), ..., bn(t) at time point t such that it expressed as

$$x(t) = Ab(t) \qquad (2.5)$$

Here A is a mixing matrix with full rank. Equation (2.4) is written as

$$y = Kx \qquad (2.6)$$

Where K = $A^{-1}$is the demixing matrix and y =y1, y2, ...,yn denotes the independent component. The task is to estimate the demixing matrix and independent components

only based on the mixed observations, which can be done by various ICA algorithms e.g. fast ICA, JADE, Infomax etc. [45]. In ICA estimation principles, the extracted components are both non-Gaussian and independent. The preprocessing steps of ICA are: centering, whitening (employing eigenvalue decomposition), and dimensionality reduction to simplify and streamline the complexity of the problem for the subsequent iterative algorithm. Here the whitening and dimension reduction can be achieved with

principal component analysis or singular value decomposition methods. ICA finds numerous applications in data analysis, source separation, and feature extraction [46].

2.4 Feature optimization

Feature optimization in machine learning refers to the process of selecting, transforming and engineering the features used to train a model in order to improve its performance. Here are some common techniques and considerations:

1) Feature Selection: In this method the relevant features of the dataset are identified to improve model performance and to reduce computational complexity. This can be achieved through different techniques like: univariate feature selection, recursive feature elimination and model-based feature selection.
2) Feature Transformation: Here, modifications of the existing features have been accomplished to make them more suitable for the model. Some common techniques include: scaling, encoding categorical variables and dimensionality reduction.
3) Feature Engineering: In this technique, new features from the existing ones have been created to improve the model's predictive power. This may involve: creating interaction terms, polynomial features and time-series features.
4) Regularization: Here overfitting protection have been done by introducing penalties on the feature coefficients during model training. L1 (Lasso) and L2 (Ridge) regularization methods reduces the model complexity and encourage the feature selection activity.
5) Cross-validation: This technique evaluates the performance of the selected feature sets to enhance the generalization activity to the unseen data.

By carefully optimizing the features of dataset, machine learning models can achieve better performance, generalization, and interpretability. Recently metaheuristic algorithms are employed to tackle complex real-world problems spanning various domains including economics, engineering, politics, and management. Here are some metaheuristic feature optimizing techniques adopted for our research work.

2.4.1 Genetic algorithm (GA)

GA is a very popular optimization tool which is inspired from biological evolution process [47] and based on natural selection concept. GA revolves around three fundamental components: chromosome representation, fitness selection, and biological inspired operators. This algorithm selects the best combination of spatial features which provide the better generalization performance. Chromosomes ,is the basic element of GA, has binary string format and each position of chromosomes has two possible alleles 0 and 1 [48]. These undergo genetic operators to iteratively replace their population. A population is formed by assembling a set of chromosomes. Each chromosome's quality is determined by evaluating it using a fitness function. The fittest chromosomes are chosen to generate new ones. During this step, two well-suited chromosomes are selected and merged through crossover to create offspring. Subsequently, mutation is introduced to the population to enhance randomness among individuals, reducing the risk of becoming trapped in local optima [49]. All steps of the standard GA can be summarized in the following way.

1) At first the population, comprised of individuals, is initialized. This may involve randomly generating a certain number of individuals, often represented by fixed-length character strings. The subsequent steps (2 to 4) are carried out iteratively until the stopping criterion is satisfied.

2) Then the each individual within the population has a chance of undergoing mutation. A particular individual may undergo slight random modifications.

3) The modified individuals undergo a process where they randomly divide and exchange these divisions with each other in pairs, generating new individuals through crossover. Consequently, the population is altered as a result of steps 2 and 3.

4) Now the fitness of each individual within the newly generated population is assessed. Depending on their fitness levels, only a portion of all individuals

proceeds to the next step i.e., step 2. Alternatively this procedure concludes if any individual achieves a satisfactory fitness level.

GA can be applied to a wide range of problem-solving area across various domains. Here are some common applications of genetic algorithms: optimization problem, function optimization, machine learning and robotics, data mining , bio-informatics, operation management, multimedia etc. [48].

2.4.2 Bat algorithm (BA)

BA is a very promising nature inspired alternatives based on the echolocation behavior of micro-bats for solving the hard optimization problem. The feature optimization and computational intelligence activity can be done by using frequency tuning technique. Each bat is defined with location (l_1^t), velocity (v_1^t),t-iteration in a s-dimensional solution space. The current location may be viewed as a vector solution to a problem. Within the population of n bats, the current best solution can be discovered during the iterative search process. During the iteration, location, velocity, and frequency of an artificial bat are modified [50-52]. According to the paper [53] the mathematical equation for updating the location (l_1^t), velocity (v_1^t) can be written as:

$$f_i = f_{min} + (f_{min} - f_{max})\delta \tag{2.7}$$

$$v_i^t = v_i^{t-1} + (l_i^{t-1} - l_*)f_i \tag{2.8}$$

$$l_i^t = l_i^{t-1} + v_i^t \tag{2.9}$$

Where $\delta \in [0,1]$ random vector is drown from uniform distribution and f_{min} is the best solution.The pivotal part bat algorithm can be outlined by the following steps.

1) The first steps start with initialization stage. Here we initialize the algorithm's parameters, generate and evaluate the initial population, and subsequently identify the best solution, l_{best} , within the population.
2) This is second step where generate the new solution. Here, virtual bats navigate through the search space according to the updating rules of the bat algorithm.
3) Third step is the local search steps. Using random walk, best solution have been improved
4) Here evaluation of new solution has been accomplished.
5) Here the conditions of best solution have been carried out.

6) This is last step where updating of best solution have been carried out.

During the search process, bats navigate through the open space by utilizing their position and velocity vectors within the search area. To address feature selection and classification challenges, a discrete variant of the bat algorithm called binary BA has been developed. In binary BA, features are represented by ones and zeros, indicating their presence or absence. Bats navigate to new positions in the search space by flipping different bits, employing a distinct strategy for adjusting their velocity and position. The BA has found applications in various fields due to its effectiveness in solving optimization problems. Some common applications of the BA include: engineering design, signal processing, image processing, data mining and robotics, feature selection [54] etc.

2.4.3 Whale optimization algorithm (WOA)

WOA is a stochastic, population-based, nature-inspired metaheuristic algorithm. It mimics the behavior of humpback whales and simulates their bubble-net hunting strategy. Optimization processes have been initiated by the spiral bubble-net feeding maneuver technique. The mathematical model of this algorithm can be constructed based on prey encirclement, spiral bubble-net feeding maneuvers, and prey searching.

The WOA is based on the hunting behavior of humpback whales and employs several key steps to iteratively search for an optimal solution to a given optimization problem. Here's the overview of the algorithm:

1) The first steps start with initialization of whales randomly within the search space. Here, we have defined parameters including the maximum iteration count, convergence criteria, and exploration/exploitation parameters.

2) Here for the update of best solution, the fitness of each whale is evaluated based on objective function and ultimately find whale with best fitness value.

3) Whale movement can be observed by: encircling prey behavior, bubble-net feeding behavior, prey searching behavior and ensuring the new position of the whale remains within the boundaries of the search space.

4) Here convergence checking has been done. Check whether the termination criteria, such as the maximum number of iterations or convergence threshold, are met or not. If met, terminate the algorithm; otherwise, continue to the next iteration.

5) Adjust algorithm-specific parameters, such as the exploration/exploitation rate, to balance exploration and exploitation in the search process for updating the parameters.

6) Repeat the steps 2 and 5 until the termination criteria are met.

This method offers the advantages of simplicity, low cost, rapid execution speed and high optimization accuracy. The performance of WOA has been validated for real and challenging optimization problems [52, 55, 56].

The WOA has been applied to a wide range of optimization problems across various domains due to its effectiveness in finding high-quality solutions. Some common applications of WOA include: signal processing, machine learning, engineering design, feature selection, wireless sensor network, data mining and robotics etc.

2.5 Regression models

Regression analysis aims to create a model based on observed data to understand the relationship between different sets of variables. This model helps to establish a connection between a dependent variable and one or more independent variables. The purpose of regression is to build a model DV=f(IV), where IV is the independent variable and DV is the dependent variable. Multivariate regression involves considering multiple predictive variables simultaneously, resulting in more accurate

modeling of the property of interest. The common multivariate regression techniques include principal component regression (PCR) and partial least squares regression (PLSR) [57].

2.5.1 Partial least square regression (PLSR)

PLSR is a linear statistical method and this is often called as component-based structural equation modeling. It utilizes a linear multivariate model to evaluate the performance of both the response matrix and the predictor matrix [58].This method can effectively model the relationships between input (independent) and output (dependent) variables, either as a regression or structural model. For exploratory prediction or modeling purposes, PLS is preferred due to its freedom from requiring any specific assumptions. PLSR is a non-parametric method known for its robustness regardless of sample size, and it does not necessitate data normalization [59]. It can be applied to a small number of samples and it demonstrates stability even in the presence of missing information. In this study, the leaveone-out cross-validation (LOOCV) method has been employed as a performance metric. It systematically removes one data point at a time and evaluates the model's performance based on the remaining dataset, maximizing the correlation between predictors and response variables. Various parametric values such as the root mean square error of validation (RMSEV), root mean square error of prediction (RMSEP), and RMSEC are utilized to optimize the number of latent variables (LV), thereby assessing the performance of the model [52].

2.5.2 Principal component regression (PCR)

PCR is a predictive data mining technique that utilizes the PCA. It effectively addresses issues of data co linearity and minimizes the number of regression variables. PCR decomposes a data matrix X into scores S and loadings matrix L. The process of decomposition process is given as

$$X = SL^T + E \qquad (2.10)$$

Where X is the data matrix, S is scores matrix,L^T is the transpose of loading matrix and E is the error matrix. The linear regression can be defined as

$$Y = Xa + e \qquad (2.11)$$

Where Y represent the dependent variable, is a regression coefficient and e is the residual matrix. Replacing the data matrix X by the score matrix S, maximum information available in the input data matrix X can be included in the predicted output Y. The resulting regression model can now be denoted as

$$Y = S(L^T)a + e_T \qquad (2.12)$$

$$Y = Sg + e_T \qquad (2.13)$$

Here g is the new regression coefficient and e_T is the error matrix. The regression coefficient for the original model is

$$a_{pcr} = Lg \qquad (2.14)$$

The model's complexity can be optimized by adjusting the number of components utilized in its construction [57, 60].

2.6 Limitation of the Techniques

Researchers proposed verities of ML techniques for the data analysis and pattern recognition but a universal technique, which may be applied to any system is still a under developing work. In many applications, the fundamental technique is modified and fused with the other model to form the hybridized model. This mixed methods or hybridized model may overcome the limitations of ML indeed [61, 62].Though ML techniques are powerful tools for data analysis and pattern recognition system, they also come with several limitations that have been addressed here:

1) Data quality: ML models rely on the quality and representativeness of large amounts of labeled data to learn patterns effectively. Poor quality data, e.g. missing values, outliers, or biased samples, can lead to inaccurate or unreliable predictions .The requirement of more data leads to better algorithmic performance [63].
2) Over fitting data: This phenomenon occurs when a model learns to capture noise or random fluctuations in the training data. This can result in poor generalization to new, unseen data. These types of limitation can be reduced by regularization techniques [64] and cross-validation method.
3) Interpretability: In complex ML models like deep neural networks, often suffer lack of interpretability which makes it difficult to predict the component. This is a very significant limitation in healthcare and finance domain [65].
4) Computational resources: Deep learning algorithm requires considerable computational resources, including processing power and memory, for training, inference and computational purpose. ML is not applicable in the resource-constrained environments [66].
5) Bias and fairness: ML models may replicate biases existing within the training data, potentially resulting in unfair or discriminatory outcomes, especially in critical areas such as hiring or lending. Achieving fairness and reducing bias in machine learning models remains an ongoing challenge [67].
6) Generalization: ML models may perform well on the training data but struggle to generalize to unseen data [68]. Changes in the data distribution may lead to

degraded performance, requiring continuous monitoring and adaptation of the models.

7) Feature selection and extraction: The performance of ML models depends on the selection and engineering of relevant features from the raw data. The dimension reduction and proper feature selection, enhances the model performance and predictive accuracy [69]. Designing informative features requires domain expertise and can be a time-consuming process.

8) Security issue: ML models can be vulnerable to adversarial attacks, where malicious inputs are crafted to fool the model into making incorrect predictions. Continuously researching and enhancing the robustness and security of machine learning models is very much imperative [70].

9) Ethical and privacy issue: ML models trained on sensitive or personal data raise ethical concerns regarding privacy, consent, and potential misuse. Ensuring ethical considerations are integrated into the development and deployment of machine learning systems is essential [71].

2.7 Conclusion

This chapter explores various conventional machine learning techniques employed within the scope of this thesis. Firstly, the characteristics and operations of two clustering algorithms such as K-Means clustering and Agglomerative clustering are briefly outlined. Subsequently, five supervised ML algorithms, including SVM, KNN, ensemble bagging tree, decision tree, discriminant analysis, and MLP, are utilized for the classification of diverse tea samples. The research involves the transformation of raw data through PCA, DCT, SVD, and ICA feature transformation methods, followed by data optimization using bio-inspired GA, BA, and WOA techniques. Prediction accuracies are assessed utilizing the PLSR and PCR algorithms for both sensor performance and proposed models. Finally, the chapter concludes with a discussion on the limitations of ML techniques.

The subsequent chapter explores a novel voltammetric electrode optimization technique deploying widely used machine learning algorithm for black tea quality assessment.

References

[1] H. Harb, A. Makhoul, D. Laiymani, A. Jaber, and R. Tawil, "K-means based clustering approach for data aggregation in periodic sensor networks," in 2014 IEEE 10th international conference on wireless and mobile computing, networking and communications (WiMob), 2014, pp. 434-441.

[2] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," Pattern recognition, vol. 36, pp. 451-461, 2003.

[3] S. Shukla and S. Naganna, "A review on K-means data clustering approach," International Journal of Information & Computation Technology, vol. 4, pp. 1847-1860, 2014.

[4] T. K. Jain, D. S. Saini, and S. V. Bhooshan, "Performance analysis of hierarchical agglomerative clustering in a wireless sensor network using quantitative data," in 2014 international conference on information systems and computer networks (ISCON), 2014, pp. 99-104.

[5] A. Bouguettaya, Q. Yu, X. Liu, X. Zhou, and A. Song, "Efficient agglomerative hierarchical clustering," Expert Systems with Applications, vol. 42, pp. 2785-2797, 2015.

[6] V. Nasteski, "An overview of the supervised machine learning methods," Horizons. b, vol. 4, pp. 51-62, 2017.

[7] H. Bhavsar and A. Ganatra, "A comparative study of training algorithms for supervised machine learning," International Journal of Soft Computing and Engineering (IJSCE), vol. 2, pp. 2231-2307, 2012.

[8] I. Muhammad and Z. Yan, "SUPERVISED MACHINE LEARNING APPROACHES: A SURVEY," ICTACT Journal on Soft Computing, vol. 5, 2015.

[9] V. Vapnik, The nature of statistical learning theory: Springer science & business media, 1999.

[10] K. S. Durgesh and B. Lekha, "Data classification using support vector machine," Journal of theoretical and applied information technology, vol. 12, pp. 1-7, 2010.

[11] K. Machhale, H. B. Nandpuru, V. Kapur, and L. Kosta, "MRI brain cancer classification using hybrid classifier (SVM-KNN)," in 2015 International Conference on Industrial Instrumentation and Control (ICIC), 2015, pp. 60-65.

[12] J. Fuli and C. Chu, "Application of knn improved algorithm in automatic classification of network public proposal cases," in 2017 IEEE 2nd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), 2017, pp. 82-86.

[13] R. Mynavathi, V. Bhuvaneswari, T. Karthikeyan, and C. Kavina, "K nearest neighbor classifier over secured perturbed data," in 2016 World Conference on Futuristic Trends in Research and Innovation for Social Welfare (Startup Conclave), 2016, pp. 1-4.

[14] B. V. Dasarathy, "Nearest neighbor (NN) norms: NN pattern classification techniques," IEEE Computer Society Tutorial, 1991.

[15] M. Manjusha and R. Harikumar, "Performance analysis of KNN classifier and K-means clustering for robust classification of epilepsy from EEG signals," in 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), 2016, pp. 2412-2416.

[16] T. G. Dietterich, "Ensemble methods in machine learning," in International workshop on multiple classifier systems, 2000, pp. 1-15.

[17] R. Polikar, "Ensemble based systems in decision making," IEEE Circuits and systems magazine, vol. 6, pp. 21-45, 2006.

[18] Z. Zhang, J. Li, H. Hu, and H. Zhou, A robust ensemble classification method analysis: Springer, 2010.

[19] S. Acharya, D. Das, T. N. Chatterjee, S. Mukherjee, R. B. Roy, B. Tudu, et al., "Voltammetric Electrode Array Optimization for Black Tea Discrimination Using Computational Intelligence Approach," IEEE Sensors Journal, vol. 21, pp. 20589-20595, 2021.

[20] A. Navada, A. N. Ansari, S. Patil, and B. A. Sonkamble, "Overview of use of decision tree algorithms in machine learning," in 2011 IEEE control and system graduate research colloquium, 2011, pp. 37-42.

[21] T. M. Mitchell, "Machine learning," ed: McGraw-hill, 1997.

[22] S. Herle, "Movement intention detection from SEMG signals using time-domain features and discriminant analysis classifiers," in 2018 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR), 2018, pp. 1-6.

[23] A. Tharwat, "Linear vs. quadratic discriminant analysis classifier: a tutorial," International Journal of Applied Pattern Recognition, vol. 3, pp. 145-180, 2016.

[24] R. Setiawan, "Quadratic Classifier from Discriminant Analysis for Classification of Multiple Attributes Data."

[25] R. Bharathi and B. Shekar, "Discriminative dct-mlp based approach for off-line signature verification," in 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2014, pp. 2309-2315.

[26] M. W. Gardner and S. Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," Atmospheric environment, vol. 32, pp. 2627-2636, 1998.

[27] S. Acharya, T. N. Chatterjee, S. Mukherjee, D. Das, R. B. Roy, B. Tudu, et al., "Selection Of Optimum number Of Sensors Of An Electronic Tongue For Efficient Classification Of Black Tea: A Combinatorial Approach Based On Discrete Cosine Transform and Artificial Neural Network," in 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), 2018, pp. 108-111.

[28] R. J. Brachman, "The process of knowledge discovery in databases," Advances in knowledge discovery and data mining, pp. 37-57, 1996.

[29] H. Liu and H. Motoda, "Feature transformation and subset selection," IEEE Intell Syst Their Appl, vol. 13, pp. 26-28, 1998.

[30] P. Langley, Elements of machine learning: Morgan Kaufmann, 1996.

[31] P. Ciosek, K. Brudzewski, and W. Wróblewski, "Milk classification by means of an electronic tongue and Support Vector Machine neural network," Measurement Science and Technology, vol. 17, p. 1379, 2006.

