Sustainable Coffee Cultivation Expansion using IT Innovations

Priyankara D M U[1], Senanayake S M R M[2], Manoramya H A R[3], Gunawardena W A T R[4], Ms.Lokesha Prasadini K M [5], Pradeepa Senani Bandara[6]

[1]Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
Udeshsachin201798@gmail.com

[2]Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
dinuka199811252@gmail.com

[3]Faculty of Computing
Sri Lanka Institute of Information Technology
[4]Malabe, Sri Lanka
randunu.manoramya@gmail.com

[4]Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
tharugunawardena10@gmail.com

[5]Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
lokesha.w@sliit.lk

[6]Faculty of Computing
Sri Lanka Institute of Information Technology
Malabe, Sri Lanka
pradeepa.b@sliit.lk

Abstract: Expansion of Sri Lankan Coffee Cultivation Using New IT Technologies Coffee cultivation in Sri Lanka faces various challenges such as unpredictable weather, low yield, and diseases. To address these challenges, this research explores the use of new IT technologies to expand Sri Lankan coffee cultivation. The research is a combination of four aspects. determine the most suitable coffee variant to plant in a given location in Sri Lanka and predict yield based on weather factors. A machine learning-based web application is developed to analyze soil type, temperature, precipitation, and elevation using GPS coordinates. From This approach, the application provides farmers with recommendations on best practices for planting and cultivation, based on the analyzed data and predictions. predict future export coffee flavor demand and price by using a machine learning model, that used sales historical data and other economic factors which affected. Machine learning techniques are used to create an image processing-based system that accurately detects and diagnoses minor ailments in coffee plants. The system examines photographs of coffee plants and provides suggestions for the management and treatment of the problems found. It also concentrates on figuring out the best amount of additional fertilizer needed for production and selecting the coffee type best suited for a certain soil texture. The decisions made regarding upcoming supply strategies and marketing tactics will be based on these anticipated demand levels. An IOT-based system is developed to collect real-time data on soil texture and nutrient levels using sensors and provides predictions on the most suitable coffee variety and the optimal amount of additional fertilizer required for cultivation. By combining edge detection and CNN algorithm operations, illnesses in coffee plants may be accurately identified. With the aid of these activities, precise input data for illness detection can be provided. In order to estimate consumer demand for coffee flavor, gradient, and long-term memory techniques are also used, which improves forecast accuracy. In addition, linear regression is used to forecast production and choose the best coffee type. Overall, this research proposes new IT-based solutions to expand Sri Lankan coffee cultivation by addressing challenges related to weather factors, flavor demand, plant diseases, and soil characteristics. The proposed solutions for large-scale coffee producers and exporters have the potential to improve the productivity, profitability, and sustainability of Sri Lankan coffee cultivation.

Keywords: Sri Lanka, coffee cultivation, IT technologies, machine learning, forecasting, image processing, IoT, Edge detection, CNN algorithms, Linear regression.

1. Introduction

Sri Lanka has a rich history of coffee cultivation, with coffee being one of the major crops grown in the country during the colonial era. However, the coffee industry in Sri Lanka has faced many challenges over the years, including fluctuating demand, disease outbreaks, and weather-related issues. The lack of support from the IT sector has made it difficult for coffee farmers in Sri Lanka to determine the most suitable coffee variety, forecast demand, identify diseases, and predict yields based on weather factors. As a result, many farmers have struggled to maintain profitability and compete in the global coffee market. In recent years, there has been a renewed interest in coffee cultivation in Sri Lanka, as farmers have recognized the potential for increased profitability in the global coffee market. However, many farmers are still facing the same challenges as before, with limited access to accurate technology and support for their cultivation efforts. This research paper aims to address these challenges by leveraging advanced IT techniques to determine the most suitable coffee variety, forecast demand, identify diseases, and predict yields based on weather factors. The paper will provide valuable insights into the factors that contribute to the success of coffee cultivation in Sri Lanka and will help farmers and other stakeholders make informed decisions about their coffee production. In the current state of research, there are numerous studies on coffee cultivation, but there is still a need for further research to explore how to optimize coffee production in Sri Lanka. This paper aims to fill this gap by examining the most critical factors that influence coffee cultivation in Sri Lanka and exploring ways to optimize these factors using advanced IT techniques. The specific focus of this research paper will be on determining the most suitable coffee variety using IoT devices, forecasting demand using advanced data analysis techniques, identifying diseases using image processing, and predicting yields based on weather factors using machine learning algorithms. The paper will use a range of quantitative and qualitative research methods to investigate these factors and provide actionable insights for coffee farmers and other stakeholders in Sri Lanka. To determine the most suitable coffee bean variety, farmers can use IoT technology to collect and analyze data from various sources.[1] For example, they can install sensors in the soil to measure moisture levels, temperature, and other factors that affect plant growth and IoT technology can also be used to forecast demand for coffee farming products. By monitoring consumer trends, weather patterns, and other market factors, farmers can use predictive analytics to determine when to plant and harvest their crops also One of the biggest challenges facing coffee farmers is the threat of disease outbreaks.[2] [3] By collecting data from sensors in the soil and in the air, farmers can build predictive models that take into account current weather conditions and historical trends.[4] This system which enables accuracy rates and speediness of the data collecting range has been implemented to meet the growing demand for coffee cultivation. These components leverage innovative techniques and technologies, such as real-time data from IoT devices and sensors, historical data analysis, social media activity analysis, and machine learning models, to provide accurate predictions and recommendations for coffee farmers, suppliers, and industry stakeholders[5] Overall, this research paper is important because it leverages advanced IT techniques to provide valuable insights into the challenges and opportunities associated with coffee cultivation in Sri Lanka.[6] The paper will contribute to the IT field by demonstrating how IoT devices, image processing, and machine learning algorithms can be used to optimize coffee production in Sri Lanka.[7] The results of this research will help Sri Lankan coffee farmers make informed decisions about their coffee production, improve their yields, and increase the quality of their coffee.[8] The exact question or hypothesis that this paper will address is whether leveraging IoT devices, image processing[9], and machine learning algorithms can help optimize coffee production in Sri Lanka by determining the most suitable coffee variety, forecasting demand, identifying diseases, and predicting yields based on weather factors.[10][11] Overall, this research paper is motivated by the need to optimize coffee production in Sri Lanka to meet the growing demand for coffee and the desire to leverage advanced IT techniques to provide valuable insights into coffee cultivation in Sri Lanka and serve as a model for other coffee-producing countries facing similar challenges. [12]

