International Journal of Basic and Applied Sciences, 14 (5) (2025) 390-402

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/3s15f186
Research paper

A Hybrid Decision Logic Framework Combining Fuzzy Systems and Machine Learning Classifier for Post-Fermentation Tea Quality Grading

C. M. Sulaikha [1] *, A. Somasundaram [2]

[1] Research Scholar, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India
[2] Assistant Professor, Department of Computer Science, Sri Krishna Arts and Science College, Coimbatore, Tamil Nadu, India
**Corresponding author E-mail: sulaikhacm1984@gmail.com*

Received: July 31, 2025, Accepted: September 5, 2025, Published: September 13, 2025

Abstract

This study addresses the need for reliable post-fermentation tea quality grading by proposing a hybrid decision logic framework that combines narrowband spectrophotometric sensing, computer vision-based fermentation indicators, and intelligent classification techniques. The proposed methodology employs low-cost sensing and digital imaging for feature extraction, followed by a two-stage modeling approach that integrates rule-based fuzzy inference system (FIS) with machine learning (ML) classification. After feature extraction, statistical and ML-based feature selection techniques were applied to identify the most informative features. The initial stage employs a FIS to identify the grade and its confidence score based on recurrent fuzzy patterns corresponding to each tea grade. In the second stage, a light-gradient boosting machine LightGBM classifier is trained to enhance predictive accuracy and generate probabilistic confidence scores. Finally, hybrid decision logic is applied to finalize the grade. The proposed hybrid model achieved an accuracy of 99.00%, precision of 99.53%, recall of 98.27%, and F1-score of 98.74%, outperforming both traditional ML and state-of-the-art models. It demonstrated robust generalization across varying sample sizes and cross-validation folds. Ablation studies confirmed the efficiency of the FIS, showing strong performance with reduced feature sets. Furthermore, the framework maintained shorter execution times than ensemble-based methods, making it suitable for real-time and resource-constrained environments.

Keywords: *Fuzzy Inference System; Tea Quality Grading; Hybrid Decision Logic; LightGBM; XGBoost; Spectrophotometric Sensing.*

1. Introduction

Owing to its distinct flavor, pleasant aroma, health-promoting qualities, and affordability, tea is the second most popular beverage in the world, after water [1]. Originating in East Asia, tea plants have significant cultural and economic importance, especially in India and China, which are major tea-producing countries. Although tea can be categorized into black, green, white, oolong, and Pu-erh based on the processing methods, black tea is particularly popular for its role in improving mental alertness [2]. Its unique oxidation and fermentation processes shape its sensory and chemical characteristics, resulting in an enhanced flavor and aroma [3]. Black tea fermentation, also known as oxidation, is an essential step in the production process during which fresh tea leaves are withered, rolled, oxidized, and dried [4]. This process alters the color and aroma of the leaves via enzymatic reactions. Several chemical compounds, such as theaflavin (TF), thearubigin (TR), theabrownin (TB), catechins (CA), and gallic acids (GA), are affected during oxidation and are closely linked to the final tea quality [5], [6].

Moreover, fermentation (oxidation) is not only essential to developing the tea's characteristic color and aroma, but also plays a major role in determining its quality, market value, and international acceptance [7]. Traditionally, tea assessments are conducted by taste experts who rely on specialized knowledge and skills to evaluate quality, color, flavor, and aroma [8]. Although their expertise is crucial to the industry, it does not encompass all aspects of tea evaluation and is limited by subjectivity, inconsistency, individual differences, and lack of scalability [9]. Additionally, tea production across smallholder farms is increasing, especially in Asia and Africa, and the growing demand from national and international markets has increased the need to develop objective, and effective tea quality grading systems to assist both experts and producers [10], [11].

The advent of real-time sensing technologies, machine learning (ML), deep learning (DL), and artificial intelligence (AI), has created new possibilities for multiple stages of tea production [12], [13]. Various digital tools such as digital or hyperspectral imaging (HSI) [14], [15], near-infrared spectroscopy (NIRS) [16], [17], computer vision systems (CVS) [18], electronic nose (e-nose) [19], UV-visible spectroscopy [20], [21], electronic eyes (e-eyes) [22], and electronic tongue (e-tongue) [23], [24] enable the measurement of physicochemical traits associated with tea, which can be used for quality assessment [17]. However, despite their improved accuracy, these advanced methods often involve expensive feature-extraction processes, complex interpretability, and limited practical viability under field conditions.

Existing systems continue to face challenges owing to their reliance on expensive equipment and highly intricate models. Although convolutional neural networks (CNNs) and other DL approaches have shown promise in visual tea leaf grading [25], issues of transparency remain due to their black-box nature. This is particularly problematic in domains where expert interpretation or validation of model outputs is essential [26]. Additionally, because of their high cost and operational complexity, systems that use aroma or near-infrared sensors are not appropriate for use in remote areas or by small-scale producers, who frequently lack the funding necessary to maintain such technology [27].

Another key limitation is that most intelligent grading systems prioritize accuracy as the primary objective while overlooking other important factors such as human interpretability, cost-effectiveness, and the integration of expert knowledge. In quality control or quality-enabled grading mechanisms, this limitation increases the risk of algorithmic opacity and hinders wider adoption [28]. Additionally, while considerable research has been conducted in China regarding tea quality assessment [29], limited attention has been paid to the Indian tea industry, despite India's significant contribution to the global tea supply [30]. As a result, hybrid decision models that bridge the gap between the statistical power of ML and the conventional reasoning used by human experts are urgently needed [31].

Thus, the objective of the study is to propose a hybrid fuzzy–ML framework for the automated grading of fermented tea quality using sensing and computer vision. The methodology comprises several stages. Initially, chi-square and Extreme Gradient Boosting (XGBoost) were used to select the most relevant and non-redundant features, which were then carried forward to the next stage. In the first stage, a fuzzy inference system (FIS) was designed to simulate expert decision logic by converting normalized visual and chemical features into qualitative scores through rule-based reasoning. In the second stage, a Light Gradient Boosting Machine (LightGBM) classifier employs all features to enhance grading accuracy. Furthermore, a confidence score is used for both fuzzy and ML predictions, based on which a hybrid decision logic is applied for the final decision.

FIS was used due to its ability to express the domain experience of tea tasters in an algorithmic format. Fuzzy logic is particularly effective at modeling subjective knowledge using interpretive rules. It enables interpretable and customizable rule-based decision-making while also supporting language-based reasoning [32]. However, since fuzzy systems alone may not handle high-dimensional data effectively or produce robust predictions, a hybrid framework is proposed that combines human-like reasoning with high-performance computational learning [33]. The ML classifiers, such as XGBoost and LightGBM, were chosen after conducting extensive experimental analysis. The primary motivation for this work is to improve the interpretability, accuracy, and cost-efficiency of automated tea grading solutions, particularly those suitable for small-scale producers. The use of fuzzy logic also enhances the transparency of decisions, aligning well with the increasing consumer demand for explainability in agro-processing systems.

The remainder of this paper is structured as follows: Section 2 reviews the current literature on the study field. Section 3 outlines the proposed methodology, including data acquisition, feature engineering, fuzzy logic rule formulation, ML classifiers, and hybrid decision logic. Section 4 presents an experimental analysis, detailing performance evaluation strategies, ablation studies, and comparisons with standard and state-of-the-art classifiers. The study's limitations and practical implications are discussed in Section 5. Finally, Section 6 concludes the study with recommendations for future research directions.

2. Related Works

Recent advancements in sensors, computer vision, and machine-learning technology have enabled innovative techniques for evaluating food and beverage quality. The following are some studies that focus on the various methods used for tea and similar items. Gill et al. (2013) employed the multilayer perceptron (MLP) technique on physical parameters using texture analysis and achieved an accuracy of 82.33% [34]. However, because the method relies solely on texture-based grey-tone spatial features, it may overlook other important visual cues, such as color or shape. To detect the aroma of beer, Shi et al. (2019) proposed a deep feature mining method using e-nose sensor data by combining a convolutional neural network (CNN) with a support vector machine (SVM), with parameters optimized through improved particle swarm optimization (PSO) [19]. Although effective, the model limits its generalizability to other beverage types and sensory modalities. Extending this idea, Yu and Gu (2021) developed a similar framework for classifying the six sub-categories of Maofeng and Maojian green tea using e-nose data [35]. Their CNN-SVM model demonstrated strong accuracy and robustness, but demanded high computational resources and expert-level tuning for efficient model training and deployment.

Similarly, Li et al. (2021a) proposed a model-based scheme using the stepwise regression method (SRM) to estimate tea quality grades by detecting 19 chemical substances that influence Huangshan Maofeng tea quality [5]. They also evaluated the elastic net (EN) method by Kelly et al. (2012) [36], which offers high classification accuracy while retaining correlated features in neural decoding. However, for the purpose of simplifying chemical detection and achieving efficient batch tea quality-grade estimation, the stepwise regression method (SRM) proved more effective in our study, demonstrating better prediction accuracy with fewer features. This contextualizes the use of EN as a comparative benchmark while highlighting the suitability of SRM for our framework. Similarly, Suhandy and Yulia (2019) applied UV-visible spectroscopy with partial least squares discriminant analysis (PLS-DA) to distinguish two grades of Indonesian black tea [20]. The PLS-DA model successfully distinguished pre-treated samples by utilizing mean normalization and moving average smoothing. However, these models require complex spectral preprocessing and expert-driven chemical profiling, which limit their scalability and practical deployment in resource-constrained environments.

