Quality Evaluation of Tea using Chemical Sensors and NIR Spectroscopy: Machine Learning Approach

Synopsis of the thesis submitted

By

Angiras Modak

Doctor of Philosophy (Engineering)

Department of Instrumentation and Electronics Engineering

Faculty Council of Engineering & Technology

Jadavpur University

Kolkata, India

2025

QUALITY OF TEA USING CHEMICAL SENSORS AND NIR SPECTROSCOPY: MACHINE LEARNING APPROACH

Introduction

The rising complexity of product quality assessment across diverse industries has placed data-driven analysis at the centre of ensuring consistency, reliability, and process optimisation. In the tea sector, where precision in quality evaluation directly influences market value, machine learning (ML) has emerged as a powerful transformative tool. Recent studies report accuracy levels exceeding 95% for tasks such as tea blending assessments, leaf-grade classification, and the evaluation of oolong and black tea varieties, highlighting the enormous potential of these analytical methods. Beyond classification, advanced techniques such as reinforcement learning and predictive modelling are now being applied to optimise key processing steps like fermentation. These systems enable real-time adjustments, reducing waste and ensuring uniform quality. Deep convolutional neural networks (DCNNs) have further demonstrated near-perfect precision in classifying fermentation levels by interpreting colour and texture features, strengthening the push towards intelligent and automated production lines.

Tea has long been valued for its rich flavours, historical significance, and well-documented health benefits [1]. According to delicate records, from the subtle taste of green tea to the robust flavour of black tea, each variety offers a unique tasting experience that promotes appreciation among enthusiasts [2]. Its global appeal continues to grow, with the global marketprojected to exceed US$1.48 trillion by 2027. Therefore, the importance of ensuring high-quality tea cannot be overstated [3]. Nutritional research consistently emphasises the health advantages of tea, chiefly attributable to its physiologically active constituents. Furthermore, constituents like L-theanine and caffeine enhance cognitive performance and provide a soothing effect, hence encouraging tea drinking for mental well-being [4].

Various factors, including as tea cultivars, geographic illnesses, processing techniques, and storage conditions, affect the intricate chemistry of tea. Tea processing encompasses critical stages including reduction, rolling, oxidation, and drying, among others [5]. Every step affects the chemical makeup and sensory attributes of the finished product, including its aroma and texture. Oxidation levels are critical. For instance, black tea experiences full oxidation, yielding a strong flavour profile, whereas green tea is subjected to minimum processing to maintain its fresh and fruity characteristics. Comprehending this process is crucial for both producers and customers seeking to recognise the subtleties of tea quality [6].

The diversity of tea kinds and their tastes complicates value assessments, however they have historically been modelled in a specific manner. Conventional methods primarily rely on the assessment of sensory input conducted by expert tasters, who evaluate various attributes such as texture, scent, and aroma [7]. While these methods introduce subjectivity and unpredictability, they also generate possible inconsistencies in evaluating tea quality. The widespread deficiency in consumers' knowledge and skill to evaluate tea quality through sensory attributes necessitates a more objective method for quality assessment [8].

Contemporary advancements in analytical technology, including high-performance liquid chromatography (HPLC) and near-infrared spectroscopy [9], have markedly enhanced the assessment of tea quality [10]. HPLC facilitates the comprehensive analysis of essential substances, such as catechins, caffeine, and other phytochemicals, through the separation, identification, and quantification of the many constituents in a tea sample [11]. NIR offers a swift, non-invasive technique for evaluating tea quality by the measurement of infrared absorption spectra, yielding information on chemical composition and moisture content, therefore rendering it suitable for regular quality assessments in commercial settings [12]. The E-Nose replicates human olfactory perception to evaluate the aroma of tea, guaranteeing uniformity in scent across manufacturing batches, while the E-Tongue measures taste characteristics, including sweetness, bitterness, and acidity. The E-Eye employs image processing to assess visual attributes, such as colour and shape, that influence market acceptance [13].

ML-driven approaches blend seamlessly with these analytical tools. Algorithms such as Support Vector Machines (SVM), Random Forests (RF), AdaBoost, and modern deep learning models including CNNs and ResNets have demonstrated exceptional capability in predicting chemical composition, identifying quality grades, and analysing sensory profiles [14]. Enhanced models, such as YOLOv8x variants equipped with attention mechanisms, have delivered precision rates above 95% in classifying fresh leaf quality [15]. Integration of these models with computer vision, spectral data, and electronic sensing systems has led to a more holistic, objective, and scalable framework for tea quality assurance [16].