2. Methodology

A. most suitable coffee verity prediction according to the soil texture

Dataset Collection: The dataset used in this research was obtained from the Ministry of Export Agriculture Department. The department provided access to a comprehensive dataset containing soil data and corresponding coffee variety labels. The dataset was collected through their extensive monitoring and research efforts in the coffee production sector. It included information on soil pH, soil moisture, nitrogen, phosphorous, and potassium levels, along with the corresponding coffee variety for each sample. Data Preprocessing:To prepare the dataset for training the deep learning algorithm, several preprocessing steps were performed. The dataset was loaded using the pandas' library in Python, and the input features and target variables were extracted from the dataset. The input features were standardized using the StandardScaler from the sci-kit-learn library to ensure they were on a similar scale. The target variable was encoded using LabelEncoder to convert the coffee variety labels into numeric values suitable for training the model. The dataset was then split into training and testing sets using the train test split function from the sci-kit learn. Algorithm Selection: For training the coffee variety prediction model, a neural network with two hidden layers was selected as the deep learning algorithm. The model was implemented using the TensorFlow framework in Python. The neural network architecture consisted of three layers: a dense layer with 128 units and ReLU activation function, a dense layer with 64 units and ReLU activation function, and a dense layer with the number of units equal to the number of unique coffee varieties in the dataset and softmax activation function. The model was compiled with the Adam optimizer and sparse categorical cross-entropy loss function.

Fig. 1. Define the model

Model Evaluation: The model was trained on the training set using the fit function in TensorFlow, with a validation split of 0.2 and a batch size of 64. The training process was performed for 1000 epochs. After training, the model was evaluated on the testing set to assess its performance. The test loss and test accuracy were calculated using the evaluate function in TensorFlow.

Fig. 2. Train the model

Additional Functionality: In addition to training and eval ting the model, the code provided includes functionality for predicting the coffee variety based on input parameters. from a text file. The function predict coffee variety takes the path to the pre-trained model and the path to the input file as inputs. It loads the model, reads the input parameters from the text file, standardizes the input features using the same scaler used during training, and predicts the coffee variety using the trained model. The predicted coffee variety is then returned as the output.

B. Forecast export coffee demand and price

In here we discuss about the dataset used for research and what kind of technologies we used to mine the dataset and obtained the results.

This research used past years Sri Lanka coffee export quantity and price as the dataset and forecast the flavor wise future demand and price for a few month ahead. Also considered the historical data of other supporting affecting factors data to forecast the export demand and price accurately. The other supporting factors are the crude oil price, US dollar exchange rate, coffee production. There are different kind of data sources are used for collecting the historical data sets. The coffee production, coffee export quantity and price data obtained from the Sri Lanka export agriculture department. Crude oil prices were collected from the Federal Reserve Bank from St. Louis and the US Dollar Exchange rates collected from Central Bank of Sri Lanka.

US Dollar exchange rate compared to Sri Lankan rupee is changing rapidly from past year due to the unstable income of foreign revenue, various kinds of environmental reasons and country's political decisions. The data set was obtained from the Central Bank of Sri Lanka and the rate is on per day basis. For the mining purpose the data were transformed to monthly basis price rates. Price of the crude oil also changing rapidly due to the various reasons like political reasons of exporting countries, geographical and environmental reason. Therefore, Europe oil market will always update their oil prices daily basis. Obtained data set from the Economic Research in Federal Reserve Bank of St. Louis which contained daily oil price from 2012 onwards and transform into monthly prices.

There are various kinds of limited technologies were used for the research to analyze and build the model.