Ding et al. (2024) employed HSI and principal component analysis (PCA) to extract features and developed ResNet-50 models for tea quality classification [15]. With transfer learning (TL) and augmentation, the model produced an accuracy of 86.15%, and PSO improved it to 89.23%, while combining both via Two-Strategy PSO (TSPSO) achieved 92.31%. However, reliance on hyperspectral equipment and extensive model adaptation may hinder their applicability in cost-sensitive or field-based grading scenarios. In a related domain, Oliveira et al. (2021) applied computer vision to extract imaging and handcrafted features from fermented cocoa beans, which are used as predictors in random decision forests (RDF) classifier [37]. The model achieved 0.93 accuracy on an unbalanced dataset and 0.92 on a balanced dataset. Although suitable for industrial classification, it may limit adaptability and generalization and is sensitive to bean variability. Liu et al. (2019) proposed a multi-task model based on a back propagation neural network (MBPNN) for grade classification and quality prediction of organic green teas using an e-nose [38]. Compared with random forest (RF) and SVM, MBPNN showed efficiency; however, relying only on e-nose data may not capture all chemical components. In their study, Zhao et al. (2024) utilized DL and image recognition to classify tea leaf grades using the YOLOv8x-SPPCSPC-CBAM model, thereby enhancing the results on scattered and stacked leaves [39]. However, heavy data augmentation may reduce the generalizability and overlook the grading variability from mechanical handling.

Li et al. (2023a) used a CVS and miniature NIRS to predict the degree of fermentation of Pu-erh tea [18]. The model used the least-squares SVM (LSSVM) for analysis. With CVS data, the LSSVM with fusion features delivered more accurate performance than NIRS features evaluated using the LSSVM with standard normal variate (SNV) and successive projections algorithm (SPA). However, environmental conditions, batch variability, and manufacturing scale can affect quality monitoring accuracy. Li et al. (2023b) employed micro-NIRS, CVS, and a colorimetric sensor array to collect data from Keemun black tea using LSSVM, an extreme learning machine, and PLS-DA to qualitatively discriminate different tea grades [40]. The LSSVM with mid-level data fusion achieved 98.57% accuracy. Although effective, integrating various sensing technologies increases the system cost and complexity, requiring careful pre-processing and alignment of heterogeneous data.

Zhou et al. (2023) used CVS and e-nose with feature/data-level fusion [41]. Their analysis reported that SVM outperformed other classifiers, but incurred higher implementation costs and reduced robustness. Xia et al. (2024) utilized an e-nose and NIRS to evaluate tea quality and found that e-nose achieved 97.00% accuracy and NIRS 99.63% [42]. When these features were fused, the SVM, K-nearest neighbors (KNN), and artificial neural network (ANN) achieved accuracies of 98.13%, 96.63%, and 97.75%, respectively. However, obtaining well-calibrated instruments and representative data in industrial settings can be challenging for obtaining accurate quantitative forecasts. Ren et al. (2024) developed a deep CNN classification model that incorporates cross-sensor multimodal fusion features, including color, texture, shape, optical, e-tongue, and e-nose [24]. With a misclassification rate of only 0.86%, the CNN algorithm outperformed the existing classification methods in evaluating the sensory and quality grades of tea samples.

Liang et al. (2025) integrated CVS and NIRS using a multimodal fusion method [43]. The fused Temporal Convolutional Network (TCN) model achieved 98.2% accuracy, although its high computational requirements may limit its real-time deployment. Wu et al. (2025) demonstrated that low- and mid-level data fusion strategies performed better than single data models and proposed an improved classification model that integrated spectral and image data [44]. The model employs the SPA, PCA, and KNN, achieving 100% accuracy for Anji white tea. Zhang et al. (2025) introduced the Grouped Convolutional Hybrid Attention Mechanism (GCHAM), which achieved 95.33% accuracy but faced power and processing limitations [45].

Despite these advancements, there are still some limitations. Many models are not feasible for scalable, real-time applications because they rely on single-modality inputs or require sophisticated tools, such as high-precision e-nose, micro-NIRS, or HSI. The adoption of these technologies is further hindered by limited generalizability due to handcrafted features, high computational demands, the need for expert intervention, and high costs. In contrast, the proposed hybrid framework directly addresses these gaps by leveraging low-cost narrowband spectrophotometry and simple computer vision indicators, which reduce hardware dependency and acquisition costs. Unlike Ding et al. [15] and Zhu et al. [25], which relies on costly hyperspectral imaging, our framework achieves higher accuracy while maintaining interpretability through FIS by using low-cost spectrophotometric sensing and smartphone imaging. Furthermore, by integrating a fuzzy inference system with LightGBM, the framework enhances interpretability and ensures computational efficiency, overcoming the limitations of CNN-based approaches that are often resource-intensive and opaque. Thus, this study introduces a practical, explainable, and resource-friendly alternative for robust tea quality grading. These enhancements directly address the limitations highlighted in prior studies, particularly the high cost of hyperspectral imaging and the lack of interpretability in CNN-based approaches.

3. Methodology

This study proposes a hybrid grading framework that integrates expert knowledge through a FIS with data-driven learning using a ML classifier. The framework is illustrated in Fig. 1 and begins with expert-labelled tea samples, followed by data pre-processing and feature selection to extract significant and meaningful features. In the training phase, the selected features are first evaluated using a fuzzy system in which rules are generated using rule-based logic. A confidence evaluation layer was then applied to assess the reliability of the fuzzy prediction. If the confidence falls below a predefined threshold, the system delegates the grading task to a LightGBM classifier trained on the full feature set. During the testing phase, both the FIS and LightGBM were applied, and the final output was determined using a hybrid decision logic to ensure transparent and accurate grading. The interpretability and adaptability of the model make it suitable for post-fermentation tea quality evaluation.

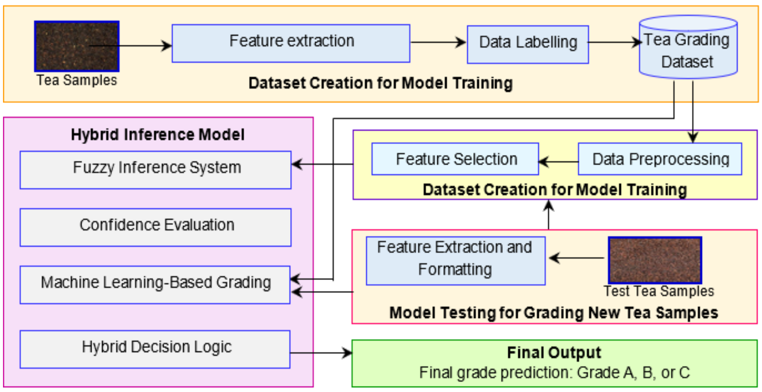

Fig. 1: Proposed Hybrid Decision Logic-Based Framework for Tea Quality Grading.

3.1. Sample Preparation

Using quality indicators that capture the chemical and physical characteristics relevant to post-fermentation tea evaluation, a comprehensive feature vector was generated for each fermented tea sample through narrowband spectrophotometric sensing and computer vision. These features reflect the biochemical transformations and visual properties necessary for accurate tea grading. Experiments were conducted on three classes of fermented tea samples, representing high, medium, and low grades. The fermented black tea samples were collected from a local tea factory in Tamil Nadu, India, over 10 days, starting from June 15, 2025, after expert grading. For each sample, 2.0 grams of fermented tea granules were taken from each package. Image analysis was performed first, followed by the extraction of sensing parameters. A total of 500 samples were analyzed, with approximately 300 high-grade, 150 medium-grade, and 50 low-grade samples. This dataset reflects batch-level variability within black tea, while future studies will include samples from multiple factories and cultivars to enhance generalizability. The data were processed, features were extracted, and the results were stored in a CSV file.

Chemical properties such as TF, TR, and TB contents were estimated using a DIY visible-light spectrometer assembled with a digital camera, a DVD-based diffraction grating, a cuvette, and a dark enclosure to reduce ambient light interference [46]. For each sample, 2.0 grams of fermented tea leaves were infused in 20 mL of hot distilled water (90–95 °C) for 5–10 minutes with gentle stirring. After cooling to room temperature, the infusion was filtered, and the clear extract was transferred to a cuvette for spectral measurements.

Absorbance values were recorded at specific wavelengths corresponding to pigment compounds: 380–385 nm for TF, 460 nm for TR, and 480–500 nm for TB. The absorbance spectra were captured using open-source tools such as Public Lab's Spectral Workbench, and three spectral scans were averaged to reduce variability. From these data, pigment ratios such as TF/TR, termed the Briskness Index (BI), and TF/(TR + TB), termed the Extended Oxidation Index (EOI), were computed to evaluate oxidation dynamics and brightness ratios [47]. Using trapezoidal numerical integration, the Polyphenol Index (PI) was determined as the area under the absorbance curve between 350 and 500 nm, and the Spectral Slope (SS) was obtained from the absorbance difference between 380 and 500 nm.

Owing to their stability and potential health benefits, catechins and gallic acid are frequently employed in tea chemistry research; however, their measurement typically requires advanced technologies such as high-performance liquid chromatography (HPLC). Instead, fermentation-induced pigment changes (TF, TR, and TB) were monitored using this low-cost, non-invasive visible-light spectrometry technique. These changes are directly associated with oxidation and can be visually identified using absorbance measurements.

After the tea infusion was cooled, the pH level was determined using a digital pH meter, which indicates the acidity and microbial stability of the liquor. Total Soluble Solids (TSS) were measured using a handheld analog refractometer and expressed in °Brix. A pin-type digital moisture meter was used to calculate moisture content (MC) in percentage by estimating the electrical resistance. To estimate flavor concentration, the TSS-to-Moisture Ratio (TSS-MR) was computed using the °Brix and moisture values.

Color and visual consistency were assessed using a digital camera positioned 25 cm above the sample, accompanied by a circular light source to ensure consistent lighting. The tea liquor was stored in a white or transparent vessel to reduce background interference. Images were acquired under constant resolution and lighting conditions and calibrated to extract color characteristics. The color index (L, a, b*) was derived from the images processed in the CIELab color space. The following derived features were then computed: the color uniformity score (CUS), defined as the standard deviation of the L* or a* values across various locations, reflecting visual consistency. The color Variance (CV) denotes the pixel-wise variance across the RGB or Lab channels, indicating intra-sample heterogeneity. The fermentation color ratio (FCR) was calculated using RGB values obtained from the digital images. These visual features reflect key sensory attributes, such as liquor clarity, brightness, and aromatic characteristics. The details are presented in Table 1.

