Abstract

Tea has been one of the most widely consumed non-alcoholic beverages worldwide since ancient times due to its distinctive flavour and health benefits. It includes components such as polyphenols, alkaloids, tannic acids and amino acids that contribute to its taste and aroma. Tea is classified into six distinct types, which are determined by the diverse processing techniques employed. These types include dark, black, oolong, yellow, green, and white teas. Among them, black and green tea are highly consumed by many people globally. However, variations in processing and manufacturing practices may result in lower-quality tea products, affecting both market reputation and consumer interests. Assessing the quality of tea is essential for both producers and consumers; however, it poses significant challenges due to the multitude of compounds and their varying impacts on tea quality.

Typically, the determination of tea quality is conducted through human sensory analysis, which offers direct and comprehensive measurements of different characteristics. The classification of tea is derived from the ratings given by seasoned tea tasters, who evaluate the perceived intensities of appearance, aroma, and flavour through their sensory perception. Though human sensory evaluation is inherently subjective, even the same individual may yield different assessments across various trials, and sensory panels can exhibit inconsistencies due to individual differences. Additionally, numerous emotional and physical factors (such as mental or bodily states) can further compromise the accuracy of human sensory evaluations. Furthermore, conventional tea quality assessment methods are often time-consuming and may lack consistency.

To address these issues of subjective sensory evaluation, objective instrumental methods are used. Near-Infrared (NIR) spectroscopy provides a rapid chemical fingerprint, while Molecularly Imprinted Polymers (MIPs) are a sensor

development technique used to create highly selective electrochemical sensors. Both approaches offer reliable, consistent alternatives to human panels for tea quality assessment. As a result, the present research aims to enhance the tea quality prediction, particularly for black and green tea, using NIR and MIP electrochemical methods and implements four proposed models using machine learning (ML) and deep learning (DL) methods.

The first proposed model employs an ML approach to assess the quality of green tea by analysing its EGCG content. Evaluation of EGCG is conducted using an MIP electrode, employing a three-electrode configuration to capture the voltammetric response. Following data pre-processing, the UMAP algorithm is applied to enhance the output by selecting features. Then, K-means clustering is used to differentiate between samples. The classification task is performed using the XGBoost algorithm, renowned for its quick convergence and minimal error. Additionally, the Epigallocatechin-3-gallate (EGCG) levels are measured using High-performance liquid chromatography (HPLC) data for each green tea sample. The effectiveness of the model is evaluated through various metrics, demonstrating its potential of classifying green tea accurately.

The second proposed model introduces a novel Particle Swarm Optimisation-based Random Forest (PSRF) regression model. The modified PSO, obtained by hyper-tuning the traditional PSO, is combined with the RF Model to perform the regression process. First, the study collects the data by incorporating multiple MIP sensors to measure Tannic Acid (TAN), Theophylline (THEO), and Caffeine (CAFF). The collected data are fused and prepared as a dataset for quality analysis. Then, a greedy-based genetic algorithm (Greedy-GA) is employed to perform feature selection, selecting the enriched feature set from the input data. The proposed PSRF is to conduct the regression process. The efficacy of the sensor fusion has been evaluated using standard metrics and demonstrates enhanced accuracy with a lower error rate in predicting black tea quality.

The third proposed framework focuses on differentiating black tea samples using NIR spectroscopy, specifically analysing the TAN content, as it is a significant factor in tea quality. The samples are examined in absorbance mode over a wavelength range of 900 to 1700 nm. Data preprocessing is conducted with the standard scalar method to standardise the responses, which are then subjected to feature selection. Data reduction techniques, including Principal Component Analysis (PCA), Uniform Manifold Approximation and Projection (UMAP), and Kernel PCA (KPCA), are utilised to further analyse the selected features. The study's effectiveness is assessed using criteria such as the Silhouette, Calinski-Harabasz, and Davies-Bouldin indices, with KPCA yielding a satisfactory Silhouette coefficient of 0.64. This indicates that KPCA effectively aids in predicting TAN content and serves as a reliable method for evaluating the quality of black tea.

The final proposed model develops an MIP electrochemical sensor to predict tannic acid content in black tea samples, enhancing prediction accuracy by integrating electrochemical responses with Near-Infrared (NIR) spectroscopic data. A regression model based on a Siamese Neural Network-Long Short-Term Memory (SNN-LSTM) architecture is employed, optimised with a Particle Swarm Optimisation-Dragonfly Optimisation (PSO-DFO) algorithm for effective feature selection and a Convolutional Neural Network (CNN) for feature extraction. Individual system responses are analysed through Canonical Correlation Analysis (CCA) and UMAP techniques to achieve effective data fusion. CCA computes canonical weights for feature projection, while UMAP assists in dimensionality reduction. The SNN-LSTM model successfully identifies complex relationships among compounds in black tea and achieves strong predictive capability through CCA, indicating strong correlation and improved performance in quantifying tannic acid content.

Overall, this thesis makes a significant contribution to the field of the tea industry by assessing the enhanced prediction of tea quality based on the chemical composition of black and green tea samples. The integration of MIP electrochemical, NIR spectroscopy, and AI techniques significantly improves the accuracy and reliability of tea quality predictions, providing producers and consumers with a robust framework for distinguishing high-quality tea, ultimately enhancing market reputation and consumer confidence.