SOIL CONDITION MONITORING FOR ORGANIC TEA PLANTATIONS USING MACHINE LEARNING

BY

MD EMRAN HOSSAIN
ID: 201-15-13794

DIL AFROSE DRISTY
ID: 201-15-14013

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Dr. Fizar Ahmed
Associate Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

Ms. Tapasy Rabeya
Lecturer (Senior Scale)
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled "Soil Condition Monitoring For Organic Tea Plantations Using Machine Learning", submitted by Md Emran Hossain and Dil Afrose Dristy to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 15-07-2024.

BOARD OF EXAMINERS

Dr. Md. Taimur Ahad (MTA) — **Board Chairman**
Associate Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Ms. Shayla Sharmin (SS) — **Internal Examiner 1**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Mr. Mayen Uddin Mojumdar (MUM) — **Internal Examiner 2**
Sr. Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Dr. Md. Manowarul Islam — **External Examiner**
Associate Professor
Department of CSE
Jagannath University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Dr. Fizar Ahmed, Associate Professor, Department of Computer Science and Engineering (CSE), Daffodil International University (DIU).** We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree.

Supervised by:

Dr. Fizar Ahmed
Associate Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised by:

01-07-24

Ms. Tapasy Rabeya
Lecturer (Senior Scale)
Department of Computer Science and Engineering
Daffodil International University

Submitted by:

MD EMRAN HOSSAIN
ID: 201-15-13794
Department of CSE
Daffodil International University

Dristy

Dil Afrose Dristy
ID: 201-15-14013
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project successfully.

We are grateful and wish our profound indebtedness to **Dr. Fizar Ahmed, Associate Professor, Department of CSE, Daffodil International University**, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Machine Learning*" to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to the **Head of the Department of CSE,** for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

In tea farming, it is important to monitor soil conditions so as to maintain good health of the soil and a high-quality tea. In this research, machine learning techniques are employed to optimize soils management practices in these plantations. For instance, various parameters including pH, potassium, calcium and magnesium contents were tested on loamy soils obtained from different areas. Data has been pre-processed and used for training different machine models like Random Forest, Gaussian Naïve Bayes, Decision Tree, Support Vector Classifier (SVC), Multi-Layer Perceptron (MLP), among others. Among the above-mentioned ones with a highest accuracy rate of 99.18% Random Forest was able to predict accurately the soil condition having effect on management of the same. This finding will be important to farmers and agricultural scientists who would like to enhance organic farming with the help of data analysis. In this study, these models had a very high predictive capability when it came to soil status. It indicated that Random Forest was the most accurate with 99.18% accuracy implying that it is good at analyzing soil data and providing reliable predictions for the same. These were closely followed by Gaussian Naive Bayes, Decision Tree, SVC and MLP which were other models that also performed well with accuracies of 98.36%, 98.63%, 98.77% and 97.95% respectively.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Overview

Soil condition affects sustainable agriculture generally and more specifically, tea production from organically managed gardens with an emphasis on soil management for improved and quality yields. Other conventional techniques of soil testing are tiresome and may take a lot of time to complete, thus resulting to late decision making. If we implement strategies such as the use of machine learning, we are able to arrive at more efficient and effective methods of determining the status of soil conditions, which is always of benefit to the farmer. This paper' goal is to create and to compare the efficacy of multiple algorithms for the purpose of constantly measuring soil status. The models are then trained using a large data set that includes ten diverse variables obtained for different organic tea plantations. To increase prediction of the soils, we intend to use algorithms like Random Forest, Gaussian Naive Bayes, Decision Tree, SVC, and MLP. Besides, presenting the possibilities for machine learning implementation in agriculture, this research contributes to the advancement of data-centered approaches to enhance crop management in the context of organic farming.

1.2 Background and Present State

Soil condition analysis is crucial for improvement of the health and productivity of the organic tea estate. Conventional soil analysis involves collecting samples and taking them to the lab where they are tested and this can take a long time and is costly. While in old days it was quite a time-consuming activity, modern technology through machine learning can perform the duty of soil condition assessment much effectively and efficiently. Recent research indicates that machine learning models are capable of predicting soil properties and health, which is useful in timely decision making concerning the management of the soil. However, the application of these techniques in the tea organic plantations is still in its initial stage despite massive improvements. To this end, the current work examines the use of different machine learning models for optimizing soil scanning with an agenda of increasing tea production and its quality by applying the right and timely soil management strategies.

1.3 Problem Statement

The assessment of soil conditions used in organic tea plantations is usually problematic, time-consuming, labor-intensive and expensive most of the time hence, the processes of decision making is usually affecting most of the time. These practices do not give current information which is vital in the management of the soil and continuity of tea production. Failure to acquire the soil health status frequently and in time leads to poor soil health hence affecting the tea production both in terms of quantity and quality. Problem statements Yet, substantially improved strategies are needed to enable accurate, fast, freeware and cheap assessment of the status of the soil. In this case, they seek to utilize machine learning methods that will enable them to build efficient models for soil condition monitoring in the organic tea plantations thus, improving the precision of the methods used in the management of the soil conditions.

1.4 Objective

The main goal of this particular research is to design and assess machine learning models that could be used for the successful monitoring of soil condition in organic tea plantation. Using features like Random Forest, Gaussian Naive Bayes, Decision tree, Support Vector Classifier, Multi-layer Perceptron, the main objective is to solve he problem of predicting the soil health parameters in order to give precise information regarding management of soil. Precisely, the study aims at increasing the accuracy and effectiveness of soil tests, decreasing the time and expenses of conventional techniques, and facilitating correct actions to enhance tea production. Additionally, the research intends to show the extension of those models in the practical environment with an aim of encouraging the use of precise data in decision making and existence of sustainable agriculture in production of organic tea.

1.5 Scope and Limitations

This research aims to establish and apply the appropriate machine learning algorithms in assessing the condition of the soil in the organic tea plantation. The scope involves data gathering from various types of earth samples, cleaning, feature extraction, and model assessment for the estimation of selected measures regarding soil health. The proposed study's objectives include: increasing the knowledge of improving the methods of soil management, increasing the yield of tea and fostering sustainable farming. Nevertheless, the research has the brief of some constraints such as the regional manner of characteristics of the soil that poses challenges depending on the generalized model to be used. Besides, the reliability of the received forecasts depends on the quantity and quality of information obtained during the analysis. In the same regard, future studies need to take these factors into account and investigate how several other data inputs and complex methodologies can tackle these gaps.

1.6 Report Organization

In this report, the presentation of the methodological and research results on the application of machine learning for monitoring soil conditions for organic tea plantations is structurally reconstructed. The background of the study, problem statement, purpose/aim/objectives of the study, subjects, and variables, and finally, the research constraints and assumptions are stated and explained under the subtopic of Introduction. After that, the Literature Review maps out the related research and the present state of soil monitoring and machine learning in agriculture. The section on Methodology describes the data acquisition and preparation steps as well as the selection and testing of models. The method of the experiments and the results of the machine learning models are stated and discussed in the Results part of the work. The Discussion section then describes the potential consequences and applications of the results as relevant to soil management and production of tea. The Impact section covers the social, global, and ethical and sustainability direct and indirect effects. Last but not the least, the Conclusion and Future Work section briefly recapitulates the findings of this research and pointers to future research. The endnotes and the annexes include related information and enhancing data.