Table 1: Summary of Extracted Tea Features, Formulas, and Corresponding Measurement or Computation Methods.

Feature	Representation	Formula	Extraction Method
Theaflavin (TF)	Brightness, briskness	$A_{380} = \log_{10}\left(\frac{I_0}{I_{380}}\right)$ I_0 - the intensity of the incident (reference) light I_{380} - the intensity of the transmitted light at 380 nm	Absorbance at 380–385 nm via DIY spectrometer
Thearubigin (TR)	Body, liquor depth	$A_{460} = \log_{10}\left(\frac{I_0}{I_{460}}\right)$ I_{460} - the intensity of the transmitted light at 460nm	Absorbance at 460 nm
Theabrownin (TB)	Over-fermentation, dullness	$A_{500} = \log_{10}\left(\frac{I_0}{I_{500}}\right)$ I_{500} - the intensity of the transmitted light at 500 nm	Absorbance at 500 nm
Briskness Index (BI)	indicates the relative dominance of brightness (TF) over color/body (TR)	TF/TR	Computed from absorbance values
Extended Oxidation Index (EOI)	Indicates the relative dominance of TF compared to the more oxidized components TR and TB	TF / (TR + TB)	Computed from absorbance values
Polyphenol Index (PI)	Total pigment activity	$\int_{380}^{500} A(\lambda)d\lambda$	Integration of absorbance at across 350–500 nm
Spectral Slope (SS)	Oxidation gradient	$\frac{(A_{500} - A_{380})}{500 - 380}$	Derived from TF and TB absorbance
CIELab	Color and visual quality	L and a* component	Extracted via smartphone colorimeter apps under uniform lighting
Color Uniformity Score (CUS)	Consistency of fermentation	$\sigma(L^*)$ or $\sigma(a)$	Std. deviation (σ) of L* or a* from multiple image regions
Color Variance (CV)	Visual inconsistency	Pixel-wise variance across Lab or RGB values	Computed from digital image analysis
Moisture Content (MC)	Residual drying efficiency	-	Digital moisture meter
Total Soluble Solids (TSS)	Concentration of dissolved solids	-	Handheld analog refractometer
TSS-to-Moisture Ratio (TSS-MR)	Flavor intensity index	$\frac{TSS}{Moisture}$	Computed from TSS and moisture readings
pH Level	Acidity, flavor, microbial stability	-	Digital pH meter
Fermentation Color Ratio (FCR)	Oxidation and fermentation extent	$FCR_{RG} = \frac{R}{G}$; $FCR_{RB} = \frac{R}{GB}$	Computed from RGB values in digital liquor images

3.2. Data Preprocessing and Feature Evaluation

A systematic dataset was compiled using fermented tea samples to assess the hybrid tea-grading framework. Each sample was characterized by a detailed feature vector comprising 17 quantitative variables that describe both chemical transformations and physical properties relevant to post-fermentation assessment. These features include: 1) chemical composition indicators (related to fermentation and oxidation), such as TF, TR, TB, and PI; 2) visual appearance attributes such as CIELab color parameters (L*, a*), the fermentation color ratio derived from RGB values, color uniformity score, and color variance; 3) physicochemical properties such as moisture content, total soluble solids

(TSS), and pH level; 4) derived indices including TF/TR ratio, TF/(TR+TB), spectral slope, and the TSS-to-moisture ratio. Grade labels, such as High, Medium, or Low, are assigned to each sample based on expert-validated reference data and available historical data, serving as the ground truth for supervised learning and fuzzy rule construction.

Upon creating the dataset, all feature values were standardized using z-score normalization, which adjusts each feature to have a mean of zero and a standard deviation of one. This normalization ensures comparability across features with different units and scales, preventing bias in models that are sensitive to feature magnitude. Identified missing values are addressed through mean imputation, wherein missing entries are replaced with the mean of observed values for that feature, maintaining dataset completeness and minimizing distribution distortion.

A dual-feature assessment was then conducted to determine the most informative and relevant features for the next phase. First, a statistical approach using the chi-square (χ^2) test was applied to evaluate the degree of association between each feature and the categorical grade labels. Features were ranked based on their χ^2 scores, where higher scores indicate stronger predictive capability for tea grade classification. This test is suitable for categorical outcomes such as grading and does not assume linearity. Second, a model-based interpretability method was employed by training an XGBoost classifier on the full feature set. The contribution of each feature to classification performance was measured by extracting feature importance scores from the trained model. To ensure a balanced evaluation, the average rank of each feature was calculated by combining the rankings from both the chi-square test and XGBoost scores. Features were then sorted based on these average ranks to reflect their overall relevance.

To address multicollinearity, a correlation matrix was computed, and redundant features were eliminated. In cases of highly correlated feature pairs (Pearson correlation coefficient > 0.85 based on empirical analysis), the feature with the lower average rank was removed through a rank-guided pruning step. This ensured that the final feature set used in the fuzzy inference system (FIS) was both informative and non-redundant. To determine the optimal number of features to retain, the Elbow method was applied. This involved training a classifier multiple times using the top-k ranked features, with k ranging from one to the total number of remaining features. The “elbow point” on the resulting performance curve was used to identify the trade-off between model simplicity and predictive power for the hybrid tea grading system. Based on this analysis, a reduced subset of seven features was selected for fuzzy rule construction.

Each selected feature was discretized into linguistic categories, such as Low, Medium, and High, and corresponding membership functions were constructed. These linguistic terms were then used to formulate fuzzy rules by combining expert insights with statistically derived thresholds. Furthermore, the ML-driven grading component employs residual features, which are those not selected for fuzzy rule construction. These features capture higher-dimensional relationships and nonlinear interactions that are better handled by data-driven techniques. By integrating statistically validated fuzzy inputs with ML-derived features, the hybrid model ensures both transparent, rule-based interpretability and adaptive classification performance, making it suitable for real-world tea grading applications.

3.3. Hybrid Grade Inference Module

3.3.1. Fuzzy Inference-Based Grading System

A FIS was established as the foundational layer of the hybrid grading framework that resembles the cognitive procedures of experienced tea graders. Owing to the lack of clear definitions for tea quality variables, such as color, flavor strength, moisture content, and fermentation depth, fuzzy logic works well in this situation. In contrast to conventional binary or threshold-based systems, fuzzy logic handles the inherent variability present in real-world tea samples and enables seamless transitions between quality levels, which are best described in words that are easy for humans to understand.

Fuzzification:

Using data-driven membership functions, each selected input feature was converted into fuzzy linguistic variables classified as Low, Medium, or High. Percentile-based triangular membership functions were employed to determine the membership range parameters, ensuring that the fuzzy partitions accurately reflected the distribution of each feature. Specifically, the triangular membership function parameters (a, b, c) for the Low, Medium, and High categories were defined using the 10th, 50th (median), and 90th percentiles of the normalized feature values, respectively. This method is more flexible than fixed ranges or min-mean-max definitions because it accounts for data skewness and outliers, which are frequently observed in chemical or sensory properties.

For example, after normalization, assume that the 10th, 50th, and 90th percentiles of the feature EOI are 0.25, 0.55, and 0.8, respectively. The triangular membership functions are defined as follows: Low: (0.0, 0.25, 0.5), Medium: (0.4, 0.55, 0.7), and High: (0.6, 0.8, 1.0). This configuration facilitates overlapping domains where a specific input value may partially belong to multiple categories. Such overlapping allows the model to represent subtle transitions more accurately than crisp thresholds do.

For instance, an EOI value of 0.65, above 0.5, has no membership in the Low category, but partially belongs to both Medium and High, more effectively capturing complex transitions than rigid boundaries. The triangular membership function is mathematically expressed as in (1):

$$\mu(x) = \begin{cases} 0 & x \le a \\ \frac{x-a}{b-a} & a < x \le b \\ \frac{c-x}{c-b} & b < x < c \\ 0 & x \ge c \end{cases} \tag{1}$$

For input values of x, fuzzy membership values were determined using the triangular membership functions defined for the three categories: Low (0.0, 0.25, 0.5), Medium (0.4, 0.55, 0.7), and High (0.6, 0.8, 1.0). Since 0.65 is greater than the upper bound of Low (0.5), its membership in Low is zero. For Medium, as 0.65 lies between 0.55 and 0.7, the membership is computed using the formula $\frac{c-x}{c-b}$, resulting $\frac{0.7-0.65}{0.7-0.55} = 0.33$. For High, since 0.65 lies between 0.6 and 0.8, the membership is calculated using the formula $\frac{x-a}{b-a}$ yielding $\frac{0.65-0.6}{0.8-0.6} = 0.25$. Therefore, at x = 0.65, the degrees of membership are 0.33 in Medium, 0.25 in High, and 0 in Low.

Fuzzy Rule Base and Inference:

The system employs a series of IF–THEN rules derived from expert knowledge and the analysis of labelled data. These rules reflect the grading reasoning that human tea evaluators usually employ. The first step in the rule-generating procedure is to examine the labelled data to identify the recurrent feature combinations associated with particular grades. These patterns were then verified and improved by experts to refine the comprehensible linguistic rules.

Every rule employs the minimum operator to calculate the degree of rule activation (firing strength) and the maximum operator to aggregate the outcomes over all applicable rules. The outcomes of each active rule are combined using the Mamdani inference method to create a fuzzy output set that represents the predicted grade. This inference is suitable for this study because of its ability to generate interpretable fuzzy outputs and conform to human decision logic [48]. Without depending on complex mathematical models, it enables domain experts to visualize and improve grading rules. Its structure facilitates the handling of linguistic variables, as well as intuitive reasoning.

