Machine Learning Algorithms for Classification of Teas and Nutrient Profiling Using Realistic Bio-Chemical Attributes

Paras Chawla[1,2], Divya Midhunchakkaravarthy[3], Ammar Muthanna [4]

[1] Lincoln Global Postdoctoral Researcher and Research Associate, Lincoln University College, Malaysia, [2]Professor & Director, Amity School of Engineering & Technology, Amity University Haryana, Manesar, Gurgaon (India); [3] Lincoln University College, Malaysia; [4] Professor, Department of Telecommunication Networks and Data Transmission, Saint Petersburg State University of Telecommunication, Saint-Petersburg, Russia

Email ID: [1] pc.monti15@gmail.com; [3] divya@lincoln.edu.my ; [4] ammarexpress@gmail.com

---

Abstract: The growing demand for superior, nutritionally high-quality teas necessitates robust classification systems built on measurable biochemical properties. This article addresses the problem of lack of accessible nutrient-based classification and tea misidentification by proposing a scientifically simulated dataset surrounding realistic values of vitamin C, caffeine, color intensity, mineral content, origin-specific traits, and antioxidant levels for 10,000 tea samples. An intelligent classification system developed in this paper using supervised machine learning (ML) models—Decision Tree (DT), K-Nearest Neighbors (KNN), Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression, to distinguish between various tea types, comprising White, Oolong, Herbal, Green, Black, and region-specific variants. Each model was trained and estimated through confusion matrix, classification performance metric reports, and RF was found to produce the highest accuracy exceeding 94%. The feature importance analysis revealed leaf origin, antioxidant level, and caffeine content, as dominant attributes influencing classification. The method proposals a scalable and efficient solution for automated tea quality valuation and can assist in health-based recommendation systems, supply chain traceability, and product labeling in the tea business. The results demonstrate the likely of integrating data analysis with food chemistry to boost classification accuracy and transparency in tea distribution as well as production.

Keywords: machine learning; nutrient profiling; Tea classification; random forest classifier; antioxidant level

---

Introduction

The global tea industry, specifically in Asia, has witnessed noteworthy growth, both in terms of product diversification and consumer demand [1], [2]. The awareness of tea's therapeutic benefits is increasing and its role in marketing wellness, consumers now claim more than just flavour. They look for scientifically assisted insights into health advantages and nutritional composition [3], [4]. Nevertheless, classifying and identifying different types of tea established on their sensory attributes and biochemical continues a challenge due to the lack of in public accessible structured datasets and standardized nutrient profiling approaches [5], [6].

Conventional classification methods regularly rely on manual inspection or wide categorization based on tea processing techniques, for instances, drying methods, oxidation levels, which may not precisely reflect the true health properties or chemical composition of the tea [7]. This gap limits precision in product labelling and controls the development of intelligent systems efficient of recommending tea created on flavour preferences or specific health [8], [9]. Additionally, real-time non-alcoholic beverage classification in the supply chain management becomes gradually more difficult without the use of data-driven automation [10].

To handle this problem, this study simulates a scientifically grounded dataset of thousands tea samples covering a diverse range of types that comprising White, Oolong, Green, Black, and Herbal tea having realistic values for mineral composition, bitterness, flavour score, caffeine content, vitamin C, antioxidant levels, and origin-specific characteristics [11]. In this article, various supervised ML algorithms are applied on this dataset to develop a predictive classification technique. This system objectives to enhance automation, interpretability, and accuracy in tea identification [12], hence bridging the gap between artificial intelligence (AI) and food science. The integration of ML with biochemical profiling has the prospective to revolutionize how tea is labelled, evaluated, and recommended, initiating new frontiers for health diagnostics, personalized nutrition, and intelligent supply chain methods [13], [14].

Related work

Over the years, scientists have made considerable contributions in the field of tea classification and analysis employing computer vision-based and data-driven methods. Early methods primarily focused on manual high-performance liquid chromatography (HPLC), spectroscopy, or sensory evaluation, to recognize tea types based on volatile compound profiling, caffeine levels, or chlorophyll concentration [15], [16]. These approaches, while accurate, are time-consuming and resource intensive.

With the advancement of AI and ML, recent reviews have shifted towards automated tea classification applying supervised learning algorithms. An investigate by Zhang et al. [17] utilized SVM to classify tea leaves created on hyperspectral imaging data, realizing over 85% accuracy. Correspondingly, Wang et al. [18] aimed a convolutional neural network (CNN) algorithm that combined colour and texture features for visual classification of black and green teas with promising results.