As research advances, the convergence of ML, advanced sensors, and traditional analytical techniques is reshaping the future of the tea industry. This combined approach not only strengthens the reliability of quality evaluations but also promotes sustainability by automating labour-intensive tasks, reducing dependence on human expertise, and minimising production variability. The momentum toward smarter, data-driven systems promises to enrich the global tea experience, uniting centuries-old tradition with modern scientific precision. Ultimately, these innovations pave the way for real-time monitoring, improved authenticity checks, and enhanced consumer confidence, setting the foundation for the next generation of tea quality evaluation and control [17].

Research Objectives

The study aimed to precisely estimate the quality of black and green tea using spectroscopic and electrochemical techniques. The research provides a strong foundation for advanced sensor-driven quality evaluation frameworks and supports the development of objective, scalable tea quality monitoring systems.

Analytical frameworks were developed to ensure accurate quality assessment:

- Evaluation of EGCG using a Molecularly Imprinted Polymer (MIP) sensor.
- Prediction of TAN, THEO, and CAFF using an array of three MIP electrodes.
- Estimation of tannic acid in black tea using NIR spectroscopy.
- Data fusion model integrating MIP sensor data with NIR for enhanced tannic acid prediction.
- A combined analytical approach using MIP sensors and NIR spectroscopy was investigated for effective tea quality assessment. HPLC was used as a benchmark technique for quantifying caffeine, catechins, and other bioactive compounds but requires extensive sample preparation and lengthy analysis.
- Coupling instrumental methods with pattern recognition algorithms which ensures rapid, accurate, and comprehensive evaluation of tea quality.

Summary of chapters

- **Chapter 1** introduces tea quality evaluation, research context, sample overview, method selection, research goals, and scope, and includes a literature review.
- **Chapter 2** outlines the data analysis approaches, summarising the machine learning (ML) and deep learning (DL) techniques used.
- **Chapter 3** introduces a novel method for assessing green tea quality using Molecularly Imprinted Polymers (MIPs) specifically designed to recognise EGCG, the key compound responsible for the tea's flavour and health benefits. By employing a three-electrode electrochemical setup with Differential Pulse Voltammetry (DPV), the study achieved precise and rapid EGCG quantification. Advanced data pre-processing techniques such as Box–Cox transformation and standard scaling further improved model accuracy, while feature selection using UMAP and K-means clustering identified the most relevant electrochemical attributes for quality classification. A robust dataset of sixteen samples, each generating eighty data points, enabled effective model training, and the XGBoost classifier delivered exceptional performance with a mean AUC of 0.99 across ten-fold cross-validation. Overall, Chapter 3 demonstrates the strong potential of combining MIP-based electrochemical sensing with advanced machine learning to provide a fast, objective, and highly reliable approach to green tea quality evaluation.
- **Chapter 4** introduces a comprehensive multi-output regression framework for evaluating black tea quality using a PSRF-based Random Forest model capable of simultaneously predicting three key chemical constituents—caffeine (CAFF), theophylline (THEO), and tannic acid (TAN)—which collectively define black tea quality. The study followed a four-step methodology, beginning with detailed data collection using three highly selective MIP sensors tailored to detect individual compounds and capture the full chemical profile of each sample. To manage the complex, non-linear relationships between sensor outputs and quality indicators, GA-based feature selection was employed, outperforming conventional methods. The PSRF model was further enhanced through PSO and ACO optimisation techniques and refined using grid-search hyperparameter tuning, enabling deeper and more effective model exploration than standard RF approaches. Delivering exceptional performance, the model achieved 98.75% accuracy and an R^2 of 0.9945, surpassing traditional methods such as PLSR. Individual MIP sensors

also demonstrated strong predictive capability, with accuracies of 97.42% for TAN, 91.3% for THEO, and 95.16% for CAFF. Importantly, data fusion significantly improved prediction accuracy, confirming the value of integrating multi-sensor outputs for a more robust and holistic assessment of tea quality. Overall, the chapter establishes the PSRF model as a precise, reliable, and advanced tool for black tea quality evaluation, showcasing the transformative impact of combining cutting-edge sensor technology with sophisticated machine learning techniques.