[32] L. Lvova, A. Legin, Y. Vlasov, G. S. Cha, and H. Nam, "Multicomponent analysis of Korean green tea by means of disposable all-solid-state potentiometric electronic tongue microsystem," Sensors and Actuators B: Chemical, vol. 95, pp. 391-399, 2003.

[33] J. Olsson, F. Winquist, and I. Lundström, "A self polishing electronic tongue," Sensors and Actuators B: Chemical, vol. 118, pp. 461-465, 2006.

[34] F. Winquist, R. Bjorklund, C. Krantz-Rülcker, I. Lundström, K. Östergren, and T. Skoglund, "An electronic tongue in the dairy industry," Sensors and Actuators B: Chemical, vol. 111, pp. 299-304, 2005.

[35] I. Batal and M. Hauskrecht, "A supervised time series feature extraction technique using dct and dwt," in 2009 international conference on machine learning and applications, 2009, pp. 735-739.

[36] B. Schwerin and K. Paliwal, "Local-DCT features for facial recognition," in 2008 2nd International Conference on Signal Processing and Communication Systems, 2008, pp. 1-6.

[37] K. R. Rao and P. Yip, Discrete cosine transform: algorithms, advantages, applications: Academic press, 2014.

[38] X. Chen and X. Chang, "Feature extraction of lung ventilation by biomedical electrical impedance tomography," in 2017 IEEE International Conference on Imaging Systems and Techniques (IST), 2017, pp. 1-5.

39] M. Narwaria and W. Lin, "SVD-based quality metric for image and video using machine learning," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 42, pp. 347-364, 2011.

[40] Y. Wang, L. Tian, and C. Li, "LBP-SVD based copy move forgery detection algorithm," in 2017 IEEE international symposium on multimedia (ISM), 2017, pp. 553-556.

[41] X. Zhang, V. Ramani, Z. Long, Y. Zeng, A. Ganapathiraju, and J. Picone, "Scenic beauty estimation using independent component analysis and support vector machines," in Proceedings IEEE Southeastcon'99. Technology on the Brink of 2000 (Cat. No. 99CH36300), 1999, pp. 274-277.

[42] N. Kwak, C.-H. Choi, and J. Y. Choi, "Feature extraction using ICA," in International Conference on Artificial Neural Networks, 2001, pp. 568-573.

[43] S.-I. Lee and S. Batzoglou, "Application of independent component analysis to microarrays," Genome biology, vol. 4, pp. 1-21, 2003.

[44] A. Kapoor, T. Bowles, and J. Chambers, "A novel combined ICA and clustering technique for the classification of gene expression data," in Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005., 2005, pp. v/621-v/624 Vol. 5.

[45] M. S. Reza and J. Ma, "ICA and PCA integrated feature extraction for classification," in 2016 IEEE 13th International Conference on Signal Processing (ICSP), 2016, pp. 1083-1088.

[46] D. Jeyabharathi and A. Suruliandi, "Performance analysis of feature extraction and classification techniques in CBIR," in 2013 International Conference on Circuits, Power and Computing Technologies (ICCPCT), 2013, pp. 1211-1214.

[47] Z. Michalewicz and M. Schoenauer, "Evolutionary algorithms for constrained parameter optimization problems," Evolutionary computation, vol. 4, pp. 1-32, 1996.

[48] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: past, present, and future," Multimedia tools and applications, vol. 80, pp. 8091-8126, 2021.

[49] P. Ghamisi and J. A. Benediktsson, "Feature selection based on hybridization of genetic algorithm and particle swarm optimization," IEEE Geoscience and remote sensing letters, vol. 12, pp. 309-313, 2014.

[50] M. W. U. Alam, "Improved binary bat algorithm for feature selection," 2019.

[51] R. Ramasamy and S. Rani, "Modified binary bat algorithm for feature selection in unsupervised learning," Int. Arab J. Inf. Technol., vol. 15, pp. 1060-1067, 2018.

[52] S. Acharya, D. Das, S. Nag, S. Mukherjee, A. K. Hazarika, S. Sabhapondit, et al., "Optimization techniques for a voltammetric signal to predict green tea quality parameters using MIP electrode," IEEE Sensors Journal, 2023.

[53] X.-S. Yang, "A new metaheuristic bat-inspired algorithm," in Nature inspired cooperative strategies for optimization (NICSO 2010), ed: Springer, 2010, pp. 65-74.

[54] I. Fister, X.-S. Yang, S. Fong, and Y. Zhuang, "Bat algorithm: Recent advances," in 2014 IEEE 15th International symposium on computational intelligence and informatics (CINTI), 2014, pp. 163-167.

[55] S. Mirjalili and A. Lewis, "The whale optimization algorithm," Advances in engineering software, vol. 95, pp. 51-67, 2016.

[56] Y.-W. Liu, H. Feng, H.-Y. Li, and L.-L. Li, "An improved whale algorithm for support vector machine prediction of photovoltaic power generation," Symmetry, vol. 13, p. 212, 2021.

[57] S. V. Suryakala and S. Prince, "Investigation of goodness of model data fit using PLSR and PCR regression models to determine informative wavelength band in NIR region for non-invasive blood glucose prediction," Optical and Quantum Electronics, vol. 51, pp. 1-20, 2019.

[58] S. Sabzi, R. Pourdarbani, M. H. Rohban, G. García-Mateos, and J. I. Arribas, "Estimation of nitrogen content in cucumber plant (Cucumis sativus L.) leaves using hyperspectral imaging data with neural network and partial least squares regressions," Chemometrics and Intelligent Laboratory Systems, vol. 217, p. 104404, 2021.

[59] S. Sabzi, Y. Abbaspour-Gilandeh, and G. García-Mateos, "A new approach for visual identification of orange varieties using neural networks and metaheuristic algorithms," Information processing in agriculture, vol. 5, pp. 162-172, 2018.

[60] S. Mahesh, D. Jayas, J. Paliwal, and N. White, "Comparison of partial least squares regression (PLSR) and principal components regression (PCR) methods for protein and hardness predictions using the near-infrared (NIR) hyperspectral images of bulk samples of Canadian wheat," Food and bioprocess technology, vol. 8, pp. 31-40, 2015.

[61] D. Patton, P. Blandfort, W. Frey, M. Gaskell, and S. Karaman, "Annotating social media data from vulnerable populations: Evaluating disagreement between domain experts and graduate student annotators," 2019.

[62] F. Cardoso, L. J. van't Veer, J. Bogaerts, L. Slaets, G. Viale, S. Delaloge, et al., "70-gene signature as an aid to treatment decisions in early-stage breast cancer," New England Journal of Medicine, vol. 375, pp. 717-729, 2016.

[63] G. Mastorakis, "Human-like machine learning: Limitations and suggestions. arXiv 2018," arXiv preprint arXiv:1811.06052.

[64] X. Ying, "An overview of overfitting and its solutions," in Journal of physics: Conference series, 2019, p. 022022.

[65] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal, "Explaining explanations: An approach to evaluating interpretability of machine learning," arXiv preprint arXiv:1806.00069, p. 118, 2018.

[66] C. Chen, P. Zhang, H. Zhang, J. Dai, Y. Yi, H. Zhang, et al., "Deep learning on computational-resource-limited platforms: a survey," Mobile Information Systems, vol. 2020, pp. 1-19, 2020.

[67] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," ACM computing surveys (CSUR), vol. 54, pp. 1-35, 2021.

[68] Y. Chung, P. J. Haas, E. Upfal, and T. Kraska, "Unknown examples & machine learning model generalization," arXiv preprint arXiv:1808.08294, 2018.

[69] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in 2014 science and information conference, 2014, pp. 372-378.

[70] Z. Guan, L. Bian, T. Shang, and J. Liu, "When machine learning meets security issues: A survey," in 2018 IEEE international conference on intelligence and safety for robotics (ISR), 2018, pp. 158-165.

[71] K. Santosh, L. Gaur, K. Santosh, and L. Gaur, "Privacy, security, and ethical issues," Artificial Intelligence and Machine Learning in Public Healthcare: Opportunities and Societal Impact, pp. 65-74, 2021.

CHAPTER 3

BLACK TEA QUALITY ASSESSMENT BY OPTIMIZING SENSOR ARRAY: EMPLOYING VOLTAMETRIC POLYMER GRAPHITE ELECTRODES

This chapter delineates a computational intelligence approach to optimize a voltammetric electrode array in a three electrode configuration, aimed at evaluating the overall quality of black tea. Nine working electrodes were meticulously fabricated through the modulation of polymer and graphite compositions. The sensor data acquired from experimentation were transformed using four feature transformation methodologies. Subsequently, five widely recognized classification algorithms, in conjunction with novel polling technique, were implemented to optimize the electrode array.

LIST OF SECTION

Contents of this chapter are based on following publication:

S Acharya, D.Das , T. N.Chatterjee, S. Mukherjee ,R. B.Roy , B .Tudu , R.Bandyopadhyay. “Voltammetric Electrode Array Optimization for Black Tea Discrimination Using Computational Intelligence Approach”, *IEEE Sensors Journal*, vol. 21, no. 18, pp. 20589-20595, 15 Sept.15, 2021, doi: 10.1109/JSEN.2021.3098036.

Chapter 3

Black Tea Quality Assessment by Optimizing Sensor Array: Employing Voltammetric Polymer Graphite Electrodes

3.1 Introduction

Black tea is a prevalent beverage of the modern world due to its myriad health benefits owing to the presence of significant antioxidants like theaflavins, thearubigins, catechins etc. [1]. Assessing the quality of black tea is utmost importance due to the dependency of a vast population on this specific beverage. Traditionally, the evaluation of tea quality by human tasters is predicated on sensory attributes such as aroma, flavor, and taste . Nevertheless, this evaluative approach is fraught with unreliability owing to the cognitive and physiological constraints inherent in human assessors [2]. Analytical instruments, including mass spectrometry, liquid chromatography, and high-performance liquid chromatography (HPLC) [1], yield precise measurements; however, these techniques are characterized by their considerable cost, complexity, and the necessity for proficient operators, as elaborated in detail in Chapter 1. To overcome these limitations, electrochemical methods combined with efficient pattern recognition systems have emerged as effective techniques for estimating tea quality.

An electrode array in an electrochemical measurement system comprises multiple electrodes, and numerous research efforts have been dedicated to optimizing these arrays for food and beverage quality analysis. For example, theaflavins and thearubigins in black tea liquor have been measured using sensor arrays with expensive noble metal electrodes [3]. Researchers have proposed carbon paste electrodes (CPEs) as a cost-effective alternative, although they exhibit limited selectivity due to non-specific nature [4-6]. Polymer graphite composite electrodes (PGE) are affordable, conductive, and rapid in their response. A batch of nine electrodes have been developed from three monomers namely acrylamide, aniline, and pyrrole. Consequently, these synthesized polymer graphite composite electrodes are immersed in liquid tea samples and subjected to cyclic voltammetry (CV) with an objective of developing an ET with sensor array. The optimization of electrode arrays has been attempted with the application of suitable machine learning techniques.

A comprehensive review of literature demonstrates that the data collected by the sensor undergo analysis through feature transformation methodologies and classification algorithms. In reference [7], the authors condensed the electronic tongue (ET) data, derived from pulse voltammetry, utilizing discrete wavelet transform (DWT). An innovative method was introduced in reference [8] for forecasting the quality of black tea using ET, in which DWT coefficients are acquired through the moving window technique and the energy from various frequency bands is utilized as the features. Another study has established the classification of black tea and the effectiveness of ET in conjunction with voltammetry by examining the PCA score plots [9]. In reference [10], electronic nose (EN) sensors are selected through different feature selection and classification algorithm methodologies, with redundant sensors being eliminated from the sensor array. Reference [11] employs rough set techniques for both black tea pattern classification and sensor optimization for EN. The concept of distinguishing teas using an electronic nose was introduced in reference [12], where the reduction of EN data dimensions is carried out through PCA, and the performance of three nonlinear classifiers and support vector machine (SVM) is compared. In various other fields of application [13], multi-sensing techniques like ET, EN, and Fourier Transform Infrared Spectroscopy (FTIR) are utilized to replicate human sensory abilities for honey testing. This involves the classification of pure and adulterated honeys using different classifiers such as K-nearest neighbors (KNN), artificial neural network (ANN), support vector machine (SVM), and probabilistic neural network (PNN). Additionally, a method combining ET and EN is used to classify five different types of strawberry juices [14]. The literature review emphasizes that the optimization of electrodes and reduction of redundancy can be accomplished through a computational intelligence approach.

Optimization of electrodes in a measurement system is crucial for the tea industry in analyzing tea quality. The study presents the development of nine different polymer graphite composite electrodes utilizing specific monomers such as acrylamide, aniline, and pyrrole. Incorporating suitable fillers in polymer composites has led to improved conduction properties compared to regular polymers, attributed to the synergistic effect at the matrix-fillers interface [15]. Graphite, a cost-effective carbonaceous material, exhibits favorable electrical and mechanical characteristics, making it a suitable base material for preparing voltammetric electrodes.For electrochemical analysis, the nine electrodes are immersed in black tea liquor and subjected to cyclic voltammetry (CV) using a triangular voltage range from -0.04 V to 0.8 V. Electrode array optimization is carried out by utilizing four feature extraction techniques: principal component analysis (PCA), discrete cosine transform (DCT), singular value decomposition (SVD), independent component analysis (ICA) and five classification algorithms such as, support vector machine (SVM), K-nearest neighbor (KNN), ensemble, decision tree, and discriminant analysis, followed by a polling process.The evaluation of black tea quality using low-cost polymer graphite composite electrodes highlights the novelty of this study. Furthermore, the optimization of an electrode array through computational methods adds to the unconventional nature of this research. A polling strategy is implemented to identify the top-performing electrodes among the nine developed ones, which effectively differentiate tea quality based on their rankings. Each electrode is scored out of 20 according to its classification accuracy rate, with the top two electrodes, WE1 and WE8, achieving average classification accuracy rates of 91.27% and 91.34%, respectively, post optimization. This underscores the importance of selecting the appropriate working electrodes following optimization. The operational flowchart illustrating the methodology employed in this study is presented in Fig. 3.1. The electrode array optimization techniques utilized in this research enhance the overall performance of the measurement system, demonstrating the potential for similar applications in other fields.

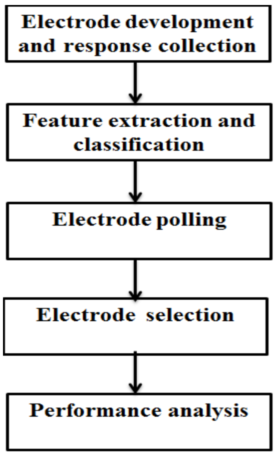


Fig. 3.1. Operational flowchart

3.2 Experimentation

3.2.1 Reagents and Standards

Aniline, acryl amide, and pyrrole are acquired from Sisco Research Laboratories Pvt. Ltd., India. The hydrochloric acid utilized as a buffer and paraffin oil acting as a binder are sourced from Merck. Both graphite powder (with a purity of 99%) and di-vinyl benzene (a cross linker) are obtained from Sigma Aldrich. E-Merck has provided benzoyl peroxide, which serves as the polymerization initiator. All the chemicals utilized are of analytical grades, and the experimental procedures are conducted utilizing Millipore water (with a resistance of 18MΩ).

3.2.2 Electrodes Preparation

Three varieties of polymer-graphite composites have been synthesized employing polyaniline (PANI), polyacrylamide (PAM), and polypyrrole (PPY). For each composite, compositions of the polymers are varied from 10% to 30%. The electrodes are denoted by the polymer's abbreviation along with the percentage of polymer-graphite composition. The chemical composition ratios of the electrodes are presented in Table 3.1. The procedures for preparing the working electrodes and composites are detailed in [16]. A fine paste of the respective polymer-graphite composites is prepared and then injected into the glass capillary tube. Copper wires are utilized for establishing the electrical connections. Before each measurement, the electrode surfaces are meticulously smoothed and cleaned with double distilled water.

3.2.3 Sample Collection

Tea samples are meticulously prepared at the Tea Research Association - Tocklai, located in Assam, India, within a carefully controlled environment. In this particular study, a total of 8 different variations of tea samples have been carefully gathered throughout the season. All of these samples undergo both voltammetry analysis and sensory evaluation by trained human tasters in order to validate their quality. The human taster assessments of the tea samples utilized in this research are documented in Table 3.2. The overall score provided represents the mean value of the quality ratings attributed to the aroma, flavor, and visual characteristics of the tea, which are of paramount importance in the tea industry. This aggregated score plays a crucial role in determining the pricing strategy of the tea product, as well as its overall commercial significance.

Table 3.1. *Chemical composition ratio of electrodes*

Working electrode	Composition ratio
WE1	PAM – Graphite = 10:90
WE2	PAM – Graphite = 20:80
WE3	PAM– Graphite = 30:70
WE4	PANI – Graphite = 10:90
WE5	PANI– Graphite = 20:80
WE6	PANI – Graphite = 30:70
WE7	PPY – Graphite = 10:90
WE8	PPY– Graphite = 20:80
WE9	PPY– Graphite = 30:70

Table 3.2. *Tea sample under test*

Samples name	Scores (1-10)				
	Leaf quality	Infusion	Aroma	Strength	Overall score
CTC TB 3213 DUST-291013	5	6.5	7	7.5	6.5
CTC TB 3513 BP-301013	7	6.5	7	7.5	7
CTC TB 3513 PF-301013	6	5	6	7	6
M+1 CUT CTC TB 3113 DUST- 291013	7.5	8	8	8.5	8
M+1CUT CTC TB 4013 PF-081113	7.5	7.5	6.5	7.5	7.5
MICRO TB 3313 BP-301013	5	5	4.5	5.5	5
MICRO TB 3313 DM-301013	9.5	9.5	9.5	9.5	9.5
MICRO TB 3313 PF-301013	8.5	9	9	9.5	9

3.2.4 Preparation of Black Tea Liquor

Tea samples are prepared utilizing conventional techniques as mentioned in reference [17]. Small amount, 0.2g of the black tea leaves are introduced into 30 ml of heated water within a sealed thermos flask. Subsequent to a 5-minute steeping period, the infusion is subsequently sieved. 20 ml of the tea infusion is allocated for the execution of the analysis.

3.2.5 Measurement Set-up

Electrochemical measurements are conducted utilizing a three-electrode configuration employing the Autolab Potentiostat/Galvanostat 101 from the Netherlands. Illustrated in Figure 3.2, this setup consists of an Ag/AgCl reference electrode, a platinum counter electrode, and a working electrode. All measurements undergo cyclic voltammetric (CV) analysis through the application of a triangular voltage ranging from -0.04 V to 0.8 V. The experimental procedures are carried out at ambient temperature (25°C ±2°C).