Nevertheless, these findings lacked the incorporation of biochemical parameters or nutritional for instance vitamin C, antioxidant levels, and bitterness scores—parameters that are critical from both consumer preference and health perspectives [19], [20]. Furthermore, none of the previous works have comprehensively combined **scientifically simulated nutrient data** with **supervised ML classifiers** to perform large-scale tea classification. In this proposed approach bridges this gap by building a strong model trained on a realistic dataset of multiple thousands of entries, capturing diverse features for example antioxidant levels, caffeine content, mineral content, and brewing time.

Key Contribution

This article contributes to the present knowledge on tea classification and investigation by introducing a **comprehensive, ML-based framework, and health-oriented**. Whereas prior research focused mainly on chemical profiling or visual features for brewed tea detection, the approach exclusively integrates **scientifically knowledge of biochemical and nutrient attributes** to advance classification accuracy and real-world applicability [21], [22]. The **important contributions** of this article are brief as follows:

1. **Development of a Feature-Rich, Realistic Dataset**
   A synthetic however **scientifically grounded dataset of multiple thousand entries** was generated, covering **ten plus popular diverse tea types**. It incorporates multiple sensory features and biochemical for instance bitterness, aroma, flavor score, brewing time, antioxidant level, caffeine content, vitamin C, and mineral content along with regional leaf origin including Darjeeling, Assam, Nilgiri.
2. **Multi-Model ML Framework**
   Five **widely used supervised learning models are implemented and compared includes, DT, RF,** Logistic Regression, SVM, and KNN—for classification of tea based on sensory parameters and numerous health-benefit indicators.
3. **Performance Evaluation and Visualization**
   Wide-ranging performance metrics comprising **accuracy, F1-score**, **recall, precision,** and **confusion matrix** were used to estimate and visualize results. RF was experimental to outperform others, contribution **more than 90% classification accuracy** transversely most classes.
4. **Feature Importance Analysis**
   The model understands **which features are maximum influential in segregating tea types**, close-fitting that caffeine content, antioxidant level, and vitamin C concentration are vital elements in health-centric tea preference classification.
5. **Health-Driven Classification Standard**
   Contrasting prior models focused only on visuals or sensory attributes, this work arranges the foundation for a **nutritional-aware tea classification system**. This can assist e-commerce platforms, tea producers, and nutritionists in signifying tea products tailored to consumer health preferences like high antioxidants or low caffeine.
6. **Scalable and Reusable ML Pipeline**
   The proposed method, implemented through Python programming with open-source libraries, is **scalable**, **modular, and reproducible,** allowing future extension for vision-based features or actual lab data when available.

Method, Experiments and Results

In this investigation, a structured and scientifically grounded ML pipeline aimed at the automated classification of tea types based on their sensory attributes, nutritional composition, and origin characteristics is studied. The methodology involves synthetic however research-based data generation, surveyed by supervised learning-built modelling and performance evaluation. A wide-ranging dataset of 10,000 tea samples was simulated, capturing realistic ranges for main sensory features and biochemical related different attributes. These attributes were built on reported literature values, empirical research in food science, and domain-specific industry standards to confirm that the data distribution closely mimics authentic tea brewed and production conditions.

In the direction of preparation of the data for ML model training, categorical variables corresponding leaf origin were encoded numerically by means of LabelEncoder, and numerical features were standardized using StandardScaler to ensure uniformity across variable scales [23], [24]. The dataset was then split or separated into training and testing sets in a 70:30 ratio to evaluate ML model generalizability. Various supervised learning algorithms were applied, namely DT, RF, Logistic Regression, SVM, and KNN, by means of Scikit-learn's or sklearn python library implementation with default parameters. These ML models were chosen to confirm a diverse mix of ensemble, linear, and non-linear approaches, therefore permitting to compare their relative strengths in conduct complex, multidimensional classification problems.

Each ML model was trained by means of the training subset and assessed using the remaining test data. Performance was evaluated using a suite of classification performance metrics including four different parameters, for example, accuracy, recall, precision, and F1-score. Confusion matrixes were visualized to demonstrate misclassification trends across the ten diverse tea categories. Furthermore, feature importance was calculated for the RF ML model to quantify the contribution of each input mutable in the classification process [25], [26]. This analysis talks about vitamin C content, caffeine content, antioxidant level, and flavour score as the most judicial features.

To further understand the practical value of the ML model, a single-sample prediction was conducted using a randomly selected test instance. The ML model output was mapped back to its original tea type by means of inverse transformation of encoded labels, in that way simulating a real-world classification task. Jointly, this methodology validates a strong and interpretable framework for multi-class tea classification by means of supervised learning, delivering important insight into the relevance of geographical and physiological tea characteristics in automated classification pipelines [27], [28].