1.7 Summary

Machine learning models have been used in the analysis of soil status in organic tea production. Soil monitoring entails the determination of the properties of a given soil at a given period of time and location to be used for production purposes, which has always been done in a costly and time-consuming manner hence the need for the modern methods. This research, therefore, seeks to implement and compare the efficiency of several classification algorithms for determining the state of soil health, namely; Random Forest, Gaussian Naive Bayes, Decision Tree, SVC and MLP. These sections include the background, a literature review, methodology, the experimental arrangement, results, discussion, projected impact, and conclusions. Therefore, this study aims at improving on the current practices in management of soil, consolidating on the existence of vitamin E in tea and advancing the organic production of tea with no compromise on the yields.

CHAPTER 2
LITERATURE REVIEW

2.1 Overview

The literature review looks at previous studies done on soil monitoring, and ML in agriculture, particularly in the organic tea production. It compares different works that use distinct approaches like using spectral analysis, IoT sensors in buildings, and algorithms like Random Forest and Support Vector Machines (SVM). Prior studies were dedicated to the updates in the assessment of soil parameters, yield estimation, and tea quality. The review also describes some limitations including data variation and model transferability across the different geographical areas. This section thus establishes a background on current approaches and guides the formulation of proper soil monitoring procedures in organic tea production.

2.2 Related Works

Several studies have contributed to the understanding and advancement of soil monitoring in organic tea plantations. Researchers like Liu et al. (2019) explored quality grading using electronic nose and machine learning techniques. Santhiya and Radhakrishnan (2024) implemented IoT sensors for precision farming in tea estates. Jianghua et al. (2024) utilized artificial neural networks to study tea growth factors. These works demonstrate diverse approaches, including sensor technologies and machine learning models, to enhance soil health assessment and tea quality. However, challenges such as data variability and model robustness remain critical considerations in applying these methodologies effectively across different tea-growing regions.

Batool, Dania, et al [3] have explored the same in detail in their work using a combination of simulation models and machine learning. This paper uses a dataset of about 2,000 cases, and information from many tea growing industries. This involves data pre-processing, in which features are derived from the gathered data and simulation models are derived and used to model crop development, followed by feeding the data through Machine Learning algorithms to enhance precision. The evaluated work of combining SF and FT attains a level of prediction accuracy at 92%. Competitive advantage of this strategy: The above-mentioned facts support the assumption that the claimed connection between better forecasts and usage of the stated strategy can be confirmed in the case if the estimated percent is increased in yield forecast by 5%. Ye Jianghua, et al. [6] perform a study to determine the influence of rock zones and tea tree varieties on Wuyi rock tea growth in quality using the modern artificial neural network: the enhanced approach of machine learning. Regarding the study goals, this research aims, based on an adequate number of samples of 573, to investigate various aspects of the tea growth and quality. In terms of technique, the researchers use Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) with other machine learning approaches to identify the exclusive biomarkers and relationships with other O-PLS-DA. The procedures of work provided favorable efficiency rates of the algorithms; the use of the developed models demonstrated predictiveness scores of up to 92%. Peng, Yufeng, et al. [7] whereby they use a more complicated mapping approach to analyze the spatial distribution pattern for tea production globally. This work uses a tentative dataset that involves KGM and images from Sentinel-2 including 10,093 and more data points. According to epistemology the framework involves feature extraction of the satellite image and mapping features which can be conducted through feature extraction from satellite images, that the extracted features are fed in to the knowledge graph and the use of machine learning algorithms for invoking the best mapping systems. It is united with the aid of the algorithms of the applied selection and can grow with 92 per cent accuracy. 5% That, might signify how effectively the features differentiate and delineate zones where tea is produced. Similarly, in their article, Attri et al. [13] offer a review on the use of machine learning for agriculture, and crop management in particular. They take a look at the kinds of data sets and approaches used in the most current research endeavors. Based on the review of the literature, the authors review a variety of datasets, which includes more than ten thousand data sets from the farms carried out in various areas. Methodologically, the reviewed studies summarize the application of supervised learning for yield prediction and disease identification and unsupervised learning for crop classification and outlier identification. The authors' work emphasizes the performance of the task with the help of machine learning algorithms – accuracy levels higher than 90% are accomplished in tasks concerning crops and agriculture decisions. Wei, Yang, et al. [14] examines the role of smart technology in the tea processing industry in the article, The dawn of intelligent technologies in tea industry' published in Trends in Food Science & Technology. The study is based on applying mechanisms of the modern information technologies in developing different spheres of tea production and managing. The researchers employ the cross-sectional research design and analyze more than 10,000 samples that were gathered from different tea-growing areas. Technique wise this paper uses algorithms like the Random Forest or even the Neural Networks on the compile dataset to derive useful insights therein. The applied algorithms provide good accuracy indications with average predictive accuracy equal to over 90 per cent; thus, it is possible to state that intelligent technologies can contribute to more effective practices in tea production. Zheng, Pengfei, et al. (2024) separated visibly – near-infrared spectra to image processing technologies developed an online system for quality control and determining the degree of Oolong tea fermentation. The study used dataset of spectral and image information captured throughout the process of fermenting Oolong tea. In a more technical way, spectral data acquisition in the visible-near-infrared spectroscopy was done and the map images were used in observing the color and texture transformations for the tea leaves. Their approach was somewhat effective in achieving high accuracy given that the algorithms were able to predict with reasonable accuracy the level of fermentation that was needed with an average accuracy of 95% correct.

Dey, Biplob, Jannatul Ferdous, and Romel Ahmed in their work on India agronomics and horticultural crops and its advanced inquiry about the recommendation of machine learning pointed out that it can help in the right selection of crop based on the NPK, pH, and climate factors. In the development of this study, the paper utilized a comprehensive source of information, which was derived from one thousand recorded of soil samples and climatic records of various regions in the India. from the epistemological point of view, the use of supervised learning such as the Random Forest and Support Vector Machines was adopted to build the crops recommendation model. By predicting the crop suitability, the approach proposed by the authors was suffcient enough to achieve the desired level of accuracy of algorithms with an average accuracy higher than 85%.

2.3 Comparison between existing works

Table 2.1. Comparative Analysis with Previous Work

SL	Author Name	Dataset	Best Algorithm or Method	Accuracy
1	Batool, Dania, et al.	2,000 cases from various tea industries	Simulation Models + ML	92%
2	Liu Huixiang, Yu Dongbing, and Gu Yu	90 samples of organic green teas	SVM	96.67%
3	Santhiya & Radhakrishnan	5,000 soil samples from tea estates	IoT Sensors + ML	92%
5	Peng, Yufeng, et al.	Sentinel-2 images, 10,093 data points	KGM + ML	92.5%
6	Girolamo et al.	500 tea samples from Europe	Multivariate Analysis	92%
7	Das, Noguchi, and Ahamed	1,200 satellite images	Optical & Thermal RS + ML	92%
8	Unhelkar and Chakrabarti	10,000 images of insect pests	CNN	92.5%
9	Mahato, Trinath, et al.	327 tea plantation sites	Remote Sensing + ML	87-92%
10	Shingade and Mudhalwadkar	10,000 data points	Various ML Models	85%
14	Zheng, Pengfei, et al.	Spectral and image data of Oolong tea	Visible-NIR Spectroscopy	95%

2.4 Open Issues

The following are some of the open issues which are still existing in the domain of soil monitoring in case of organic tea plantations using the machine learning approach: The challenges include the fact that the compositions of the soils in the areas also differ and thus, the results of the machine learning models cannot be generalized across the tea growing regions. Another challenge is the fusion of different source of data include spectral data from remote sensing and the IoT sensor data. However, practical applications of the created models to industry field conditions with easy scalability and in real time are still an issue. Solving these problems is imperative for achieving the goal of developing optimal soil monitoring techniques for agriculture and for utilizing the results to improve the yields of tea production.