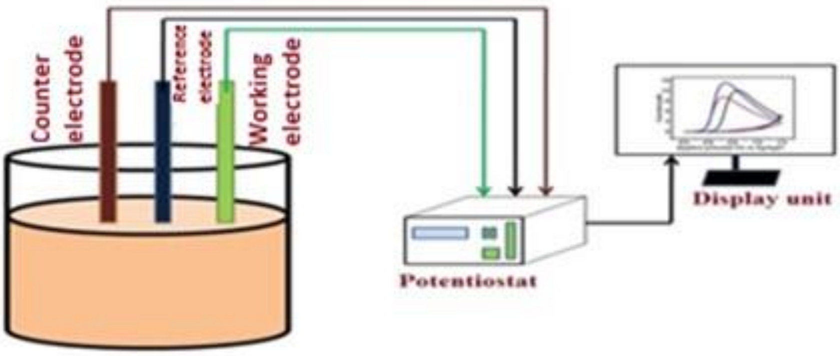


Fig. 3.2. Electrochemical measurement setup

3.3 Data Analysis

Ten replicated CV measurements were conducted using each of the synthesized electrodes. The dimensions of the raw data matrix obtained from each individual working electrode were [698×80]. Dimension reduction, feature transformation and extraction of raw data, are the fundamental assignments of a pattern recognition system. Feature extraction technique reduces computational complexity of the system. A set of feature extraction and transformation technique has been explored to address the problem. Data points of different classes are predicted by the classification technique. Five types of classification algorithms are deployed for the multi-class data analysis. Section 2.2, along with its subsection of chapter 2, provides the concise overview of the all feature transformation techniques and classification algorithm employed in the present work.

3.4 Result and Discussion

The optimization process of the voltammetric electrode array in black tea using a computational intelligence approach is thoroughly discussed within this section. Each tea sample undergoes ten repeated CV readings, leading to an [698×80] order matrix for the eight tea samples on each electrode. The utilization of MATLAB 2016a is noted in this study for computational purposes. Various techniques for feature extraction and classification algorithms have been applied for the analysis of data. Through experimentation, it has been observed that the initial fifteen arranged feature vectors of the transformed data yield the highest correct classification accuracy rate, consequently eliminating redundant data. The assessment of liquid tea samples is conducted separately using a three-electrode system, followed by a detailed analysis of electrode responses. By subjecting working electrode WE1 to a triangular voltage ranging from -0.04 V to 0.8V, the CVs for eight black tea samples are acquired, as illustrated in Fig.3.3. An optimized scan rate of 0.05 V/s is determined for the experiment. The data sets acquired before and after the optimization process are (698×9)×80 and (698×2)×80, respectively, and are considered for further analysis. The effectiveness of these classifiers is then showcased through the 10-fold cross-validation technique.

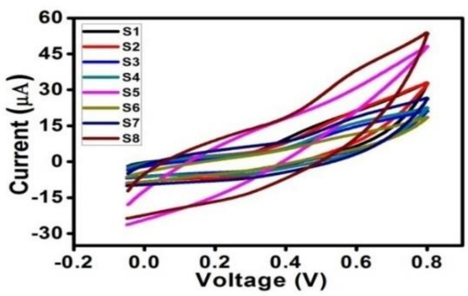


Fig.3.3. Cyclic voltammogram (CV) of WE1

3.4.1 Performance Analysis

The performances of nine operational electrodes are assessed based on the accuracy rates for correct classification, as illustrated in Table 3.3. The electrodes utilized for each classifier and feature extraction methods are organized in a descending manner according to their respective accuracy rates. On the right side of the electrode abbreviation WE, the accuracy rate for correct classification is displayed. The presentation of the average classification accuracy rate is provided in Table 3.4.

3.4.2 Electrode Polling

A methodology for polling has been implemented to enhance the optimization of the electrode array. Each electrode's value can be determined based on its ranking or placement within the array. Within Table 3.3, the initial five top-ranked electrodes are designated a value of "1," while the remaining electrodes are designated "0." Consequently, the total polling score for each electrode amounts to "20," due to the utilization of four distinct feature extraction methods and five diverse classifiers for optimizing the voltammetric electrodes. The polling scores for the working electrodes, which involve the use of five classifiers and various feature extraction techniques, are displayed in Table 3.5. The highest rank or position value assigned to each working electrode is limited to "5." The collective polling scores for the nine working electrodes are presented in Table 3.6. The final scores for each working electrode are calculated by summing the individual scores obtained from Table 3.5. The highest achievable score for each working electrode in Table 3.6 is considered up to "20." Upon examination of Table 3.6, it is evident that WE1 and WE8 exhibit the highest polling scores, followed by WE2, WE4, WE5, and WE7, respectively. Consequently, the two most effective working electrodes within the array for evaluating the overall quality of black tea are identified as WE1 and WE8.

Table 3.3. *Percentage of correct classification accuracy rate*

Feature	Classifiers				
	Cubic SVM	QDA	Medium Decision Tree	Ensemble Bagging Tree	KNN
PCA	**WE1=96.3**	**WE8=95.0**	**WE8=88.8**	**WE8=95.0**	**WE1=95.0**
	WE5=93.8	WE1=86.3	WE1=80.0	WE5=91.5	WE2=93.8
	WE8=92.5	WE5=86.3	WE6=73.8	WE1=91.3	WE8=91.3
	WE2=92.5	WE6=81.3	WE9=67.5	WE2=87.5	WE5=91.3
	WE6=91.3	WE3=78.8	WE5=65.0	WE9=85.0	WE3=91.3
	WE9=90.0	WE9=75.0	WE3=61.3	WE6=83.8	WE6=90.0
	WE3=87.5	WE2=72.5	WE4=58.8	WE3=82.5	WE9=90.0
	WE4=78.8	WE4=71.3	WE2=53.8	WE4=80.0	WE4=83.8
	WE7=70.0	WE7=56.3	WE7=48.8	WE7=63.7	WE7=73.8
SVD	**WE8=95.0**	**WE1=96.3**	**WE7=95.0**	**WE1=97.5**	**WE1=95.0**
	WE2=92.5	WE5=95.0	WE1=93.8	WE7=95.0	WE7=95.0
	WE1=91.3	WE2=92.5	WE2=93.8	WE2=93.8	WE2=93.8
	WE9=91.3	WE4=92.5	WE4=92.5	WE4=93.8	WE4=92.5
	WE4=91.3	WE9=92.5	WE8=91.3	WE8=91.3	WE8=92.5
	WE3=88.8	WE8=91.3	WE5=90.0	WE5=90.0	WE5=91.3
	WE7=87.5	WE7=91.3	WE9=88.8	WE9=88.8	WE6=90.0
	WE5=86.3	WE3=90.0	WE3=86.3	WE3=86.3	WE9=90.0
	WE6=86.3	WE6=87.5	WE6=86.3	WE6=86.3	WE3=88.8

Feature	Classifiers				
	Cubic SVM	QDA	Medium Decision Tree	Ensemble Bagging Tree	KNN
DCT	**WE7=97.5**	**WE5=95.0**	**WE6=96.3**	**WE7=97.5**	**WE7=98.8**
	WE2=92.5	WE3=95.0	WE7=95	WE2=96.3	WE2=93.8
	WE4=92.5	WE7=92.5	WE2=92.5	WE1=95	WE6=93.8
	WE5=91.3	WE4=91.3	WE9=91.3	WE6=93.8	WE1=91.3
	WE8=91.3	WE6=91.3	WE3=91.3	WE8=93.8	WE8=91.3
	WE1=90.0	WE8=90.0	WE1=90.0	WE4=92.5	WE4=90.0
	WE6=90.0	WE1=90.0	WE8=88.8	WE3=91.3	WE3=88.8
	WE3=88.8	WE2=88.8	WE5=87.5	WE5=90	WE5=88.8
	WE9=86.3	WE9=88.0	WE4=87.5	WE9=88.8	WE9=87.5
ICA	**WE8=90.0**	**WE9=96.3**	**WE5=96.3**	**WE6=100**	**WE2=92.5**
	WE1=87.5	WE7=96.3	WE7=96.3	WE4=100	WE1=91.3
	WE2=86.3	WE6=95.0	WE8=87.5	WE3=98.8	WE8=86.3
	WE3=83.8	WE5=95.0	WE1=85.0	WE1=97.5	WE4=85.0
	WE4=82.5	WE8=92.5	WE3=85.0	WE2=97.5	WE3=85.0
	WE9=81.3	WE4=85.0	WE9=85.0	WE9=96.3	WE7=85.0
	WE7=80.0	WE1=85.0	WE2=83.8	WE5=96.3	WE9=82.5
	WE6=75.0	WE2=65.0	WE6=78.8	WE7=96.3	WE6=81.3
	WE5=72.5	WE3=53.8	WE4=71.3	WE8=91.3	WE5=72.5

Table 3.4. *Percentage of average classification accuracy*

Working Electrodes	Accuracy (%)
WE1	91.27
WE2	87.77
WE3	85.22
WE4	85.64
WE5	88.28
WE6	87.59
WE7	85.51
WE8	91.34
WE9	87.11

Table 3.5. *Electrode polling score employing five classifiers and each feature extraction technique*

WE / Feature	1	2	3	4	5	6	7	8	9
PCA	5	3	2	0	5	3	0	5	2
SVD	5	5	0	5	1	0	3	4	2
DCT	2	4	2	2	2	4	5	3	1
ICA	4	3	4	3	2	2	2	4	1

Table 3.6. *Total electrode polling scores employing four feature extraction and five classifiers technique*

Name of working electrodes	Total score of working electrodes
WE1,WE8	16
WE2	15
WE4,WE5,WE7	10
WE6	9
WE3	8
WE9	6

3.4.3. Data Clustering Using PCA

Two dimensional PCA plot of before and after optimization of electrode array are illustrated in Fig.3.4. and Fig. 3.5. Analysis of the plots reveals the formation of distinct clusters representing eight variants of black tea samples. The cumulative contribution of the first two principal components (PCs) in Fig.3.4 accounts for 98.44% of the total variance, with PC1 and PC2 explaining 95.766% and 2.68% of the variance, respectively. Similarly, in Figure Fig.3.5, the first two PC loadings contribute to 99.22% of the total variance in samples, where PC1 and PC2 represent 98.04% and 1.17% of the variance, respectively.

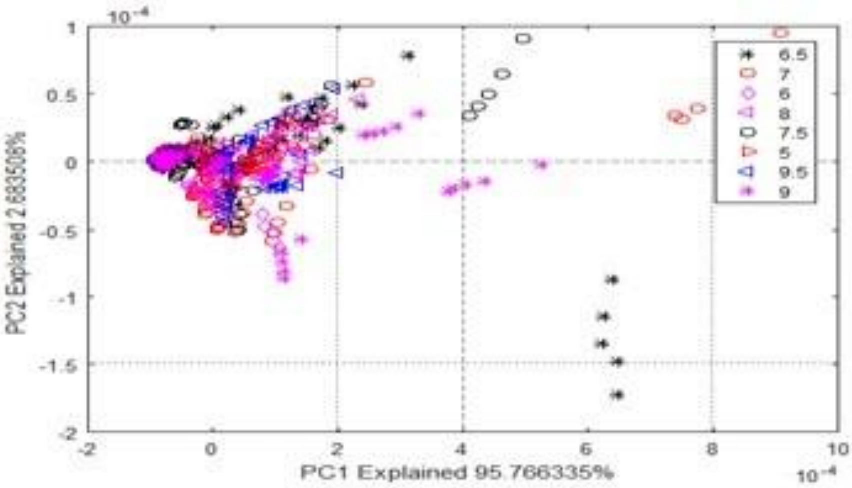


Fig.3.4. PCA score plot of eight varieties of tea samples, before optimization of working electrodes

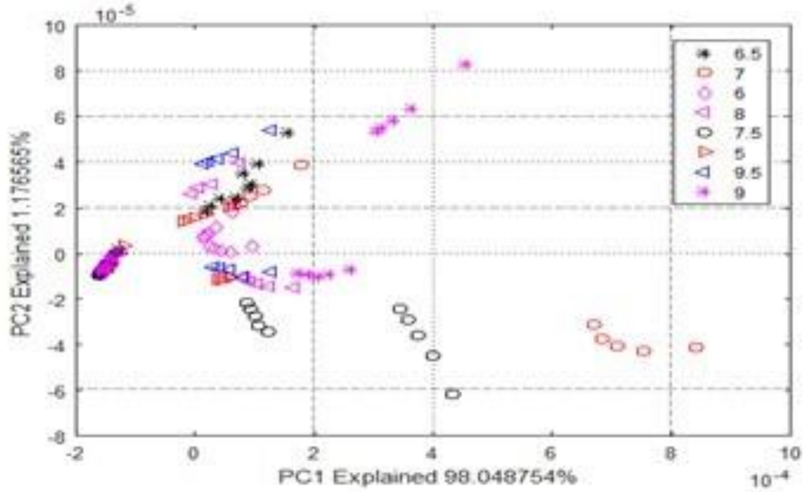


Fig.3.5. PCA score plot of eight tea samples, after optimization of working electrodes

3.5 Conclusion

The present study has focused on development of nine polymer graphite composite electrodes that were produced utilizing monomers such as acrylamide, aniline, and pyrrole, while varying their concentrations from 10% to 30% in relation to graphite. These synthesized voltammetric electrodes exhibit the capability to effectively differentiate the overall quality of black tea samples. The optimization of the electrode array within a voltammetric measurement system stands as a crucial phase to ensure the economic stability and feasibility of the system. In the present treatise, the optimization of the electrode array is accomplished through the application of machine learning and artificial intelligence methodology. The dataset derived from the cyclic voltammetry analysis is intelligently manipulated utilizing four distinct feature extraction techniques namely PCA, DCT, SVD, ICA, and five classification methods including SVM, KNN, ensemble, decision tree, and discriminant analysis. Among the electrodes, WE1 and WE8 demonstrate an average correct classification accuracy rate exceeding 91%. A polling strategy has been implemented for the optimization of the electrode array, assigning individual scores out of 20 to each electrode. This process facilitates the identification of the top two voltammetric electrodes, WE1 and WE8, for the comprehensive qualitative assessment of black tea samples. Consequently, the removal of redundant working electrodes is anticipated to enhance the system's speed of response, reproducibility, effective cost, and sensitivity. However, the data from the optimized sensor array remains large and contains redundant information, which could affect the overall performance of the measurement system. Therefore, data optimization, which involves refining and improving data for enhanced performance, accuracy, and efficiency, is essential for machine learning, data analysis, and computational systems.

The sensor array developed and used here for the tea quality estimation are non-specific sensors. The next chapter introduces a technique to improve prediction quality through feature optimization on target-specific MIP sensor data.

References

[1] A. Ghosh, B. Tudu, P. Tamuly, N. Bhattacharyya, and R. Bandyopadhyay, "Prediction of theaflavin and thearubigin content in black tea using a voltammetric electronic tongue," Chemometrics and Intelligent Laboratory Systems, vol. 116, pp. 57-66, 2012.

[2] R. B. Roy, B. Tudu, L. Shaw, A. Jana, N. Bhattacharyya, and R. Bandyopadhyay, "Instrumental testing of tea by combining the responses of electronic nose and tongue," Journal of food engineering, vol. 110, pp. 356-363, 2012.

[3] A. Ghosh, P. Tamuly, N. Bhattacharyya, B. Tudu, N. Gogoi, and R. Bandyopadhyay, "Estimation of theaflavin content in black tea using electronic tongue," Journal of food engineering, vol. 110, pp. 71-79, 2012.

[4] A. Ghosh, T. Chatterjee, P. K. Borah, D. Sing, B. Tudu, P. Tamuly, et al., "Multi-frequency large amplitude pulse voltammetric electronic tongue to assess key compounds imparting the taste of briskness to finished black tea liquor," in Proceedings of the 2nd International Conference on Perception and Machine Intelligence, 2015, pp. 252-257.

[5] W. d. J. R. Santos, M. Santhiago, I. V. P. Yoshida, and L. T. Kubota, "Electrochemical sensor based on imprinted sol–gel and nanomaterial for determination of caffeine," Sensors and Actuators B: Chemical, vol. 166, pp. 739-745, 2012.

[6] J. Manjunatha and C. Raril, "Cyclic voltammetric investigation of caffeine at methyl orange modified carbon paste electrode," Biomed. J. Sci. Tech. Res., vol. 9, pp. 1-6, 2018.

[7] M. Palit, B. Tudu, P. K. Dutta, A. Dutta, A. Jana, J. K. Roy, et al., "Classification of black tea taste and correlation with tea taster's mark using voltammetric electronic tongue," IEEE Transactions on Instrumentation and Measurement, vol. 59, pp. 2230-2239, 2009.

[8] P. Saha, S. Ghorai, B. Tudu, R. Bandyopadhyay, and N. Bhattacharyya, "A novel technique of black tea quality prediction using electronic tongue signals," IEEE Transactions on Instrumentation and Measurement, vol. 63, pp. 2472-2479, 2014.

[9] P. Ivarsson, C. Krantz-Rülcker, F. Winquist, and I. Lundström, "A voltammetric electronic tongue," Chemical Senses, vol. 30, pp. i258-i259, 2005.

[10] P. Saha, S. Ghorai, B. Tudu, R. Bandyopadhyay, and N. Bhattacharyya, "Optimization of sensor array in electronic nose by combinational feature selection method," Sensing Technology: Current Status and Future Trends II, pp. 189-205, 2014.

[11] A. K. Bag, B. Tudu, J. Roy, N. Bhattacharyya, and R. Bandyopadhyay, "Optimization of sensor array in electronic nose: A rough set-based approach," IEEE Sensors Journal, vol. 11, pp. 3001-3008, 2011.

[12] Q. Chen, J. Zhao, Z. Chen, H. Lin, and D.-A. Zhao, "Discrimination of green tea quality using the electronic nose technique and the human panel test, comparison of linear and nonlinear classification tools," Sensors and Actuators B: Chemical, vol. 159, pp. 294-300, 2011.

[13] H. Maamor, F. Rashid, N. Zakaria, A. Zakaria, L. Kamarudin, M. Jaafar, et al., "Bio-inspired taste assessment of pure and adulterated honey using multi-sensing technique," in 2014 2nd International Conference on Electronic Design (ICED), 2014, pp. 270-274.

[14] S. Qiu, J. Wang, and L. Gao, "Discrimination and characterization of strawberry juice based on electronic nose and tongue: Comparison of different juice processing approaches by LDA, PLSR, RF, and SVM," Journal of agricultural and food chemistry, vol. 62, pp. 6426-6434, 2014.

[15] K. Wakabayashi, C. Pierre, D. A. Dikin, R. S. Ruoff, T. Ramanathan, L. C. Brinson, et al., "Polymer−graphite nanocomposites: effective dispersion and major property enhancement via solid-state shear pulverization," Macromolecules, vol. 41, pp. 1905-1908, 2008.

[16] T. N. Chatterjee, R. B. Roy, B. Tudu, S. Biswas, R. Bandyopadhyay, P. Pramanik, et al., "Discrimination of black tea grades by means of cyclic voltammetry using polyacrylamide/exfoliated graphite composite electrode," in 2016 2nd International Conference on Control, Instrumentation, Energy & Communication (CIEC), 2016, pp. 60-63.

