ARDUINO AND MACHINE LEARNING DRIVEN TEA-TASTING SYSTEM FOR BLACK TEA.

P. S.M. Samaraweera[1*], N.S. Weerakkody[1] and G. Anthonys[2]
[1]Department of Agricultural and Plantation Engineering,
[2]Department of Electrical and Computer Engineering,
The Open University of Sri Lanka, Sri Lanka

Sri Lanka is one of the leading tea producers in the world and tea tasting is one of the crucial tasks in the industry. This paper aims to explore an Arduino uno-based tea-tasting methodology with the help of machine learning. Image classification and taste sensor values (pH, TDS, RGB etc.) were employed to taste teas based on tea-tasting terminology, thus eliminating the need for professional tea tasters. Three models were used in image classification, and two models were used in identifying the acceptance of tea liquors - whether they were high fired or not. For image classification, two pre-trained models—MobileNetV2 and DenseNet201—were utilized, alongside a manually trained Convolutional Neural Network (CNN). Random Forest (RF) and Artificial Neural Network (ANN) models were used for other datasets. Images of dry leaves, infused leaves, and liquor, as well as taste sensor data, were used to train and test the models in real-time with the aid of Arduino. Max-min normalization and data augmentation techniques were applied to enhance image contrast. The experimental results indicated that the proposed DenseNet201 was 93.60%, 96.77%, and 97.20% accurate for dry, infused, and liquor images, respectively. Also, the ANN model showed 95% accuracy for the data set. With high accuracy and a favourable F1 score, the newly designed DenseNet201 CNN architecture and ANN model could be a helpful decision-making tool in tea tasting with the help of Arduino Uno.

Keywords: Arduino, Artificial Neural Network (ANN), Convolutional Neural Network (CNN), DenseNet, MobileNet, Tea Tasting.

[*]*Corresponding Author: psmsamaraweera@gmail.com*

ARDUINO AND MACHINE LEARNING DRIVEN TEA-TASTING SYSTEM FOR BLACK TEA.

P. S.M. Samaraweera[1*], N.S. Weerakkody[1] and G. Anthonys[2]
[1]Department of Agricultural and Plantation Engineering,
[2]Department of Electrical and Computer Engineering,
The Open University of Sri Lanka, Sri Lanka

INTRODUCTION

Tea is the second most widely consumed beverage worldwide, with Sri Lanka being the world's fourth-largest tea producer. The tea market is booming, and its primary quality characteristics include flavor, fragrance, color, and strength (Bhattacharyya et al., 2008). The taste of tea is mostly influenced by the chemical components present in the tea leaves, such as catechins, tannins, and flavonoids. Flavonol-O-glycosides, tannins and galloylated catechins are the main astringent components while caffeine and non-galloylated catechins contribute to the bitterness of the tea. Furthermore, L-theanine, succinic acid, gallic acid and theogallin contribute to the umami taste. A sweet aftertaste is a characteristic perception of green tea, which is linked to the hydrolysis of galloylated catechins and the tea grade (Zhang et al., 2020). The ability to taste tea is crucial in determining its quality. Tea tasters utilise a specialised vocabulary (Tea Tasting Terminology) devised by the tea business to assess numerous aspects such as flavour, leaf colour, size, and shape to determine overall quality.

However, traditional analytical methods like gas chromatography (GC), high-performance liquid chromatography (HPLC), and capillary electrophoresis (CE) are time-consuming, labor-intensive, and expensive (Patil et al 2010). The accuracy of the assessment of tea quality needs to be improved due to the complex contributions of the various compounds found in tea. Training a tea tester for a tea factory is time-consuming and costly, and most people are now reluctant to be tea testers due to side-effects on health, including dental health (Patil et al., 2021). To address these problems, alternative methods for evaluating tea quality need to be developed. Products built on the Arduino platform are a viable solution due to their low cost, wide availability, and ease of use. By exploiting the capabilities of the Arduino platform, it is possible to construct low-cost, portable equipment that can precisely and quickly analyze significant chemical components in tea samples. This research aims to investigate and develop Arduino-based devices made exclusively for tea quality testing (Tea tasting) (Patil et al 2010). These devices would measure and analyze the relevant chemical properties in tea samples using the correct sensors and algorithms, enabling a quick and unbiased assessment of tea quality. This efficient and cost-effective remedy to the current limitations in tea quality evaluation methodologies would benefit tea producers and consumers. Hence, the main objective of this study is to develop an Arduino and Machine Learning Driven tea-tasting methodology for black tea. Also, the study aims explicitly to collect a benchmark dataset for assessing black tea using sensory mechanisms of Arduino based on the tea-tasting terminology. Accurate machine learning algorithms will be developed using random forest classifier (RF) and Artificial Neural Networks (ANN), to identify the black tea samples to identify whether they are high fired or not and Convolutional Neural Network (CNN), MobileNetV2 and Densenet201 to identify different taste profiles according to the Tea Tasting Terminology in different tea grades. A linear regression model will be developed to determine Brix from TDS (ppm) (Total Dissolved Solids) value and identify any correlation for sugar-based adulterations in black tea based on different tea blends.