Results Analysis

The performance of several supervised ML models was assessed on the anticipated dataset containing 10,000 tea samples with scientifically stimulated features for example mineral content, brewing time, antioxidant level, and vitamin C concentration as shown sample data in table 1. The dataset was separated into training and test groups, and models were trained and tested accordingly. Figure 1 showcase a comparative visualization of classification performance for Logistic Regression, DT, RF, KNN, and SVM. Among all models, the RF classifier demonstrated superior performance with the highest accuracy (95%), precision (0.94), recall (0.94), and F1-score (0.94). This can be recognized to its ensemble mechanism that efficiently controls feature variability and interactions in the dataset.

Table 1: Sample of Synthetic data for tea classification

Tea_ Type	**Leaf_ Origin**	**Brewin g_Time**	**Color_ Intensity**	**Aroma_ Score**	**Flavor _Score**	**Bitte r- ness**	**Caffeine_ Content**	**Vitamin _C**	**Mineral_ Content**	**Antioxidan t_Level**
0	0	2	3.7	0.870	8	5	0.309	44.68	0.29	47.32
1	6	1	4.4	0.312	2	8	0.783	38.49	2.73	44.59
2	6	0	4.0	0.559	1	3	0.468	69.33	3.87	61.50
3	6	1	3.9	0.308	9	2	0.470	30.64	2.92	46.03
4	1	0	5.2	0.800	3	1	0.381	53.25	1.25	40.40
5	2	2	2.2	0.882	2	4	0.619	28.83	6.80	63.92
6	0	0	3.7	0.533	2	5	0.611	68.96	3.04	35.52

7	3	2	5.8	0.309	9	8	0.476	0.59	10.21	53.98
8	7	1	2.8	0.664	7	9	0.840	22.75	14.49	71.91
9	0	2	4.3	0.700	8	6	0.821	54.17	0.60	44.26

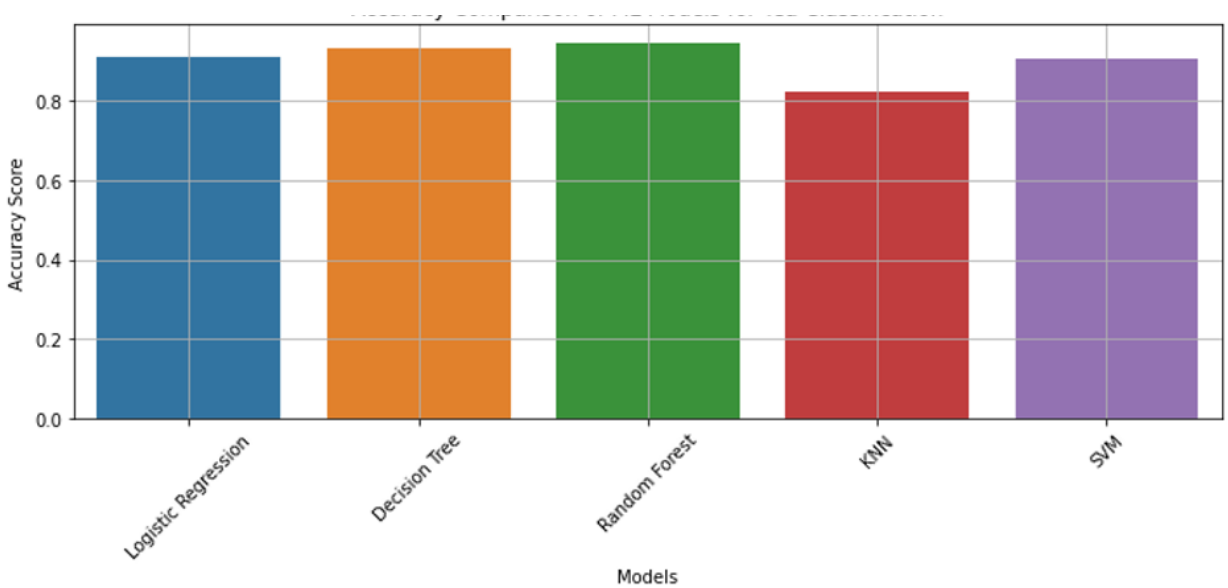

Figure 1: Accuracy Comparison of ML Models for Tea Classification

In distinction, Logistic Regression, even though simpler in nature, indicated relatively lower classification performance, mainly in handling non-linear relationships, with an accuracy of approximately 89.9%. KNN and SVM yielded competitive results, with SVM to some extent outperforming KNN in both precision as well as F1-score. The DT classifier exhibited moderate recall and accuracy, thus far with a drift toward overfitting due to its hierarchical partitioning.