The firing strength (α_r) for Rule 1 was then calculated as min(0.25, 0.60, 0.70, 0.80, 0.90, 0.85) = 0.25, showing that Rule 1 contributes to Grade A with a firing strength of 0.25. After calculating the firing strengths for every derived rule, the results were aggregated for each grade. The grades were mapped as [A, B, C] → [3, 2, 1] using Mamdani inference, and the total firing strengths were calculated as [0.25, 0.75, 0.60].

Defuzzification:

Since, there are three categories in the fuzzy output variable "Grade": Grade A, Grade B, and Grade C, triangular membership functions across the numerical range of 1 to 3 are used to depict these grades. Grade A peaks at 3, Grade B at 2, and Grade C at 1. Using the centroid defuzzification approach, the final crisp grade was determined as in (2):

$$\text{Defuzzified Grade} = \frac{\sum_i \mu(G_i).G_i}{\sum_i \mu(G_i)}, G_i \in \{1,2,3\} \tag{2}$$

where $\mu(G_i)$ is the aggregated firing strength for each grade.

Thus, for the above example with the total firing strengths [0.25, 0.75, 0.60], defuzzified grade is computed as $\frac{(1\times0.60)+(2\times0.75)+(3\times0.25)}{0.60+0.75+0.25} = 1.78$ This is then rounded to the nearest whole number to assign the final crisp grade as 2, corresponding to Grade B.

The fuzzy grading system effectively modeled the ambiguous and overlapping characteristics of tea quality perceptions. It provides a structured and flexible approach to replicate the decision-making process of expert graders. Because precise thresholds for variables, such as color, pH, and oxidation, are difficult to define, fuzzy logic allows experts to formulate approximate but consistent rules. The use of Mamdani inferences significantly improves interpretability. Rules can be audited, adjusted, or expanded by domain experts, without requiring deep statistical or programming knowledge. This makes the fuzzy grading module technically sound and practically usable in tea evaluation contexts.

3.3.2. Confidence Evaluation

Even when fuzzy systems are used, there are situations where they are unclear or overlap, such as when a sample is almost evenly in Grade B or Grade C, or when its color and moisture levels are on the boundary line. To address these uncertainties, fuzzy systems have a confidence evaluation layer that measures the reliability of their predictions.

The fuzzy system determines a confidence score after generating a grade prediction based on the strength of the rules that support the predicted grade. For instance, if the system predicts Grade B, it evaluates the number of Grade B rules triggered and the aggregate strength of their activations and compares them with the total activation of all triggered rules. Thus, for each fuzzy-inferred grade G_i , the confidence score is computed as in (3).

$$\text{Conf}\left(G_{\text{fuzzy}_i}\right) = \frac{\sum_{r\in R_i} \alpha_r}{\sum_{r\in R} \alpha_r} \tag{3}$$

Where:

- R_i is the set of fuzzy rules suggesting grade G_i.
- α_r is the firing strength of rule r, which is calculated as the minimum of its input membership value.
- R is the set of all triggered rules regardless of grade.

A threshold τ (0.60), calculated empirically, was used to evaluate fuzzy reliability. If the predicted grade confidence level $\text{Conf}\left(G_{\text{fuzzy}_i}\right)$ is less than the confidence threshold τ, the sample is automatically sent to a supervised ML classifier for secondary evaluation.

Thus, if the total rule activation (i.e., the sum of the firing strengths of all triggered rules) is 1.60 and the sum of the firing strengths of rules that support Grade B is 0.75, then the confidence for Grade B is computed as 0.75 / 1.60 = 0.469. Because this value is lower than the threshold τ=0.60, the confidence of the fuzzy system is considered low, and the sample is regarded as borderline. Consequently, it was sent to a ML classifier for further analysis.

3.3.3. Machine Learning-Based Grading

As outlined in Section 3.3.2, if the confidence of the fuzzy system is inadequate, the ML-based grading module provides a secondary decision path. A LightGBM-based ML model was employed to handle borderline grading scenarios identified using the FIS. LightGBM was selected for its superior classification accuracy, efficiency in processing structured data, ability to capture nonlinear feature interactions, and scalability to large datasets with moderate dimensionality [49]. The use of histogram-based decision tree learning and regularization techniques mitigates overfitting while enabling rapid parallelizable training. The model was trained using a consistent 17-dimensional feature vector extracted from each fermented tea sample, reflecting the chemical, visual, sensory, and physicochemical indicators which were assigned with an expert-labelled classification: Grade A (3), Grade B (2), or Grade C (1). During classification, for a new sample x, the LightGBM generates a probability vector P = [P_A, P_B, P_C], where P_A is the likelihood of grade A, P_B denotes the chance of grade B, and P_C indicates the probability of grade C. The predicted grade $\hat{y}$ was selected as in (4).

$$\hat{y} = \arg\max(P_A, P_B, P_C) \tag{4}$$

The maximum value among these probabilities (P_A, P_B, and P_C) was used as the model's internal confidence score $\text{Conf}(G_{\text{ML}_i})$ for the prediction grade G_i. The LightGBM classifier functions as a supplementary evaluation tool and is activated only when the confidence level of the FIS falls below a predefined threshold.

3.3.4. Grade Integration and Hybrid Decision Logic

The final grade for each sample was ascertained using a hybrid decision fusion methodology that integrates the outputs from the FIS and LightGBM. When properly implemented, this integration improves the accuracy of the system without compromising interpretability.
Let G_{fuzzy} be the grade predicted by the fuzzy system, Conf(G_{fuzzy}) be its corresponding fuzzy system confidence score, G_{ML} be the grade predicted by the LightGBM model, and Conf(G_{ML}) be the highest probability assigned by LightGBM to the predicted grade. The hybrid logic is defined as in (5).

$$G_{final} = \begin{cases} G_{fuzzy} \text{ if } Conf(G_{fuzzy}) \geq \tau \\ G_{ML} \text{ if } Conf(G_{fuzzy}) < \tau \end{cases} \quad (5)$$

where τ is the fuzzy confidence threshold, empirically set (0.60).
If the predictions from the fuzzy and ML models differ by a single-grade level (e.g., B versus C), a tie-breaking procedure is implemented. In such cases, if the LightGBM exhibits high confidence (Conf(G_{ML})≥0.90), its output is deemed acceptable. This scenario is specifically considered since a difference of one grade could arise from borderline features or sensor fuzziness, making it reasonable to resolve the conflict using confidence-based logic. In other cases, if the ML model's confidence is low, the fuzzy prediction is preferred. For larger disagreements (i.e., more than one grade apart), the final decision defaults to the ML output.
All thresholds, including the fuzzy confidence threshold (τ) and LightGBM probability cutoffs, were calibrated using a validation set to align the hybrid system with expert grading judgments. The final output, G_{final}, was returned as the predicted tea grade for post-fermentation quality assessment.

4. Experimental Analysis and Results

4.1. Experimental Setup

The proposed hybrid tea grading system was implemented on a 64-bit Windows 10 operating system with an Intel(R) Pentium(R) 6405U CPU @ 2.40GHz and 64 GB RAM. After extracting data samples, as explained in Section 3.1, all experiments were conducted using Python 3.8.20 within the Jupyter Notebook environment configured via Anaconda Navigator 2.6.4. Core general libraries, including pandas, numpy, and scipy were used for data preprocessing, z-score normalization, imputation, sklearn for statistical analysis, feature selection, and model validation, and skfuzzy for implementing the FIS. For ML, XGBoost and LightGBM classifiers were developed using the xgboost and lightgbm Python packages, respectively. All visualizations and result analyses were performed using matplotlib, seaborn, and plotly.
A FIS was developed using a scikit-fuzzy (skfuzzy) library. Each selected feature was mapped to linguistic terms (Low, Medium, High) using triangular membership functions, with parameters determined based on the percentile values (25th, 50th, and 75th) from the dataset. Mamdani-type fuzzy inference was applied and rule activation was calculated using the minimum operator. A fuzzy rule base was constructed based on expert knowledge and prior grading data. For each sample, the firing strengths of all applicable rules were computed and the fuzzy outputs were aggregated. Defuzzification was performed using the centroid of area method, producing a crisp numerical Grade A rounded off to assign Grade A (3), B (2), or C (1).
Model training for the proposed framework was performed using 10-fold cross-validation to ensure generalization, and the hyperparameters were optimized through a grid search to improve the predictive performance. Thus, in XGBoost, a model used for feature importance analysis, the number of classes (num_class) was defined as three and the target was set to 'multi:softprob'. Other parameters included a learning_rate of 0.3, max_depth of 6, min_child_weight of 1, gamma set to 0, and both subsample and colsample_bytree set to 1.0. The regularization terms included lambda (L2) at 1 and alpha (L1) at 0. The evaluation metric is "mlogloss," and the tree booster that is employed is "gbtree." For the multiclass classification models used in this study, the following hyperparameters were employed for LightGBM, which produced a softmax-based probability distribution over grades, with the highest probability determining the predicted grade and serving as the model's internal confidence score. The model was set with objective as 'multiclass,' boosting_type as 'gbdt,' and the required num_class specified. Other parameters included a learning_rate of 0.1, num_leaves of 31, max_depth of -1, min_data_in_leaf of 20, feature_fraction and bagging_fraction both at 1.0, and bagging_freq of 0. The regularization parameters were set as lambda_l1 = 0.0 and lambda_l2 = 0.0, and the evaluation metric was 'multi_logloss.'

4.2. Feature Selection and Analysis

Feature selection was performed using both statistical and model-driven techniques. The Chi-square test and the XGBoost classifier were applied to evaluate the relevance of the feature to the target class. The importance scores obtained using both methods are summarized in Table 2. The results indicated that EOI, TB, BI, FCR_RG, pH Level, TR, TF, TSS-MR, and CV were highly influential in classifying tea grades, justifying their inclusion in the FIS.