METHODOLOGY

2.1 Proposed Electronic Device for Tea Tasting

pH Value Detection Sensor Control Board Module with BNC Electrode Probe, TDS sensor with Signal Transmitter Board, DS18B20 Waterproof Digital Temperature Sensor, RGB sensor (TCS34725) and ESP 32 Camara module were connected with the Arduino Uno R3. The proposed Arduino-based electronic device was interfaced with a PC/ laptop using Universal Serial Bus (USB3.0) communication for data acquisition (Figure 1). A calibration process was carried out to improve and ensure the accuracy of the readings of the pH, TDS sensors and the ESP-32 camera. An ESP-32 camera was positioned 20 cm above the sample with D65 standard light source and it was linked to a PC or laptop via a Wi-Fi connection. The camera settings were as follows: manual mode, fixed focus manually adjusted, no zoom, no flash, and photographs were shot at a size of 640x480 pixels, saved in 24-bit JPG format. The camera's white balance was set to auto mode to adjust to different lighting conditions. To ensure constant illumination, the whole equipment was housed inside a controlled atmosphere.

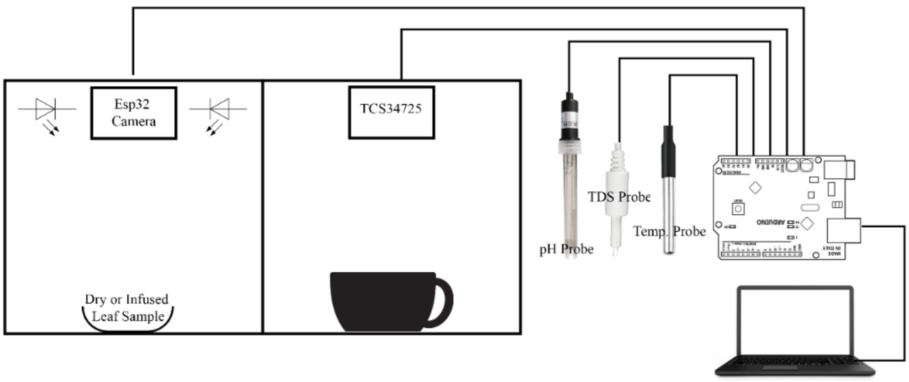

Figure 1: Proposed electronic device for tea tasting

2.2 Data Collection and Preparation

In the research study, the analysis of dry leaf data involved three datasets: train, test, and validation, consisting of 905, 453, and 452 images, respectively. A total of 1810 dry tea leaf images from orthodox teas were classified into five terminological categories (Ragged, Brown, Blackish, Leafy, Flaky) describing dry leaves. Additionally, 1860 infused leaf images from orthodox teas were categorized into five terminological categories (Dull, Fair, Dull & Green, Dull & Dark, Mixed or Uneven) describing infused leaves. The three datasets contained 930 (50% from the infused leaf dataset), 465, and 465 images for training, testing, and validation, respectively. Furthermore, the dataset comprised 1709 tea liquor images collected from orthodox teas, categorized into five terminological categories describing liquor. In this case, the three datasets comprised 854, 428, and 427 images for training, testing, and validation in the liquor dataset.