[17] T. N. Chatterjee, R. B. Roy, B. Tudu, P. Pramanik, H. Deka, P. Tamuly, et al., "Detection of theaflavins in black tea using a molecular imprinted polyacrylamide-graphite nanocomposite electrode," Sensors and Actuators B: Chemical, vol. 246, pp. 840-847, 2017.

CHAPTER 4

PREDICTION OF GREEN TEA QUALITY USING FEATURE OPTIMIZATION TECHNIQUE AND MIP VOLTAMMETRIC ELECTRODES

This chapter explores a machine learning model designed to enhance the overall prediction accuracy of a three-electrode electrochemical system using optimized dataset. Two specifically targeted molecularly imprinted polymers (MIPs) electrodes, identified as Q-IPG and CuO NPs, have been synthesized for the detection of epicatechin (EC) and gallic acid (GAL) in green tea. The sensor response data were processed utilizing the DCT technique, and three metaheuristic optimization methodologies were implemented to enhance the data optimization process. Two widely recognized regrssion algorithms have been employed for the evaluation of prediction accuracy.

LIST OF SECTION

Contents of this chapter are based on following publication:

S.Acharya ,D. Das , S.Nag , S. Mukherjee , AK. Hazarika , S. Sabhapondit , B.Tudu ,R. Bandyopadhyay , R.B.Roy ,“Optimization techniques for a voltammetric signal to predict green tea quality parameters using MIP electrode”, *IEEE SensorsJournal*,OnlineISSN:1558-1748,2023,doi:10.1109/JSEN.2023.3297140.

Chapter 4

Prediction of Green Tea Quality Using Feature Optimization technique and MIP Voltammetric Electrodes

4.1 Introduction

Green tea contains a variety of health-promoting compounds, including gallic acid (GAL), catechin, epigallocatechin gallate, caffeine, and epicatechin (EC). GAL, as the most prevalent phenolic compound in tea, plays a significant role in various health benefits such as antibacterial, anti-inflammatory, and anti-cancer effects [1-4]. Different methods traditionally used to measure GAL levels in food and beverages consist of reversed-phase high-performance liquid chromatography [5], flow injection analysis [6], chemiluminescence [7], thin-layer chromatography [8], and chronoamperometry [9]. The utilization of electrochemical techniques for GAL detection has been suggested due to their inherent selectivity [10-12]. The concept of molecularly imprinted polymers (MIPs) involves the selective adsorption of a specific molecule within a complex mixture [13]. The application of MIPs as a specialized recognition material has gained recent attention, where target analytes are incorporated into a network of monomers and cross linkers through molecular imprinting. Following the removal of the template, a molecularly imprinted polymer matrix is left behind. Molecular recognition is a crucial aspect of the adsorption process for all compounds [14, 15]. Various MIP-based electrodes have been designed to detect compounds present in tea, such as GAL, caffeine, EC, theaflavin, and catechins [16-18]. The dataset obtained from these different MIP electrodes is large and complex, negatively affecting performance and decision-making processes. Data optimization is essential for ensuring the efficiency, accuracy, and cost-effectiveness in data-driven processes. Different metaheuristic algorithms [19-25] have been used for optimization problems. In this study, a model based on metaheuristic approaches, using genetic algorithm (GA), bat algorithm (BA), and whale optimization algorithm (WOA) was proposed to address the optimization problem effectively.In the literature [19],the optimization of process parameters in a fluidized bed drier utilized for the final drying of tea leaves has been achieved through the amalgamation of an artificial neural network (ANN) with genetic algorithms (GA). A real-time quick measuring method [20] was employed to construct the GA-SVM model by utilizing the best attributes of multiple tea samples. Through the utilization of a GA-optimized dataset, there has been an enhancement in the overall classification accuracy of the model. Another study identified the grades of Keemun black tea using a cognitive spectroscopy approach [21]. Various screening techniques, such as the successive projections algorithm (SPA), competitive adaptive re-weighted sampling (CARS), and shuffled frog leaping algorithm (SFLA), have been introduced to identify the key features of different tea samples. The primary objective of these screening techniques is to enhance the predictive accuracy of the system. Bio-inspired algorithm (BA) capitalizes on the echolocation features of microbats and offers numerous advantages including rapid classification, swift convergence through the transition from exploration to exploitation, automatic zooming, and parameter control adaptability to diverse application domains [22]. In a separate study, BA, Whale Optimization Algorithm (WOA), and Intense Weed Optimization Algorithm (IWO), all drawing inspiration from nature, were harnessed for optimization due to their distinct features and advantages [23]. WOA is a straightforward, rapid, and highly accurate algorithm that requires minimal coding. Scholars [24] developed a symmetric model based on photovoltaic power to ensure accurate performance under varying weather conditions, and the identification of optimal features was proposed through two binary WOA algorithms [25]. When dealing with highly correlated input data, discrete cosine transformation (DCT) exhibits superior energy compaction properties compared to discrete Fourier transformation (DFT). In a particular investigation [26], an artificial neural network (ANN) classifier followed by the DCT technique was used to optimize five polymer-graphite composite electrodes of the electronic tongue. In this work, the MIP-GAL and Q-IPG electrodes were fabricated following methodologies outlined in [17] and [18]. Subsequent to the fabrication of the MIP-GAL electrode, it was immersed in green tea samples and underwent DPV analysis. The resulting data was captured and processed utilizing the DCT feature transformation technique. Different data optimization techniques were employed to obtain a low-dimensional feature dataset. These optimized feature sets were then utilized in the development of prediction models using partial least square regression technique (PLSR) and principal component regression (PCR). Calibration models were established based on the results of HPLC analysis. Similar methodologies were applied to Q-IPG electrodes for the purpose of EC detection in green tea. The root mean square error of calibration (RMSEC) achieved using PLSR was 0.253 for GAL and 0.094 for EC detection. As for the RMSEC values obtained from PCR, they were 0.239 and 0.088 for GAL and EC detection, respectively. The details of the research work are presented in Fig. 4.1.The novelty in this study lies in the development of a metaheuristic-based model aimed at enhancing the overall predictive accuracy of an electrochemical system through the utilization of optimized and condensed datasets. The primary objective of this research was to enhance predictive accuracy by employing optimized reduced datasets, as opposed to raw datasets as mentioned in [17]. The outcomes demonstrate that the generalized predictive model established in this study exhibited a superior predictive accuracy of 96.24% for GAL detection compared to the 88.97% reported in [17]. To assess the validity of the model, the dataset from EC detection [18] was utilized, applying the same techniques for feature transformation and data optimization.

4.2 Material and Methods

4.2.1Chemical reagents and material

GAL, ethylene glycol dimethyl acrylate (EGDMA), graphite powder (99%), itaconic acid (IA), acrylamide (AAM), maleic acid (MA), catechin (CAT), and ascorbic acid (AA) were procured from Sigma Aldrich, India. Copper chloride (CuCl2), ethanol, cetylpyridinium chloride (C21H38NCl), sodium hydroxide (NaOH), and paraffin oil were obtained from Merck and Co, India. Benzoyl peroxide (BP) was acquired from Sisco Research Laboratories Pvt. Ltd, India. Deionized water (Resistance 18 MΩ), used for cleaning the electrodes was taken from the Millipore water purification system.

4.2.2 Synthesis of CuO NPs, GAL imprinted polymer (MIP) and non-imprinted polymer (NIP)

Synthesis of CuO NPs was carried out in the laboratory by the sol-gel method. The step-wise synthesis of the CuO NPs, MIP, and NIP is described in the work [17, 18]

4.2.3 Synthesis of quercetin imprinted electrode (Q-IPG) and non-imprinted polymer (NIP)

The synthesis of Q-IPG material was carried out in the laboratory. The detailed study and the fabrication of the electrode synthesis are elucidated in [18].

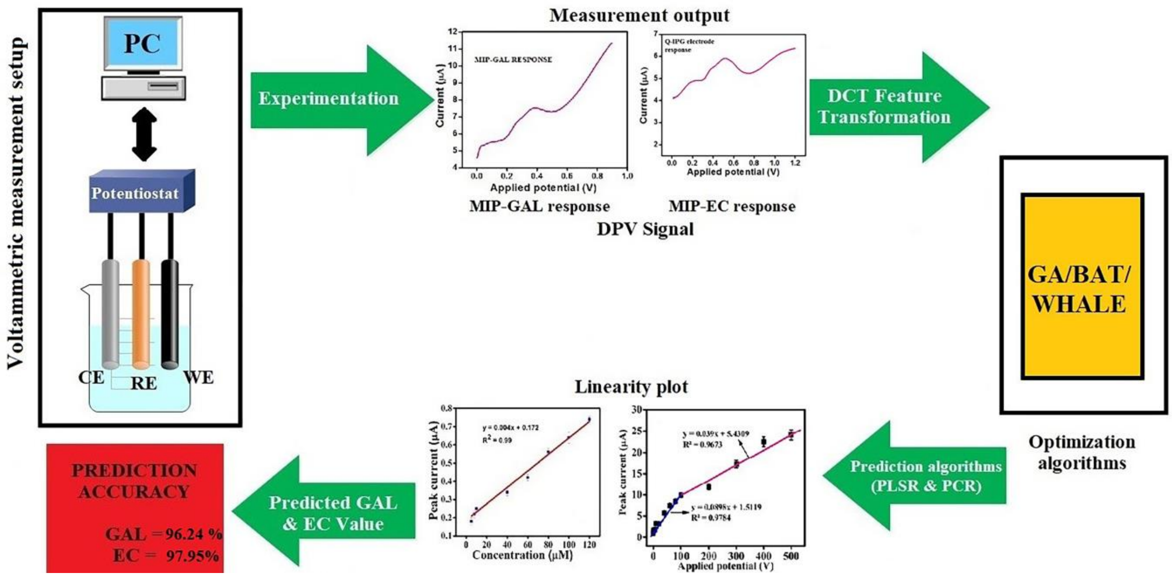


Fig.4.1. Presenting the work flow diagram of the work.

4.3 Data Analysis

The DPV responses were obtained using the MIP-GAL electrode by varying the voltage from 0 V to 1 V. The study yielded 179 data points corresponding to the current values in microamperes (μA) as the voltage varied within this range. For each tea sample, eight repetitive DPV responses were recorded, resulting in a dataset size of 179 × 8 for a single tea variant. Consequently, the size of the raw data matrix for ten samples using the MIP-CuO electrode with the GAL samples was [179 × 8 × 10]. This data was then transformed using the discrete cosine transformation (DCT) technique. A set of feature optimization techniques, namely GA, BA, and WOA, were explored. The predictive ability of the synthesized electrodes was evaluated using the partial least square regression (PLSR) and principal component regression (PCR) models

4.3. 1 Feature transformation

In this work, an orthonormal time series Discrete Cosine Transformation (DCT) technique has been applied to sensor data. This transformation uses real numbers and maintains the original vector length. The computational complexity of this method is very low. A detailed overview of this transformation is provided in subsection 2.3.2 of Chapter 2.

4.3. 2 Feature optimization

Feature optimization involves reducing the dimensionality of input variables when developing predictive models. This reduces computational costs and enhances model performance. This chapter explores three distinct nature-inspired feature optimization techniques: Genetic algorithm (GA), Binary Bat algorithm (BA), and Whale optimization algorithm (WOA) in section 2.4 of chapter 2.

4.3.3 Calibration and prediction analysis

Partial Least Squares Regression (PLSR) and Principal Component Regression (PCR) are both regression techniques used for analyzing relationships between predictors (independent variables) and a response (dependent variable). Both PLSR and PCR prove to be beneficial in situations when large numbers of predictors are their or predictors are highly correlated. The detail of these prediction models have been discussed in section 2.5 of chapter 2. Table 4.1 shows the numbers of features after implementation of optimization technique and Table 4.2 shows the prediction of GAL and EC concentration in unknown samples.

Table 4.1. *Number of features after implementation of optimization technique*.

Compounds	Feature Numbers	Number of features obtained after optimization		
		GA	**BA**	**WOA**
GAL	179	**83**	35	19
EC	239	**129**	33	14

Table 4.2. *Prediction of GAL and EC concentration of unknown samples.*

Compound	Method	LV	Calibration RMSEC	Calibration R_C^2	Validation RMSEC	Validation R_V^2	Prediction RMSEP	Prediction R_P^2
GAL	PLSR	**11**	0.253	0.93	0.02	0.92	0.019	**0.92**
	PCR	14	0.239	0.89	0.023	0.91	0.011	0.90
EC	PLSR	15	0.094	0.99	0.045	0.95	0.0386	**0.98**
	PCR	15	0.088	0.99	0.039	0.96	0.0435	**0.98**

4.4 Result and Discussion

The methodology for developing a meta-heuristic model to enhance the efficacy of the synthesized MIP electrode utilized in the classification of green tea is elucidated in this section. The model considers two synthesized MIP electrodes: MIP-GAL for GAL detection and Q-IPG electrode for EC detection in green tea as part of the case study. Specifically, two phenolic compounds found in green tea, namely GAL and EC, are identified using the voltammetric technique. The DPV profiles of the two synthesized electrodes are generated by applying specific potential windows, (0V – 1V) with a scan rate of 0.01-0.4 Vs-1 and (0V - 1.2V) with 0.025 - 0.3 Vs−1, respectively. These DPV profiles are illustrated in Fig. 4.1 (a) and Fig. 4.1 (b). The resulting matrix dimensions for the two synthesized electrodes and ten variations of green tea samples, each with eight replicates, are [179×8×10] and [239×8×10], respectively and are taken into account for further analysis using chemometrics [26, 27].

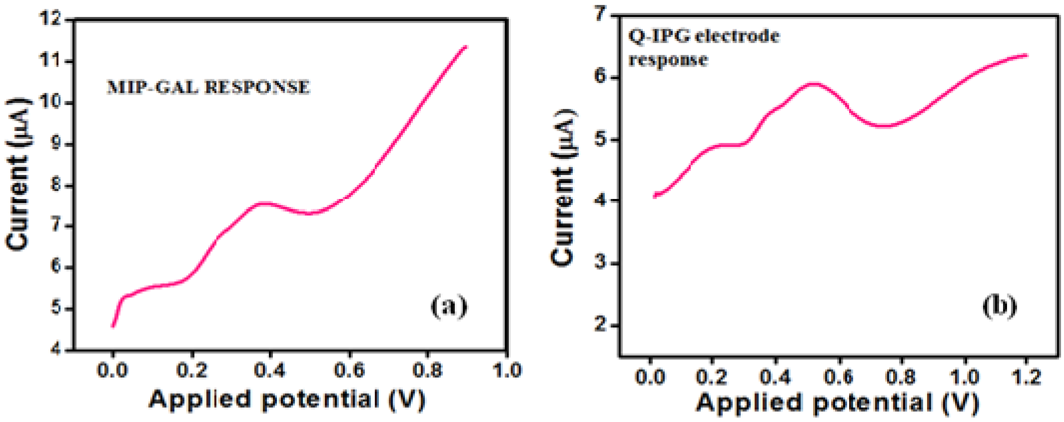


Fig.4.2. Response in green tea samples of (a) MIP-GAL electrode (b) Q-IPG electrode

4.4.1 Feature transformation and data optimization

The dataset of dimensions [179×8×10] has undergone a transformation utilizing the Discrete Cosine Transformation (DCT), a robust technique that utilizes the cosine function exclusively, as opposed to a combination of sine and cosine functions. The coefficients of the DCT are of a real nature and exhibit enhanced capability for energy compaction, particularly in scenarios where the dataset displays high levels of correlation. The process involved optimization of data and reduction of dimensions, achieved through the application of popular bio-inspired meta-heuristic algorithms such as GA, BA and WOA on the features transformed using DCT.

4.4.1.1 Data optimization using GA

GA is employed on the DCT transformed dataset of size [179×8×10]. Here we use 0.1 as a mutation rate (μ), 0.7 as a crossover percentage (pc), and 0.3 as a mutation percentage (pm) for the data optimization. The cost function is taken into account and optimized features are found from the best solution. Among 179 features only 83 optimized feature vectors are obtained using 20 courses of iteration. GA is again applied on the dataset of size [239×8×10] obtained using a Q-IPG electrode. In this case, 129 data points have been optimized among the 239 data points for the detection of EC in green tea samples.

4.4.1.2 Data optimization using BA

The BA is a data optimization algorithm that operates without supervision, specifically designed for the transformed datasets of dimensions [179×8×10] and [239×8×10]. Within the scope of this study, 20 bat instances, which have undergone optimization through iterative methods, are taken into account for the purpose of optimization. The resultant optimized feature data consists of 35 features from the set of 179 feature vectors and 33 features from the set of 239 feature vectors.

4.4.1.3 Data optimization using WOA

The WOA has been employed on the transformed dataset of dimensions [179×8×10] and [239×8×10] concurrently to identify GAL and EC in ten variations of green tea samples. A total of 100 iterations and 10 whale individuals were considered for optimal performance on the transformed dataset. Upon implementation of this technique, 19 and 14 feature vectors were fine-tuned for the identification of GAL and EC. A comparison presented in Table I illustrates the varying number of features acquired post-optimization using different meta-heuristic approaches. The table demonstrates a reduction in data size. The primary aim of this study is to construct a model that achieves the highest prediction accuracy with the least number of feature vectors; hence, GA-optimized features were favored over BA and WOA. It is noteworthy that BA and WOA exhibit superior prediction accuracy compared to prior studies for detecting GAL and EC with minimal feature vectors.

4.4.2 Calibration and prediction analysis

4.4.2.1 PLSR analysis

The empirical relevance of the synthesized electrodes has been assessed by the PLSR model in conjunction with the LOOCV technique. The DPV response and HPLC data pertaining to GAL and EC were partitioned into two distinct subsets, specifically the training and testing sets, using an 80:20 ratio. The evaluation of this model's performance has been based on the correlation coefficient (R2) and root mean square error (RMSE) between the predicted values and the empirical data. The selection of 11 latent variables (LV) for GAL and 15 for EC aimed at achieving the lowest RMSEC values of 0.253 and 0.094 as presented in Table 4.2. Consequently, this model provides predictions for GAL and EC concentration (mg/g) in green tea with an average prediction accuracy of 92.24% and 97.95%, as shown in Table 4.3.

4.4.2.2 PCR analysis

The PCR model customized the relationship between PCs of the data matrix of size [83×8×10] and [129×8×10] for GAL and EC with HPLC data. The latent variables (LV) 14 for GAL and 15 for EC have been considered for yielding the lowest RMSEC values 0.239 and 0.088. The statistical metrics for PLSR and PCR regression model have been shown in Table 4.2 which includes the calibration coefficient(R_C^2) , validation coefficient (R_V^2), and RMSEC values. The average prediction accuracy of GAL and EC has been measured to be 96.24% and 97% which have been shown in Table 4.3.