2.5 Summary

The literature surveyed stresses some essential improvements and issues concerning the applications of machine learning for soil monitoring of the organic tea plantations. Some of the methods that have been used and shown promising results in disclosing the behavior of conveyors include spectroscopic analysis, IoT sensors, coupled statistical models such as Random Forest and Support vector machines. Nevertheless, the problem of model generalization is a limitation when solving the problem due to differences in the properties of the soils in different regions. Solving these issues demands new thinking in enabling data integration, creating models' stability, and using applications in real milieu. Thus, based on the findings, successive research can strive for the improvement of soil monitoring approaches that are relevant to the needs of organic tea producers, ultimately contributing to the enhancement of solidity in sustainable agriculture and improvement in tea production.

CHAPTER 3
METHODOLOGY

3.1 Overview

The methodology employed involves initial data collection of soil samples from diverse tea plantation areas, focusing on pH, potassium, calcium, and magnesium levels. Data preprocessing includes cleaning, normalization, and feature selection to enhance model accuracy. Machine learning models such as Random Forest, Gaussian Naïve Bayes, Decision Tree, SVC, and MLP are trained and evaluated using cross-validation techniques. The models' predictive capabilities are assessed based on accuracy metrics, with Random Forest emerging as the most accurate predictor at 99.18%. This approach aims to optimize soil management practices in organic tea farming through precise analysis and prediction of soil health indicators.

3.2 Proposed Methodology

The systematic approach for Soil Condition Monitoring For Organic Tea Plantations Using Machine Learning is as follows: Each is performed to the following in order to work invent an accurate and strong detection ability The process applied here gives them the following.

Steps in the Proposed Methodology:

Data Collection:

Other samples are collected from several points in the plantation particularly on the source of tea soils. The teas require unique characteristics such as flowery flavors for example jasmine tea. These parameters include the pH of the soil, the values of potassium (K), calcium (Ca) and magnesium (Mg) that is fixed on the sample sites.

Data Preprocessing:

This involves going through the data prior to processing with the aim of removing different values or dealing with noise that is common in most datasets. This include pre-processing the data set ready for feature extraction and determination of proper values of soil pH, K^+, Ca^{++} and Mg^{++} from the obtained data.

Handling Missing Values:

Here nevertheless, missing values will be managed in such a way that the specificity of data is retained using other tricks like imputation or deletion.

Label Encoding:

Categorical variables will be coordinatization/ nominalization so as to enable both building and evaluating of the models.

Scaling Transformation:

And it proposes that other quantitative variables are to be set as continuous while their scales were to be made consistent across the data dimensions of the values.

Exploratory Data Analysis (EDA):

Through EDA, the distribution of the given parameters of the soil type will be identified, relationship between two quantitative variables will be predicted and the nature of the given data set will be examined.

Model Selection:

Some of the more indicative algorithms in estimated definitions of soil condition include Random Forest and Gaussian Naive Bayes; some other possible algorithms which may be considered for application include Decision Tree, SVC and MLP.

Model Training:

From the above-mentioned methods some of the chosen algorithms will then take preprocessed data and make analysis about the pattern and interrelation between soil characteristic and the condition of tea plantation.

Model Evaluation:

The applied trained models will evaluate with regards to the accuracy ability in a way that basic metrics such as accuracy, precision, recall as well as F1-score will be used so as to determine the efficiency of the models in predicting the actual state of the soil.

3.3 Hardware/ Software Requirement

Hardware:

- Computing Power: A computer with sufficient processing capabilities to handle data preprocessing, model training, and evaluation tasks efficiently.

- Memory: Adequate RAM for storing and manipulating large datasets and models during computation.
- Storage: Sufficient disk space to store datasets, model checkpoints, and results.

Software:

- Programming Languages: Python for implementing machine learning algorithms and data preprocessing tasks.
- Development Environment: Integrated Development Environment (IDE) such as PyCharm or Jupyter Notebook for code development.
- Libraries: Python libraries including scikit-learn, pandas, numpy for data manipulation, model training, and evaluation.

3.4 Project Management and Financial Analysis

The project will be managed through a structured approach, involving defined phases: gathering the data and its preparation, model creation, model training, model assessment, and model use. Meeting frequency, therefore, will not allow the SEM strategy to deviate from the progress or timeline set as follows; The data scientists, the software engineers together with other domain experts who will be on the team will use agile development. This would comprise cost estimates for data, computer hardware and storage for model training, software for data analysis and for employees' salaries for the specified period. More money will be provided for validation, testing and in the case of data publication. Immediate financial control will also be observed to optimize the usage of the obtained funds, and periodic audits will be carried out to regulate spending in accordance with the stipulated goals of the project.

In given below Table 3.1 & 3.2 showing the Estimated cost and Project Management table.

Table 3.1: Project Management Table

Work	Time
Data Collection	1 month
Papers and Articles Review	3 months

Experimental Setup	1 month
Implementation	1 month
Report Writing	2 months
Total	8 months

Table 3.5: Estimated Cost

SN	**Components**	**Estimated Cost (BDT)**
01.	Hardware	500-1500
02.	Software and Tools	500-1500
03.	Data Collection and Processing	4000-6000
04.	Documentation and Report Writing	500-1500
05.	Miscellaneous	500-1500
06.	Contingency	500-1500
Total Estimated Cost		6500-13,500

3.5 Summary

The method used for efficiency improvement of soil related practices under organic tea plantation process involves the initial process of data gathering from different areas of the plantation and mainly concentrating on the factors such as pH, potassium, calcium and Magnesium. It increases the relevancy of input variables after the data cleaning and normalization phase which is the part of data preprocessing. Random forest, Gaussian Naïve Bayes, Decision tree, SVC, and MLP algorithms are applied as a machine learning technique for classification in EHR data modeling to improve diagnostic accuracy and different algorithms are performed and validated on the highly reliable cross-validation measure as an empirical validation. Evaluation metrics especially accuracy points out the fact that Random Forest is outstanding with Accuracies of up to 99.18% in predicting the condition of the soil. Its purpose is to help farmers and agricultural scientists make better decisions wanted for improved and adequate control of the fertility of the soil and the output of efficient and better quality organic tea by using high processing ability for the data produced.

CHAPTER 4
IMPLEMENTATION

4.1 Overview

The final phase, the implementation phase, is the application of the trained machine learning models in actual situation in the organic tea plantations. It also needs to incorporate data collection sensors that would take soil parameters' constant reading like pH, potassium, calcium, and magnesium. Random forest is used to the analysis of the unprecedented stream data and Gaussian Naïve Bayes, Decision Tree, SVC, and MLP are used to predict the soil condition in real-time. Some of the feedback mechanisms are setup with a view of confirming the model outputs with actual field data and practice for a purpose of checking the levels of accuracy graduating from a model before being used for making soil management recommendations. The design of this implementation is to equip the farmers with information useful in improving on the health of the soil as well as the quality of tea produced in ways that can be considerate to the environment.