The second dataset consisted of 13 attributes and included 1145 rows, with 51.2% labeled as Not Accept and 48.8% labeled as Accept based on whether they were high-fired or not. To predict the acceptance of a given tea (liquor) sample, the Random Forest and ANN model (100 epochs and 32 batch sizes) were employed in this experiment. The third dataset contained two attributes (Brix, pH) and comprised 181 rows for the linear regression model to predict the Brix value of the tea from a given TDS value in BOPF (Broken Orange Pekoe Fannings) and BP1 (Broken Pekoe 1) tea grades.

Tea samples from various Orthodox categories in Sri Lanka were prepared by boiling tea for three minutes and then removing it to prevent oversteeping. The samples were stored at 40°C until ready for hot liquor tasting. Infused leaf samples were placed on porcelain lids, while dry leaves were distributed in red containers. Professional tea tasters (Mercantile Tea Brockers (PVT) Ltd in Colombo) gathered judgments on dry leaves, infused leaves, and liquor. Data for dry, infused, and liquor tasting was collected from tea samples based on attributes such as tea grade, elevation, and quality type (According to the tea-tasting terminology) for image identification. Data pre-processing was crucial for model accuracy and precision. Images were downsized to 224*224 for CNN, MobileNetV2, and DenseNet201 models, with RGB color mode used. Kaggle (T4 x 2 GPU runtime architecture) was used for data analysis and model building, using CNN models with transfer learning techniques (MobileNetV2 and DenseNet201) for dry leaves, infused leaves, and liquor analysis. Type of Tea Grade (BOPF, BP1), Elevation (High Grown, Medium Grown, Low Grown), pH Level of Water, TDS Level of Water, Tasting Temperature, Brewing Time, pH Value of the Liquor, TDS Value of the Liquor, RGB Value (Colour Values of the Liquor), Taste Type (Based on the Tea Tasting Terminology), Final Acceptance of the Tea (High fired or not) parameters were used for Liquor Tasting from RF, and ANN.

2.3 Converting TDS (ppm) to Brix (%)

There is no specific sensor to measure the Brix value in Arduino. Therefore, a linear regression model was computed for converting %TDS to %Brix of given tea grade. Simple linear regression equation often takes the form (Gómez, 2019)

y = mx + c or Brix (%) = m * TDS (ppm) + c (1)

Where: Brix (%) was the Brix measurement want to be calculated, TDS (ppm) is the TDS measurement in parts per million, "m" is the slope of the regression line, and "c" is the y-intercept of the line. Meanwhile, a linear regression model was used to convert %TDS to %Brix of a given tea grade. Data was collected and sorted according to tea grade BP1 in CTC (Crush, Tear, Curl) production and BOPF in Orthodox production. The analysis was conducted using Python's "sklearn.linear_model" library, specifically for the "LinearRegression()" model. This approach revealed correlations between TDS and pH of tea. During the experiment data cleansing, feature selection, model training, evaluation, and hyperparameter tuning steps were involved. Therefore, the most appropriate model offered perceptions of feature importance and judgement of the tea tasting based on the terminology and the acceptance of a liquor sample. K-fold cross-validation was used for the evaluation of the classification performance. The prediction result was based on True positive (TP), False positive (FP), False negative (FN) and True negative (TN) values (A. B. Patil et al., 2021). The quality measures are defined as below.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (2)$$

$$\text{Sensitivity (Recall)} = \frac{(TP)}{(TP + FN)} \quad (3)$$

$$\text{Precision} = \frac{(TP)}{(TP + FP)} \quad (4)$$

$$\text{F1} - \text{score} = \frac{2 * (\text{Precision}) \text{ x } (\text{sensitivity})}{(\text{Precision} + \text{Sensitivity})} \quad (5)$$

RESULTS AND DISCUSSION

3.1 Dry Leaf Recognition

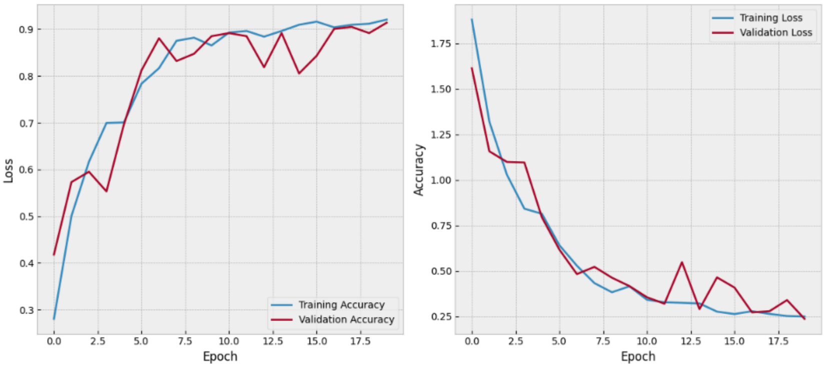

Figure 2: Graph representing model accuracy and model loss for training and validation set using the CNN approach for Dry Leaf.