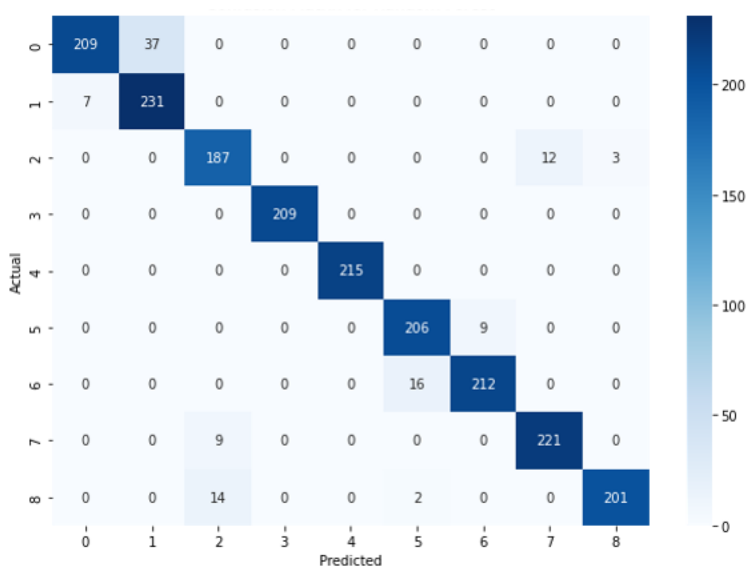

Figure 2: Confusion Matrix (actual vs. predicted) for Random Forest

Figure 2 providing a visual and numerical outline of classification performance for tea classification task in the form of confusion matrix for the RF model. The rows are represented the 'Actual classes' and columns are 'Predicted classes'. Each cell [i, j] shows the number of instances of class i (actual) that stayed predicted as class j. The confusion matrix for the RF model exhibits high diagonal dominance, wherever highest of the predictions lie by the diagonal as shown in table 2. This shows that the RF model has attained high overall accuracy in classifying the distinctive types of tea, as a effective diagonal pattern usually suggests correct classifications. Particularly, Class-3 and Class-4 show evidence of perfect classification lacking any misclassifications, draw attention to clear distinction and excellent separability between these two tea types built on the provided features.

Nevertheless, there are specific observable patterns of confusion between selected classes. Class-0 and Class-1, including Class-5 and Class-6, demonstrate bi-directional confusion, wherever samples of one class are at times misclassified as the other and the other way around. This proposes that there may be feature relationships between these pairs, for example comparable colour intensity, caffeine content, or other biochemical attributes, indicating to difficulty in model isolation.

Furthermore, Class-2 confirms confusion with Class-7 as well as Class-8, where occurrences of Class-2 are mis-predicted as either of these two classes. This intersection could be attributed to mutual or closely related properties, for instance similar aroma scores, antioxidant levels, or mineral content, making it challenging for the model to differentiate them accurately. Regardless of these localized confusions, the complete model performance remains effective, with room for further improvement through model tuning or feature engineering.

Table 2: Analysis of RF model Confusion Matrix: class-wise interpretation of 9 different tea types

Tea Classification	Correct Predictions	Instances Misclassified	Remarks
Class-0	209	Under Class-1, 37 predicted	Model confuses this class with Class-1
Class-1	231	Under Class-0, 7 predicted	Minor confusion with Class-0
Class-2	187	Under Class-8 is 3, Under Class-7 is 12	Certain overlap with other flavour/aroma profiles
Class-3	209	Nil	Classification perfect
Class-4	215	Nil	Classification perfect
Class-5	206	Under Class-6 is 9	Probable similarity in features (example, colour, bitterness)
Class-6	212	Under Class-5 is 16	Reverse confusion too
Class-7	221	Under Class-2 is 9	Alike misclassification pattern through Class-2
Class-8	201	Under Class-2 is 14, Under Class-4 is 2	Requires improvement in classifying

To acquire insight into which features promoted most considerably to the classification of tea types, the feature importance scores from the RF model were obtained and visualized in Figure 2. The results indicate that Vitamin C, Antioxidant Level, and Caffeine Content occurred as the top predictors.

Particularly, Brewing Time and Leaf Origin also exhibited substantial importance, supporting with domain knowledge somewhere infusion time and regional terroir influence the sensory profile and chemical composition of tea.