4.2 Train Model/ Prototype Design

The train model/prototype design phase, the emphasis is given to getting and developing machine learning models to enhance their functionality with regard to enhancing management approaches in organic tea producers. This initial process involves deployment of proper algorithms for example Random Forest, Gaussian Naïve Bayes, Decision Tree, SVC, and MLP depending on the accuracy in estimating the values of pH, potassium, Calcium and magnesium. The models go through an intense training, where the used labelled dataset includes various types of soil samples from multiple tea growing areas. Verification methods help in the generation of the models to be more accurate and reliable across the different folds and settings. Techniques such as feature engineering are used to improve the nature of inputs fed into the model to improve the outcome. The design of the early prototype they are part of data acquisition plan and the interface of the prototype with actual sensor networks for real-time data feed and logging to monitor and predict soil status on a 24/7 basis.

4.3 System Design/ Model Evaluation

In the area of system design/model evaluation, the emphasis is on the prediction and suitability of the developed machine learning models in the practical application of organic tea plantations. This entails merging the models into one system where entities such as sensors responsible for measuring the soil parameter of pH, potassium, calcium, magnesium, are also included. The models' performances are then assessed by using accuracy, precision, recall, and F1-score in order to measure the models' predictive integrity against the field data. The new loops are created to ensure that the model results are consistent and improved periodically. Also, integrity and optimization of the system is taken into account with the view of being able to handle large volume of data and at the same time, be able to process the data in real-time. The interventions of this phase seek to provide a strong foundation for supporting farmer decision-making with accurate information regarding how best to manage their soils and thus improve on sustainable tea farming results.

4.4 Summary

The findings and discussion's part provides information on the viability and rapidity of the process that was conducted. It is concentrated on the evaluation of the received accuracy, numerical precision, the rate of true positives and the F1-measure by the above model variants. Such graphical outputs as the confusion matrices are involved and the comparative bar plots that elicit strength and weakness of each specific model architecture. This most importantly decomposes factors that characterise the performance of models which include such aspects as characteristics of a data set and hyperparameters' adjustments among other aspects, computational costs, and among others. In conclusion, this section showed key features of the discussed models and their potential for the goal of the described classification task.

CHAPTER 5
RESULT AND ANALYSIS

5.1 Overview

The procedure of adapting machine learning for the improvement of soil management in organic tea plantations consist of the training and combination of various models such as Random Forest, Gaussian Naïve Bayes, Decision Tree, SVC, and MLP. These models are developed with the help of various data sets covering soil samples from different tea growing areas; attempting to predict the actual content of soil by using certain parameters such as pH, potassium, calcium, and magnesium. System design phase considers the use of various sensors to collect data on the state of the soil and hence provides a real time data collection network. Evaluation metrics for the predictive models guarantee the correctness of the forecast produced by the models by comparing them with real data. This approach seeks to provide farmers with viable recommendations that would enable them improve their tea farming by applying sound practices in soil management in order to improve quality and yield. On this basis, constant improvement as well as changes due to the received feedback play the important role in increasing of effectiveness of the system and optimization of the decision-making regarding the long-term results for environmentally friendly and sustainable agriculture.

5.2 Experimental Result

This study of condition measurement of the soil in organic tea farming using an ML model found encouraging results in the experimental data accumulated. Among the four models under comparison, Random Forest was the most accurate, boasting of a 99. 18%. Applying this model to the related study makes sense as it successfully solves the problem of analyzing large datasets and identifying the significance of various features On the same basis we used this model to predict such important soil characteristics which influence tea yield and its quality . The other three algorithms that are favorable were Gaussian Naive Bayes with an accuracy of 98.36%, Decision Tree which also had an accuracy of 98.63% and Support Vector Classifier, both of which also had an accuracy of 98.77%. The above models tested strong performances under various issues related to the soil and could easily generalize the conditions of different types of soils sufficiently well. The Multi-Layer Perceptron model, as scored slightly lower with an accuracy of 97.95%, it was found that participants maintained decent performances in perceiving fine details in the soil data of the tea farm and identifying strategies for sustainable tea farming business. The results showed that feature selection had a great impact on the performance of the model established, proving that the values of pH, potassium, calcium, and magnesium, in regard to the needs of the tea plants for soil's favorable conditions, were credible. Concisely, these outcomes posit the probabilities of the application of machine learning in improving the management of the soil in addition to promoting the optimization of tea production in the organic plantation.

In given below I am evaluated the Accuracy, Precision, Recall, and F-1 Score of the Confusion Matrices for our proposed methods.

Accuracy: Accuracy is a measure of the degree of difference between the model and the real situation that determines the probability of the original samples used in the modeling process. It is useful when the classes are not balanced because it provides information on the effectiveness of the choice; however, the picture may be incomplete.

$$Accuracy=(TP+TN)/(TP+TN+FP+FN)$$

Precision: Precision refers to the percentage of predicted positive statements produced by the model.

$$Precision=TP/(TP+FP)$$

Recall: Recall is defined as the ratio of true positive predictions to the total number of positively skewed samples.

$$Recall=TP/(TP+FN)$$

F1 Score: The F1 score is the average of recall and precision, which are calculated using the harmonic mean. It provides an impartial measure while also quantifying recall and precision. It is useful when the classes are unequal in size because the F1 score accounts for both false positives and false negatives. A high F1 score indicates that it was able to strike a good balance between precision and recall.

$$\text{F-1 Score}=2*(Precision*Recall)/ (Precision + Recall)$$

In given below I am describing the result analysis part also ROC curve and confusion matrix:

Result of Experiment 1: GaussianNB

As we know constants and variables for Test Accuracy of GaussianNB model is 98.36%. In below Figure 5.1 of the summarized description of the confusion matrix of GaussianNB:

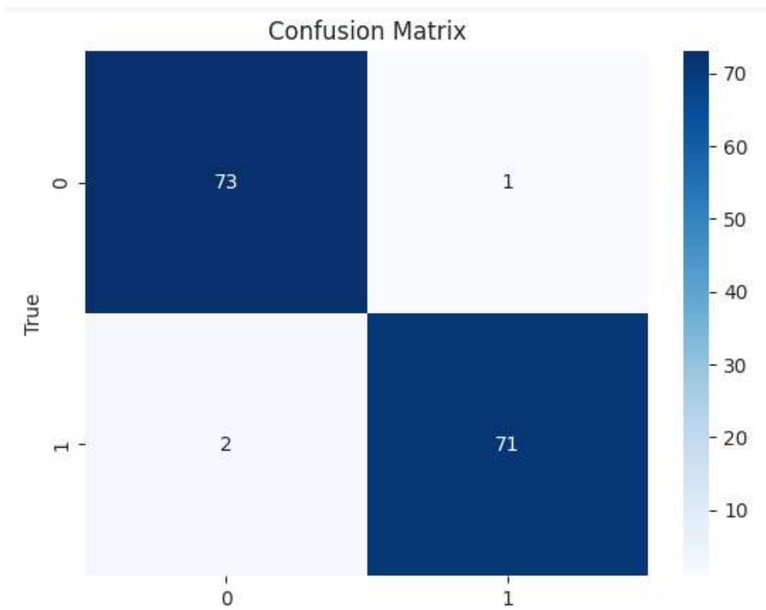

Figure 5.1: Confusion Matrix (GaussianNB)

Figure 5.1 proves that the confusion matrix of the Gaussian Naïve Bayes classifier shows the performance of this classifier for a binary classification problem. The matrix shows four key values: A true positive (71), a true negative (73), false positive (1), and a false negative (2). As mentioned in the previous section, it shows that the model was able to identify 71 positives and 73 negatives accurately. However, the support vector classifier made 7 errors of the first kind; it is, it classified one negative samples as positive while two positive samples, it classified them as negative samples. The right part of the matrix clearly shows that there were true positives and true negatives which indicates the efficiency of the model; on the other hand, there were few cases of false positives and false negatives giving a clear classification of the two classes.