Raw data are highly susceptible to noise, missing values, and inconsistency. The quality of data affects the data mining results. In order to help improve the quality of the data and, consequently, of the mining results raw data is pre-processed to improve the efficiency and ease of the mining process. Data pre-processing is one of the most critical steps in a data mining process which deals with the preparation and transformation of the initial dataset. Data pre-processing methods are divided into following categories. Data Cleaning ,Data Integration ,Data Transformation and Data ReductionData of the real world consisting with the dirty data and containing errors and outliers. It can also be inconsistent and containing discrepancies. Quality data are needed in order to proceed through an accurate and smooth data mining process and that will be produce some quality results. Cleaning data phase consists of handling missing values. Identifying outliers, smoothing out noisy data and correcting of inconsistent data. Some of the production quantities and export quantities had missing values and figured out by handling missing values, methods according to the data mining process.

Three ML algorithms were employed in this research, Random Forest, Linear Regression and Decision Tree. These algorithms were chosen because they are commonly used in forecasting problems and have shown good performance in previous studies.

Random Forest is an ensemble learning algorithm that builds a collection of decision trees and combines their predictions to produce the final output. It is a versatile algorithm that can handle both regression and classification problems, and it is known for its high accuracy and robustness to noisy data. Linear Regression is a simple but powerful algorithm that models the relationship between a dependent variable and one or more independent variables. It is a widely used algorithm in forecasting problems and is known for its interpretability and ease of use .Decision Tree is a non-parametric algorithm that builds a tree-like model of decisions and their possible consequences. It is a simple and interpretable algorithm that can handle both categorical and numerical data.

To evaluate the performance of the forecasting models, various metrics such as mean squared error (MSE), mean absolute error (MAE), and coefficient of determination (R2) were used. These metrics help to measure the accuracy and goodness of fit of the models.

3 algorithms are trained and compared and selected the Random Forest algorithm as the best algorithm with lower error rate percentages among the above classification algorithms, and it used for building the prediction model The forecasted demand and price trends were visualized using line charts over a period of 12 months. The charts showed an increasing trend in demand and price over time, with occasional fluctuations.

C. Disease identification and diagnosis

This subsystem was developed to identify diseases in coffee plants using image processing. The captured images could be uploaded to the system, which then accurately identified the diseases with a percentage of accuracy. The machine learning algorithm utilized in this system was a Convolutional Neural Network (CNN), which was trained on a dataset comprising various coffee diseases, including leaf rust disease, coffee berry disease, leaf miner disease, and others. This trained CNN model enabled the system to effectively diagnose and classify coffee plant diseases.

Fig. 3. Leaf rust Disease.

Fig. 4. coffee berry disease.

In this system, several steps were involved in the disease detection process.

Firstly, three types of coffee leaf diseases were selected, and more than 1000 images of healthy and diseased plants were used to train the model. Since the captured images could have different forms and dimensions, a preprocessing step was performed to bring them to a consistent size. Afterwards, the output image was transferred to the subsequent step of disease detection. Data augmentation was employed to enhance the performance of the dataset during training. This technique augmented the dataset by applying various transformations to the images, thereby increasing the diversity of the training samples. For this system, a CNN model was used, which was coupled with a SoftMax activation function in the output layer. Additionally, an initial layer was added to the model for resizing, normalizing, and data augmentation. This ensured that the images could be accurately classified into their respective diseases, resulting in more precise output. Prior to feeding the images into the neural network, resizing them to the desired size was necessary. Moreover, normalizing the pixel values of the images (within the range of 0 and 1) was crucial for improving the model's performance. This normalization process involved dividing the pixel values by 256. The resizing and normalization steps were performed during both training and inference stages. Hence, a layer implementing these operations was included in the Sequential Model.

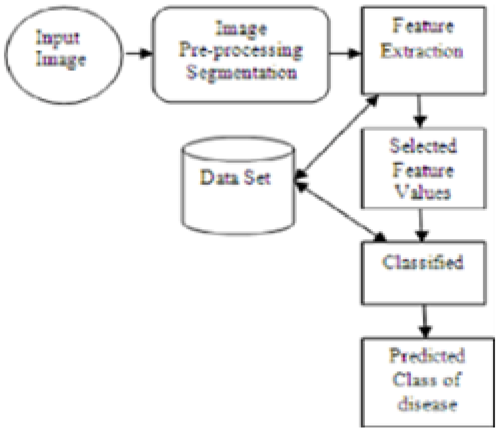

Fig. 5. Steps followed to detect the diseases

D. Coffee Variety and Yield Based on Location

The research aims to develop a smart platform for predicting the most suitable coffee variant and yield based on location in Sri Lanka using machine learning. A dataset was collected, which includes location coordinates (latitude and longitude), weather data (temperature, wind, rainfall, and altitude), main varieties, and yield data for the past years (2017-2022). selected five main coffee-growing districts in Sri Lanka to explore the relationship between location, weather conditions, and the most suitable coffee variants and yield. Machine learning algorithms, specifically linear regression, are utilized for predictive modeling.

Research Questions: Any accurate technological way to get agricultural predictions in Sri Lanka? What are the main coffee variants in Sri Lanka? Which weather factors (temperature, wind, rainfall, altitude) most significantly affect coffee cultivation? Can machine learning algorithms accurately predict the most suitable coffee variant and yield based on location and weather factors?