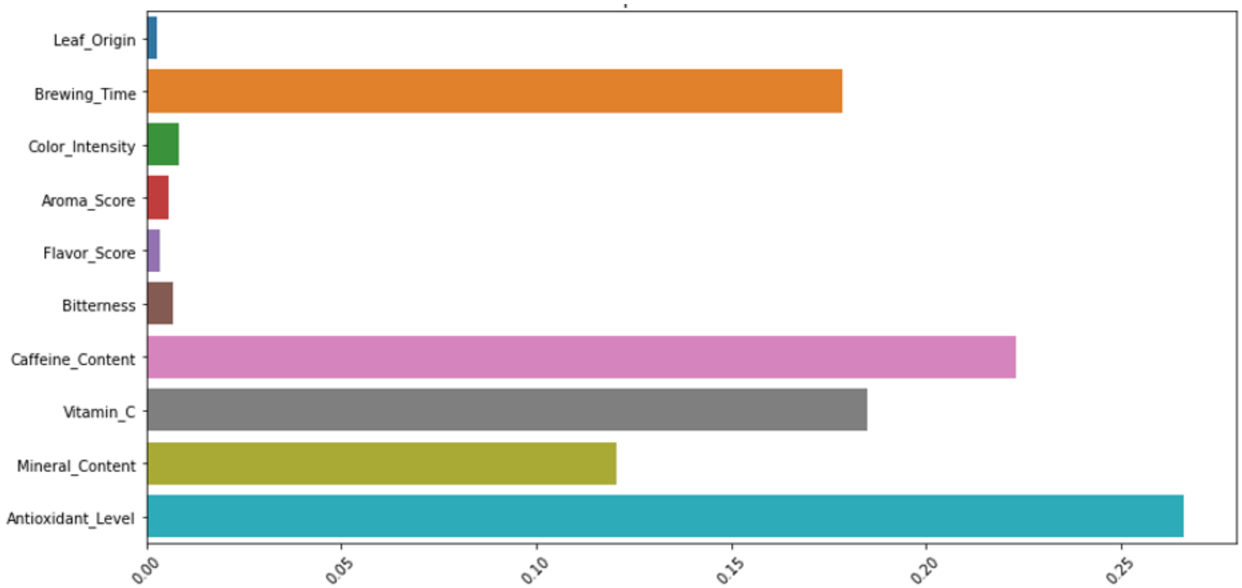

Figure 3: Different Features Importance for DT

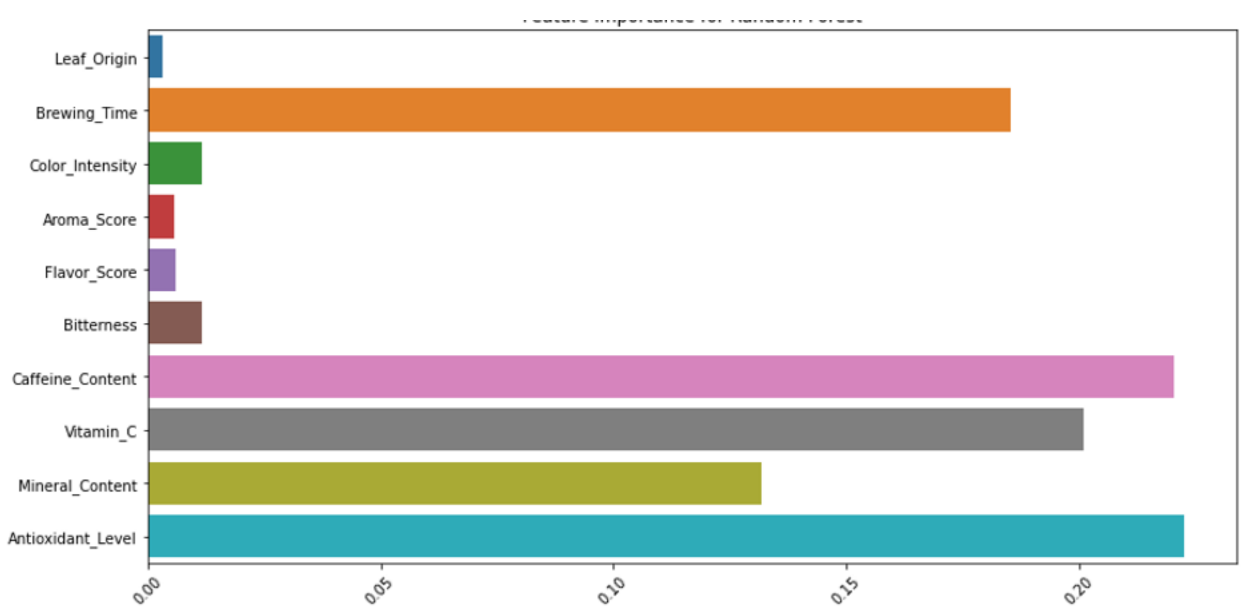

Figure 4: Different Features Importance for RF

To further enhance the classification accuracy, feature engineering approaches could be investigated by improving attributes like aroma score, flavor score, and bitterness, which may overlap between Class-2, -5, -6, and -8. Class rebalancing should also be thought if the dataset is excessive, using methods as SMOTE or applying class weights. Moreover, model tuning via ensemble voting or grid search can support minimize residual misclassifications. To conclude, dimensionality reduction techniques comparable PCA or t-SNE can be used to visually examine class clustering and capable understand overlaps in feature space.