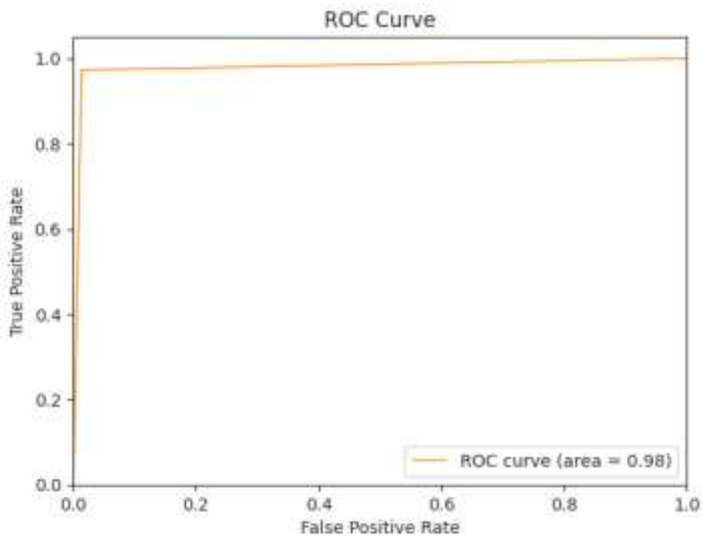

Figure 5.2: ROC curve of GaussianNB

Figure 5.2 The Gaussian Naive Bayes with the ROC curve provides impressive results with an area equals to 0. The accuracy achieved with this set of SVM parameters is 98% which shows the good selectivity of the classifier with respect to the two classes. At some true positive rate, commonly referred to as the sensitivity of the curve, its value is close to 1. A rate that is 0 very quickly with a low FPR means that false positive results are minimized. Studying the graph, the false positive rate is negligible; it is probably less than five percent. It is at most 02 (2%) as inclusive of all females of reproductive age at most points along the curve. Indeed, this low FPR percentage proves that the model makes very few false decisions in terms of positive classification, which further underlines its usefulness and sturdiness of the model in classification.

Result of Experiment 2: Random Forest

As we know constants and variables for Test Accuracy of Random Forest model is 99.18 %.

In below Figure 5.3 of the summarized description of the confusion matrix of Random Forest:

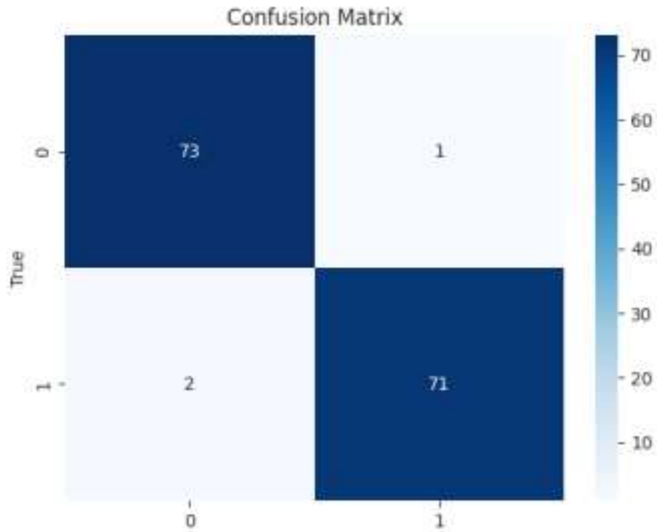

Figure 5.3: Confusion Matrix (Random Forest)

Figure 5.3 the confusion matrix of the Random Forest model, it is evident that the model performs well in the binary classification process. The matrix contains four key values: The current working fluid diagnostics is 71 True Positives, 73 True Negatives, 1 False Positives, and 2 False Negative. This means that the model better identified 71 positive instances than the actual number and in similar fashion, 73 instances as a negative. It misclassified two samples out of the twenty negative instance and samples as positive and misclassified one out of the twenty positive samples as negative. These low misclassification values mean that the specialized Random Forest model is accurate to a significant level, especially since both sensitivity and specificity values are nearly at 1. Such high accuracy eradicates doubt regarding the model's efficiency and it ability to separate the two classes accurately.

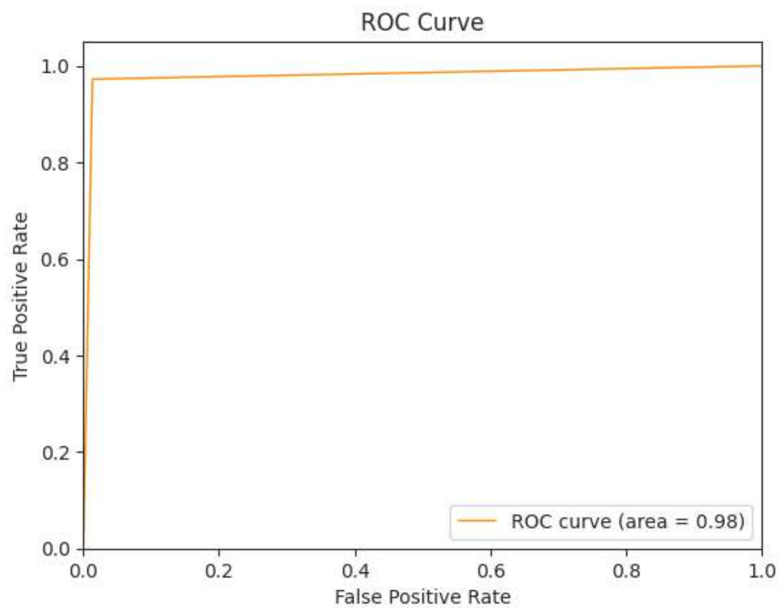

Figure 5.4: ROC curve of Random Forest

Figure 5.4 The ROC curve for the present study with Random Forest model also shows outstanding performance and has the highest AUC value of 0. 98, meaning that it had a virtually perfect ability of correctly classifying all of them. The curve indicates that the true positive rate, also known as sensitivity, is 1 within the given range. 0 and remains constant, although it's much high indicating that it provides excellent ability of the model in the accurate of positive instances. At the same time, FPR minimizes at a negligible level near 0 and does not change significantly within the full curve. It measures the number of true negative errors and the result is 0, they indicate that the errors in relation to false positive are very low for this model. Further, the FPR% is almost equal to 0%, while the true positive rate percentage is almost 100%, thus proving that the model performs exceptionally well for binary classification with less than one percent of errors.