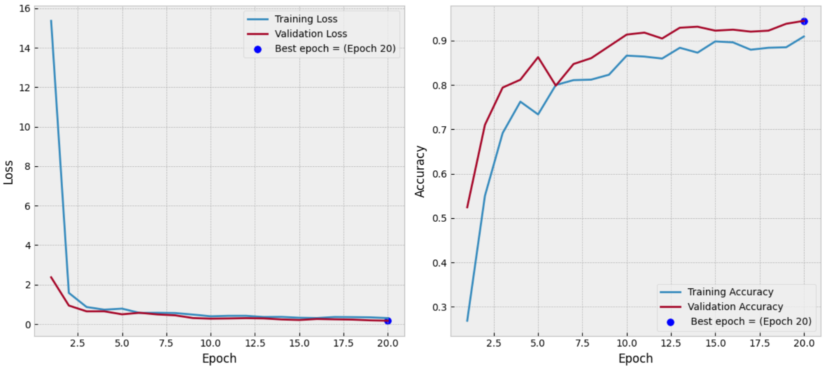

Figure 3: Graph representing model accuracy and model loss for training and validation set using the MobileNetV2 approach for Dry Leaf.

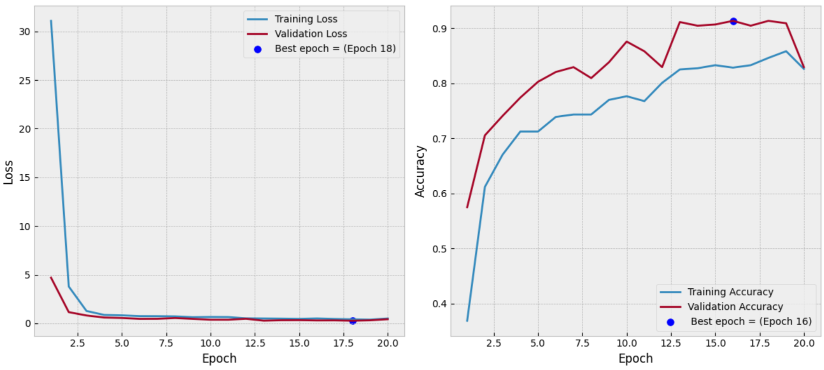

Figure 4: Graph representing model accuracy and model loss for training and validation set using the DenseNet201 approach for Dry Leaf.

In the CNN model for dry leaves, initial validation accuracy is below 0.4, but after one epoch, the validation accuracy suddenly increases to nearly 0.8 after a slight zigzag variation between 0.5 and 0.6. Similarly, the initial validation loss is above 1.50, but after five epochs, the loss decreases below 0.75. As shown in Figure 2, there is a positive trend toward improving accuracy and reducing loss. At first, training accuracy is low, but it progressively improves to almost 91.16 percent. The transfer learning part of the measure was accomplished on the MobileNetV2 and the DenseNet201 model, shown in Figure 3 and Figure 4, respectively.

3.2 Infused Leaf Recognition

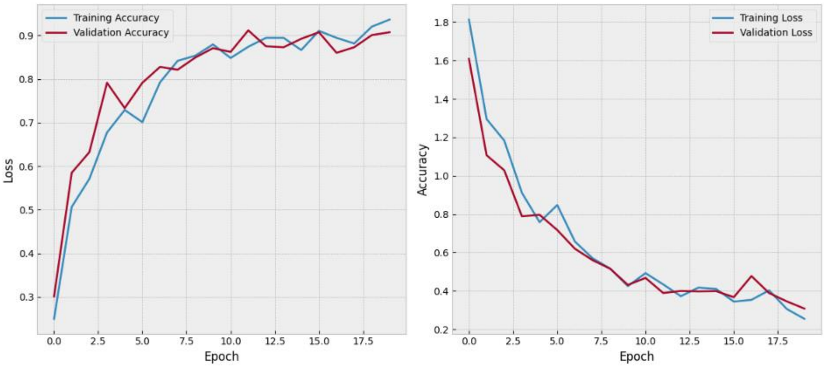

Figure 5: Graph representing model accuracy and model loss for training and validation set using the CNN approach for Infused Leaf.