- **Chapter 5** presents an advanced approach for predicting tea quality by integrating NIR spectroscopy with electrochemical data from Molecularly Imprinted Polymer (MIP) sensors, addressing the key challenge of developing prediction models that harness the complementary strengths of multiple analytical techniques. Central to this work is the use of Canonical Correlation Analysis (CCA) for feature fusion, which effectively identifies and maximises correlations between the two data modalities while reducing dimensionality and retaining the most informative features for prediction. The integrated SNN-LSTM model demonstrated exceptional performance, achieving an R^2 value of 0.964109, along with significantly improved error metrics—an MAE of 0.311486 and an RMSE of 0.456471—when using fused features. This represents a substantial enhancement over individual approaches, where MIP data alone yielded an MAE of 2.891555 and NIR data alone an MAE of 0.009685. The comparison clearly highlights the advantage of data fusion, as the combined model benefited from NIR's broad chemical coverage and the high specificity of MIP sensors. Beyond improving tannic acid prediction in black tea, the study offers deeper insight into how multi-modal analytical strategies can capture complementary attributes of tea quality, enabling a more holistic evaluation than any single method. The adaptable SNN-LSTM architecture paired with CCA-based feature fusion shows strong potential for broader application in assessing other tea quality markers and agricultural products. Overall, Chapter 5 demonstrates how integrating NIR and MIP data within a sophisticated deep learning framework can significantly enhance prediction accuracy, laying a strong foundation for future multidimensional quality assessment models in the tea industry and related fields.
- **Chapter 6** concludes with key findings, contributions, future research directions, and final remarks.

Conclusion

This study marks a major leap forward in tea quality evaluation by tackling long-standing gaps in traditional assessment methods and creating a unified, objective, and highly reliable framework for determining quality. It began by examining the fundamental attributes that define tea quality and the shortcomings of existing testing systems, then advanced through the development of next generation sensing technologies and finally culminated in multimodal data integration using machine learning. Each stage contributed fresh insights and strengthened the foundation for a more accurate and science driven approach. The use of MIPs as artificial receptors for key tea components enabled the creation of ultra-selective and sensitive sensors which, when paired with electrochemical techniques and ML models, delivered precise measurements of important indicators like EGCG in green tea and tannic acid in black tea. Simultaneously, NIR spectroscopy offered a non-destructive way to capture broad chemical information, complementing the specificity of MIP based sensing. The integration of these methods through advanced AI and ML techniques, including fuzzy logic, deep learning models such as SNN LSTM, and explainable AI, proved highly effective in extracting subtle patterns, improving prediction accuracy, and increasing trust in model outputs. Most notably, fusing NIR and MIP data using innovative feature integration strategies produced prediction models with exceptional accuracy, addressing the inherently complex nature of tea quality far better than any single technique. The industrial implications of this work are significant: the framework enables consistent, real time quality monitoring, greater production efficiency, reduced wastage, and improved product communication through quantifiable links between chemical composition and sensory attributes. Beyond tea, the methodology provides a scalable template for evaluating a wide range of foods, agricultural products, and beverages, with potential applications even extending into environmental monitoring and medical diagnostics. In essence, this thesis shows how advanced sensing, machine learning, and data fusion can reshape the way quality is assessed, offering transformative potential for the tea industry and setting the stage for future innovations in food quality science.

Future Scope

Future research should prioritise technological refinement, expanded applications, and interdisciplinary collaboration.

- On the technological side, developing next-generation MIP sensors with improved selectivity and sensitivity will be essential, supported by advancements in polymer design, imprinting techniques, and nanostructured materials. Creating multi-analyte MIP sensors capable of detecting several markers at once would allow more comprehensive assessments with minimal preparation.
- Integrating additional spectroscopic techniques such as Raman, Terahertz, or NMR with NIR may provide deeper insights into tea composition, especially when combined through advanced data fusion strategies.
- At the same time, exploring more powerful deep learning architectures including Graph Neural Networks for analysing compound relationships, attention-based models for highlighting key indicators, and Transformer-based designs for handling sequential data may substantially strengthen predictive performance. Hybrid models that combine the strengths of multiple architectures could further improve reliability.
- Miniaturisation and portability of sensing systems will also be vital for practical use in the tea industry. Compact devices that merge MIP sensors, miniaturised spectrometers, and embedded ML models could deliver real-time evaluation from field to packaging. Achieving this will require improvements in energy-efficient hardware, low-power computing, and durable system design.
- Standardising procedures for sample preparation, data acquisition, preprocessing, and validation, along with constructing large reference databases that include chemical, sensory, and quality attributes across tea types and processing conditions, will support reproducibility and strong ML model development.
- Expanding practical applications should include evaluating additional chemical markers such as theanine, volatile compounds, catechin ratios, and broader polyphenol profiles. Building predictive models for sensory characteristics including aroma, astringency,

umami, and bitterness based on instrumental measurements would help bridge analytical data with consumer experience.