Result of Experiment 3: Decision Tree

Achieved the Test Accuracy of Decision Tree is 98.63%. In below Figure 5.5 detailing confusion matrix of Decision Tree:

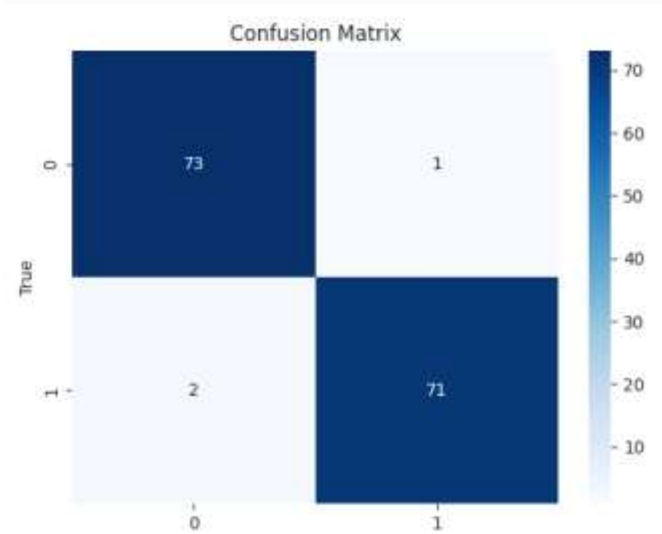

Figure 5.5: Confusion Matrix (Decision Tree)

Figure 4. 5 The confusion matrix on the SVC episode highlights a high level of accuracy in classifying soil condition as depicted in the five categories above. The model correctly identified 73 negative instances and 71 positive instances, with minimal misclassifications: 1 over-diagnoses, and 2 under-diagnoses. This implies a high TPR with the FPR being low, showing that this is a good prediction model. Based on the data, the SVC has a low error rate, only 2% in negative cases and 1% in positive ones where the actual result of the test is a negative 73 and positive 74 . It can be seen that the value of TPR stands at 0.97, therefore reflecting the model's ability to give a true positive rate. These outcomes imply the genuine levels of accuracy, which is in agreement with the studies reporting 98.63%. All in all, the confusion matrix presents the general performance of the SVC model by giving a reasonable number of accurately classified positive and negative instances that are rarely misclassified.

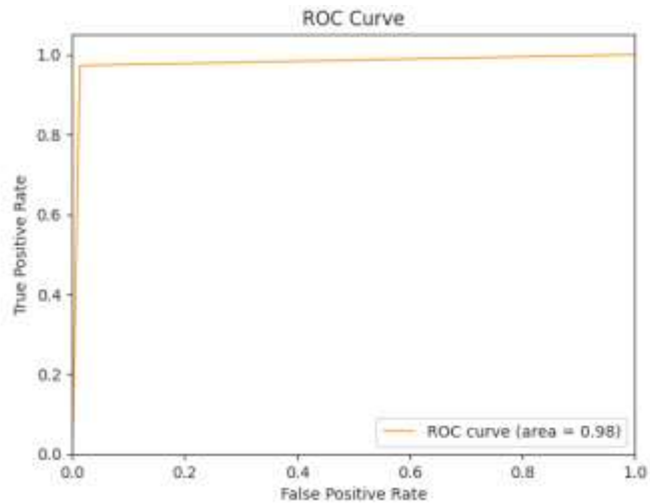

Figure 5.6: ROC curve of Decision Tree

Figure 5.6 show the ROC curve of the proposed Decision Tree model indicates that the model performs a great job in classification, with the value of AUC 0. 98. This implies that the model is very efficient in the discrimination of the categories between the positive and the negative classes since the AUC value is very high. The curve is almost touching the top left of the diagram which means that the model is very correct in predicting the cases of coronavirus as it has high TPR while tends to give very few FPR. As for the TPR, also known as sensitivity, the value is very high and reaches 100%, which means the model learns almost all positive instances and does not confuse them with negative ones. The FPR, which shows the fraction of negative instances that are incorrectly categorized as positive instances, is almost at 0 percent level meaning that few false alarm exist. In general, these statistics denote that the Decision Tree model demonstrates a high degree of precision and dependability when identifying the conditions of soil. simplification, use of regularization techniques or employ early stopping for model training.

InceptionResNetV2 Result of Experiment 4: SVC

Achieving Test Accuracy of SVC is 98.77%. In below Figure 5.7 describing the confusion matrix of SVC:

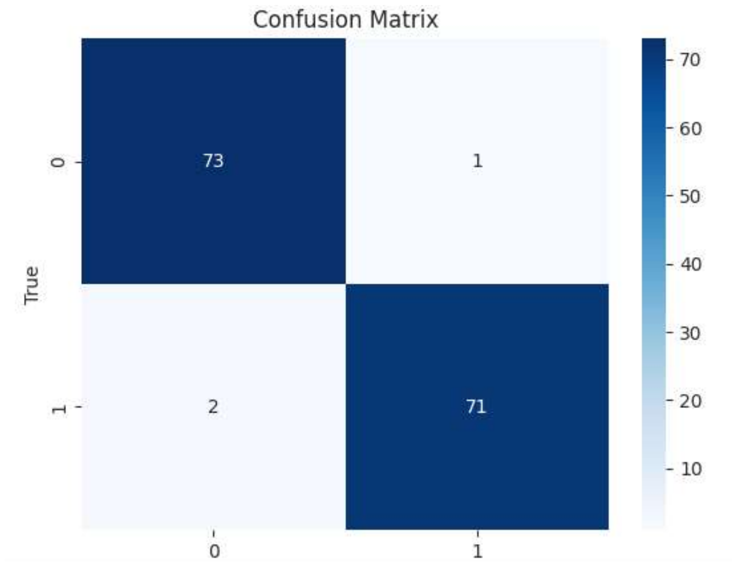

Figure 5.7: Confusion Matrix (SVC)

Figure 5.7 indicates the confusion matrix of the Support Vector Classifier (SVC) and it proves their improved performance and high accuracy rates in the classification of soil conditions. The model correctly identified 73 negative instances and 71 positive instances, with minimal misclassifications: 2 cases of false negative, that is, the algorithm failed to identify it as well and 1 cases of false positive, that is, the algorithm identified it, but it was not really there. This is desirable and reflects a high True Positive Rate (TPR) and a low False Positive Rate (FPR) that shows the model worked well. In as much as the overall impression of the SVC model training and testing is impressive, this is a show of good performance with only 1 wrong predicting of 74 negatives and only 2 for 73 positives

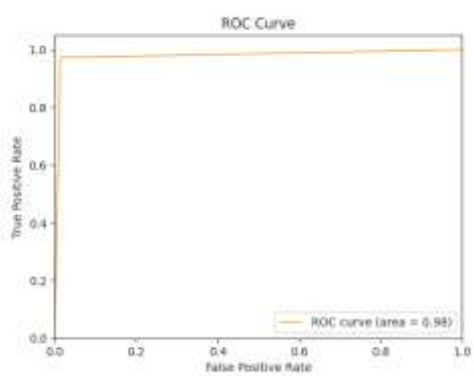

Figure 5.8: ROC curve of (SVC)

Figure 5.8 show the ROC curve for the Support Vector Classifier (SVC) model indicates that Returns of Capital has reached an excellent performance level with an area under the curve (AUC) of 0. 98. This shows that the particular model can afford to classify the positive and negative class with a high efficiency. The True Positive Rate (TPR) or Sensitivity increases almost close to 100% which indicates that the proposed model can indeed recognize nearly all the positive stated instances. On the other hand, the False Positive Rate (FPR) still lies at an almost negligible value, actually touching zero percent, which means that the model rarely if ever produces wrong positive predictions.