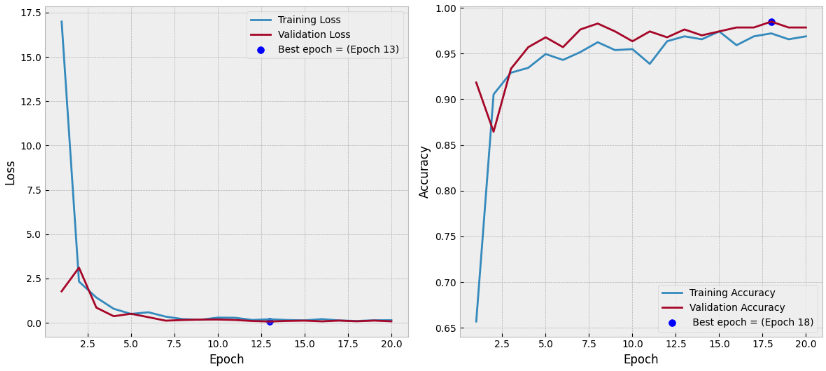

Figure 6: Graph representing model accuracy and model loss for training and validation set using the MobileNetV2 approach for Infused Leaf.

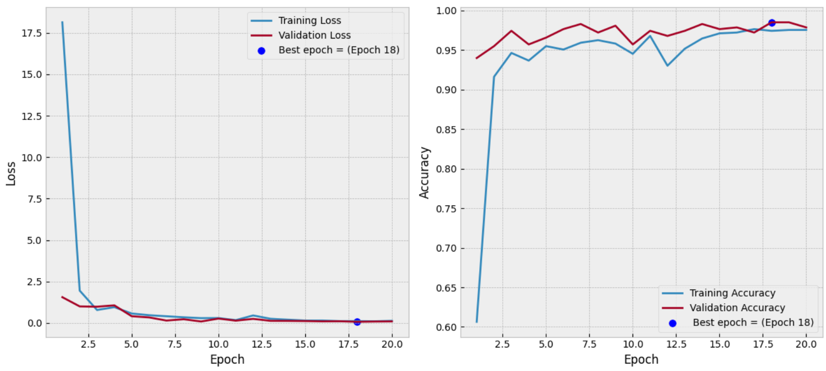

Figure 7: Graph representing model accuracy and model loss for training and validation set using the CNN approach for Infused Leaf.

The CNN model for infused leaves showed an initial validation accuracy of 0.3, which increased to nearly 0.8 before five epochs. The initial validation loss is above 1.80 but decreased near 0.8 after five epochs. At first, testing accuracy is low, but it progressively improves to almost 89.24 percent. The transfer learning part of the measure was accomplished on the MobileNetV2 and the DenseNet201 models, which are shown in Figure 6 and Figure 7, respectively.

3.3 Liquor Recognition

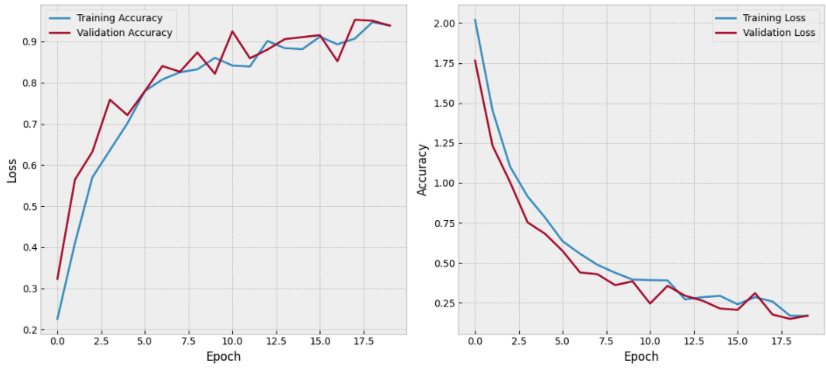

Figure 8: Graph representing model accuracy and model loss for training and validation set using the CNN approach for Liquor.