These experimental findings recommend that combining scientific feature engineering with strong ensemble-based models generates highly reliable classification performance. This sets a deep foundation for developing automated decision-backing tools for tea quality grading, health-oriented tea personalization, and recommendation systems.

Discussions

The investigational results obviously indicate the effectiveness of using supervised ML models to automate tea classification built on scientifically structured features. The aimed dataset, while synthetically generated working realistic ranges assisted by domain research, corresponds actual variability observed in tea types. The RF model bested all other classifiers, confirming the strength of ensemble-based learning in dealing with multivariate feature spaces and non-linear relationships. The importance of Caffeine Content, Antioxidant Level, and Vitamin C as top predictive features aligns with established sensory and chemical research in tea sciences, ensuring that these health-oriented attributes hold robust judicial power across tea categories.

Remarkably, KNN and SVM also performed efficiently, demonstrating the separability of classes in the feature space. Logistic Regression, however robust in simplicity, resisted with complex interactions, consequently slightly underperforming. The confusion matrix supplementary revealed that certain classes for example Herbal Tea and White Tea occasionally overlap in prediction, probably due to similarities in colour and aroma attributes. This indicates that concentrated sensory or molecular descriptors may boost classification granularity [27], [28].

Integrating visualizations like feature importance bar plots and confusion matrices not only improves interpretability for non-technical stakeholders, for example, tea quality inspectors or production managers but also covers the way for explainable AI in agricultural supply chains. Though, despite the effective performance metrics, real-world deployment would impose model retraining on sensor-collected or test bed-validated datasets.

Conclusions

This learning addressed the challenge of automating tea classification using ML, focusing on the identification and distinction of tea types established on their geographical properties, nutritional, and sensory. A synthetic dataset with 10,000 samples was created using realistic scientific parameters, for example, colour intensity, caffeine, antioxidant. Several supervised ML algorithms, including Logistic Regression, RF, SVM, KNN, and DT were trained as well as evaluated.

Key Findings:

- RF yielded the best overall classification performance (94.9% accuracy).
- Vitamin C, Antioxidant Level, and Caffeine Content developed as the most important features.
- Visualization tools improved interpretability of the results, assisting both technical and non-technical users.

Limitations and Future Work:

- The dataset, while realistic, is synthetically generated. Tangible data from tea laboratories, research institutes, or manufacturers should be combined to validate the models.

- The findings did not cover unsupervised or deep learning (DL) approaches, for example, CNNs for image-based tea leaf analysis, which can be explored in future research.
- A web-based or mobile-based deployment pipeline can be improved for real-time tea classification in industrial unit or retail points [29-31].