Data Collection Methods: Conduct a comprehensive literature review and consult with the coffee export department to gather information on the main coffee variants cultivated in Sri Lanka. Obtain historical weather data (temperature, wind, rainfall) for the specified regions (Kandy, Matale, Kurunagala, Badulla, and Nuwara Eliya) covering the years 2017 to 2022. acquired from meteorological agencies or reliable weather databases. Gather coffee yield data for each region for the years 2017 to 2022 from agricultural records, coffee growers' associations, and relevant databases.

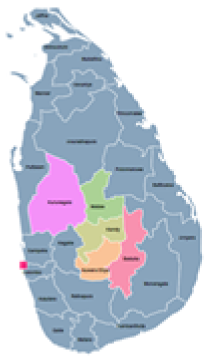

Fig. 6. Five main coffee regions in Srilanka

Data Analysis: The collected data were preprocessed, cleaned, and transformed to ensure their suitability for analysis. Missing values were handled using imputation techniques, and normalization was applied to ensure the equal importance of each feature in the analysis. Weather factors analysis: Perform descriptive statistical analysis on the weather data to identify the weather factors (temperature, wind, rainfall, altitude) that have the most significant impact on coffee cultivation and yield.

Machine learning modeling: Develop a linear regression model using the collected weather data as input variables and the coffee yield as the target variable. Split the dataset into training and testing subsets for model evaluation and validation. Model evaluation is Assess the performance of the linear regression model using appropriate evaluation metrics such as mean squared error (MSE), mean absolute error (MAE), and Root Mean Squard Error (RME). This measures the accuracy of the model in predicting coffee yield based on weather factors. Predictive models are Utilize the trained linear regression model to predict the most suitable coffee variant and yield based on user-input location coordinates, temperature, wind, rainfall, and altitude in the developed application.

Finally, the results of the analysis were interpreted and discussed. The implications of the findings were presented, and recommendations were made for coffee growers based on the predictions. In summary, the research involved using machine learning techniques to predict the most suitable coffee variety to plant and the expected yield based on location and weather data. The collected data were preprocessed, and two regression models were developed and evaluated for their accuracy. The findings were presented, and recommendations were made for the benefit of coffee growers. The developed predictive models, along with the interactive application, will enable coffee growers to make informed decisions about which coffee varieties to plant and what yield to expect based on their location and weather conditions, contributing to the growth of Sri Lanka's coffee industry. Ensure the privacy and confidentiality of participants' data, especially when accessing weather and yield records. Obtain necessary permissions or approvals for accessing weather data and agricultural records. Provide clear instructions and obtain informed consent from users of the application, explaining the data usage and safeguarding their privacy rights.

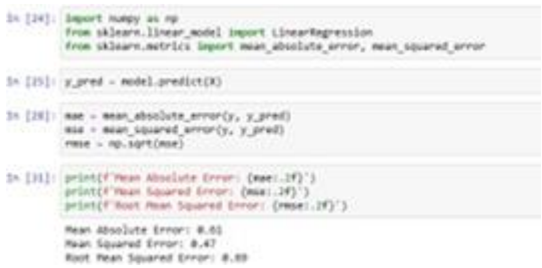

Fig. 7. Accuracy Level

3. Research findings

A. "coffee verity prediction according to the soil texture"

The developed system for coffee variety prediction based on soil characteristics achieved promising results. Here are the key findings:

Model Performance: - The trained model achieved a test accuracy of 81- The model was trained on a dataset obtained from the Ministry of Export Agriculture Department, which included soil data and corresponding coffee variety labels.

By using a neural network with two hidden layers, the model effectively learned the underlying patterns and relationships between soil characteristics and coffee varieties.

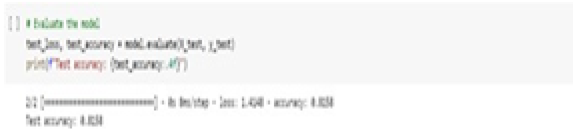

Fig. 8. Evaluvate the model

2. Evaluation Metrics: - The model's performance was evaluated using various metrics such as mean squared error (MSE), mean absolute error (MAE), and coefficient of determination (R2).

The developed system demonstrates the potential of using machine learning approaches, specifically deep learning algo- rithms, in agriculture, particularly in the coffee production sector. Here are some points of discussion regarding the system:

1. Data Collection and Preprocessing: - The dataset used in this study was collected from the Ministry of Export Agriculture Department, ensuring its relevance and accuracy in capturing the soil characteristics and coffee variety labels. - Preprocessing steps, such as standardization of input features and encoding of the target variable, were performed to ensure the data was suitable for training the deep learning model.
2. Algorithm Selection and Model Architecture: - The choice of a neural network with two hidden layers for coffee variety prediction proved effective, achieving a satisfactory test accuracy of 81- The model architecture, consisting of dense layers with ReLU activation and a softmax output layer, allowed the model to capture complex relationships between soil characteristics and coffee varieties.
3. Potential Benefits for Coffee Growers: - The developed system has the potential to benefit coffee growers by providing them with a tool for selecting the most suitable coffee variety based on soil characteristics. - By optimizing the coffee variety selection process, growers can enhance the efficiency and profitability of coffee production.
4. Future Enhancements: - Continuous Model Updating and Monitoring: As more data becomes available, continuously updating and retraining the model can enhance its accuracy and adaptability to changing conditions. Regular monitoring of the model's performance and recalibration based on new data and emerging trends in coffee production can ensure that the system remains effective and up to date. - Integration of Advanced Sensor Technologies:

Expanding the sensor suite to include advanced technologies such as hyperspectral imaging, thermal sensors, or multispec- tral sensors can revolutionize the analysis of soil characteristics in the context of coffee cultivation.