Table 2: Ranking of Tea Quality Features by Chi-Square and Xgboost for Classification

Feature Name	Chi2 Score	Chi2 Rank	XGBoost Importance	XGBoost Rank	Average Rank
Extended Oxidation Index	122.1194	1	0.1099	3	2
Theabrownin	60.8833	4	0.3576	1	2.5
Briskness Index	63.4206	3	0.2785	2	2.5
Fermentation Color Ratio_RG	15.1245	5	0.0819	4	4.5
pH Level	11.3404	6	0.0751	5	5.5
Thearubigin	8.7078	7	0.0231	6	6.5
Theaflavin	67.8988	2	0.0043	13	7.5
TSS-to-Moisture Ratio	3.5622	8	0.0088	9	8.5
Color Variance	0.9920	12	0.0212	7	9.5
Fermentation Color Ratio_RB	1.8809	10	0.0061	10	10
Moisture Content	1.8943	9	0.0015	14	11.5
Spectral Slope	1.5963	11	0.0049	12	11.5

Color Uniformity Score	0.1739	16	0.0208	8	12
CIELab_L	0.2034	15	0.0053	11	13
Total Soluble Solids	0.5600	14	0.0005	15	14.5
Polyphenol Index	0.6103	13	0.0003	16	14.5
CIELab_a	0.1224	17	0.0003	17	17

The correlations among the various features were computed, and the findings are presented in Fig 2. Because of the strong correlation (exceeding 0.85) between BI and TF with EOI, these two features were eliminated to reduce redundancy. The decision to maintain EOI while excluding BI and TF was based on their average importance scores, with EOI possessing a superior total ranking. Following correlation-based pruning, the Elbow method was applied to the sorted and refined feature subsets. In this phase, a decision tree classifier is iteratively trained using the top-k features, with k ranging from one to the total number of remaining features. Cross-validated accuracy, a performance statistic, was calculated for every k using stratified 5-fold cross-validation. The elbow point on the curve, which marks where adding more features results in minimal performance gain, is used to determine the ideal number of features. Accordingly, using seven features yielded maximum accuracy in this study. Consequently, EOI, TB, FCR_RG, pH Level, TR, TSS-MR, and CV were the top seven features selected for further analysis. These features collectively encapsulate the essential sensory dimensions of tea quality, including taste, texture, aroma, and visual appearance, and closely align with the criteria generally assessed by human specialists during manual grading.

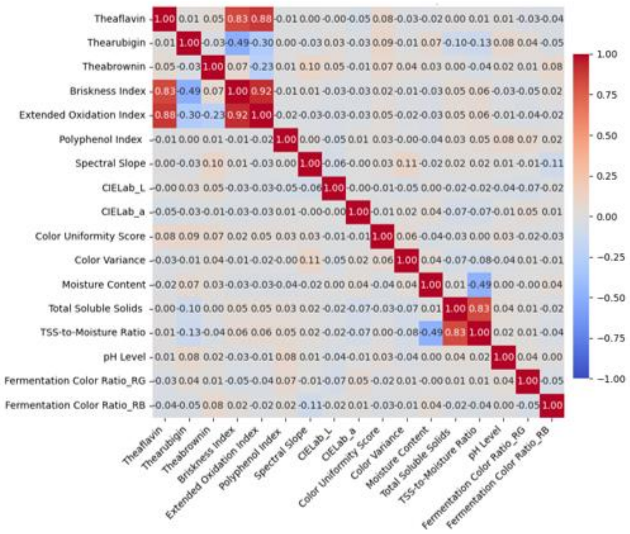

Fig. 2: Feature Correlation Matrix Highlighting Relationships Among Tea Quality Indicators.

Ablation research was performed to assess the efficacy of the FIS grading utilizing a decreasing number of top-ranked features from 17 to 3. Fig 3 illustrates the results, which show a consistent improvement in classification performance as redundant or less useful features were eliminated. The model obtained only 2.4% accuracy with all 17 features and an F1-score of 7.5%, suggesting that using all features could increase noise and make generalization more difficult. Both the accuracy and F1-score have considerable improvements as the number of features decreased. When the top seven features were included, there was a noticeable improvement in performance, as seen in the 88.4% accuracy and highest F1-score in the analysis (83.6%). Remarkably, the most effective accuracy (91.3%) was achieved with only three features; nevertheless, the F1-score was marginally lower at 80.9%, indicating that there may be some imbalance in the prediction. These results demonstrate that improved fuzzy classification results can be achieved through careful feature selection. Compact and high-quality feature subsets are more valuable than exhaustive feature inclusion, as demonstrated by performance trends.

As the number of features (k) increases, there are apparent trends in the performance of the model, as determined by comparing the accuracy and F1-measure of various feature selection techniques. As shown in Fig. 4, the proposed approach, which combines fuzzy inference with a feature selection method that identifies the most relevant and least redundant features, consistently outperforms conventional techniques such as Chi-square, Mutual Information, F-statistic, Correlation, ReliefF, Random Forest, Logistic Regression, Decision Tree, Extra Trees, and Gradient Boosting. The proposed method achieved the highest accuracy of 91.3% at k = 3, outperforming all the other methods. From k = 4 to k = 6, this trend continued, with the proposed approach maintaining high accuracy scores above 89%, indicating its effectiveness in the early selection of highly informative features. Although Gradient Boosting reaches an accuracy of 94.6% at k = 6, the proposed strategy remains more stable and competitive across higher values of k.

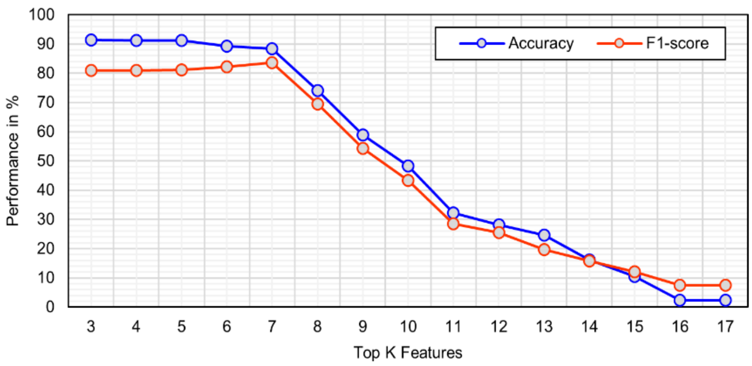

Fig. 3: FIS Performance Using Top-K Feature Subsets for Classification Accuracy and F1-Score.

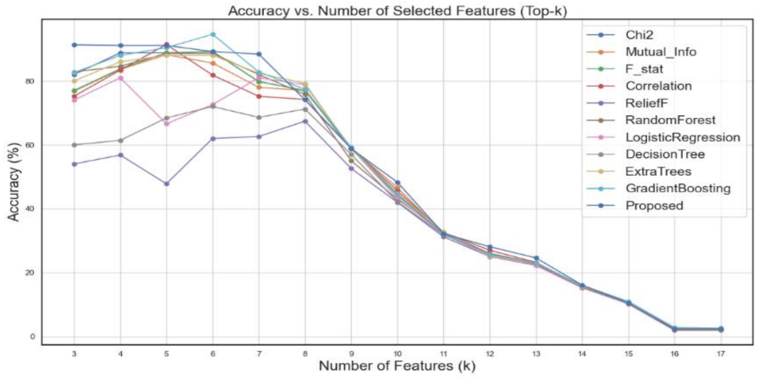

Fig. 4: Accuracy Comparison of Hybrid Model Versus Conventional Feature Selection Techniques.

4.3. Performance Analysis of Proposed Classification Model

This section assesses the effect of increasing sample size on the classification efficacy of the proposed model. Essential parameters, such as accuracy, precision, recall, F1-score, and runtime, have been reported for datasets with sample sizes ranging from 100 to 500. Table 3 presents the results. The performance measures clearly showed an upward trend as the sample size increased from 100 to 500. With more data, the accuracy increased from 0.800 for 100 samples to 0.990 for 500 samples, suggesting an improved classification reliability. Additionally, precision and recall also steadily improved, increasing from 0.7719 to 0.9953 and from 0.5983 to 0.9827, respectively. Consequently, there was a notable improvement in the F1-score from 0.6428 to 0.9874, indicating enhanced model generalization. Although the runtime increases from 1.39 to 3.89 seconds, the performance improvements justify the additional computational expense. This indicates that larger sample sizes enhance the model accuracy and stability. The proposed framework achieved an average accuracy of 93.1%, precision of 91.44%, recall of 84.8%, and F1-score of 87.05% with an average runtime of 2.51 seconds across all sample sizes.

Table 3: Model Performance Across Sample Sizes Showing Accuracy, F1-Score, and Runtime Scalability

Sample Size	Accuracy	Precision	Recall	F1-score	Runtime (s)
100	0.8000	0.7719	0.5983	0.6428	1.3904
200	0.9350	0.8906	0.8395	0.8615	1.8587
300	0.9600	0.9506	0.8925	0.9173	2.2843
400	0.9700	0.9636	0.927	0.9434	3.1393
500	0.9900	0.9953	0.9827	0.9874	3.8976

To examine the scalability and resilience of the proposed LightGBM hybrid model, cross validation was performed with different fold counts. This assessment helps determine the consistency and generalization ability of the model across various data partitions. As the number of folds in cross-validation increases, the proposed hybrid model exhibits a steady improvement in both accuracy and F1-score, indicating enhanced generalization performance owing to more extensive validation. The F1-score begins at 0.9272 with 2 folds and consistently increases, reaching a maximum of 0.9877 with 10 folds. The trend for the accuracy was similar, increasing from 0.958 to 0.990. Although the F1-score variance fluctuated across different folds, it remained within a reasonable range, indicating a consistent performance without notable overfitting. For the final model evaluation, the 10-fold cross validation is a suitable choice as it offers an optimal balance between predictive accuracy and resilience, even though its variance may be slightly higher than that observed with some mid-range folds. A visualization of the F1-score stability with different K-fold cross-validation settings is shown in Fig 5. Here, the blue line with dots indicates the actual F1-scores, and the light blue shaded area represents the range of the F1-score variability (±1 standard deviation) at each fold count. Thus, a rising blue line indicates the performance trend, whereas a narrow-shaded area reflects low variance and stable results, especially at K = 10.