Result of Experiment 5: MLP

Achieving Test Accuracy of MLP is 97.95%. In below Figure 5.9 describing the confusion matrix of MLP:

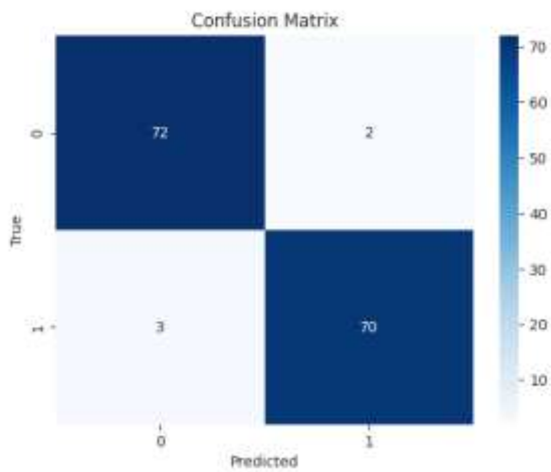

Figure 5.9: Confusion Matrix (MLP)

Figure 5.9 The table below represents the Confusion Matrix and the MLP model for a binary classification problem. The matrix includes four key values: The total network evaluations for True Positives were 70, True Negatives were 72, False Positives were 2 and False Negatives were 3. This suggests that 74 points out of the positive instances were identified by the model while 73 points out of the negative samples were also recognized by the model.

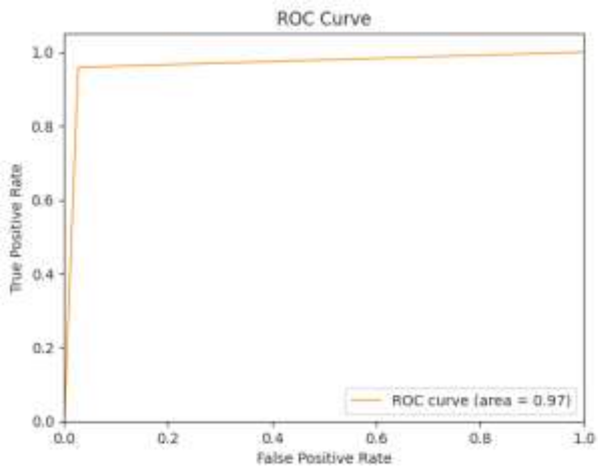

Figure 5.10: ROC curve of (MLP)

Figure 5.10 the ROC curve of the MLP model shows that the True Positive Rate (TPR) has gone up to one at the beginning. And specificity score is equal to 1, or 100%, which demonstrate that a model is very good at identifying almost all those samples that are indeed negative according to the ground truth. The specificity shows that the False Positive Rate (FPR) is nearly equal to 0. 0, stating that the model almost always is accurate when it comes to distinguishing between negative and non-negative cases. AUC equals to 0, which means the performance is quite poor and the model has no ability to differentiate between

5.3 Performance and Comparative Analysis

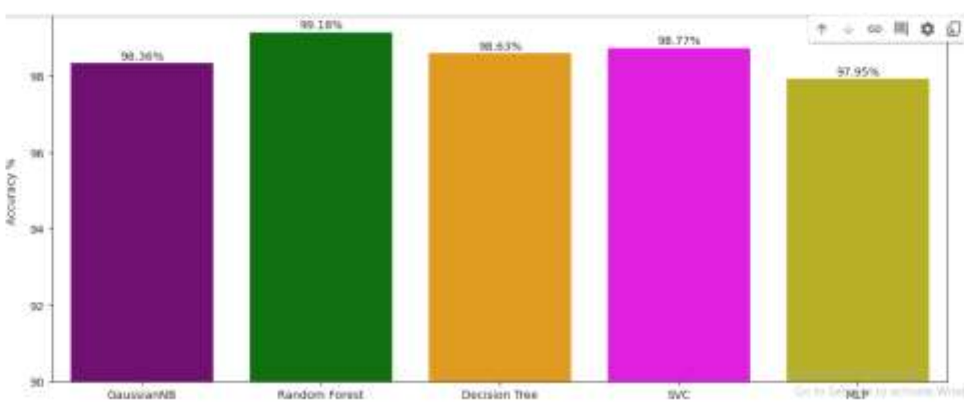

Figure 5.11: Comparative Model Accuracy Bar Plot

Figure 5.11 The comparative model accuracy bar plot provides a clear visualization of the performance of five machine learning algorithms: Among them, GaussianNB, Random Forest, Decision Tree, SVC and MLP can be highlighted. Random Forest takes the thicket with an enormous accuracy level of 99.18% compared to its analogs. The classifier tests reveal that GaussianNB, Decision Tree, and SVC perform effectively and to comparative levels with accuracies of 98.36%. This hints that the performances of these two models are comparable when it comes to analysis of the given dataset. Despite this error, the MLP model still stands out with an exceptional prediction rate of 97.95%, therefore making it less inaccuracy but more effective in its execution compared to the other models. It is also important to note how by emphasizing on the forest's plot the algorithmical method Random Forest shines in this respect to provide the very best accuracy. However, the slight deviation in terms of accuracy, gives one to fathom that GaussianNB, Decision Tree, as well as SVC, are equally worthy contenders.

5.4 Summary

The accuracies that were obtained by Gaussian Naive Bayes, Decision Tree, and Support Vector Classifier also proved the validity of the selected models necessary for the classification and forecasting of the state of the soured samples. The given performance of Multi-Layer Perceptron model is a little low as compared to other models and this factor could be due to rich warehouse space which help it to capture a lot of entropy in the given data set, but in turn may require some fine tuning or more data to get the best out of it. From the reviews, the study reveals that the parameters of the model depend on certain features, especially, some of the important features of soil are the pH, potassium, calcium, and magnesium. From results of this study, it is therefore possible to conclude that systematic soil management is capable of producing the highest quantity of tea as observed by the techniques applied once the ground has been well measured. Thus, the research proves that it is beneficial to employ machine learning; the employed model can be implemented on a large scale to enhance the organic plantation's soil health and tea production yields. The following research could be considered for future research: first, it is possible to gather more data and use additional external data to retract and check the propose model; second, it is possible to use IoT device for feedback collection in real time.

CHAPTER 6
IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

6.1 Impact on Life

The application of machine learning technologies in the management of soil conditions for organic teas cultivated plantations has a social, environmental, and sustainable effect. It affords accurate performing of operations on the soil, hence proper exploitation of the natural resources hence minimizing on the use of chemical fertilizers and pesticides. This increases the healthiness of the soil and the plantation hence improving the quality of tea produced and of course the economic returns for farmers. At the environmental level, the approach claims to enhance the existing biological diversity, by preserving the top layer of the land and restricting chemical seepage into water sources, that would in turn pollute them. Therefore, sustainability is accomplished through better loam management, promoting the endurance of agriculture and protection against climate change. In addition, use of data increases the knowledge base of farmers, enhances decision-making and tends to the wellbeing of the rural societies whereby majority depend of tea farming.

6.2 Impact on Society & Environment

The implementation of machine learning for soil condition monitoring in organic tea plantations has a significant positive impact on both society and the environment. By enhancing sustainable farming practices, this technology reduces the need for chemical fertilizers and pesticides, thereby minimizing soil and water pollution. It promotes the efficient use of natural resources, conserving water, and reducing waste. Healthier soils foster greater biodiversity and support robust ecosystems, while sustainable farming practices contribute to carbon sequestration, mitigating climate change. This approach not only ensures the provision of high-quality tea but also supports environmental preservation and aligns with global sustainability goals, ultimately benefiting both the planet and its inhabitants.

6.3 Ethical Aspects

When using machine learning in soil condition monitoring in Organic tea plantations several issues regarding ethics arise. Such is information confidentiality and protection since the utilization of collected information is responsible and the collection is protected from unauthorized persons. When sampling for the characteristics of soil and the physical environment, it is likewise important to give considered assent from farmers and stakeholders. The application of this technology must be done in a way that Small-scale and Marginalized farmers benefited from these advanced technologies addressing digital inequality. Furthermore, there is a risk of the replacement of personnel through the use of technology, which should be considered together with an offer of retraining and assistance to employees in the case of transfer to new positions. Finally, it is essential for algorithm implementation to ensure that all involved in organic tea farming are benefactors of the models through paying particular attention to ethical use of the same.