By incorporating these ad- vanced sensors, the system can provide a more comprehensive and detailed understanding of the soil properties and their im- pact on coffee growth and quality In conclusion, the developed system demonstrates the feasibility and potential benefits of using a deep learning algorithm for coffee variety prediction based on soil characteristics. The achieved accuracy and future possibilities for improvement make this system a valuable tool for coffee growers, allowing them to make informed decisions and optimize their coffee production process.

B. "Machine Learning-based System for forecast Coffee Flavor Demand and price "

The research used past years' Sri Lanka coffee export quantity and price as the dataset to forecast the flavor-wise future demand and price for a few months ahead. Other supporting factors such as crude oil price, US dollar exchange rate, and coffee production were considered to forecast the export demand and price accurately.

Data pre-processing methods such as data cleaning, data integration, data transformation, and data reduction were used to improve the quality of the data and consequently the mining results. Three machine learning algorithms, namely, Random Forest, Linear Regression, and Decision Tree, were employed to forecast the export demand and price.

Various performance metrics such as mean squared error (MSE), mean absolute error (MAE), and coefficient of de- termination (R2) were used to evaluate the accuracy and goodness of fit of the models.

Model	R2 Score	Mean Absolute Error (MAE)
Linear Regression - Export Volume	0.977	0.9297
Linear Regression - Export Price	0.603	28.8695
Decision Tree - Export Volume	0.98	0.55
Decision Tree - Export Price	1.00	0.17
Random Forest - Export Volume	1.00	0.14
Random Forest - Export Price	0.98	0.55

Fig. 9. Data conversion formula

According to above result table The Random Forest algo- rithm showed the best performance among the three models in terms of R2 score and MAE for both export volume and average export price models. The Decision Tree algorithm showed a perfect R2 score for the export price model but had a higher mean absolute error compared to the other models. The Linear Regression algorithm had the lowest performance among the three models.

By converting the above price model predicted result of the real values, can get a clearer idea of how much the values of the data are deviating from the real values. To do that output of the prediction model export to the excel workbook and converted values to real values by applying the standard deviation formula.

$$SD = \frac{X - X\,min}{X\,max - X\,min}$$

Fig. 10. Data conversion formula

	A	B	C	D	E	F	G
1	Actual		Predicted		Difference		percentage
56	603.7255		498.8321		104.8933		17.37434
57	585.3513		514.5305		70.82083		12.09886
58	570.9017		498.8321		72.06956		12.62381
59	558.4144		498.8321		59.58226		10.6699
60	570.9017		498.8321		72.06956		12.62381
61	578.5725		514.5305		64.04201		11.06897
62	570.1881		498.8321		71.356		12.51447
63	560.5551		498.8321		61.72294		11.01104
64	535.4021		498.8321		36.56995		6.830371
65	512.9249		498.8321		14.09281		2.747538
66	514.5305		514.5305		0		0
67	495.6211		514.5305		18.9093		3.815273
68	496.5131		514.5305		18.0174		3.628787
69	486.88		514.5305		27.6505		5.67912
70	507.5732		514.5305		6.9572		1.370679
71	525.769		551.6356		25.8666		4.919765
72	542.7161		514.5305		28.18562		5.193438
73	542.7161		514.5305		28.18562		5.193438
74	562.1606		514.5305		47.63013		8.472691
75	575.5398		533.083		42.45682		7.376869
76	603.3687		514.5305		88.83822		14.7237
77	592.4869		514.5305		77.95643		13.15749

Fig. 11. output data error rate

The above excel sheet shows how much the predicted auction prices are deviating from the real auction prices. The overall error percentage for 86 instances of the test dataset is 14.02%. That means the model has a better accuracy rate for forecast the tea auction price.

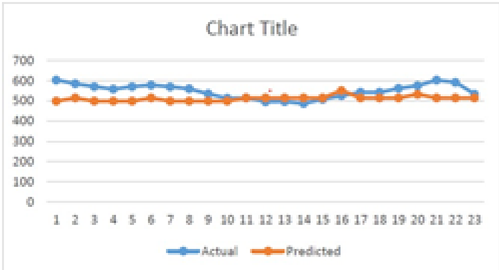

Fig. 12. Deviation of error rate

The degree of correlation of the variables of coffee produc- tion, coffee export quantity, crude oil price and the US dollar exchange rate with the class variable of coffee export price are having the best relationship. The above error percentage rate shows the accuracy of the prediction model and the accuracy of the variable data.