- Applying the developed approaches across a wider range of teas such as white, oolong, yellow, and pu-erh, and studying the effects of processing steps like withering, rolling, fermentation, and drying, could enable more precise process control and optimisation. Integrating these methods into comprehensive quality-management frameworks is also important. A complete platform that unites real-time sensing, ML-based predictions, automated process adjustments, traceability tools, and supply-chain support could transform production efficiency and consistency.
- Predictive models of consumer acceptance based on instrumental data could enhance product positioning and segmentation.
- Interdisciplinary collaboration and the adoption of emerging technologies will play a major role in future progress. Integrating genomics and metabolomics with quality assessment can help connect genetic variation and metabolic pathways to quality traits, providing insights for breeding and cultivation strategies. Combining metabolomics data with ML methods may support more accurate predictions and targeted improvements. Blockchain technology offers secure traceability and authenticity verification, creating transparent farm-to-cup records and reducing issues related to mislabelling. IoT networks capable of monitoring environmental conditions, processing variables, and quality indicators can provide continuous data streams, and when combined with ML forecasting, can support automated control and real-time optimisation. Pursuing greener, more sustainable sensing approaches that reduce waste, lower energy consumption, and eliminate hazardous chemicals will align these methods with environmental priorities. Investigating how sustainable farming and processing practices influence quality, as measured through the techniques developed in this work, will further strengthen environmentally responsible production. Finally, accessible decision-support systems and user-friendly interfaces will be essential to encourage adoption. Simplified dashboards that merge human expertise with ML recommendations can provide reliable guidance to producers and evaluators with varying levels of experience.

REFERENCES

[1] L. Fu and W. Ying, "Advancements in the application of carbon nanotubes for tea quality and safety assessment," *Fullerenes, Nanotubes and Carbon Nanostructures,* vol. 31, no. 11, pp. 1007-1018, 2023, doi: https://doi.org/10.1080/1536383X.2023.2246162.

[2] H. S. A. Huda *et al.*, "Exploring the ancient roots and modern global brews of tea and herbal beverages: A comprehensive review of origins, types, health benefits, market dynamics, and future trends," *Food Science & Nutrition*, vol. 12, no. 10, pp. 6938-6955, 2024, doi: https://doi.org/10.1002/fsn3.4346.

[3] J. Ouyang *et al.*, "Insights into the flavor profiles of different grades of Huangpu black tea using sensory histology techniques and metabolomics," *Food Chemistry: X,* vol. 23, p. 101600, 2024, doi: https://doi.org/10.1016/j.fochx.2024.101600.

[4] M. T. El-Saadony *et al.*, "Chemistry, bioavailability, bioactivity, nutritional aspects and human health benefits of polyphenols: A comprehensive review," *International journal of biological macromolecules,* p. 134223, 2024, doi: https://doi.org/10.1016/j.ijbiomac.2024.134223.

[5] J. Moreira *et al.*, "Tea Quality: An Overview of the Analytical Methods and Sensory Analyses Used in the Most Recent Studies," *Foods,* vol. 13, no. 22, p. 3580, 2024, doi: https://doi.org/10.3390/foods13223580.

[6] I. Ben Alaya, G. Alves, J. Lopes, and L. R. Silva, "Use of Encapsulated Polyphenolic Compounds in Health Promotion and Disease Prevention: Challenges and Opportunities," *Macromol,* vol. 4, no. 4, pp. 805-842, 2024, doi: https://doi.org/10.3390/macromol4040048.

[7] S. S. Rodrigues, L. G. Dias, and A. Teixeira, "Emerging Methods for the Evaluation of Sensory Quality of Food: Technology at Service," *Current Food Science and Technology Reports,* vol. 2, no. 1, pp. 77-90, 2024, doi: https://doi.org/10.1007/s43555-024-00019-7.

[8] A. I. de Almeida Costa, M. J. P. Monteiro, and E. Lamy, *Sensory Evaluation and Consumer Acceptance of New Food Products: Principles and Applications*. Royal Society of Chemistry, 2024.

[9] S. Jakabová, J. Árvay, M. Šnirc, J. Lakatošová, A. Ondejčíková, and J. Golian, "HPLC-DAD method for simultaneous determination of gallic acid, catechins, and methylxanthines and its application in quantitative analysis and origin identification of

green tea," *Heliyon*, vol. 10, no. 16, 2024, doi: https://doi.org/10.1016/j.heliyon.2024.e35819.