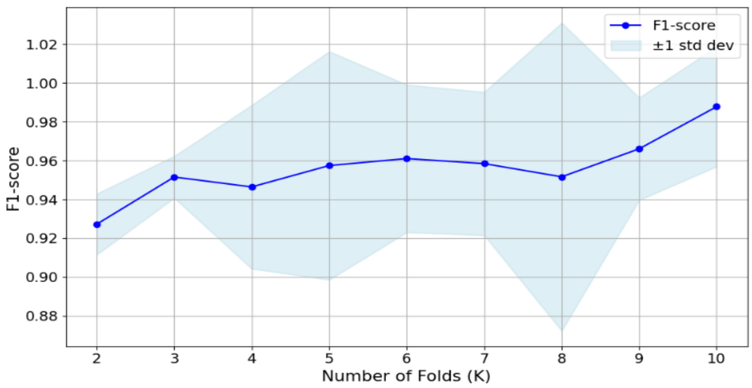

Fig. 5: F1-Score Stability Across K-Fold Cross-Validation Splits for Model Robustness.

4.4. Comparison with Standard Classifiers

A detailed evaluation was carried out using four important assessment metrics: accuracy, precision, recall, and F1-score, in order to compare the performance of different classification models on the tea grade prediction task. These measures offer a comprehensive perspective on each model's predictive efficacy, particularly in a multiclass classification context, where both overall accuracy and class-specific balance (precision and recall) are crucial.

Fig 6 demonstrates that the hybrid model attained the highest F1-score (0.9877), along with exceptional accuracy (0.990), precision (0.9952), and recall (0.9827), signifying its superior performance across all metrics. It consistently outperforms conventional classifiers, such as Logistic Regression, Ridge Classifier, and KNN, which exhibit comparatively lower results across all four metrics. CatBoost, LightGBM, and XGBoost are ensemble-based models that perform well in terms of both precision and recall, trailing the hybrid model only by a narrow margin. Tree-based models, such as Random Forest and Decision Tree, also demonstrate high accuracy and precision, although their recall is marginally lower than that of gradient boosting techniques.

Fig 7 shows the training time comparison between the models in terms of execution efficiency. Although the proposed hybrid model offers superior predictive performance, it remains suitable for real-time applications because of its moderate training time of 3.89 seconds. In contrast, CatBoost, with an execution duration of 157.79 seconds, delivers high accuracy, but may not be as suitable for time-sensitive deployments. Although simpler models, such as Naive Bayes and Logistic Regression, provide faster training times, they often result in lower predictive accuracy. Overall, the visual comparison demonstrates that without imposing excessive computational overhead, ensemble and hybrid learning algorithms offer the most dependable and balanced performance for high-stakes classification tasks such as tea grade prediction.

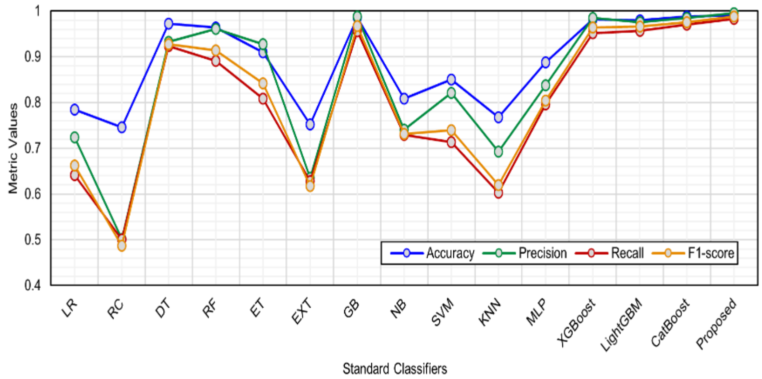

Fig. 6: Comparison of Hybrid FIS-LightGBM with Standard Classifiers on Classification Metrics.

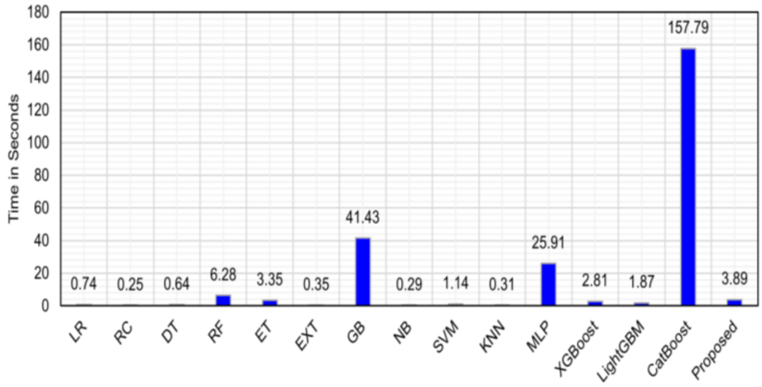

Fig. 7: Runtime Comparison of Hybrid FIS-LightGBM Against Standard Classifiers for Efficiency Evaluation.

4.5. Complexity Analysis

To evaluate suitability for deployment in resource-constrained environments, we analyzed both the computational and memory requirements of the proposed hybrid framework. The fuzzy inference system (FIS) operates with a complexity of $O(n \cdot m)$, where n is the number of features (17) and m is the number of fuzzy rules (≈20). This process requires only simple arithmetic operations, including membership evaluations and min–max aggregations. For borderline samples, the LightGBM model is invoked, which performs decision-path evaluations across shallow trees (depth ≤ 7) with approximately 100 leaves. This results in a per-sample complexity of $O(d \cdot t)$, where d is the average tree depth and t is the number of trees (≈150).

The computational cost of the framework was quantified in terms of floating-point operations (FLOPs) and memory usage. The FIS primarily involves rule evaluations, requiring fewer than 2.5×10^3 FLOPs per sample. LightGBM requires tree-based evaluations with complexity proportional to the number and depth of trees, resulting in about 1.1×10^5 FLOPs per sample. Compared to deep learning models, even compact convolutional networks such as MobileNet demand nearly 3.5×10^8 FLOPs per image, underscoring that the proposed approach is several orders of magnitude more efficient.

The total memory footprint of the trained model (fuzzy rule base + LightGBM) is less than 20 MB, enabling deployment on Raspberry Pi, mobile platforms, or IoT-based field units. Runtime analysis further supports this efficiency, with 1.39–3.89 seconds required per batch of 10 samples on a mid-range CPU.

Beyond speed and storage, the framework is also energy-efficient, an important factor for portable and embedded systems. Both FIS and LightGBM scale linearly with sample size and feature dimensionality, ensuring that even as datasets expand to include diverse tea types, fermentation practices, or adulterated samples, the computational burden remains manageable. This predictable scaling behavior allows the framework to support large-scale deployments while maintaining responsiveness and interpretability.

5. Implications and Limitations

The proposed hybrid decision logic ensures that the fuzzy system adeptly manages interpretable cases, whereas uncertain or marginal samples are assessed by a data-driven model capable of capturing complex nonlinear interactions and subtle patterns, such as total soluble solids, moisture content, or color variance, which are challenging to demonstrate through expert-defined rules. This approach enhances the classification accuracy in ambiguous cases by combining human-like reasoning with the adaptive learning capabilities of ML, resulting in a robust, interpretable, and scalable grading system for real-world tea evaluation.

Several important factors enhance the efficacy of the proposed method. This is achieved using DIY visible-light spectrometry and smartphone-based image processing, which eliminates the need for costly devices such as e-tongue systems or HSI and allows for low-cost feature acquisition. It also makes use of fermentation-specific indicators, such as spectral slope, polyphenol index, color uniformity score, color variance, and fermentation color ratio, with an emphasis on oxidation-related pigment compounds such as theaflavin, thearubigin, and theabrownin. These features are frequently overlooked in the existing techniques and are extremely important for tea grading. Third, it provides a clear and understandable pipeline that combines ML and rule-based logic, ensuring transparency and classification accuracy. As data augmentation techniques have the potential to alter the natural visual and chemical properties of fermented tea samples, the model refrains from using them to avoid producing inaccurate results. Rather than depending on artificial data, models were carefully selected for their ability to handle unbalanced datasets. Finally, its adaptability and scalability, which eliminates the need for complex laboratory equipment or aroma sensors, make it a good option for small-scale or resource-constrained tea production settings.

Practically, the suggested methodology offers a scalable and easily accessible method for evaluating the quality of tea in real time at smallholder or decentralized processing facilities. Without the need for specialized knowledge, the system allows producers to make fast and consistent grading decisions by utilizing easily accessible sensing technologies and interpretable decision logic. This eliminates the need for arbitrary human judgement, encourages process homogeneity, and supports fair market pricing. Additionally, the system's modular design and low weight enable its incorporation into mobile or embedded platforms, making it appropriate for use in cooperative processing units or field settings with minimal technological resources. This modularity also positions the framework within the broader trajectory of intelligent food assessment, where multimodal sensor fusion (e.g., combining vision, spectroscopy, and aroma data) and edge computing (e.g., lightweight deployment on mobile or IoT devices) are emerging as key trends. By being both interpretable and computationally efficient, the framework is well-suited for future integration with these developments, extending its utility beyond the current scope.