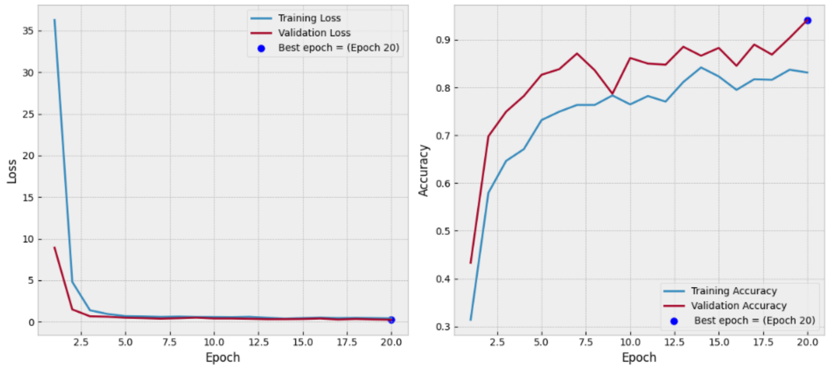

Figure 9: Graph representing model accuracy and model loss for training and validation set using the MobileNetV2 approach for Liquor.

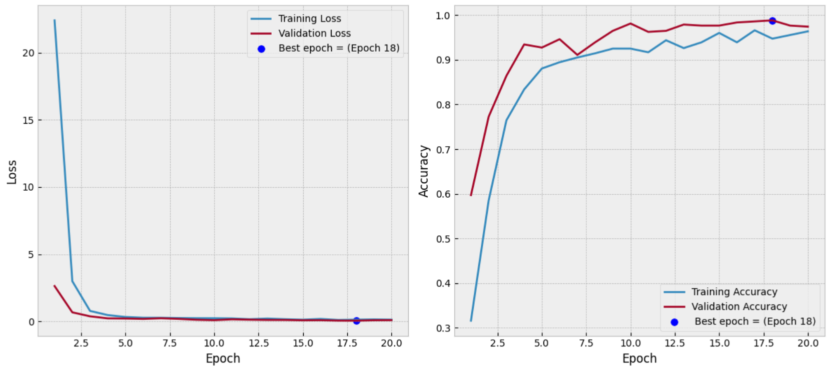

Figure 10: Graph representing model accuracy and model loss for training and validation set using the DenseNet201 approach for Liquor.

The CNN model for tea liquor showed an initial validation accuracy above 0.3, which increased to nearly 0.75 after 2.5 epochs. The initial validation loss is close to 1.75 but decreased considerably after 7.5 epochs. At first, testing accuracy is low, but it progressively improves to almost 94.86

percent. The transfer learning part of the measure was accomplished on the MobileNetV2 and the DenseNet201 model, which are shown in Figure 9 and Figure 10, respectively.

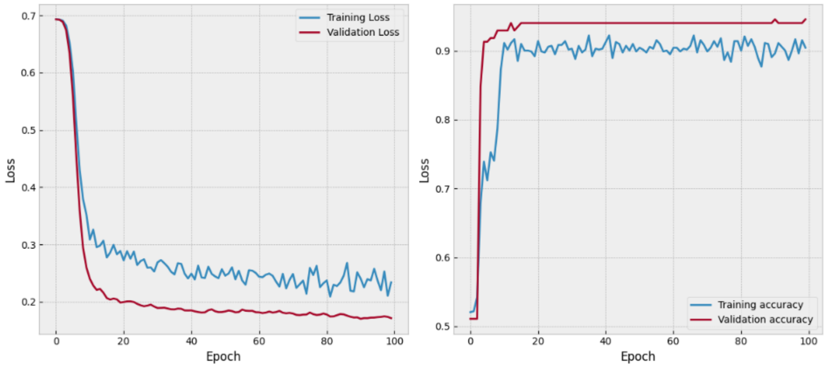

Figure 11: Graph representing model accuracy and model loss for training and validation set using ANN model.

3.4 Random Forest, and ANN Classifier Models for Acceptance Based on High Fired or Not.

The present study showed that the ANN model is suitable for predicting the final acceptance of teas based on 13 attributes. It showed that this model has more accuracy, 95%, compared to the RF model, which has 92.13%. The results of this study demonstrate the potential of machine learning models in predicting the acceptability of tea liquors based on their pH of water, TDS of water, brewing temperature, pH of tea, TDS of tea, and RGB color properties, which were collected using an Arduino Uno.