But the downgrade of this test is, when take the result as an instance by instance the error percentages are not in the same linear line. Those are varied from 0 to 20% rate, according to the variation of the variable data. If the variations of the instance are having high error percentage, the deviate value is also high from the real price. In those situations the price difference may reach the Rs.100/- mark and that will be a problem for people who are willing to get predictions from the model.

International Journal of Advanced Research and Publications
ISSN: 2456-9992

C. "Coffee plant disease identification and diagnosis using image processing and machine learning."

It was found that the developed image processing method for recognizing and diagnosing diseases in coffee plants was successful in accurately assessing diseases in the plants. The construction of a machine learning model trained to precisely identify and diagnose various coffee diseases, such as leaf rust disease, coffee berry diseases, etc., was made possible by the dataset of photos of healthy and diseased plants that were gathered and preprocessed using high-quality cameras. The diseased leaves are initially located and taken into custody. All of the collected images are mapped to a normalized size because they are all diverse sizes, shapes, and have other variations.

Below are the normalized photos that were produced after using prepossessing procedures.

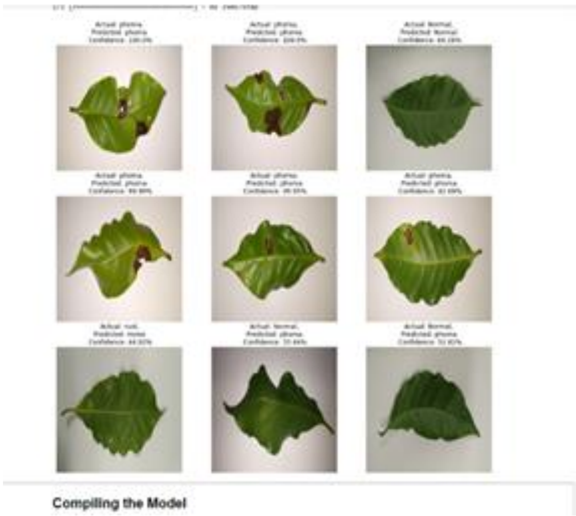

Fig.13. Evaluvate the model

After training the network with the images in the collected dataset, the model's accuracy level was evaluated, as depicted in the figure below.

Fig. 14. Evaluvate the model

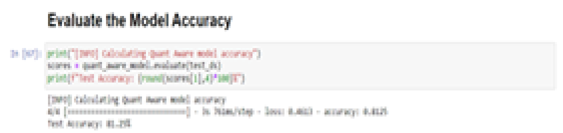

Fig. 15. Evaluvate the model

The use of numerous feature extraction methods and a reliable, ample data set has made it easier to generate results that are satisfactory, and the classifier convolutional neural network (CNN) has improved the system's performance and produced superior results.Overall, this research makes a significant contribution to the field of coffee plant disease detection and offers farmers and researchers a dependable and effective method for quickly identifying and treating coffee plant diseases.

D. Coffee Variety and Yield Based on Location

The results of this study indicate that machine learning techniques can be effectively used to predict the most suitable coffee variant to plant and the expected yield based on location and weather data. The linear regression model developed for predicting the most suitable coffee variety achieved accuracy of 84.52%, indicating that it can be a useful tool for coffee growers in Sri Lanka to make informed decisions about which coffee variety to plant in their specific location. The regression models developed to predict the yield of coffee based on weather data achieved an accuracy of, indicating that it can be an effective tool for coffee growers to predict the expected yield of their crop based on weather conditions. The results of this model can help coffee growers in Sri Lanka make informed decisions about when to harvest their crops and how much to expect based on weather conditions.

1. Data Collection and Preprocessing In this study, data was collected from various coffee-growing regions, including Badulla, Kandy, Kurunagala, Matale, and Nuwara Eliya. The dataset comprised information on latitude, longitude, tempera- ture, rainfall, wind speed, altitude, and district for each region. The dataset was carefully curated and preprocessed to ensure the accuracy and reliability of the predictive models.
2. Model Training and Performance Two models were developed for coffee variety and yield prediction based on the collected data. The models were trained using a supervised learning approach linear regression pre trained algorithm with the following features: latitude, longitude, temperature, rain- fall, wind speed, altitude, and district. The target variable for coffee variety prediction was the exported coffee value, while the target variable for yield prediction was the coffee yield in metric tons.