Notwithstanding these advantages, it is important to recognize several study limitations. To retain affordability, the system purposefully omits aroma-based elements, which are crucial for a thorough assessment of tea quality, but require gas sensors or e-nose systems. Pigment measurements rely on a DIY spectrophotometer, which may not be as accurate as commercial spectrophotometers, particularly when it comes to identifying subtle concentration variations; however, it is useful for relative comparisons. Additionally, the model may not be directly applicable to early-stage processing or unfermented tea because it is primarily intended for the evaluation of the fermentation stage. Furthermore, the performance might be impacted in settings with fluctuating illumination and varying environmental conditions such as humidity and temperature, since the system relies on static imaging under controlled conditions. The current framework has also not been tested for robustness against intentionally adulterated or blended tea samples, which may present different spectral or visual signatures. Although the framework exhibited better performance on the gathered dataset, additional external validation is required to guarantee wider generalizability across a variety of tea types and production situations.

6. Conclusion

To categorize fermented tea samples across commercial grades, this study introduces a unique hybrid decision logic-based tea-grading framework that integrates computer vision-derived indicators with spectrophotometric sensing. A unified feature vector, composed of pigment absorbance values, spectral indices, and image-based fermentation indicators, was used to assess the samples. The framework combines rule based and ML based decision logic. It employs FIS that translates biochemical and visual cues into interpretable IF-THEN rules and LightGBM classifier to enhance both the confidence scoring and predictive accuracy. Feature selection for the FIS was performed using a combination of statistical and ML techniques to ensure feature relevance while reducing redundancy.

The proposed framework was comprehensively evaluated through classifier comparisons, feature-selection benchmarking, and ablation studies. According to the ablation analysis, using only three features yielded the highest accuracy, whereas selecting the top seven features significantly improved the F1-score to 83.6%. This indicates that compact and relevant feature subsets improve classification balance and reliability more effectively than using the entire feature set. The customized feature selection method, which uses statistical and machine-learning techniques to pick relevant and non-redundant features, consistently outperformed conventional methods in both accuracy and F1-score, especially when using smaller feature sets. This hybrid framework eliminates the need for expensive hyperspectral systems or aroma sensors by employing a DIY visible light spectrometer and smartphone-based imaging under controlled lighting. In terms of classification performance, it surpassed several existing approaches that rely on high-end instrumentation or multi-sensor fusion, achieving an accuracy of 99.00% and an F1-score of 98.76%. This makes the framework a scalable, interpretable, and cost-effective solution suitable for small-scale, resource-constrained tea production settings.

Scalability was further validated by analyzing datasets with increasing sample sizes from 100 to 500. The results showed consistent improvements in F1-score, recall, and precision. A modest increase in runtime, from 1.39 to 3.89 seconds, demonstrated the efficiency of the algorithm. The 10-fold cross-validation produced the highest accuracy (0.990), highest F1-score (0.9877), and lowest variance (0.0308), confirming the robustness of the model. The proposed framework also delivered a better runtime-to-performance ratio than many standard models while maintaining superior F1-scores compared to conventional classifiers. This study provides a solid foundation for the broader application of intelligent decision systems in food processing and agriculture. This represents a significant advancement in the practical, scalable, and interpretable evaluation of the quality of fermented tea.

Future studies should consider integrating low-cost gas sensor arrays to capture aroma-based attributes, thereby enhancing the sensory completeness of the quality evaluation. Real-time deployment through android-based mobile applications or Raspberry Pi platforms can support field-level grading for small producers. Incorporating advanced feature fusion techniques such as transformer-based encoders, graph neural networks, or attention-driven models may improve generalization across various lighting conditions, humidity, and temperature variations. A key research direction is optimizing these models to handle noisy field data and unpredictable environmental disturbances while maintaining robust performance. The model should also be assessed for its suitability in detecting adulterants or blended tea products that may present distinct spectral or visual signatures, raising the challenge of designing reliable classifiers that remain effective across heterogeneous mixtures. IoT-based cloud connectivity can be employed for centralized quality auditing, batch traceability, and supply chain integration, with research attention directed toward ensuring data privacy and achieving scalability in rural production ecosystems.

Additionally, expanding the dataset to include diverse tea types and region-specific fermentation practices will improve the adaptability of the model. An important research question here is how effectively the system can generalize beyond black tea to varieties such as green, oolong, or white teas, each with distinct biochemical and visual markers. Finally, embedding guided image capture, rule-based grading, and feedback mechanisms into mobile applications can empower artisanal and small-scale producers to perform data-driven quality control autonomously. This will enable small-scale producers and non-expert users to adopt AI-powered grading tools in their tea production processes, making intelligent quality control both accessible and practical across various production contexts. This presents both the technical challenge of user interface design for non-expert operators and the practical challenge of adoption in resource-constrained settings.

References

[1] Farag, M. A., Elmetwally, F., Elghanam, R., Kamal, N., Hellal, K., Hamezah, H. S., Zhao, C., & Mediani, A. (2023). Metabolomics in tea products; a compile of applications for enhancing agricultural traits and quality control analysis of Camellia sinensis. *Food Chemistry*, *404*, 134628. https://doi.org/10.1016/j.foodchem.2022.134628.

[2] Erukainure, O. L., Chukwuma, C. I., Nambooze, J., Tripathy, S., Salau, V. F., Olofinsan, K., Ogunlakin, A. D., Ebuehi, O. A. T., & Unuofin, J. O. (2025). Tea Consumption and Diabetes: A Comprehensive Pharmacological Review of Black, White, Green, Oolong, and Pu-erh Teas. *Plants*. 14, 1-48. https://doi.org/10.3390/plants14131898.

[3] Hilal, Y., & Engelhardt, U. (2007). Characterisation of white tea – Comparison to green and black tea. *Journal of Consumer Protection and Food Safety, 2*, 414–421. https://doi.org/10.1007/s00003-007-0250-3.

[4] Sarmah, P. P., Deka, H., Sabhapondit, S., Chowdhury, P., Rajkhowa, K., & Karak, T. (2025). Black tea: Manufacturing and composition. In *Tea in Health and Disease Prevention* (pp. 11-20). Academic Press. https://doi.org/10.1016/B978-0-443-14158-4.00081-6.

[5] Li, M., Pan, T., & Chen, Q. (2021a). Estimation of tea quality grade using statistical identification of key variables. *Food Control*, *119*, 107485. https://doi.org/10.1016/j.foodcont.2020.107485.

[6] An, T., Wang, Z., Li, G., Fan, S., Huang, W., Duan, D., Zhao, C., Tian, X., & Dong, C. (2023). Monitoring the major taste components during black tea fermentation using multielement fusion information in decision level. *Food Chemistry*: X, 18, 1-11. https://doi.org/10.1016/j.fochx.2023.100718.

[7] Wang, T. S., Liang, A. R. D., Ko, C. C., & Lin, J. H. (2022). The importance of region of origin and geographical labeling for tea consumers: The moderating effect of traditional tea processing method and tea prices. *Asia Pacific Journal of Marketing and Logistics*, *34*(6), 1158-1177. https://doi.org/10.1108/APJML-02-2021-0121.

[8] Moreira, J., Aryal, J., Guidry, L., Adhikari, A., Chen, Y., Sriwattana, S., & Prinyawiwatkul, W. (2024). Tea quality: An overview of the analytical methods and sensory analyses used in the most recent studies. *Foods*, *13*(22), 1-21. https://doi.org/10.3390/foods13223580.

[9] Bhargava, A., Bansal, A., Goyal, V., & Bansal, P. (2022). A review on tea quality and safety using emerging parameters. *Journal of Food Measurement and Characterization*, *16*(2), 1291-1311. https://doi.org/10.1007/s11694-021-01232-x.

[10] Food and Agriculture Organization (FAO) of the United Nations. (2022). International tea market: Market situation, prospects and emerging issues. https://www.fao.org/documents/card/en/c/ cc0238en/.

[11] Arhin, I., Li, J., Mei, H., Amoah, M., Chen, X., Jeyaraj, A., Li, X., & Liu, A. (2022). Looking into the future of organic tea production and sustainable farming: a systematic review. *International Journal of Agricultural Sustainability*, *20*(5), 942-954. https://doi.org/10.1080/14735903.2022.2028398.

[12] Wang, H., Gu, J., & Wang, M. (2023). A review on the application of computer vision and machine learning in the tea industry. *Frontiers in Sustainable Food Systems*, *7*, 1-16. https://doi.org/10.3389/fsufs.2023.1172543.

[13] Choudhary, M., & Kaur, P. (2025). Integrating AI with Machine Learning (ML) for Real-Time Detection in Food Industry. In *Artificial Intelligence in the Food Industry* (pp. 297-315). CRC Press. https://doi.org/10.1201/9781032633602-15.

[14] Yin, Y., Li, J., Ling, C., Zhang, S., Liu, C., Sun, X., & Wu, J. (2023). Fusing spectral and image information for characterization of black tea grade based on hyperspectral technology. *Lwt*, *185*, 1-9. https://doi.org/10.1016/j.lwt.2023.115150.

[15] Ding, Y., Zeng, R., Jiang, H., Guan, X., Jiang, Q., & Song, Z. (2024). Classification of tea quality grades based on hyperspectral imaging spatial information and optimization models. *Journal of Food Measurement and Characterization*, *18*(11), 9098-9112. https://doi.org/10.1007/s11694-024-02862-7.

[16] Ding, Y., Yan, Y., Li, J., Chen, X., & Jiang, H. (2022). Classification of tea quality levels using near-infrared spectroscopy based on CLPSO-SVM. *Foods*, *11*(11), 1-10. https://doi.org/10.3390/foods11111658.

[17] Yang, J., Wang, J., Lu, G., Fei, S., Yan, T., Zhang, C., Lu, X., Yu, Z., Li, W., & Tang, X. (2021). TeaNet: Deep learning on Near-Infrared Spectroscopy (NIR) data for the assurance of tea quality. *Computers and Electronics in Agriculture*, *190*, 106431. https://doi.org/10.1016/j.compag.2021.106431.