***Table4.3**. Actual and predicted GAL and EC content from the LOOCV based PLSR and PCR model.*

Compound name	The actual content(HPLC value in mg/g)	Predicted content(mg/g)						Prediction accuracy (%)					
		GA		BAT		WOA		GA		BAT		WOA	
		PLSR	PCR	PLSR	PCR	PLSR	PCR	PLSR	PCR	PLSR	PCR	PLSR	PCR
GAL	0.11	0.12	0.11	0.11	0.11	0.11	0.12	90.90	100	100	100	100	90.90
	0.08	0.08	0.08	0.08	0.08	0.08	0.08	100	100	100	100	100	100
	0.05	0.05	0.05	0.07	0.07	0.07	0.07	100	100	60	60	60	60
	0.06	0.07	0.06	0.05	0.06	0.06	0.06	83.30	100	90	100	100	100
	0.10	0.10	0.09	0.10	0.10	0.09	0.09	100	99	100	100	90	90
	0.12	0.11	0.11	0.09	0.09	0.09	0.09	91.66	91.66	80.83	75	75	79.16
	0.10	0.09	0.09	0.11	0.13	0.07	0.08	90	90	90	70	70	84
	0.13	0.12	0.13	0.12	0.12	0.13	0.13	92.30	100	90.90	92.30	100	100
	0.11	0.10	0.09	0.10	0.10	0.09	0.09	90.90	81.81	90	90.90	81.80	87.20
	0.06	0.07	0.06	0.06	0.07	0.06	0.06	83.33	100	100	83.33	100	100
Average prediction accuracy(%)								92.24	**96.24**	90.17	87.15	87.68	89.13
Standard deviation								6.22	6.32	12.40	14.50	15.04	12.71
EC	0.81	0.80	0.79	0.77	0.74	0.83	0.84	98.76	97.53	95.06	91.35	97.53	96.29
	1.10	1.06	1.02	0.93	0.94	0.86	0.88	96.36	92.72	92.72	93.63	78.18	80.00
	0.97	0.97	0.96	0.96	0.97	1.05	1.05	100	98.96	98.96	100	91.75	91.75
	0.81	0.81	0.80	0.92	0.93	0.87	0.86	100	98.76	86.41	85.18	92.59	93.82
	1.08	1.09	1.09	0.98	0.96	1.01	1.01	99.07	99.07	90.74	88.88	93.51	93.51
	1.05	1.10	1.01	1.09	1.09	1.04	1.00	95.23	96.19	96.19	96.19	99.04	95.23
	0.96	1.02	1.01	1.15	1.09	1.06	1.06	93.75	94.79	80.20	86.45	89.58	89.58
	1.21	1.22	1.22	1.17	1.18	1.08	1.06	99.17	99.17	96.69	97.52	89.25	87.60
	1.15	1.13	1.22	1.14	1.14	1.15	1.17	98.26	93.91	99.13	99.13	100	98.26
	1.90	1.88	1.88	1.83	1.81	1.77	1.79	98.94	98.94	96.31	95.26	93.15	94.21
Average prediction accuracy(%)								**97.95**	97.00	93.24	93.35	92.45	92.04
Standard deviation								2.11	2.43	5.99	5.21	6.24	5.24

4.4.3 Statistical significance of the result

The average accuracy achieved using GA is better than BA and WOA feature selection algorithm which has been shown in Table 4.3. Validation of the statistical significance of the predictive accuracy findings necessitated the execution of a Mann-Whitney U Test, a nonparametric alternative to the two-sample t-test, comparing GA against BA and GA against WOA individually. In both, the case Mann-Whitney U Test compares two populations with accurate results. The null hypothesis posits equality between the

results in both instances. The outcomes of the Mann-Whitney U test comparing GA versus BA and GA versus WOA produced p and z values of 0.034 (i.e. < 0.05), 2.1169, and 0.018 (i.e. < 0.05), 2.3575, respectively, indicating a significant disparity. Unlike BA and WOA, GA exhibited the lowest standard deviation in the PLSR and PCR models. As per Table 4.3, the optimal standard deviations were determined as 6.22 and 6.32 for GA, and 2.11 and 2.43 for EC. Consequently, it can be inferred that the outcomes derived from utilizing GA are markedly distinct and superior to those obtained from BA and WOA.

4.5 Improvement in prediction accuracy using the proposed model

The examination of existing literature demonstrates that the MIP-CuO electrode exhibited an 88.97% accuracy in predicting the presence of GAL in samples of green tea. This level of accuracy was attained through the utilization of electrodes, resulting in the acquisition of a comprehensive data set comprising 179 data points. The primary objective of the study was to improve predictive accuracy by employing a refined set of features. A comparative analysis presented in Table IV explores the findings of prior research with those of the current study. In this context, an accuracy rate of 96.24% was realized by leveraging 83 optimized data points, as opposed to the complete dataset, for GAL detection. This outcome underscores the development of a predictive model that enhances the accuracy of the electrochemical system. The model's validity was confirmed through its application to the electrode data collected during EC detection in an earlier study. Various techniques for data manipulation and enhancement were incorporated into the model. As a result, the model achieved a heightened prediction accuracy of 97.95% using 129 optimized data sets, surpassing the 94.54% accuracy obtained with 239 data sets in the previous research. Consequently, this model holds promise for delivering superior prediction accuracy with an optimized data set, thereby reducing the computational burden on the system.

***Table4.4**. Comparison of previous literature with present work.*

Compound	Number of features	Prediction accuracy (%)	Related work
GAL	179	88.97	[17]
	83	**96.24**	**Proposed work**
EC	239	94.54	[18]
	129	**97.95**	**Proposed work**

4.6 Conclusion

The present treatise is dedicated to the creation of a metaheuristic-driven model aimed at enhancing the overall predictive precision of an electrochemical system through the utilization of optimized and reduced datasets than that obtained using raw datasets. In this work, synthesized electrode responses for the detection of GAL and EC in green tea samples were transformed by the DCT technique. Then the transformed features were optimized independently using three metaheuristic optimization methods viz., GA, BA, and WOA to reduce the size of the feature dataset. The predictive performance of two synthesized electrodes, MIP-GAL and Q-IPG datasets, was assessed utilizing PLSR and PCR models. It was observed that the GA-optimized dataset demonstrated the highest prediction accuracy, achieving 96.24% for GAL detection and 97.95% for EC detection. This notable improvement in predictive accuracy was attained by employing 83 and 129 features, respectively, instead of the entire dataset. In essence, this study proposed a universal predictive model for the examination of voltammetric electrode data for qualitative analysis of tea and other relevant domains, serving as a versatile methodology. The outcomes obtained from this study suggest that the universal predictive model introduced herein exhibited higher prediction accuracy compared to the findings reported in a previous study [26]. To validate the model, the dataset from a prior study[18] was utilized, with the same feature transformation and data optimization strategies being implemented.The advancement of selective, stable, economically viable, and reusable sensing devices is highly advantageous within the food and beverage sector. The following chapter investigates the development of an innovative molecularly imprinted polymer (MIP)-tethered capacitive sensor specifically designed to effectively quantify target analytes in green tea.

References

[1] R. van Lith and G. A. Ameer, "Antioxidant polymers as biomaterial," in Oxidative stress and biomaterials, ed: Elsevier, 2016, pp. 251-296.

[2] P. Carloni, L. Tiano, L. Padella, T. Bacchetti, C. Customu, A. Kay, et al., "Antioxidant activity of white, green and black tea obtained from the same tea cultivar," Food research international, vol. 53, pp. 900-908, 2013.

[3] H. G. Valery, A. Chtaini, and B. Loura, "Voltammetric sensor based on electrodes modified by poly (vinyl alcohol)-natural clay Film, for the detection of gallic acid," Portugaliae Electrochimica Acta, vol. 37, pp. 327-333, 2019.

[4] S. Sarafraz, H.-A. Rafiee-Pour, M. Khayatkashani, and A. Ebrahimi, "Electrochemical determination of gallic acid in Camellia sinensis, Viola odorata, Commiphora mukul, and Vitex agnus-castus by MWCNTs-COOH modified CPE," Journal of Nanostructures, vol. 9, pp. 384-395, 2019.

[5] L. Wang, M. S. Halquist, and D. H. Sweet, "Simultaneous determination of gallic acid and gentisic acid in organic anion transporter expressing cells by liquid chromatography–tandem mass spectrometry," Journal of Chromatography B, vol. 937, pp. 91-96, 2013.

[6] W. Ma, D. Han, S. Gan, N. Zhang, S. Liu, T. Wu, et al., "Rapid and specific sensing of gallic acid with a photoelectrochemical platform based on polyaniline–reduced graphene oxide–TiO 2," Chemical Communications, vol. 49, pp. 7842-7844, 2013.

[7] X. Wang, J. Wang, and N. Yang, "Flow injection chemiluminescent detection of gallic acid in olive fruits," Food chemistry, vol. 105, pp. 340-345, 2007.

[8] K. Dhalwal, V. Shinde, Y. Biradar, and K. Mahadik, "Simultaneous quantification of bergenin, catechin, and gallic acid from Bergenia ciliata and Bergenia ligulata by using thin-layer chromatography," Journal of food composition and analysis, vol. 21, pp. 496-500, 2008.

[9] R. de Queiroz Ferreira and L. A. Avaca, "Electrochemical determination of the antioxidant capacity: the ceric reducing/antioxidant capacity (CRAC) assay," Electroanalysis: An International Journal Devoted to Fundamental and Practical Aspects of Electroanalysis, vol. 20, pp. 1323-1329, 2008.

[10] S. Nag, S. Pradhan, H. Naskar, R. B. Roy, B. Tudu, P. Pramanik, et al., "A simple nano cerium oxide modified graphite electrode for electrochemical detection of formaldehyde in mushroom," IEEE Sensors Journal, vol. 21, pp. 12019-12026, 2021.

[11] S. Nag, S. Pradhan, D. Das, B. Tudu, R. Bandyopadhyay, and R. B. Roy, "Fabrication of a molecular imprinted polyacrylonitrile engraved graphite electrode for detection of formalin in food extracts," IEEE Sensors Journal, vol. 22, pp. 42-49, 2021.

[12] S. Nag, D. Das, H. Naskar, B. Tudu, R. Bandyopadhyay, and R. B. Roy, "Detection of metanil yellow adulteration in turmeric powder using nano nickel cobalt oxide modified graphite electrode," IEEE Sensors Journal, vol. 22, pp. 12515-12521, 2022.

[13] M. Kahl and T. D. Golden, "Electrochemical determination of phenolic acids at a Zn/Al layered double hydroxide film modified glassy carbon electrode," Electroanalysis, vol. 26, pp. 1664-1670, 2014.

[14] N. Leibl, K. Haupt, C. Gonzato, and L. Duma, "Molecularly imprinted polymers for chemical sensing: A tutorial review," Chemosensors, vol. 9, p. 123, 2021.

[15] J. W. Lowdon, H. Diliën, P. Singla, M. Peeters, T. J. Cleij, B. van Grinsven, et al., "MIPs for commercial application in low-cost sensors and assays–An overview of the current status quo," Sensors and Actuators B: Chemical, vol. 325, p. 128973, 2020.

[16] D. Das, T. N. Chatterjee, R. B. Roy, B. Tudu, A. K. Hazarika, S. Sabhapondit, et al., "Titanium oxide nanocubes embedded molecularly imprinted polymer-based electrode for selective detection of caffeine in green tea," IEEE Sensors Journal, vol. 20, pp. 6240-6247, 2020.

[17] D. Das, D. Biswas, A. K. Hazarika, S. Sabhapondit, R. B. Roy, B. Tudu, et al., "CuO nanoparticles decorated MIP-based electrode for sensitive determination of gallic acid in green tea," IEEE Sensors Journal, vol. 21, pp. 5687-5694, 2020.

[18] D. Das, S. Nag, S. De, A. K. Hazarika, S. Sabhapondit, B. Tudu, et al., "Electrochemical detection of epicatechin in green tea using quercetin-imprinted polymer graphite electrode," IEEE Sensors Journal, vol. 21, pp. 26526-26533, 2021.

[19] M. S. H. Kalathingal, S. Basak, and J. Mitra, "Artificial neural network modeling and genetic algorithm optimization of process parameters in fluidized bed drying of green tea leaves," Journal of Food Process Engineering, vol. 43, p. e13128, 2020.

[20] M. Yao, G. Fu, T. Chen, M. Liu, J. Xu, H. Zhou, et al., "A modified genetic algorithm optimized SVM for rapid classification of tea leaves using laser-induced breakdown spectroscopy," Journal of Analytical Atomic Spectrometry, vol. 36, pp. 361-367, 2021.

[21] G. Ren, Y. Wang, J. Ning, and Z. Zhang, "Highly identification of keemun black tea rank based on cognitive spectroscopy: Near infrared spectroscopy combined with feature variable selection," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, vol. 230, p. 118079, 2020.

[22] X.-S. Yang and X. He, "Bat algorithm: literature review and applications," International Journal of Bio-inspired computation, vol. 5, pp. 141-149, 2013.

[23] M. Deepak and R. Rustum, "Review of latest advances in nature-inspired algorithms for optimization of activated sludge processes," Processes, vol. 11, p. 77, 2022.

[24] Y.-W. Liu, H. Feng, H.-Y. Li, and L.-L. Li, "An improved whale algorithm for support vector machine prediction of photovoltaic power generation," Symmetry, vol. 13, p. 212, 2021.

[25] M. Mafarja and S. Mirjalili, "Whale optimization approaches for wrapper feature selection," Applied Soft Computing, vol. 62, pp. 441-453, 2018.

[26] S. Acharya, T. N. Chatterjee, S. Mukherjee, D. Das, R. B. Roy, B. Tudu, et al., "Selection of optimum number of sensors of an electronic tongue for efficient classification of black tea: A combinatorial approach based on discrete cosine transform and artificial neural network," in 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), 2018, pp. 108-111.

[27] O. Valencia, M. Ortiz, S. Ruiz, M. Sánchez, and L. Sarabia, "Simultaneous class-modelling in chemometrics: A generalization of Partial Least Squares class modelling for more than two classes by using error correcting output code matrices," Chemometrics and Intelligent Laboratory Systems, vol. 227, p. 104614, 2022.

CHAPTER 5

GREEN TEA QUALITY ANALYSIS BY DEVELOPING A NOVEL CAPACITIVE SENSOR AND EXPLORING CLUSTERING APPROACH

This chapter presents the development and performance evaluation of a stable, selective, reusable, and cost-effective capacitive sensor, designed using molecular imprinted polymer (MIP) technology for detecting epicatechin (EC) in green tea. The sensor was fabricated by depositing a molecularly imprinted polydopamine-polyethylene glycol (PDA-PEG) composite onto a copper-clad FR-4 plate. Characterization of the synthesized materials was carried out using scanning electron microscopy (SEM) and X-ray diffraction (XRD) techniques. Sensor capacitance (C_P) response was analyzed as a function of frequency (MIP-EC@C). Its performance was further assessed through the application of two popular clustering algorithms: K-Means and Agglomerative clustering.

LIST OF SECTION

Contents of this chapter are based on following publication:

- **S.Acharya** ,S.Nag, D. Bandyopadhyay, D. Das, A. Mandal , RB Roy. "A Molecular Imprinted Polymer Tethered Capacitive Sensor for Epicatechin Detection in Green Tea". *IEEE Sensors Journal*. 2023 Dec 28.
- **S.Acharya** ,S.Nag, D. Bandyopadhyay, S. Mukherjee , D. Das , RB. Roy . "Unveiling the capacitive sensor performance developed for Epicatechin detection in Green tea: A Clustering Approach." In2024 IEEE 3rd International Conference on Control, Instrumentation, Energy &Communication (CIEC) 2024 Jan 25 (pp. 175-179). IEEE.

Chapter 5

Green Tea Quality Analysis By Developing A Novel Capacitive Sensor and Exploring Clustering Approach

5.1 Introduction

This chapter explores the creation of a stable, selective, reusable, and economical capacitive sensor using molecular imprinted polymer (MIP) technology for the detection of epicatechin (EC) in green tea. The sensor performance is assessed using two widely-used clustering algorithms: K-Means and Agglomerative Clustering. Epicatechin (EC) is a naturally occurring bioactive compound with potent antioxidant properties and is one of the most abundant flavonoids found in green and black tea.

The quality of green tea is largely influenced by its EC content, which provides several benefits, including the removal of heavy metal toxicity, reduction of bacterial inflammation, enhancement of disease resistance, and scavenging of free radicals [1-3]. The structure of EC ($C_{15}H_{14}O_6$), shown in Fig. 5.1, includes a hydroxyl group at C-3 and a dihydropyran ring.

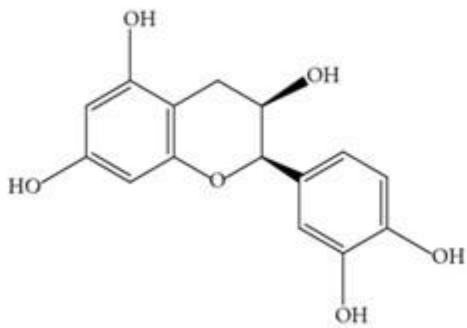


Fig.5.1. Chemical structure of EC.

A comprehensive literature survey reveals various instrumental methods viz. Liquid chromatography [4], infrared (IR) spectroscopy, Raman spectroscopy [5], nuclear magnetic resonance spectroscopy (NMR)[6], thin-layer chromatography (TLC) [7],spectrophotometry [8], and electrophoresis were employed for detection of EC in past few decades. However, these techniques are not only expensive and complex but also cumbersome and require highly skilled professionals. In contrast, electrochemical detection methods are comparatively easy, simple, low cost and fast responding [9-13]. In some of the literature, electrochemical detection of EC had been reported using different voltammetry techniques viz. adsorption stripping voltammetry, cyclic voltammetry (CV), and square wave voltammetry developing various polymer modified electrodes [13-15]. Due to their inherently high selectivity and reusability, MIP technology has emerged in recent years [16]. In [1] and [2] two electrodes were developed for the detection of EC using MIP technology. In the first investigation, the target molecule itself was used as a template, and then in the second work, dummy template quercetin was used for imprinting to achieve the low-cost electrode. From the literature survey, it can be found that MIP-based electrochemical detection techniques had been already implemented for the detection of theaflavin, catechins (CAT), methyl-xanthines, and phenolic acid in tea [17-21]**.** However, these methods necessitate the current flow through the electrolyte and sophisticated computer-based systems. The electrochemical sensors are still limited in scope viz., short shelf life, and cross-sensitivity. Moreover, electrode material preparation and measurement of electrochemical systems are somewhat tedious and time-consuming. Thus, it is imperative to develop an economic, susceptible, and rugged sensor for the qualitative measurement of food and beverages.

On the contrary, capacitive sensors are characterized by high stability, sensitivity, reproducibility, and low power consumption. A capacitive sensor coated with reduced graphene oxide (RGO) and polymethylmethacrylate (PMMA) on copper has been proposed for the detection of formaldehyde in beverages like milk and water [22]. The phase angle of this sensor changes with frequency when immersed in pure milk and milk adulterated with formaldehyde. Conversely, impedance and phase exhibit frequency-dependent variations in water. In a separate study [23], researchers introduced a parallel plate capacitive sensor for detecting 2-Furfuraldehyde (2-FAL) in transformer oil. This sensor was created by applying polydimethylsiloxane (PDMS) on the copper electrode. The coating of MIP on the sensor provides specific recognition sites for target molecules, facilitating adsorption and capacitance changes in the sensing layer. The sensor sensitivity was evaluated by testing different concentrations of 2-FAL in transformer oil. In another investigation [24], a MIP-based capacitive sensor was developed on a gold-coated silicon electrode for detecting sulphanilamide (SN) in water and milk.