The study aimed to create a reliable model for classifying tea samples based on tea-tasting terminology, with the final acceptance of liquors based on whether they are highly fired or not. Convolutional neural networks, transfer learning models, Random Forests, and Artificial Neural Networks were used on datasets containing dry and infused tea leaves and liquor. Results are presented in Tables 1 and Table 2 for each model.

Table 1: Comparison of performance among different deep-learning-based techniques

Data Set	Model	Precision	Recall	F1-Score	Accuracy
Dry Leaf	CNN	92.66	92.49	92.48	92.49
	MobileNetV2	86.66	84.55	83.16	84.55
	DenseNet201	93.78	93.60	93.53	93.60
Infused Leaf	CNN	89.07	87.96	87.77	87.96
	MobileNetV2	95.81	95.70	95.67	95.70
	DenseNet201	96.92	96.77	96.75	96.77
Tea Liquor	CNN	94.64	94.63	94.56	94.63
	MobileNetV2	91.21	90.42	90.32	90.42
	DenseNet201	97.40	97.20	97.22	97.20

Table 2 : Comparison of performance between ANN and RF model

Data Set	Model	Precision	Recall	F1-Score	Accuracy
13 Attributes	ANN	96.00	95.50	96.00	95.50
Data Set	Random Forest	91.00	92.50	91.50	92.50

The results showed that the MobileNetV2 dry leaf and liquor model have lower accuracies than the CNN models built from scratch. Table 2 shows that in the dry leaf, infused leaf, and liquor datasets, DenseNet201 outperforms the other two models regarding precision, recall, and F1-score. In the DenseNet201 mode, the infused leaves and liquor all have high precision, recall, and F1-score values. Moreover, denseNet201's overall performance in dry leaf could be better than expected. The research also evaluated the acceptability of tea liquors based on whether they were high-fired or not, using Random Forest Classifier and ANN models. The study found that ANN achieved high levels of accuracy of 95%.

3.5 Linear Regression Model for Brix and TDS Value

The linear regression analysis of the data showed that the slope (m) and intercept (c) values were 0.005 and -0.149, respectively. The R-squared value (R^2) of 0.286 suggests that only 28.6% of the variation in Brix value can be explained by the linear relationship with the TDS value in tea. This means there is a significant amount of unexplained variance, and the model might not accurately predict the Brix value based on the given TDS value. Also, the Mean Squared Error (MSE) and the Root Mean Squared Error (RMSE) values are 0.0095 and 0.0975, respectively.

The adjusted R^2 of 0.9992 (Gómez, 2019) suggests a robust linear relationship between Brix and TDS values in coffee. In contrast, the present study ($R^2 = 0.286$) indicates a fragile linear association for tea. According to the findings, there is a weak positive trend between the Brix value and TDS value of tea; the R^2 value and error metrics suggest that the model does not capture the relationship very well. There might be other factors influencing Brix value that the model does not account for. The study also revealed that the TDS values varied according to the tea grades. These findings provide valuable insights into the properties and characteristics of low-grown tea samples, which can be helpful for the tea industry in product development and quality assurance.

3.6 Performance Evaluation

The study compares various machine learning algorithms performed by other researchers, revealing improved accuracy in detecting tea flavor perception. Transfer learning models, particularly

DenseNet201, can reliably classify tea samples based on leaf and liquor properties, proving useful in tea-tasting terminology, according to Tables 3 and 4.

Table 3: Comparison of performance among proposed DenseNet201 model and other techniques

Authors	Model	Data Set	Accuracy
(Tharanga et al., 2023)	ANN	180 Tea Liquor Image Data set	95.00%
(Zheng et al., 2022)	CMTP-CNN	SA-402B e-tongue (Insent Inc., Japan) Data Set	96.00%
Proposed Method	DenseNet201	Liquor Image Dataset	97.20%