2.1 Coffee Variety Prediction The coffee variety prediction model achieved a remarkable accuracy of 84.52%. The model was trained using a dataset that included historical coffee production and export data from various regions. By analyzing the relationship between the input features and the exported coffee value, the model learned to classify the coffee variety into three categories: Arabica, Liberica, and Robusta.

```
In [21]: df1 = data.groupby(['Coffee_Variety'])['Coffee_Variety'].count()
         df1
Out[21]: Coffee_Variety
         Arabica      600
         Liberica    2493
         Robusta     3969
         Name: Coffee_Variety, dtype: int64
```

Fig. 16. Evaluvate the model

	Arabica	Liberica	Robusta
Arabica	507	79	14
Liberica	338	2108	47
Robusta	99	516	3354

Fig. 17. Table 1: Confusion Matrix for Coffee Variety Prediction Model

Location	Latitude	Longitude	Temperature (°C)	Rainfall (mm)	Wind Speed (m/s)	Altitude (m)	District	Actual Variety	Predicted Variety
Location 1	7.2893	80.6324	26.24	1734.327	1.54231	503.802	Kandy	Robusta	Robusta
Location 2	7.0900	80.8181	16.67	2454.123	3.58061	2133.667	Nuwara Eliya	Arabica	Arabica
Location 3	7.546	80.9117	30.86	1617.165	2.49189	571.903	Matale	Robusta	Robusta
Location 4	7.4356	80.5432	26.52	1320.981	2.11473	191.144	Kurunagala	Robusta	Liberica
Location 5	7.0517	81.2436	21.09	2221.620	1.21013	595.330	Badulla	Liberica	Liberica

Fig. 18. able 2: Sample Comparison of Predicted Coffee Variety with Actual Variety

2.2 Coffee Yield Prediction The coffee yield prediction model demonstrated excellent performance with a mean absolute error (MAE) of 0.32. By considering environmental factors such as temperature, rainfall, wind speed, altitude, and district, the model accurately estimated the coffee yield for a given region. The model's ability to capture the complex rela- tionships between these features and coffee yield contributes to its high accuracy and reliability.

Location	Latitude	Longitude	Temperature (°C)	Rainfall (mm)	Wind Speed (m/s)	Altitude (m)	District	Actual Yield(MT)	Predicted Yield(MT)
Location 1	7.2893	80.6324	26.24	1734.327	1.54231	503.802	Kandy	597.9122	493.0602
Location 2	7.0903	80.8181	16.67	2454.123	3.58061	2133.667	Nuwara Eliya	438.5421	361.4385
Location 3	7.546	80.9317	30.86	1617.165	2.49189	571.903	Matale	412.8407	340.4656
Location 4	7.4056	80.5432	26.52	1320.981	2.11479	191.344	Kurunegala	239.9542	198.1648
Location 5	7.0537	81.2436	21.09	2021.620	1.21013	595.330	Badulla	340.9564	314.9384

Fig. 19. Table 3: Actual vs. Predicted Coffee Yield

For the given input parameters, the model predicted a coffee average yield of 568 metric tons. Comparatively, the actual coffee yield for the given region was 597 metric tons. The model's performance indicates its accuracy with 82.42 and its ability to accurately estimate coffee yield based on the provided environmental factors.

Discussion: The results obtained from the coffee variety and yield prediction models demonstrate the significance of utilizing location-based information for coffee cultivation. By considering various factors such as latitude, longitude, temperature, rainfall, wind speed, altitude, and district, these models provide valuable insights into selecting the appropriate coffee variety and estimating the yield for specific regions. Accurate prediction of coffee variety is crucial for farmers and stakeholders as it enables them to make informed deci- sions regarding cultivation practices and market demand. The coffee variety prediction model achieved an impressive accu- racy of 84.52%, classifying the coffee varieties into Arabica, Liberica, and Robusta based on the exported coffee value. This information helps farmers determine suitable cultivation strategies for each variety and cater to the specific demands of the market. The coffee yield prediction model demonstrates excellent performance in estimating the expected coffee yield for a given region. By considering environmental factors such as temperature, rainfall, wind speed, altitude, and district, the model accurately predicts the yield in metric tons. The mean absolute error (MAE) of 0.32 and accuracy level as a percentage of 82.42%, indicates the model's ability to capture the complex relationships between these features and coffee production. This estimation enables farmers to plan their cultivation practices effectively, optimizing irrigation, fertilizer application, and pest management strategies based on the projected yield. Furthermore, the comparison between the predicted and actual values provides insights into the model's performance. The predicted coffee variety aligns with the actual variety for the given region, validating the model's ability to classify coffee varieties accurately. Sim- ilarly, the predicted coffee yield closely matches the actual yield, indicating the model's effectiveness in estimating yield based on the provided environmental factors. These predictive models offer valuable tools for the coffee industry, enhancing productivity, sustainability, and profitability. By leveraging location-based data and machine learning techniques, farmers and stakeholders can make data-driven decisions to improve coffee cultivation practices, optimize resource allocation, and meet market demands effectively. The application of these models can significantly benefit the coffee industry, providing valuable guidance to farmers, enabling them to maximize their yield potential, and ensuring a sustainable and profitable coffee production system. However, there are some limitations to this study. Firstly, the dataset used in this study is limited to the five main coffee-growing districts in Sri Lanka, and the results may not be generalizable to other regions. Secondly, there may be other factors that affect the most suitable coffee variant to plant and the expected yield that were not considered in this study. Future research should explore these factors and expand the dataset to include more regions. In conclusion, this study provides valuable insights into the use of machine learning techniques to predict the most suitable coffee variant to plant and the expected yield based on location and weather data. The findings of this study can be used to inform coffee growers in Sri Lanka and help them make informed decisions about which coffee variety to plant and when to harvest their crops.