In recent times, there has been a growing trend towards the utilization of bio-mimetic materials in the development of various engineering devices. The widespread adoption of dopamine (DA), 3,4-dihydroxyphenylalanine (DOPA) derivatives, and macromolecular forms like polydopamine (PDA) as surface coatings on diverse substrates has been observed [25-29]. PDA, derived from the oxidation of DA, exhibits remarkable adhesive properties due to the presence of ligands (phenyl, catechol, and amine) that engage with substrates via hydrogen bonds, hydrophobic interactions, and electrostatic forces. The efficacy of PDA as a surface modifier has been validated in numerous studies [30-34]. To enhance the performance and modify the wettability of PDAs, some researchers have employed PDA co-deposition with polymers such as polyethylene glycol (PEG) [29].

In this proposed work, the investigation focuses on the characteristics of sensitivity and selectivity of the produced capacitive sensor, utilizing MIP technology. According to the existing literature, research on EC selective MIP-based capacitive sensors for green tea is limited. The current study introduces a novel MIP-tethered capacitive sensor designed to be stable and specific for detecting EC in green tea. The sensing material was created using a composite of molecularly imprinted polydopamine-polyethylene glycol (PDA-PEG), and the sensor's linear performance was assessed within the 5 to 20 ppm range. This work investigates the variations in electrical properties, specifically capacitance, impedance, and phase angle, with frequency for different concentrations of EC in a test solution. Given the focus on capacitive sensors in this study, changes in capacitance (C_p) were chosen for detailed analysis. A comparative assessment of capacitance (C_p) variation with frequency between molecularly imprinted (MIP) and non-imprinted (NIP) sensors is presented. The sensor's stability was monitored for over a hundred days at intervals exceeding thirty days. Furthermore, the paper delves into the relationships between sensor impedance, phase angle, and varying frequencies. The proposed method stands out due to the superior selectivity, reproducibility, and cost-efficiency of the synthesized sensor for recognizing EC in green tea. Additionally, the performance of fabricated sensor is assessed using two popular clustering methods K-Means and Agglomerative clustering algorithms. These two algorithms were employed to distinguish between ten varieties of green tea based on their varying EC content. The extensive investigation revealed maximum Silhouette scores of 0.8568 and 0.8465, along with minimum Davies-Bouldin scores of 0.2016 and 0.2013, respectively, when the optimal cluster value was set at 10. These findings highlight the sensor's capability in differentiating green tea samples with distinct EC content, as demonstrated by the peak Silhouette and minimum Davies-Bouldin scores achieved at the specific K value for both algorithms. Notably, the sensor can be directly immersed in any infusion without the need for pre-treatment before capacitive measurements. Predictive capabilities of the fabricated capacitive sensor were explored using partial least squares (PLS) and principal component (PC) regression models. The newly developed sensor is capable for evaluating EC levels in green tea.

5.2 Experimental

5.2.1 Proposed EC sensor

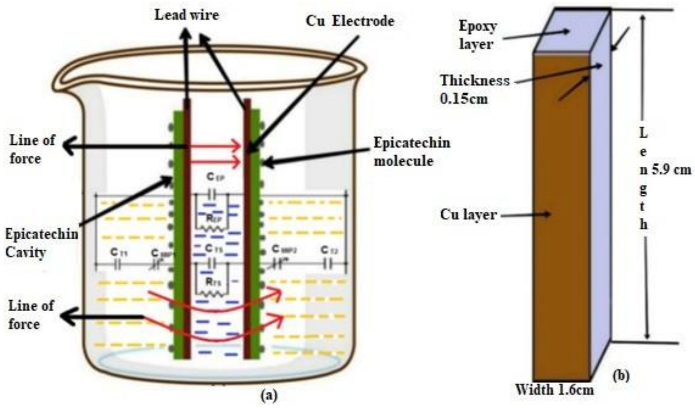


Fig.5.2. (a) EC-MIP coated Cu strip dipped in test solution with the equivalent circuit of the sensor (b) Schematic of Cu strip.

5.2.1.1 Materials used for the sensor fabrication

3-Hydroxytyramine hydrochloride, also known as Dopamine hydrochloride (DA-HCl), was acquired from Tokyo Chemical Industry (TCI) under the CAS RN: 62-31-7. The utilization of PDA, an oxidized derivative of DA, serves as a surface coating agent in this study, showcasing remarkable chemical characteristics due to the existence of ligands such as phenyl, catechol, and amine. These ligands interact with Cu substrates through hydrogen bonds, hydrophobic interactions, and electrostatic forces. A 0.15 cm thick Copper cladded FR-4 sheet was procured from local vendors, while Industrial-rated EC (>97%, CAS RN: 490-46-0) was obtained from TCI. The coating and cleaning processes involve the use of Chemical grade polyethylene glycol (PEG 400) from Merck Millipore and Millipore water. Cleaning of the Cu surface was carried out using Zero-size sandpaper.

5.2.1.2 Procedure of fabrication

The dimension of the proposed sensor is 5.9 cm × 1.6 cm × 0.15 cm (Fig. 5.2(b)), composed of copper-cladded FR-4 material, which is a glass epoxy resin commonly utilized in printed circuit boards (PCB), filed on the edges and smoothed with sandpaper. Following ultra-sonication in acetone for 15 minutes at 60^{o}C, the strips underwent a rinse with Millipore water. Subsequently, a thermal-controlled oven was employed for drying for 10 minutes at 70^{o}C. Copper wires were soldered on both sides of the Cu strip to establish the electrical connections. In the preparation of the MIP material, DA base functions as the functional monomer while EC serves as the template molecule. A mixture of 2 mg mL^{-1} dopamine hydrochloride (DA-HCl) and 8mg EC in 90mL Millipore water was prepared. By gradually introducing tris-ammonium buffer solution to the mixture, the pH was adjusted to 9. Cu plates coated with PEG (MW 400) were immersed in the solution and agitated for 48 hours using a magnetic stirrer. The formation of PDA via a single-step self-polymerization process from dopamine hydrochloride was evident, depositing on both sides of the copper strip within 48 hours. The synthesized MIP-EC@C sensor was subsequently dried in a furnace at 40^{o}C for 30 minutes and immersed in Millipore water for approximately five minutes to eliminate loose particles and EC molecules, resulting in cavities on the polymer surface. Following another round of drying at 40^{o}C for 30 minutes, the sensor was deemed prepared for experimental evaluation. The NIP capacitive sensor material was prepared in a similar manner to the MIP, with the exclusion of EC in its preparation process.

5.2.1.3 Experimental set-up

To perform the experimental study, a 100 mL Borosil beaker was filled with the test sample. The sensor was dipped into the solution with a sensor dip length of 4.4cm. The experimental setup in the laboratory is shown in Fig.5.3. The top end of the sensor was soldered with lead wire on both sides. The contacts of the sensor were connected to the impedance analyzer (KEYSIGHT, E4990A, 20 Hz-50MHz). Five test samples with varying EC concentrations (1 ppm, 5 ppm, 10 ppm, 20 ppm, and 50 ppm) have been made and tested at room temperature. For each test sample, eight repetitive readings have been collected for each parametric viz., capacitance, impedance, and phase angle measurements. After every measurement, the sensor was cleaned by dipping it into millipore water for about 5 minutes. Each experiment was performed after 10 minutes of incubation in the test solution. The washing time is optimized and chosen to be 5 minutes as a minimum error (0.32%) was obtained after comparing before testing and after testing maximum capacitances in different time intervals (Table 5.1). Since the temperature and the beaker size were kept the same for all the experiments, these parameters did not affect the measured capacitance during the experiments.

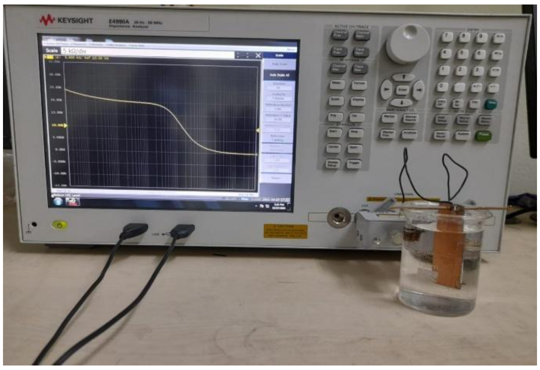


Fig.5.3. Experimental setup connected with impedance analyzer during testing.

***Table 5.1**. Optimization of washing time.*

Sl No.	Before Testing (Max. Capacitance (nF))	After Testing		Error (%)
1		**Washing Time (min)**	**Max. Capacitance (nF)**	0
2	223.17	0	117.83	47.20
3		2	178.29	20.11
4		**5**	**222.46**	**0.32**
5		10	222.39	0.35
6		15	222.38	0.35

5.2.2 Electrical equivalent model of proposed sensor

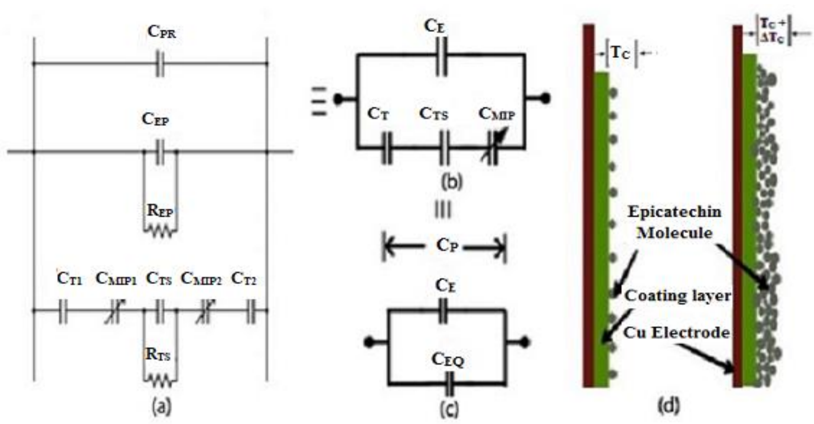


Fig.5.4.(a) Electrical equivalent circuit (b) and (c) simplified equivalent circuit (d) MIP coated Cu strips showing deposition of EC layer on the electrode surface.

5.2.2.1 Equivalent electrical model of the proposed sensor

The electrical circuit equivalent to that illustrated in Figure 5.2(a) of the suggested sensor can be derived based on the prior investigation [23]. Within the sensor are copper electrodes that enclose FR4 fiberglass epoxy. Various circuit components such as C_{EP}, R_{EP}, C_{TS}, R_{TS}, C_{MIP}, and C_T are included in the electrical equivalent circuit of the sensor, as depicted in Figure 5.4(a). In the sensor's operation, the electric field lines take two routes, one passing directly through the FR4 epoxy and the other through the surrounding test solution (TS). The direct path through the epoxy (EP) FR4 is linear, resulting in C_{EP} and R_{EP} between the two copper electrodes. Additionally, the curved electric field path originates from one copper electrode to the other through the TS surrounding the sensor, due to fringing fields at the sensor's edge. The alternative route involves capacitance and resistance from the TS, as well as coating capacitances C_T, which stem from the coating layers on both sides of the copper electrode (C_{T1} and C_{T2}). Two variable capacitances are created due to the EC molecules' adsorption onto MIP sensor layers on both sides of the device, as shown in Figure 5.4(a). R_{TS} is maintained open due to its high values [35]. The simplified circuit configurations are displayed in Figure 5.4(b) and Figure 5.4(c). The MIP coating thickness (T_{MC}) remains consistent on both sides of the copper electrodes regardless of EC concentration fluctuations in TS. The adsorption layer thickness fluctuates with EC concentration, leading to variable C_{MIP} on both sides of the sensor (C_{MIP1} and C_{MIP2}). It is important to highlight that C_{PR} represents the parasitic capacitance, while C_E signifies the combined capacitance of C_{PR} and C_{EP}. The comprehensive electrical model equivalent to the sensor is depicted in Figure 5.4.

$$\mathrm{C_P = C_E + C_{EQ}}\text{ , where }\mathrm{C_E = C_{PR} + C_{EP}}$$

$$\mathrm{C_{EQ}} = \frac{\mathrm{C_T \times C_{TS} \times C_{MIP}}}{\mathrm{(C_T \times C_{TS}) + (C_{TS} \times C_{MIP}) + (C_{MIP} \times C_T)}}$$

$$= \frac{\mathrm{C_T \times C_{TS}}}{\mathrm{C_T + C_{TS}} + \left(\frac{\mathrm{C_T \times C_{TS}}}{\mathrm{C_{MIP}}}\right)} \tag{5.1}$$

$$\text{Where } \mathrm{C_{MIP}} = \frac{\mathrm{C_{MIP_1} \times C_{MIP_2}}}{\mathrm{C_{MIP_1} + C_{MIP_2}}} = \frac{\varepsilon_0 \varepsilon_{EC}\,A}{2\mathrm{T_C}} = f\left(\frac{1}{\mathrm{T_C}}\right) \tag{5.2}$$

Here, T_C= Thickness of adsorbed EC molecular layer on the MIP surface and ε_{EC}= Dielectric constant of EC

$$C_T = \frac{C_{T_1} \times C_{T_2}}{C_{T_1} + C_{T_2}} = \frac{\varepsilon_0 \varepsilon_M A}{2T_{MC}} = \text{Constant} \tag{5.3}$$

where T_{MC}= Thickness of MIP coating on both sides of Cu electrode

ε_0 = Dielectric constant of vacuum

ε_M = Dielectric constant of MIP coating,

and A = Surface area of the electrode

Therefore,

$$C_P = C_E + C_{EQ} = (C_{PR} + C_{EP}) + \frac{C_T \times C_{TS}}{C_T + C_{TS} + \frac{C_T \times C_{TS}}{f(\frac{1}{T_C})}} \tag{5.4}$$

$$\text{Concentration } (\uparrow) \Rightarrow T_C(\uparrow) \Rightarrow C_{MIP}(\downarrow) \Rightarrow C_P(\downarrow) \tag{5.5}$$

$$\text{Concentration } (\downarrow) \Rightarrow T_C(\downarrow) \Rightarrow C_{MIP}(\uparrow) \Rightarrow C_P(\uparrow) \tag{5.6}$$

5.2.2.2 Dependence of Sensor Capacitance on EC concentration

The detecting mechanism of the proposed sensor depends on the molecular imprinting technique. In this method, cavities resembling the template are formed in the polymer layer where the template molecules present in the sample may be adsorbed. As soon as the sensor is dipped in the solution, adsorption of the EC commences into the template cavity, thereby creating an EC layer on the electrode surface as depicted in Fig.5.4 (d). The thickness of the adsorption layer depends on the concentration of EC in the test solution and dipping time. Surface density varies depending on the extent of molecular adsorption near the solid-liquid interface based on dipping time which might affect the overall thickness of the corresponding interface. A thinner adsorption layer (T_C) is formed for a lower concentration of EC, while a thicker layer (T_C + ΔT_C) is formed for a higher EC concentration. This change in layer thickness gives rise to variable capacitances C_{MIP1} and C_{MIP2} on both sides of the sensor as shown in Fig. 5.4(a). With the rise in EC concentration (in the ppm range) in the solution, adsorption layer thickness increases leading to a decrease in capacitance values of C_{MIP1}, C_{MIP2}, and vice-versa [36]. Hence, it is quite obvious that the equivalent capacitor (C_{MIP}) is variable and depends on T_C (equation 5.2), which in turn depends on the concentration of EC.As seen from equation 5.4, the capacitance C_P is dependent on C_{PR}, C_{EP}, C_T, C_{TS}, and C_{MIP}. Among these capacitances, C_{PR}, C_{EP}, C_T, and C_{TS} stay unaltered while the C_{MIP} capacitance value changes when the sensor is dipped in the test solution. Thus, there is a change in C_P in terms of C_{MIP}, which depends on the concentration of EC in the test solution. Since C_{MIP} decreases with the increase in the concentration of EC, the capacitance C_P also decreases which is consistent with equations 5.5 and 5.6.

5.3 Result and Discussion from the Perspective of Sensor Fabrication

5.3.1 SEM Analysis

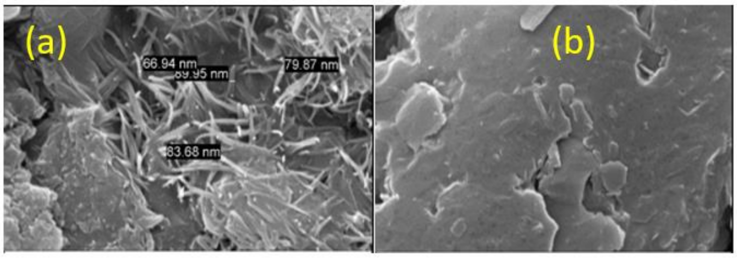


Fig.5.5. SEM images of (a) MIP material and (b) NIP material.

SEM images of MIP-EC@C and NIP@C materials are illustrated in Figure 5.5. Figure 5.5(a) displays a surface with increased roughness in comparison to that of Figure 5.5(b), suggesting the removal of EC molecules through washing and consequently hindering the smooth surface of the PDA/PEG polymer in the MIP material.

5.3.2 XRD Analysis

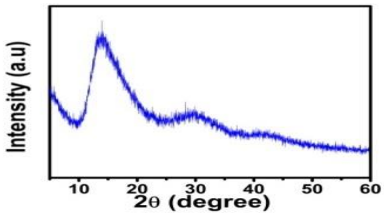


Fig.5.6. X-ray diffraction pattern of synthesized MIP-EC material.

The XRD pattern of the prepared MIP-EC material reveals the presence of two distinct broad peaks at approximately 13.9° and 29.2° as illustrated in Fig.5.6. These peaks suggest a strong bonding within the polymer chain and the formation of repeated units of the DA ring. The broad peaks observed in Fig.5.6 are indicative of the formation of an amorphous molecularly imprinted polydopamine-polyethylene glycol (PDA-PEG)[37].

5.3.3 Capacitance variation for MIP-EC@C and NIP@C

During the process of material synthesis, the interaction between DA and PEG leads to the formation of a PDA/PEG composite, where the ether group of PEG engages in hydrogen bonding with the catechol group of DA. The resulting PDA polymerization process yields a composite material with PDA aggregates evenly dispersed alongside PEG. In the specific scenario of depositing PDA/PEG onto Copper (Cu) surfaces, it is crucial to highlight that PDA not only forms coordination bonds with metal ions through its catechol and amino groups but also has the ability to chemically convert these ions (such as Cu2+ in this particular case) into their metallic form under alkaline conditions.Moreover, since Cu^{2+} exhibits a fair amount of affinity towards amine groups; the amino group from PDA is capable of coordinating with the Cu^{2+}ions. Precisely, PDA/PEG was found to be demonstrating satisfactory adhesive capability on Cu [37]. For target analyte EC detection, MIP technology was deployed in this work. The capacitance (C_P) variation as a function of frequency was depicted in Fig.5.7 for MIP-EC@C and NIP@C. The presence of molecular recognition sites in MIP-EC@C leads to a higher affinity for the EC analytes. This phenomenon gives rise to higher EC analyte adsorption. More molecular adsorption on the sensor interface produces a change in T_C as described in equation 5.2 thereby yielding low C_P. On the contrary, due to the absence of cavities on the sensor interface of NIP@C, very few molecules get adsorbed on the surface thereby yielding a negligible change in TC and high C_P. Hence, as expected high capacitance was observed for NIP@C compared to that of the MIP-EC@C sensor. Consequently, as anticipated, the capacitance observed for NIP@C was approximately five times greater than that of the MIP-EC@C sensor.