Table 4: Comparison of performance among proposed ANN model and other techniques

Authors	Model	Data Set Attributes	Accuracy
(Zheng et al., 2022)	RF	SA-402B e-tongue (Insent Inc., Japan) Data Set	88.00%
(Ravi Kumar et al., 2020)	Support Vector Machine (SVM)	Caffeine, Polyphenolic, Catechins, Aroma, Appearance, Briskness, Types, Touch, Tea quality	91.00%
(Zheng et al., 2022)	RBFNN	SA-402B e-tongue (Insent Inc., Japan) Data Set	
Proposed Method	ANN	Tea Grade, Elevation, Water pH, Water TDS, Temperature, Brew Time, Liquor pH, TDS, RGB, Taste Type, Final Acceptance	95.50%

CONCLUSIONS/RECOMMENDATIONS

This research is an attempt to introduce an Arduino-based tea-tasting system in black tea with the help of Arduino Uno plus Machine learning based on the tea-tasting terminology in Sri Lanka and whether it is high-fired or not. Also, this study focused on determining the relationship between the pH value and TDS value of black tea according to linear regression. Arduino-based tea-tasting systems can be essential due to the drawbacks of human tasters and the need for more professional tea tasters. Based on the data (images and readings) gathered from sensors, it can be done with the combination of professional tea tasters' judgments. Transfer learned DenseNet201 is a better solution for implementing machine learning for a tea tasting with the help of the Arduino platform. The design of DenseNet encourages the efficient use of parameters by repeating feature maps from earlier levels. This can be beneficial in contexts like Arduino, where processing power may be restricted. Relying on human tasters to identify different tea types is laborious and time-consuming, especially for small teams. This can be a burden on the tea industry. The model and methodology proposed in this study could be a major breakthrough as it has the potential to significantly improve efficiency by automating tea identification, streamlining the process, and reducing the workload of tea tastes. This would be especially helpful if many tea samples were handled.

The acceptance based on the firing state of the tea not only affects to the final acceptance in some situations like fungi affected samples. Also, to identify sugar based adulterations A Near infrared spectrometer (NIRS) would be implemented. This study does not focus on those gaps and needs to be improved in the future.

Keywords: Arduino Uno, Artificial Neural Network (ANN), Convolutional Neural Network (CNN), DenseNet201, Tea Tasting.

References

- Bhattacharyya, N., Bandyopadhyay, R., Bhuyan, M., Tudu, B., Ghosh, D., Jana, A., 2008. Electronic nose for black tea classification and correlation of measurements with “Tea taster” marks. IEEE Trans Instrum Meas 57, 1313–1321. https://doi.org/10.1109/TIM.2008.917189
- Patil, A., Bachute, M., Kotecha, K., 2021a. Identification and classification of the tea samples by using sensory mechanism and arduino uno. Inventions 6. https://doi.org/10.3390/inventions6040094
- Patil, A.B., Bachute, M.R., Kotecha, K., 2021. Artificial Perception of the Beverages: An In-Depth Review of the Tea Sample. IEEE Access. https://doi.org/10.1109/ACCESS.2021.3086038
- He, W., Hu, X., Zhao, L., Liao, X., Zhang, Y., Zhang, M., Wu, J., 2009. Evaluation of Chinese tea by the electronic tongue: Correlation with sensory properties and classification according to geographical origin and grade level. Food Research International 42, 1462–1467. https://doi.org/10.1016/j.foodres.2009.08.008

- Tharanga, S., Pulasinghe, K., Marikar, F., 2023. ARTIFICIAL NEURAL NETWORK BASED INTELLIGENT TEA TASTER-REVIEW. Automation of technological and business processes 15, 7. https://doi.org/10.15673/atbp.v15i1.2491
- Zheng, W., Shi, Y., Xia, X., Ying, Y., Men, H., 2022. A data processing method for electronic tongue based on computational model of taste pathways and convolutional neural network. Measurement (Lond) 205. https://doi.org/10.1016/j.measurement.2022.112150
- Ravi Kumar Tata, B.S.V.N.P.S.V.S.S.M.V.R.M., 2020. APPLICATION OF MACHINE LEARNING TECHNIQUES IN QUALITY ANALYSIS OF TEA. Journal of critical reviews 7. https://doi.org/10.31838/jcr.07.14.62
- Gómez, O.S., n.d, 2019. Converting Brix to TDS-An Independent Study Converting Brix to TDS An Independent Study. https://doi.org/10.13140/RG.2.2.10679.27040