[18] Li, T., Lu, C., Huang, J., Chen, Y., Zhang, J., Wei, Y., Wang, Y., & Ning, J. (2023a). Qualitative and quantitative analysis of the pile fermentation degree of Pu-erh tea. Lwt, 173, 1-9. https://doi.org/10.1016/j.lwt.2022.114327.

[19] Shi, Y.; Gong, F.; Wang, M.; Liu, J.; Wu, Y.; Men, H. (2019). A deep feature mining method of electronic nose sensor data for identifying beer olfactory information. J. Food Eng. 263, 437–445. https://doi.org/10.1016/j.jfoodeng.2019.07.023.

[20] Suhandy, D., & Yulia, M. (2019, April). Potential application of UV-visible spectroscopy and PLS-DA method to discriminate Indonesian CTC black tea according to grade levels. In *IOP Conference Series: Earth and Environmental Science* (Vol. 258, No. 1, p. 012042). IOP Publishing. https://doi.org/10.1088/1755-1315/258/1/012042.

[21] Jin, S., Li, M., Liu, Z., Liu, R., Li, Y., Zhu, Y., Yuan, Y., Li, P., Li, P., Chen C., Sun, Y., & Sun, Y. (2024). Study on the correlation between color and taste of beauty tea infusion and the pivotal contributing compounds based on UV–visible spectroscopy, taste equivalent quantification and metabolite analysis. *Food Chemistry: X*, *21*, 101192. https://doi.org/10.1016/j.fochx.2024.101192.

[22] Wang, L., Xie, J., Wang, Q., Hu, J., Jiang, Y., Wang, J., Tong, H., Yuan, H., & Yang, Y. (2024). Evaluation of the quality grade of Congou black tea by the fusion of GC-E-Nose, E-tongue, and E-eye. *Food Chemistry: X*, *23*, 1-9. https://doi.org/10.1016/j.fochx.2024.101519.

[23] Peng, Q., Li, S., Zheng, H., Meng, K., Jiang, X., Shen, R., Xue, J., & Xie, G. (2023). Characterization of different grades of Jiuqu hongmei tea based on flavor profiles using HS-SPME-GC-MS combined with E-nose and E-tongue. *Food Research International*, *172*, 113198. https://doi.org/10.1016/j.foodres.2023.113198.

[24] Ren, G., Wu, R., Yin, L., Zhang, Z., & Ning, J. (2024). Description of tea quality using deep learning and multi-sensor feature fusion. *Journal of Food Composition and Analysis*, *126*, 105924. https://doi.org/10.1016/j.jfca.2023.105924.

[25] Zhu, Y., Chen, S., Yin, H., Han, X., Xu, M., Wang, W., Zhang, Y., Feng, X., & Liu, Y. (2024). Classification of oolong tea varieties based on computer vision and convolutional neural networks. *Journal of the Science of Food and Agriculture*, *104*(3), 1630-1637. https://doi.org/10.1002/jsfa.13049.

[26] Kumar, P., & Sharma, M. (2022). Data, machine learning, and human domain experts: none is better than their collaboration. *International Journal of Human–Computer Interaction*, *38*(14), 1307-1320. https://doi.org/10.1080/10447318.2021.2002040.

[27] Palta, P., Kumar, A., & Palta, A. (2024). Leveraging Dielectric Properties, Remote Sensing, and Sensor Technology in Agriculture: A Perspective on Industry and Emerging Technologies. *Industry 5.0 and Emerging Technologies: Transformation Through Technology and Innovations*, 89-109. https://doi.org/10.1007/978-3-031-70996-8_5.

[28] Harshvardhan, FNU, "From Data to Decisions: Machine Learning for Enterprise Demand Forecasting. " PhD diss., University of Tennessee, 2025. https://trace.tennessee.edu/utk_graddiss/12366.

[29] Pan, S. Y., Nie, Q., Tai, H. C., Song, X. L., Tong, Y. F., Zhang, L. J. F., ... & Liang, C. (2022). Tea and tea drinking: China's outstanding contributions to the mankind. *Chinese medicine*, *17*(1), 1-40. https://doi.org/10.1186/s13020-022-00571-1.

[30] Banerjee, S., & Tyagi, P. K. (2024). Exploring the booming tea tourist industry and unconventional tourism through the ritual of drinking tea in India. *Journal of Ethnic Foods*, 11(1), 1-19. https://doi.org/10.1186/s42779-023-00215-1.

[31] Alijoyo, F. A., Janani, S., Santosh, K., Shweihat, S. N., Alshammry, N., Ramesh, J. V. N., & El-Ebiary, Y. A. B. (2024). Enhancing AI interpretation and decision-making: Integrating cognitive computational models with deep learning for advanced uncertain reasoning systems. *Alexandria Engineering Journal*, *99*, 17-30. https://doi.org/10.1016/j.aej.2024.04.073.

[32] Dey, A., & Ashok, S. D. (2024). Fuzzy logic based qualitative indicators for promoting extended producer responsibility and sustainable food packaging waste management. *Environmental and Sustainability Indicators*, *24*, 1-12. https://doi.org/10.1016/j.indic.2024.100534.

[33] Widayat, I. W., Arsyad, A. A., Mantau, A. J., Adhitya, Y., & Köppen, M. (2024). Fuzzy Methods in Smart Farming: A Systematic Review. *Informatica*, *36*(2), 453-489. https://doi.org/10.15388/24-INFOR579.

[34] Gill, G. S., Kumar, A., & Agarwal, R. (2013). Nondestructive grading of black tea based on physical parameters by texture analysis. *Biosystems Engineering*, 116(2), 198-204. https://doi.org/10.1016/j.biosystemseng.2013.08.002.

[35] Yu, D., & Gu, Y. (2021). A machine learning method for the fine-grained classification of green tea with geographical indication using a MOS-based electronic nose. *Foods*, 10(4), 1-17. https://doi.org/10.3390/foods10040795.

[36] Kelly, J. W., Degenhart, A. D., Siewiorek, D. P., Smailagic, A., & Wang, W. (2012, August). Sparse linear regression with elastic net regularization for brain-computer interfaces. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 4275-4278). IEEE. https://doi.org/10.1109/EMBC.2012.6346911.

[37] Oliveira, M. M., Cerqueira, B. V., Barbon Jr, S., & Barbin, D. F. (2021). Classification of fermented cocoa beans (cut test) using computer vision. *Journal of Food Composition and Analysis*, *97*, 1-8. https://doi.org/10.1016/j.jfca.2020.103771.

[38] Liu, H., Yu, D., & Gu, Y. (2019). Classification and evaluation of quality grades of organic green teas using an electronic nose based on machine learning algorithms. *IEEE Access*, *7*, 172965-172973. https://doi.org/10.1109/ACCESS.2019.2957112.

[39] Zhao, X. Y., He, Y. X., Zhang, H. T., Ding, Z. T., Zhou, C. A., & Zhang, K. X. (2024). A quality grade classification method for fresh tea leaves based on an improved YOLOv8x-SPPCSPC-CBAM model. *Scientific reports*, *14*(1), 4166. https://doi.org/10.1038/s41598-024-54389-y.

[40] Li, L., Chen, Y., Dong, S., Shen, J., Cao, S., Cui, Q., Song, Y., & Ning, J. (2023b). Rapid and comprehensive grade evaluation of Keemun black tea using efficient multidimensional data fusion. *Food Chemistry: X*, 20, 1-10. https://doi.org/10.1016/j.fochx.2023.100924.

[41] Zhou, Q., Dai, Z., Song, F., Li, Z., Song, C., & Ling, C. (2023). Monitoring black tea fermentation quality by intelligent sensors: Comparison of image, e-nose and data fusion. *Food Bioscience*, 52, 102454. https://doi.org/10.1016/j.fbio.2023.102454.

[42] Xia, H., Chen, W., Hu, D., Miao, A., Qiao, X., Qiu, G., Liang, J., Guo, W., & Ma, C. (2024). Rapid discrimination of quality grade of black tea based on near-infrared spectroscopy (NIRS), electronic nose (E-nose) and data fusion. *Food Chemistry*, *440*, 138242. https://doi.org/10.1016/j.foodchem.2023.138242.

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

[44] Wu, L., Xu, Q., Su, C., Yin, X., Huo, X., Zhao, X., Zhou, Y., & Huang, J. (2025). Classification of quality grading of Anji white tea using hyperspectral imaging and data fusion techniques. *Journal of Food Composition and Analysis*, 142, 107563. https://doi.org/10.1016/j.jfca.2025.107563.

[45] Zhang, Y., Li, X., Zhu, L., Xing, X., & Du, Y. (2025). A portable tea aroma detection stick system for identifying tea grades. *Journal of Food Measurement and Characterization*, 1-13. https://doi.org/10.1007/s11694-025-03348-w.

[46] Li, L., Wang, Y., Jin, S., Li, M., Chen, Q., Ning, J., & Zhang, Z. (2021b). Evaluation of black tea by using smartphone imaging coupled with micro-near-infrared spectrometer. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *246*, 118991. https://doi.org/10.1016/j.saa.2020.118991.

[47] Xia, L., & Zhu, P. (2022, August). Application of Various Spectral Techniques in Quantitative Analysis of Tea Quality. In *2022 2nd International Conference on Frontiers of Electronics, Information and Computation Technologies (ICFEICT)* (pp. 248-258). IEEE. https://doi.org/10.1109/ICFEICT57213.2022.00053.

[48] Pourjavad, E., & Mayorga, R. V. (2019). A comparative study and measuring performance of manufacturing systems with Mamdani fuzzy inference system. *Journal of Intelligent Manufacturing*, *30*(3), 1085-1097. https://doi.org/10.1007/s10845-017-1307-5.

[49] Bidwe, R., Mishra, S., Bajaj, S., & Kotecha, K. (2025). Leveraging hybrid model of ConvNextBase and LightGBM for early ASD detection via eye-gaze analysis. *MethodsX*, *14*, 1-17. https://doi.org/10.1016/j.mex.2025.103166.