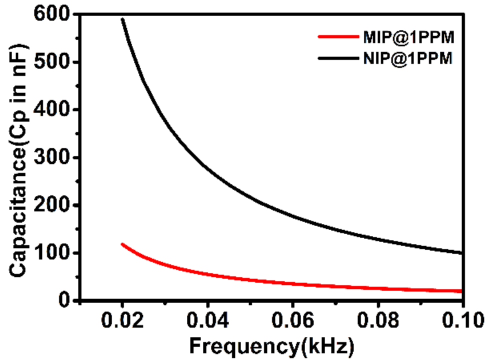


Fig.5.7. Capacitance (C_p) variations with frequency for MIP-EC@C and NIP@C at 1ppm EC.

5.3.4 Effect of EC concentration variation on capacitance (C_P) at different frequency

The variation in frequency of the capacitance is illustrated in Figure5. 8(a). It has been noted that the capacitance response diminishes as the frequency increases and stabilizes at higher frequencies, specifically beyond 5 kHz. When subjected to a strong electric field at a lower operating frequency, the capacitance could reach full charge; however, at a higher operating frequency, the capacitance may not fully charge within a single cycle [38]. Consequently, at higher operating frequencies, the alteration in capacitance is relatively insignificant compared to the low-frequency range. This observation serves to characterize capacitance in relation to frequency, as depicted in Figure 5.8(a). Additionally, the sensor capacitances decrease as the concentration of EC increases, aligning well with the proposed electrical model outlined in section 5.2 and in reference [23]. Equation 5.4 establishes the correlation between C_P, C_{MIP}, and the concentration of the test solution. The data plots are restricted to 100Hz to depict variations in capacitance values in response to changes in EC concentration.

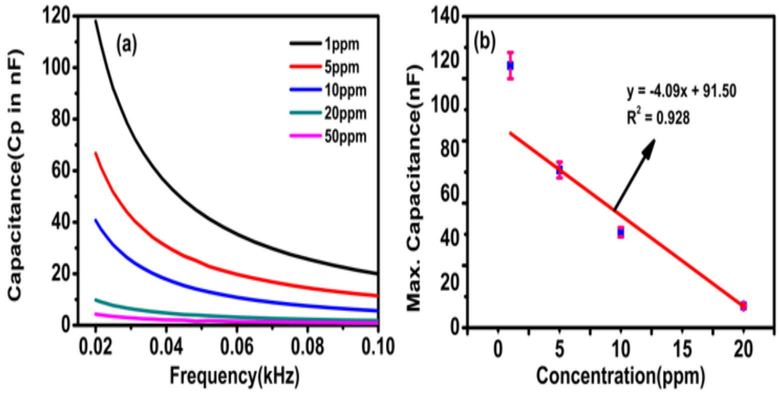


Fig.5.8.(a).Change in value of capacitance as a function of EC concentration in the test solution at different frequencies (b).Change of maximum Capacitance with concentration.

5.3.5 Concentration variation and detection limit

As shown in Fig.5. 8(b) the relationship of maximum capacitance with the EC concentration variation is linear. The calibration curve in Fig.5.8 (b) for EC indicates a linear segment from 5ppm to 20ppm with the regression equation as

$$y = -4.09\,x + 91.50, R^2 = 0.92 \qquad (5.7)$$

The limit of detection (LOD) of the MIP-EC@C was obtained as 1.07 ppb using the equation LOD= 3Sy/x/m where Sy/x is the standard deviation and m represents the slope of the calibration curve [9]. This LOD value is considerably lower than reported in [1]. Moreover, a high R-squared (R^2) value signifies the better fitting of the measurement data with the model thereby affirming the satisfactory linearity of the sensor.

5.3.6 Effect of EC concentration variation on sensor impedance at different frequencies

The plot displaying the impedance of the sensor (measured in kilo-ohms) as a function of frequency, covering various electrolyte concentrations, is illustrated in Fig. 5.9. It can be deduced that an increase in electrolyte concentration leads to a reduction in sensor impedance. Moreover, it is evident that the impedance demonstrates minimal fluctuations with frequency at a constant electrolyte concentration

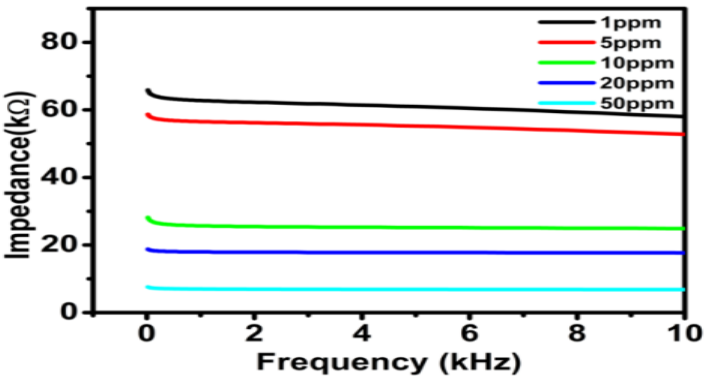


Fig.5.9.Sensor impedance plot as a function of EC concentration in the test solution at the different frequencies.

5.3.7 Effect of EC concentration variation on Phase angle at different frequencies

The modification in the impedance phase of the sensor occurs in relation to the frequency modulation corresponding to alterations in the analyte concentration within the solution [22]. The investigation presented in this study illustrates the fluctuations in phase angles across frequencies up to 10 kHz as shown in Fig.5.10. It is evident from the result that, at a specific concentration level, the phase angle diminishes proportionally with the rise in frequency. Moreover, it is notable that within a consistent frequency spectrum, the phase angle escalates with an increase in the EC concentration found within the test sample.

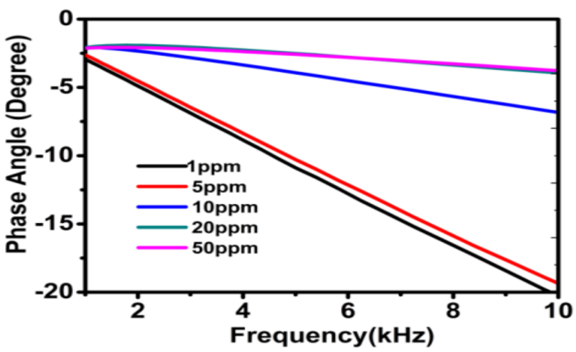
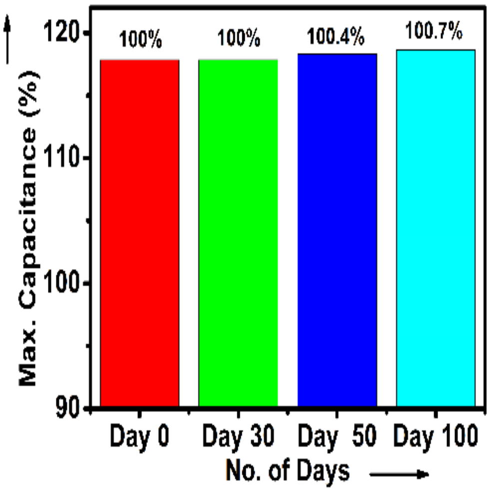


Fig.5.10. Phase angle plot of various EC concentrations in the test solution at the different frequencies.

5.3.8 Stability Analysis of MIP-EC@C Capacitive Sensor

In order to assess the stability of the sensor, a series of tests and measurements were conducted over a 100-day period in a randomized manner. Results from the experimentation indicated that the maximum capacitance exhibited by the MIP-EC@C sensor remained relatively constant during the initial month, experiencing only a minimal increase of 0.7% after the full 100 days (refer to Fig. 5.11). The molecularly imprinted capacitive sensors were enveloped in tissue paper and stored in a sealed zip bag at room temperature to prevent potential contamination from the surrounding atmosphere.



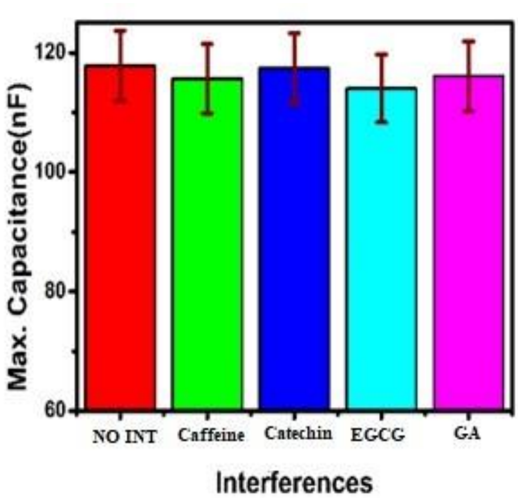


Fig. 5.11. Stability results of the MIP-EC@C Sensor. *Fig.5.12. Behaviour of MIP-EC@C Capacitive sensor in the presence of various interfering species.*

5.3.9 Interference study

Investigating the alteration in the electrical properties of the MIP-EC@C sensor when exposed to various interfering substances involved immersing it in a 1ppm concentration of EC solution following the introduction of five times the amount of interference compounds such as catechin and Gallic Acid (GA), as well as ten times the quantity of caffeine and EGCG. The impact of specific interfering agents in the presence of EC on the maximum capacitance is illustrated in Fig. 12. An in significance variation in capacitance is evident within a range of ±5% tolerance. Consequently, it can be deduced that the MIP-EC@C sensor displayed significant specificity towards EC despite the existence of other interfering substances.

5.3.10 Real sample study

The MIP-EC@C was utilized for the determination of electrical conductivity (EC) in four different varieties of green tea sourced from distinct gardens at the Tocklai Tea Research Institute in Jorhat, Assam, India. Specifically, 1gm of tea leaf was utilized for infusion in 200mL of water, following a methodology akin to that referenced in [6]. A total of eight responses were recorded for each sample within a frequency range of 20-100Hz, resulting in the acquisition of a data matrix sized [24 × 8 × 4] through the immersion of the MIP-EC@C electrode in the four respective test samples. Subsequent data analysis was conducted utilizing Partial Least Squares (PLS) and Principal Component Regression (PC) techniques in MATLAB R2017b, with the findings presented in Table 5.2 and Table 5.3. Evaluation of prediction accuracies involved a comparison between the standard High Performance Liquid Chromatography (HPLC) data and the data derived from the PLSR and PCR models. In Table 5.3, the statistical metrics have been summarized for PLSR and PCR regression models. Rc^2, $Rv^{2,}$ and Rp^2 represent the calibration, validation, and prediction coefficients respectively, while RMSEV and RMSEP represent the root mean square error value for validation and prediction. The higher coefficients value close to unity signifies excellent calibration, validation, and prediction results.

Table 5.2. *EC content from PLSR and PCR model.*

Sl No.	**Actual EC(mg/g) (HPLC)**	**Predicted EC(mg/g)**		**Prediction Accuracy (%)**	
		PLSR	**PCR**	**PLSR**	**PCR**
1	0.81	0.80	0.80	99.38	99.57
2	1.11	1.09	1.09	98.91	98.91
3	0.97	0.97	0.96	100	99.89
4	0.81	0.80	0.80	99.75	99.87
	Average prediction accuracy			**99.51**	**99.56**

Table 5.3. *Prediction of EC concentration in unknown samples.*

Technique used	**Calibration**		**Validation**		**Prediction**	
	RMSEC	**Rc^2**	**RMSEV**	**Rv^2**	**RMSEP**	**Rp^2**
PLSR	0.0030	1	0.0025	0.99	0.0024	0.99
PCR	0.0030	1	0.0022	0.99	0.0021	0.99

5.3.11 Comparative Study of the Present Technique with Previous works on EC detection

A comparative analysis of observations documented in various publications regarding the identification of EC is delineated in Table 5.4. Within this investigation, a diminished limit of detection (LOD) was identified, exhibiting a lower magnitude in comparison to findings from preceding studies [13-15], [1]. Nonetheless, the methodologies utilized in prior research endeavors for the synthesis and evaluation processes are known to be time-intensive. The exploration of an MIP capacitive sensor for EC detection in this research emerges as a viable and cost-effective solution for potential industrial applications. The notably robust and discriminating MIP-EC@C sensor demonstrated effectiveness in the examination of green tea samples. Traditionally, the minimum EC concentration in 100mL of green tea is approximated at 17.24 ppm (0.06 µM)[39] ,a value falling within the established linear range of the developed sensor, which spans from 5 to 20 ppm (0.02-0.07µM). Consequently, industries may consider adopting this reliable, discriminative, and economically feasible MIP-EC@C sensor for the detection of EC in green tea samples.

***Table 5.4**. Comparison of the existing techniques with the present work.*

Electrode/ Sensor	Technique s used	Linear range(µM)	LOD (µM)	PLSR/ PCR (%)	Refs
Pt	Ad-SV	0.69-8.6	0.34	----	[43]
GC	SWV	0.01-10	4.27	----	[44]
Pt	DPV	50-300	1.80	----	[45]
MIP-Q@G	CV and DPV	1-100 and 100-500	0.33	94.54	[2]
MIP-EC	CV and DPV	1-30 and 30-300	0.05	93.58	[1]
MIP-EC@C	**Capacitive Sensor**	**0.02-0.07**	**0.036**	**99.56**	**Present work**

5.4 Data analysis

Ten different types of green tea samples were gathered from a variety of gardens at the Tocklai Tea Research Institute in Jorhat, Assam, India. The process of making tea adhered to the method described in citation [20]. A total of 201 data points were gathered based on the sensor's reactions. Every sample was subjected to ten repeated measurements, resulting in a dataset of size [201×10]. As a result, the extensive dataset contained information from all ten green tea samples, resulting in a size of [201×10×10]. All statistical procedures were carried out utilizing Python version 3.11.3. In order to visually display the data, a principal component analysis (PCA) was performed on the sensor data. The visualization was produced using MATLAB version 2021a and is shown in Figure 4.

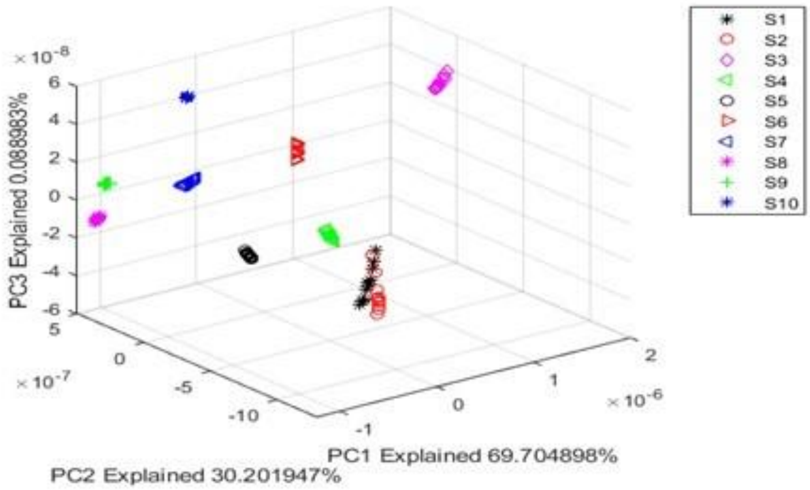


Fig.5.13. Principal component analysis (PCA).

5.4.1 Application of Clustering Algorithm

The sensor response, derived from the MIP tethered capacitive sensor was analyzed using two popular clustering algorithms: K-Means and Agglomerative clustering. This study shows the performance of the synthesized sensor. The detail overview of the two clustering algorithms K-Means and Agglomerative haven discussed in the section 2.2.

5.5 Result and discussion from the Perspective of Data Analysis

The experimentation encompassed a frequency range extending from 20 Hz to 50 MHz, corresponding to the capacitance function of the sensor. The response of both the MIP and the non-imprinted polymer (NIP) parallel plate capacitance at a frequency of 5 ppm EC concentration solution is depicted in Figure 5.14. Up to 100 Hz, the frequency range has been examined to illustrate capacitance variations resulting from the adsorption of EC molecules on the surfaces of the MIP-modified parallel plate capacitance. Due to the higher adsorption of EC molecules by the MIP-tethered sensor compared to the NIP sensor, the overall capacitance of the MIP sensor decreased in comparison to the NIP sensor. This study focuses solely on capacitance changes attributed to frequency variations at normal room temperature.

For further data analysis, two well-established clustering algorithms, K-Means and Agglomerative, were utilized. Silhouette scores achieved by these algorithms at different K values are presented in Table 5.5. Interestingly, the peak Silhouette scores for both algorithms were observed at K = 10. Additionally, Table 5.6 displays the Davies-Bouldin scores for the same algorithms at various K values. Notably, the lowest Davies-Bouldin score was also observed at K = 10. Figures 5 and 6 offer a visual representation of the relationship between K values and the Silhouette scores, as well as the Davies-Bouldin scores, for both algorithms.

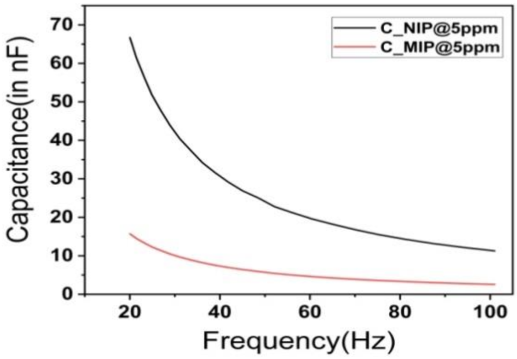


Fig.5.14. Capacitance variation responses with frequency for C_MIP@5 ppm and C_NIP@5 ppm sensors

Table 5.5*. Cluster vs. Silhoute scores.*

K value	K-Means Clustering	Agglomerative Clustering
2	0.5233	0.5233
3	0.6774	0.6774
4	0.6722	0.6722
5	0.7325	0.7325
6	0.8103	0.8103
7	0.8322	0.8452
8	0.7993	0.8011
9	0.8246	0.8264
10	**0.8568**	**0.8465**
11	0.8388	0.8299
12	0.8261	0.8217
13	0.8154	0.8212
14	0.8208	0.7937
15	0.7896	0.7857

***Table 5.6**. Cluster vs. Davies-Bouldin scores.*

K value	K-Means Clustering	Agglomerative Clustering
2	0.7642	0.7642
3	0.4545	0.4545
4	0.4892	0.4892
5	0.3634	0.3634
6	0.2856	0.2856
7	0.2053	0.2053
8	0.2882	0.2857
9	0.2417	0.2395
10	**0.2016**	**0.2013**
11	0.2260	0.2283
12	0.2531	0.2572
13	0.2814	0.2624
14	0.2782	0.3012
15	0.3138	0.3170

5.6 Conclusion

In this work, a novel, cost-effective, and reusable capacitive sensor was fabricated for EC detection, and its clustering performance was evaluated using two clustering algorithms. PDA/PEG-deposited copper plates were used to develop the capacitive sensor. The sensor's performance was tested at EC concentrations of 1 ppm, 5 ppm, 10 ppm, 20 ppm, and 50 ppm. The sensor exhibited a linear change in capacitance for the EC concentration range of 5-20 ppm, and the limit of detection (LOD) was calculated to be 1.07 ppb, which is considerably lower than the EC levels reported in previous studies.