[10] I. Ziani, H. Bouakline, A. El Guerraf, A. El Bachiri, M.-L. Fauconnier, and F. Sher, "Integrating AI and advanced spectroscopic techniques for precision food safety and quality control," *Trends in Food Science & Technology*, p. 104850, 2024, doi: https://doi.org/10.1016/j.tifs.2024.104850.

[11] H. R. Tan and W. Zhou, "Metabolomics for tea authentication and fraud detection: Recent applications and future directions," *Trends in Food Science & Technology*, p. 104558, 2024, doi: https://doi.org/10.1016/j.tifs.2024.104558.

[12] J. Zhang, X. Wu, C. He, B. Wu, S. Zhang, and J. Sun, "Near-Infrared Spectroscopy Combined with Fuzzy Improved Direct Linear Discriminant Analysis for Nondestructive Discrimination of Chrysanthemum Tea Varieties," *Foods*, vol. 13, no. 10, p. 1439, 2024, doi: https://doi.org/10.3390/foods13101439.

[13] L. Wang *et al.*, "Evaluation of the quality grade of congou black tea by the fusion of GC-E-nose, E-tongue, and E-eye," *Food Chemistry: X*, p. 101519, 2024, doi: https://doi.org/10.1016/j.fochx.2024.101519.

[14] X. Zhao, L. Cai, P. Huang, and C. J. A. a. S. Cui, "Enzymatic Synthesis and Sensory Evaluation of N-Cinnamoyl-L-Phenylalanine as a Novel Flavor Enhancer: Impact on Taste Perception and Mechanistic Insights," *Food Chemistry*, 2025, doi: https://doi.org/10.1016/j.foodchem.2025.144542.

[15] J. Liang, J. Guo, H. Xia, C. Ma, and X. J. F. C. Qiao, "A black tea quality testing method for scale production using CV and NIRS with TCN for spectral feature extraction," *Food Chemistry*, vol. 464, p. 141567, 2025, doi: https://doi.org/10.1016/j.foodchem.2024.141567.

[16] M. S. Kabir, M. A. Pathan, K. Tanvir, and D. J. Gomes, "Smart Agriculture Using Advancing Tea Leaf Quality Assessment With Deep Learning," in *Intelligent Internet of Everything for Automated and Sustainable Farming*: IGI Global Scientific Publishing, pp. 149-186, 2025, doi: https://doi.org/10.4018/979-8-3373-0020-7.ch005.

[17] U. Yadav, S. Khurana, K. Rawal, B. Arora, and A. Yadav, "Integrating AI with Electronic Tongue (E-Tongue) for Real-Time Detection in the Food Industry," in *Artificial Intelligence in the Food Industry*: CRC Press, pp. 277-296, 2025,.doi: https://doi.org/10.1201/9781032633602-14.

List of Publications

Journals

[1] **A. Modak**, D. Das, T. N. Chatterjee, B. Tudu and R. Banerjee Roy, "Regression-Based EGCG Detection in Green Tea Employing MIP Electrodes," in *IEEE Sensors Journal*, vol. 24, no. 11, pp. 18090-18097, 1 June 1, 2024, doi: https://doi.org/10.1109/JSEN.2024.3390438.

Conference Proceedings

[1] **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," 2018 Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), Kolkata, India, 2018, pp. 112-117, doi: https://doi.org/10.1109/ICRCICN.2018.8718696.

[2] **A. Modak**, D. Das and R. B. Roy, "Classification of Green Tea Samples Based on Epigallocatechin-3-Gallate Content Using Molecular Imprinting Polymerization Technique," 2024 IEEE 3rd International Conference on Control, Instrumentation, Energy & Communication (CIEC), Kolkata, India, 2024, pp. 355-360, doi: https://doi.org/10.1109/CIEC59440.2024.10468449.

[3] **Modak, A.**, Moulick, M., Roy, R.B. (2025). Efficient Near Infrared Spectroscopy Based Feature Selection of Tannic Acid for Black Tea Evaluation. In: Singh, J.P., Singh, M.P., Singh, A.K., Mukhopadhyay, S., Mandal, J.K., Dutta, P. (eds) Computational Intelligence in Communications and Business Analytics. CICBA 2024. Communications in Computer and Information Science, vol 2367. Springer, Cham. doi: https://doi.org/10.1007/978-3-031-81339-9_13.

Runu Banerjee

Angiras Modak

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