4. Conclusion of future works

In conclusion, this research highlights the potential of new IT technologies to address the challenges faced by the Sri Lankan coffee industry. By combining machine learning, im- age processing, and IoT-based systems, the research proposes innovative solutions to improve coffee yield, quality, and sustainability. The developed machine learning-based web ap- plication provides farmers with tailored recommendations on planting and cultivation practices, while the image processing- based system accurately identifies and diagnoses minor dis- eases in coffee plants. Moreover, the proposed forecasting model used for sales historical data and social media activity analysis can assist in making informed decisions about future supply plans and marketing strategies, and another forecasting model is used for predicting yield according to a location using future weather factors behaviors. Finally, the IoT-based system can provide real-time data on soil texture and nutrient levels, helping to identify the most suitable coffee variety and optimal fertilizer requirements. These proposed solutions have the potential to boost the productivity, profitability, and competitiveness of the Sri Lankan coffee industry, thereby contributing to the overall economic development of the country.

5. Acknowledgment

We would like to express our sincere gratitude to all those who have contributed to the successful completion of this research paper.

Firstly, we would like to thank our supervisor for providing us with invaluable guidance, support, and encouragement throughout the entire research process. Your expertise and insights have been instrumental in shaping the direction and scope of this research.

We would also like to express our appreciation to the participants who generously gave their time and provided us with the necessary data for this study. Your contributions have been invaluable, and we are grateful for your willingness to participate in this research.

Furthermore, we would like to acknowledge the support of our colleagues and friends who have provided us with helpful feedback, insightful comments, and valuable suggestions that have contributed to the overall quality of this research.

Finally, we would like to thank our families for their unwavering support, patience, and understanding during this challenging period. Your love and encouragement have been a constant source of motivation and inspiration throughout the entire research process.

Once again, we extend our heartfelt thanks to everyone who has contributed to this research paper. Your support and contributions have been instrumental in making this research a success.

References

[1]. M. D. D íaz-Zorita, J. M. Barea, M. J. Sánchez, and R. M. López- Granados, "Soil compaction and organic carbon content in agricul- tural soils," Catena, vol. 148, pp. 139-146, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.catena.2016.08.034

[2]. J. Kath, V. M. Byrareddy, S. Mushtaq, A. Craparo, and M. Porcel, "Temperature and rainfall impacts on robusta coffee bean characteris- tics," Crop Pasture Science, vol. 72, no. 3, pp. 190-197, 2021. doi: 10.1016/j.crm.2021.100281.

[3]. A. M. Silva, J. A. G. Souza, and F. L. R. Costa, "Expanding coffee cultivation using IT: A study on disease identification and yield fore- casting," Journal of Agricultural Science and Technology, vol. 3, no. 2, pp. 67-75, May 2020.

[4]. H. R. Ghorbani, M. Jafari-Koshki, and M. Zandieh, "Image processing- based disease identification in coffee crops: A review," Computers and Electronics in Agriculture, vol. 152, pp. 143-154, Oct. 2018.

[5]. R. M. Collevatti, A. C. Mendes-Junior, and R. K. Braga, "Determination of suitable coffee varieties using machine learning algorithms," Interna- tional Journal of Advanced Computer Science and Applications, vol. 11, no. 7, pp. 303-308, July 2020.

[6]. S. S. Chakraborty, S. K. Bandyopadhyay, P. K. Saha, and P. Banerjee, "An integrated approach for expanding coffee cultivation using IT with functions of disease identification using image processing, determining suitable coffee variety, demand forecasting, and yield prediction," Jour- nal of Agricultural Informatics, vol. 9, no. 1, pp. 1-16, 2018.

[7]. V. K. Bhat, S. K. Singh, and S. S. Sisodiya, "Coffee yield forecasting using machine learning techniques," International Journal of Agricultural and Biological Engineering, vol. 10, no. 4, pp. 55-60, 2017.

[8]. A. B. de Oliveira, L. R. de Oliveira, and J. L. B. do Amaral, "A decision support system for coffee variety selection using a fuzzy logic approach," Computers and Electronics in Agriculture, vol. 96, pp. 25-32, 2013.

[9]. K. S. Sreekumar, A. K. Singh, and M. P. Reddy, "Disease detection in coffee using image processing techniques," Journal of Agricultural Informatics, vol. 7, no. 2, pp. 1-13, 2016.

[10]. T. R. Mohan, K. Ramakrishnan, and K. V. Anbumani, "Forecasting of coffee yield using artificial neural networks," International Journal of Applied Engineering Research, vol. 10, no. 55, pp. 1329-1335, 2015

[11]. T. van Klompenburg, A. Kassahun, and C. Catal, "Crop yield prediction using machine learning: A systematic literature review," Computers and Electronics in Agriculture, vol. 179, p. 105709, Dec. 2020. doi: 10.1016/j.compag.2020.105709.

[12]. S. Sittipod, E. Schwartz, L. Paravisini, and D. G. Peterson, "Identi- fication of flavor modulating compounds that positively impact coffee quality," Food Chemistry, vol. 299, p. 125250, Dec. 2019. doi: 10.1016/j.foodchem.2019.125250.