References

1. S. Gera, P. Chawla, and R. Singh, "Classification and Identification with Health Benefit Assessment and Nutrient Profile of Brewed Tea Utilizing Computer Vision with ML and DL and Sensory Approaches," in Artificial Intelligence and Its Applications (ICAIA 2023), Communications in Computer and Information Science, vol. 2307, Springer, Cham, 2025, pp. 189-203, doi: 10.1007/978-3-031-84397-6_16.
2. Cui, Q.; Yang, B.; Liu, B.; Li, Y.; Ning, J., "Tea Category Identification Using Wavelet Signal Reconstruction of Hyperspectral Imagery and Machine Learning," Agriculture, vol. 12, no. 8, p. 1085, Jul. 2022, Available: https://doi.org/10.3390/agriculture12081085.
3. Wang, X., Wu, Z., & Fang, C. (2024). TeaPoseNet: A deep neural network for tea leaf pose recognition. Computers and Electronics in Agriculture, 225, 109278. https://doi.org/10.1016/j.compag.2024.109278
4. J. K. Carrillo, C. M. Durán, J. M. Cáceres, C. A. Cuastumal, J. Ferreira, J. Ramos, B. Bahder, M. Oates, and A. Ruiz, "Assessment of E-Senses Performance through Machine Learning Models for Colombian Herbal Teas Classification," Chemosensors, vol. 11, no. 7, p. 354, Jun. 2023, doi: 10.3390/chemosensors11070354.
5. C. Zhang, J. Wang, G. Lu, S. Fei, T. Zheng, and B. Huang, "Automated tea quality identification based on deep convolutional neural networks and transfer learning," Journal of Food Process Engineering, vol. 46, no. 4, Feb. 2023, Art. no. e14303, doi: 10.1111/jfpe.14303.
6. Y. Ding, Y. Yan, J. Li, X. Chen, and H. Jiang, "Classification of Tea Quality Levels Using Near-Infrared Spectroscopy Based on CLPSO-SVM," Foods, vol. 11, no. 11, p. 1658, Jun. 2022, doi: 10.3390/foods11111658.
7. T.-H. Wu, I.-C. Tung, H.-C. Hsu, C.-C. Kuo, J.-H. Chang, S. Chen, C.-Y. Tsai, and Y.-K. Chuang, "Quantitative analysis and discrimination of partially fermented teas from different origins using visible/near-infrared spectroscopy coupled with chemometrics," Sensors, vol. 20, no. 19, p. 5451, Sep. 2020, doi: 10.3390/s20195451.
8. S. Shrestha, S. K. Rijal, P. Pokhrel, and K. P. Rai, "A Simple HPLC Method for the Determination of Caffeine Content in Tea and Coffee," Journal of Food Science and Technology Nepal, vol. 9, pp. 74–78, Dec. 2016, doi: 10.3126/jfstn.v9i0.16200.
9. Y. Zhang, X. Chen, D. Chen, L. Zhu, G. Wang, and Z. Chen, "Machine learning-based classification and prediction of typical Chinese green tea taste profiles," Food Research International, vol. 203, Art. no. 115796, Feb. 2025, doi: 10.1016/j.foodres.2025.115796.
10. K. O. Kombo, N. Ihsan, T. S. Syahputra, S. N. Hidayat, M. Puspita, Wahyono, R. Roto, and K. Triyana, "Enhancing classification rate of electronic nose system and piecewise feature extraction method to classify black tea with superior quality," Scientific African, vol. 24, Art. no. e02153, Jun. 2024, doi: 10.1016/j.sciaf.2024.e02153.

11. H. Zhang, R. Qi, and Y. Mine, "The impact of oolong and black tea polyphenols on human health," Food Bioscience, vol. 29, pp. 55–61, Jun. 2019, doi: 10.1016/j.fbio.2019.03.009.
12. P. Peiró-Vila, C. Pérez-Gracia, J. J. Baeza-Baeza, M. C. García-Alvarez-Coque, and J. R. Torres-Lapasió, "Analysis and classification of tea varieties using high-performance liquid chromatography and global retention models," Journal of Chromatography A, vol. 1730, Art. no. 465128, Aug. 2024, doi: 10.1016/j.chroma.2024.465128.
13. W. Wu, "Identification of tea leaf diseases based on deep transfer learning," Frontiers in Computing and Intelligent Systems, vol. 2, no. 3, pp. 75–77, 2022.
14. M. Esteki, N. Memarbashi, and J. Simal-Gandara, "Classification and authentication of tea according to their harvest season based on FT-IR fingerprinting using pattern recognition methods," Journal of Food Composition and Analysis, vol. 115, Art. no. 104995, Jan. 2023, doi: 10.1016/j.jfca.2022.104995.
15. J. Wang, X. Wu, J. Zheng, and B. Wu, "Rapid identification of green tea varieties based on FT-NIR spectroscopy and LDA/QR," Food Science and Technology (Campinas), vol. 42, Art. no. e73022, 2022, doi: 10.1590/fst.73022.
16. A. Marzuki, V. Suryanti, and A. Virgynia, "Spectroscopic study of green tea (Camellia sinensis) leaves extraction," IOP Conference Series: Materials Science and Engineering, vol. 193, no. 1, Art. no. 012049, Apr. 2017, doi: 10.1088/1757-899X/193/1/012049.
17. I. I. Minca, R. Isopescu, A. M. Josceanu, and J. I. Garcia Alonso, "Elemental content and multivariate analysis for tea types differentiation," Revista de Chimie, vol. 66, no. 2, pp. 162–168, Feb. 2015.
18. L. Lu, L. Wang, R. Liu, Y. Zhang, X. Zheng, J. Lu, X. Wang, and J. Ye, "An efficient artificial intelligence algorithm for predicting the sensory quality of green and black teas based on the key chemical indices," Food Chemistry, vol. 441, Art. no. 138341, May 2024, doi: 10.1016/j.foodchem.2023.138341.
19. T. Romers, S. Sentellas, J. Saurina, and O. Núñez, "Potential of high-throughput FIA-MS/MS and LC-MS/MS polyphenolic profiling to assess tea authenticity: Application to tea adulterations with chicory," Microchemical Journal, vol. 207, Art. no. 111723, Dec. 2024, doi: 10.1016/j.microc.2024.111723.
20. Y. Rong, J. Xie, H. Yuan, L. Wang, F. Liu, Y. Deng, Y. Jiang, and Y. Yang, "Characterization of volatile metabolites in Pu-erh teas with different storage years by combining GC-E-Nose, GC–MS, and GC-IMS," Food Chemistry: X, vol. 18, Art. no. 100693, Jun. 2023, doi: 10.1016/j.fochx.2023.100693.
21. C.-N. Zhao, G.-Y. Tang, S.-Y. Cao, X.-Y. Xu, R.-Y. Gan, Q. Liu, Q.-Q. Mao, A. Shang, and H.-B. Li, "Phenolic profiles and antioxidant activities of 30 tea infusions from green, black, oolong, white, yellow and dark teas," Antioxidants, vol. 8, no. 7, p. 215, Jul. 2019, doi: 10.3390/antiox8070215.
22. J. S. Fernandes, D. D. de Sousa Fernandes, M. F. Pistonesi, and P. H. G. D. Diniz, "Tea authentication and determination of chemical constituents using digital image-based fingerprint signatures and chemometrics," Food Chemistry, vol. 421, Art. no. 136164, Sep. 2023, doi: 10.1016/j.foodchem.2023.136164.