The average prediction accuracies of the MIP-EC@C sensor, determined using PLSR and PCR models as shown in Table 5.2, were 99.51% and 99.56%, respectively, which are significantly higher than those reported in the literature. The experimental results confirmed a notable decrease in sensor capacitance with a gradual increase in EC concentration, consistent with the electrical equivalent circuit model of the sensor. The variation in capacitance (C_P) as a function of frequency for MIP-EC@C and NIP@C was also analyzed, revealing that the NIP@C capacitance was approximately five times higher than that of the MIP-EC@C sensor. Stability analysis showed a 0.7% increase in maximum capacitance after 100 days, which is acceptable according to the authors' knowledge. The study also presented the sensor impedance plot and phase angle plot, indicating the potential of the proposed MIP capacitive sensor for detecting EC in solutions.The effectiveness of clustering was assessed using Silhouette and Davies-Bouldin scores. Silhouette scores of 0.8568 and 0.8465, obtained from the two algorithms, indicate that objects align well with their respective clusters and not with neighboring clusters. This score is crucial for determining the optimal number of clusters (K) in the algorithm. Davies-Bouldin scores of 0.2016 and 0.2013, respectively, highlight the enhancement of differences between clusters by minimizing their average similarity. Lower Davies-Bouldin scores indicate superior cluster separation. Both algorithms effectively explore the optimal number of clusters for maximal clustering performance. Consequently, it can be inferred that the developed sensor adeptly distinguishes epicatechin content in green tea samples.

The following chapter is the final one of this thesis, summarizing the key findings, future prospects, and conclusions of this research.

References

[1] D. Das, S. Nag, A. Adaval, A. K. Hazarika, S. Sabhapondit, A. R. Bhattacharyya, et al., "Amine functionalized MWCNTs modified MIP-based electrode for detection of epicatechin in tea," IEEE Sensors Journal, vol. 22, pp. 10323-10330, 2022.

[2] D. Das, S. Nag, S. De, A. K. Hazarika, S. Sabhapondit, B. Tudu, et al., "Electrochemical detection of epicatechin in green tea using quercetin-imprinted polymer graphite electrode," IEEE Sensors Journal, vol. 21, pp. 26526-26533, 2021.

[3] H. Wang, G. J. Provan, and K. Helliwell, "Tea flavonoids: their functions, utilisation and analysis," Trends in Food Science & Technology, vol. 11, pp. 152-160, 2000.

[4] P. R. Machonis, M. A. Jones, B. T. Schaneberg, and C. L. Kwik-Uribe, "Method for the determination of catechin and epicatechin enantiomers in cocoa-based ingredients and products by high-performance liquid chromatography: single-laboratory validation," Journal of AOAC International, vol. 95, pp. 500-507, 2012.

[5] J. Xia, D. Wang, P. Liang, D. Zhang, X. Du, D. Ni, et al., "Vibrational (FT-IR, Raman) analysis of tea catechins based on both theoretical calculations and experiments," Biophysical Chemistry, vol. 256, p. 106282, 2020.

[6] I. Berregi, J. I. Santos, G. Del Campo, and J. I. Miranda, "Quantitative determination of (−)-epicatechin in cider apple juices by 1H NMR," Talanta, vol. 61, pp. 139-145, 2003.

[7] Y. Jaiswal, P. Tatke, S. Gabhe, and A. Vaidya, "Rapid high performance thin layer chromatographic method for quantitation of catechin from extracts of cashew leaves-a short report," Polish Journal of Food and Nutrition Sciences, vol. 63, 2013.

[8] T. Dias, M. R. Silva, C. Damiani, and F. A. da Silva, "Quantification of catechin and epicatechin in foods by enzymatic-spectrophotometric method with tyrosinase," Food Analytical Methods, vol. 10, pp. 3914-3923, 2017.

[9] S. Nag, S. Pradhan, H. Naskar, R. B. Roy, B. Tudu, P. Pramanik, et al., "A simple nano cerium oxide modified graphite electrode for electrochemical detection of formaldehyde in mushroom," IEEE Sensors Journal, vol. 21, pp. 12019-12026, 2021.

[10] S. Nag, S. Pradhan, D. Das, B. Tudu, R. Bandyopadhyay, and R. B. Roy, "Fabrication of a molecular imprinted polyacrylonitrile engraved graphite electrode for detection of formalin in food extracts," IEEE Sensors Journal, vol. 22, pp. 42-49, 2021.

[11] S. Nag, D. Das, H. Naskar, B. Tudu, R. Bandyopadhyay, and R. B. Roy, "Detection of metanil yellow adulteration in turmeric powder using nano nickel cobalt oxide modified graphite electrode," IEEE Sensors Journal, vol. 22, pp. 12515-12521, 2022.

[12] S. Nag, H. Naskar, S. Pradhan, R. Chatterjee, V. Sharma, B. Tudu, et al., "Formalin detection using platinum electrode-based electrochemical system," Journal of The Institution of Engineers (India): Series B, vol. 103, pp. 1159-1165, 2022.

[13] L. Pigani, R. Seeber, A. Bedini, E. Dalcanale, and M. Suman, "Adsorptive-stripping voltammetry at PEDOT-modified electrodes. Determination of epicatechin," Food analytical methods, vol. 7, pp. 754-760, 2014.

[14] I. Novak, M. Šeruga, and Š. Komorsky-Lovrić, "Square-wave and cyclic voltammetry of epicatechin gallate on glassy carbon electrode," Journal of Electroanalytical Chemistry, vol. 631, pp. 71-75, 2009.

[15] S. T. Duran, "Preparation of poly (pyromellitic dianhydride-co-thionin) modified voltammetric sensor for the determination of epicatechin," Journal of the Turkish Chemical Society Section A: Chemistry, vol. 5, pp. 1021-1028, 2018.

[16] N. Leibl, K. Haupt, C. Gonzato, and L. Duma, "Molecularly imprinted polymers for chemical sensing: A tutorial review," Chemosensors, vol. 9, p. 123, 2021.

[17] T. N. Chatterjee, D. Das, R. B. Roy, B. Tudu, S. Sabhapondit, P. Tamuly, et al., "Molecular imprinted polymer based electrode for sensing catechin (+ C) in green tea," IEEE Sensors Journal, vol. 18, pp. 2236-2244, 2018.

[18] T. N. Chatterjee, D. Das, R. B. Roy, B. Tudu, A. K. Hazarika, S. Sabhapondit, et al., "Development of a nickel hydroxide nanopetal decorated molecular imprinted polymer based electrode for sensitive detection of epigallocatechin-3-gallate in green tea," Sensors and Actuators B: Chemical, vol. 283, pp. 69-78, 2019.

[19] D. Das, T. N. Chatterjee, R. B. Roy, B. Tudu, A. K. Hazarika, S. Sabhapondit, et al., "Titanium oxide nanocubes embedded molecularly imprinted polymer-based electrode for selective detection of caffeine in green tea," IEEE Sensors Journal, vol. 20, pp. 6240-6247, 2020.

[20] T. N. Chatterjee, R. B. Roy, B. Tudu, P. Pramanik, H. Deka, P. Tamuly, et al., "Detection of theaflavins in black tea using a molecular imprinted polyacrylamide-graphite nanocomposite electrode," Sensors and Actuators B: Chemical, vol. 246, pp. 840-847, 2017.

[21] D. Das, D. Biswas, A. K. Hazarika, S. Sabhapondit, R. B. Roy, B. Tudu, et al., "CuO nanoparticles decorated MIP-based electrode for sensitive determination of gallic acid in green tea," IEEE Sensors Journal, vol. 21, pp. 5687-5694, 2020.

[22] S. Biswas, M. Chakraborty, and K. Biswas, "Detection of formaldehyde by a rgo/pmma coated sensor," in 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2020, pp. 1-6.

[23] M. M. Nezami, S. A. Wani, S. A. Khan, N. Khera, and S. Sohail, "An MIP-based novel capacitive sensor to detect 2-FAL concentration in transformer oil," IEEE Sensors Journal, vol. 18, pp. 7924-7931, 2018.

[24] A. Kumar Prusty and S. Bhand, "Molecularly imprinted polyresorcinol based capacitive sensor for sulphanilamide detection," Electroanalysis, vol. 31, pp. 1797-1808, 2019.

[25] J. H. Waite, "Adhesion a la moule," Integrative and comparative biology, vol. 42, pp. 1172-1180, 2002.

[26] H. G. Silverman and F. F. Roberto, "Understanding marine mussel adhesion," Marine biotechnology, vol. 9, pp. 661-681, 2007.

[27] H. Lee, S. M. Dellatore, W. M. Miller, and P. B. Messersmith, "Mussel-inspired surface chemistry for multifunctional coatings," science, vol. 318, pp. 426-430, 2007.

[28] J. Yu, Y. Kan, M. Rapp, E. Danner, W. Wei, S. Das, et al., "Adaptive hydrophobic and hydrophilic interactions of mussel foot proteins with organic thin films," Proceedings of the National Academy of Sciences, vol. 110, pp. 15680-15685, 2013.

[29] S.-C. Chou, W.-A. Chung, T.-L. Fan, Y. Dordi, J. Koike, and P.-W. Wu, "Polydopamine and its composite film as an adhesion layer for Cu electroless deposition on SiO2," Journal of the Electrochemical Society, vol. 167, p. 042507, 2020.

[30] J.-l. Wang, B.-c. Li, Z.-j. Li, K.-f. Ren, L.-j. Jin, S.-m. Zhang, et al., "Electropolymerization of dopamine for surface modification of complex-shaped cardiovascular stents," Biomaterials, vol. 35, pp. 7679-7689, 2014.

[31] W. Wang, Y. Jiang, S. Wen, L. Liu, and L. Zhang, "Preparation and characterization of polystyrene/Ag core–shell microspheres–a bio-inspired poly (dopamine) approach," Journal of colloid and interface science, vol. 368, pp. 241-249, 2012.

[32] W. Wang, A. Zhang, L. Liu, M. Tian, and L. Zhang, "Dopamine-induced surface functionalization for the preparation of Al–Ag bimetallic microspheres," Journal of the Electrochemical Society, vol. 158, p. D228, 2011.

[33] C. Xu, M. Tian, L. Liu, H. Zou, L. Zhang, and W. Wang, "Fabrication and properties of silverized glass fiber by dopamine functionalization and electroless plating," Journal of the Electrochemical Society, vol. 159, p. D217, 2012.

[34] Z. Tapsir and S. Saidin, "Synthesis and characterization of collagen–hydroxyapatite immobilized on polydopamine grafted stainless steel," Surface and Coatings Technology, vol. 285, pp. 11-16, 2016.

[35] A. Bose and K. Biswas, "Perfomance study of urease-PMMA-based aqueous urea sensor," IEEE Sensors Journal, vol. 17, pp. 6850-6858, 2017.

[36] J.-L. Gong, F.-C. Gong, Y. Kuang, G.-M. Zeng, G.-L. Shen, and R.-Q. Yu, "Capacitive chemical sensor for fenvalerate assay based on electropolymerized molecularly imprinted polymer as the sensitive layer," Analytical and bioanalytical chemistry, vol. 379, pp. 302-307, 2004.

[37] M. Maruthapandi, M. Natan, G. Jacobi, E. Banin, J. H. Luong, and A. Gedanken, "Antibacterial activity against methicillin-resistant Staphylococcus aureus of colloidal polydopamine prepared by carbon dot stimulated polymerization of dopamine," Nanomaterials, vol. 9, p. 1731, 2019.

[38] D. Das, F. A. Kamil, K. Biswas, and S. Das, "Evaluation of single cell electrical parameters from bioimpedance of a cell suspension," RSC Advances, vol. 4, pp. 18178-18185, 2014.

[39] E. Jówko, "Green tea catechins and sport performance," SPORT NUTRITION, vol. 123, 2015.

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

This chapter provides a review of the work and a summary of the findings. It also offers key recommendations for the developed methodology and the instrument designed for various industrial applications. Additionally, future research directions are discussed, along with concluding remarks highlighting the significance and impact of the research presented.

LIST OF SECTION

Chapter 6

Conclusion and future scope

6.1. Introduction

This thesis introduces a machine learning approach to optimize electrode arrays and data for improving a system that assesses tea quality. Traditionally, tea tasting depends on human experts or chemical analysis, but industries now prefer faster analytical methods. Despite their potential, limitations of analytical instruments restrict their use. Integrating sensing technology with machine learning has brought advancements across multiple sectors.

By using optimized sensor arrays and efficient data analysis, the system can extract patterns and trends from datasets. An electronic tongue, combined with data analysis, delivers qualitative tea assessments, and can be applied to other liquids. Key benefits include improved prediction accuracy, a straightforward sensor optimization method, and the development of a cost-effective, fast, and selective MIP-tethered capacitive sensor. This sensor has also been used to measure epicatechin in tea, demonstrating its effectiveness through a clustering algorithm.

6.2. Summary of Major Findings

6.2.1 An innovative technique for optimizing electrode arrays aimed at the comprehensive evaluation of black tea utilizing non-specific polymer graphite electrodes

Developing low-cost, reusable sensors for comprehensive quality assessment of tea and identifying a novel machine learning technique to optimize the sensor array are crucial for the tea industry. This work focuses on:

(i) Fabricating nine polymer graphite composite electrodes using monomers: acrylamide, aniline, and pyrrole, with concentrations ranging from 10% to 30% relative to graphite.

(ii) Optimizing the electrode array using a novel machine learning technique.

The details are presented in Chapter 3, where the real-time application of the synthesized electrodes is demonstrated using eight variants of black tea. The resulting dataset was processed through four feature transformation techniques (PCA, DCT, SVD, ICA) and five classification methods (SVM, KNN, ensemble, decision tree, and discriminant analysis).

Electrodes WE1 and WE8 achieved an average correct classification accuracy of over 91% among the nine electrodes. A polling strategy was used to fine-tune the electrode array, and a PCA model was created to visualize the data before and after electrode optimization. This research highlights WE1 and WE8 as the key electrodes for the comprehensive qualitative evaluation of black tea. As a result, eliminating redundant working electrodes is expected to improve system response speed, reproducibility, cost-efficiency, and sensitivity.

6.2.2 Development of a soft computing model to enhance prediction accuracy for green tea quality estimation using feature optimization and target-specific MIP electrodes

The current work elucidates a technique for developing a metaheuristic-driven model designed to improve the overall predictive accuracy of an electrochemical system by utilizing optimized and condensed datasets compared to raw data. The detailed of the study has been illustrated in the Chapter 4 of this thesis. Within this study, electrode responses synthesized for GAL and EC detection in green tea samples were transformed using the DCT technique. Subsequently, the modified features underwent individual optimization through three metaheuristic techniques namely GA, BA, and WOA to reduce the dimensionality of the feature set. The predictive performance of two synthesized electrodes, MIP-GAL and Q-IPG datasets, was evaluated using PLSR and PCR models. Results indicated that the GA-optimized dataset exhibited the highest prediction accuracy, achieving 96.24% for GAL detection and 97.95% for EC detection. This significant enhancement in predictive accuracy was achieved by utilizing 83 and 129 features, respectively, instead of the entire dataset. Essentially, this research proposes a universal predictive model for analyzing voltammetric electrode data to qualitatively assess tea and related areas, offering a versatile approach.

6.2.3 Development of a target specific capacitive sensor for detecting epicatechin in green tea with data analysis using a clustering approach

In this study, a novel, cost-effective, and reusable capacitive sensor was synthesized for detecting EC. The clustering performance of the sensor was assessed using widely known clustering methods. Details of the fabrication process, experimentation, and data analysis techniques are provided in Chapter 5 of this thesis. Sensor performance was evaluated at varying EC concentrations ranging from 1 ppm to 50 ppm, with a linear capacitance response observed between 5-20 ppm. The limit of detection (LOD) was calculated at 1.07 ppb, significantly lower than previous studies on EC. The mean prediction accuracies of the MIP-EC@C sensor, determined via PLSR and PCR models, were exceptionally high at 99.51%

and 99.56% respectively. Stability tests showed only a 0.7% increase in maximum capacitance over 100 days, considered acceptable based on current knowledge. The study also explored variations in sensor impedance and phase angle relative to EC concentration. Clustering efficiency was evaluated using Silhouette and Davies-Bouldin scores. Silhouette scores of 0.8568 and 0.8465 from two algorithms indicated strong alignment of objects with their respective clusters, essential for determining the optimal number of clusters (K). Davies-Bouldin scores of 0.2016 and 0.2013 demonstrated improved cluster separation, with lower scores reflecting better differentiation. Both algorithms effectively identified the ideal number of clusters, demonstrating the sensor's proficiency in distinguishing EC content in green tea samples.

6.3. Recommendations

From this research, a few recommendation may be made for the food and pharmaceutical industries in its current form.

a. Number of electrodes in an electronic tongue may be optimized using machine learning methods in order to minimize redundancy.

b. Meta-heuristic-driven model design may be performed to enhance the overall predictive accuracy of an electrochemical system.

c. The MIP-tethered capacitive sensor may be used in the food and pharmaceutical industries to detect the non-volatile specific contents in liquid sample for better diagnosis.

6.4. Future Scope of this work

The following is list of areas where the suggested technique covered in the earlier part could be improved:

a) Deep learning can be utilized for large datasets, while low-dimensional data restricts its use.

b) The sensor was tested with a limited number of tea samples provided by Tocklai Tea Research Institute, Jorhat, Assam, India. A larger number of samples is required to develop an industry-compatible sensor.

c) Multimodal sensors may be used to collect different types of data simultaneously, leading to a more comprehensive understanding of the test samples.

d) Transfer learning may be used to improve the efficiency, performance and reduce the training time.

e) Two regression models, PLSR and PCR, have been used to predict the constituents in the tea samples. Other AI powered regression models could also be explored.

f) Generative AI can be used for data augmentation, pattern detection, noise reduction and prediction purpose.

g) Low-cost and disposable printed sensors could be developed

6.5. Conclusion

In order to extract the information from the dataset, we need to utilize advanced pattern recognition and machine learning algorithms. These advanced techniques can produce high-quality results when the data is fed into the model. Although the current work is in its early stages, it paves the way for data analysis using conventional machine learning algorithms. The development of sensing technology combined with advanced machine learning systems can extend the use of the electronic tongue beyond the tea and food industry to the pharmaceutical sector as well.