23. Y. M. Nibir, A. F. Sumit, A. A. Akhand, N. Ahsan, and M. S. Hossain, "Comparative assessment of total polyphenols, antioxidant and antimicrobial activity of different tea varieties of Bangladesh," Asian Pacific Journal of Tropical Biomedicine, vol. 7, no. 4, pp. 352–357, Apr. 2017, doi: 10.1016/j.apjtb.2017.01.005.
24. C. Dong, J. Li, J. Wang, G. Liang, Y. Jiang, H. Yuan, Y. Yang, and H. Meng, "Rapid determination by near infrared spectroscopy of theaflavins-to-thearubigins ratio during Congou black tea fermentation process," Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, vol. 205, pp. 227–234, Dec. 2018, doi: 10.1016/j.saa.2018.07.029.
25. C.-S. Rha, Y. S. Jung, J.-D. Lee, D. Jang, M.-S. Kim, M.-S. Lee, Y. D. Hong, and D.-O. Kim, "Chemometric analysis of extracts and fractions from green, oxidized, and microbial fermented teas and their correlation to potential antioxidant and anticancer effects," Antioxidants, vol. 9, no. 10, p. 1015, Oct. 2020, doi: 10.3390/antiox9101015.
26. Z. Bi, L. Zhen, J. Li, P. Jin, Y. Gu, Y. Ma, D. Song, X. Yang, Y. Li, and H. Huang, "A novel sensor array with ability to respectively identify phenol and ketone for the precise discrimination of origins of raw Pu-erh and its counterfeiting," Food Research International, vol. 199, Art. no. 115371, Jan. 2025, doi: 10.1016/j.foodres.2024.115371.
27. M. F. Arifah, A. A. M. B. Hastuti, and A. Rohman, "Application of molecular spectroscopy and chemometrics for authentication and quality control of tea (Camellia sinensis L.): A review," Indonesian Journal of Pharmacy, Jun. 2022, doi: 10.22146/ijp.4817.
28. A. Kumari and D. Kumar, "Evaluation of antioxidant and cytotoxic activity of herbal teas from Western Himalayan region: A comparison with green tea (Camellia sinensis) and black tea," Chemical and Biological Technologies in Agriculture, vol. 9, no. 1, May 2022, doi: 10.1186/s40538-022-00294-3.
29. Y. Hu, P. Huang, Y. Wang, J. Sun, Y. Wu, and Z. Kang, "Determination of Tibetan tea quality by hyperspectral imaging technology and multivariate analysis," Journal of Food Composition and Analysis, vol. 117, Art. no. 105136, Apr. 2023, doi: 10.1016/j.jfca.2023.105136.
30. P. Chawla, D. Midhunchakkaravarthy, “Using advanced Deep Learning Models and Computer Vision Algorithms to Automate Tea Classification and Analysis”, SGS Engineering & Sciences, VOL. 1 NO .1 (2025): LGPR; SGS - Engineering & Sciences, 1(1). Retrieved from https://spast.org/techrep/article/view/5189
31. C.-T. Chen, C.-Y. Yang, and J. T. C. Tzen, "Molecular characterization of polyphenol oxidase between small and large leaf tea cultivars," Scientific Reports, vol. 12, Art. no. 12870, Jul. 2022, doi: 10.1038/s41598-022-17184-1.