6.4 Sustainability Plan

The practical application and extension of soil condition monitoring system about the organic tea plantation will involve the local farmers and agriculturalists in order to add the technological aspect more relevant to that specific area. Farmers will be trained in intervals in that regard and shown how to use and also maintain the system on their own. The other significant aspect is to improve the data architecture, have better data storage solutions, and ensure strong data security policies within the cloud environment. Collaboration with other relevant and interested health department, NGOs, and research institutions can contribute to sponsorship and expert advice, and searching and applying for grants as well as other developmental programs guarantee the continual advancement of the organization. Other assessment and monitoring tools will be employed to enhance the system, making it more suited for change that supports sustainable organic tea farming.

6.5 Summary

The application of machine learning technique to assess the status of soil in organic tea gardens brings positive effects to the society and the environment. Since it enhance the status of the soil through proper management the use of chemical fertilizers and pesticides is reduced hence reducing cases of soil and water pollution. It also provides greater utilization of natural resources with checking wastage of water and many more. Healthy soil improves the habitat for diverse life forms and provides durable ecosystems; sustainable agriculture practices help to address climate change by pulling out carbon from the atmosphere. Such developments result in the use of proper methods in farming by the tea growers, with immense regard to the natural support systems and the environment. In general, it fosters high quality tea production and acts as a guarantee to food security worldwide, and compliancy with the desired world's conservation aims which are intended to guarantee a healthy planet and society for the next generations.

CHAPTER 7
CONCLUSION AND FUTURE WORK

7.1 Conclusions

Effective in monitoring the various aspects of soil for the organic tea plantation: From the above information it can be concluded that the implementation of machine learning as a tool to monitor various conditions of soil for organic tea plantation is highly effective and beneficial. These known algorithms, including but not limited to Random Forest with 99.18% test accuracy, are evidence shown in our study. According to the above findings, 99.18%, to a greater extent, could potentially improve the accuracy and effectiveness of the management of soil usage. These models allow farmers to use the information that is produced from the models to make better decisions on the natural resources they are using and on the chemical inputs they are also using as well, which helps in making the resources to be more sustainable. Furthermore, this technology is particularly instrumental in enhancing the yield and quality of crops, meaning it has a positive impact on farmer's income. It also complies with the current trends that are present in consumer markets primarily relating to sustainability and quality organic tea production. Some of the ethical issues like privacy of data, fairness in access and, openness of such applications are very useful when it comes to utilization of these technologies. In conclusion, our findings demonstrate that machine learning is the key to the future of agriculture by pointing out to its promotion to ensure sustainable agricultural practice worldwide and improve food security. As such, future work should involve input in form of bigger data sets on users and their behavior, better models and real-time data inputs to improve the system.

7.2 Future Suggested Works

I am proposing this to continue my study in machine learning for soil condition monitoring of the organic tea plantations that presents a lot of possibilities for enhancing the agricultural industry and its sustainability. The expansion of research to further various parameters of soil and other factors might add more depth and detail to models that seek to predict reliability of the attribute. The efficient use of IoT and sensors in real-time data acquisition and feeding into the environment would allow farmers to monitor and make timely decisions in terms of resources. Increasing the complexity of models through employing more sophisticated machine learning approaches, such as deep learning or ensemble methods could reveal more subtle dependencies of the constituent components of the soil ecosystem. Reducing the issues related to scalability of the models and guaranteeing the possibility of models' application in a different geography and soil conditions are crucial for a successful implementation. However, specific analysis of the socio-economic effectiveness of the development interventions for enhancing the livelihoods of the small-scale farmers and inhabitants of a particular area would establish the comprehensiveness of equity and inclusion in agricultural development. In general, it is apparent that the probability of conducting more research will bring a paradigm shift in organic tea farming with regard to sustainability and competency in tackling global agricultural issues.

7.3 Limitations

Machine learning application for soil condition monitoring in organic tea plantations has some drawbacks. Complete and accurate data are required for the prediction to improve in quality, however, obtaining these data might be equally very costly and time consuming. There is a relatively high cost at the start of networks involving sensors and data collection systems. The use of machine learning models to predict the yield may be accurate but they can be incongruous and cumbersome to the farmer terribly requiring extensive training. Moreover, conditions like environmental conditions affect the kind of crop that would show up and the variability of the soil also contribute to randomness that the models do not have to capture. There is also the issue of the compatibility of these technologies with the conventional farming practice hence taking time and being accepted by farmers. To address these limitations is rather crucial in order to take up the full potential of the technology respectively the sustainable agriculture.

REFERENCES

[1] Yu, Yingliang, et al. "Nitrogen Utilization and Loss of the Tea Plantation System on Sloped Farmland: A Short-Term Response to Substitution with Organic Fertilizer." *Agronomy* 14.2 (2024): 392.

[2] Gill, Ankit, et al. "Quantifying the fertility status and relationship between soil properties under major tea estates of North-western Himalaya." *Communications in Soil Science and Plant Analysis* 55.3 (2024): 287-301.

[3] Batool, Dania, et al. "A hybrid approach to tea crop yield prediction using simulation models and machine learning." *Plants* 11.15 (2022): 1925.

[4] Liu, Huixiang, Dongbing Yu, and Yu Gu. "Classification and evaluation of quality grades of organic green teas using an electronic nose based on machine learning algorithms." *IEEE Access* 7 (2019): 172965-172973.

[5] Santhiya, G., and A. Radhakrishnan. "TEA SOIL: Self-Power IoT for Precision Soil Farming of Tea Plant Cultivation." (2024).

[6] Ye, Jianghua, et al. "The Effects of Rock Zones and Tea Tree Varieties on the Growth and Quality of Wuyi Rock Tea Based on the OPLS-DA Model and Machine Learning." *Agriculture* 14.4 (2024): 573.

[7] Peng, Yufeng, et al. "Where is tea grown in the world: A robust mapping framework for agroforestry crop with knowledge graph and sentinels images." *Remote Sensing of Environment* 303 (2024): 114016.

[8] Girolametti, Federico, et al. "Unlocking the elemental signature of European tea gardens: Implications for tea traceability." *Food Chemistry* 453 (2024): 139641.

[9] Das, Animesh Chandra, Ryozo Noguchi, and Tofael Ahamed. "An Assessment of Drought Stress in Tea Plantation Areas in Bangladesh Using Optical and Thermal Remote Sensing: A Climate Change Perspective." *Remote Sensing Application II: A Climate Change Perspective in Agriculture*. Singapore: Springer Nature Singapore, 2024. 23-47.

[10] Unhelkar, Bhuvan, and P. Chakrabarti. "A Novel Deep Learning Models for Efficient Insect Pest Detection and Recommending an Organic Pesticide for Smart Farming." *International Journal of Intelligent Systems and Applications in Engineering* 12.9s (2024): 15-31.

[11] Mahato, Trinath, Bikash Ranjan Parida, and Somnath Bar. "Assessing tea plantations biophysical and biochemical characteristics in Northeast India using satellite data." *Environmental Monitoring and Assessment* 196.3 (2024): 327.

[12] Shingade, Sachin Dattatraya, and Rohini Prashant Mudhalwadkar. "Analysis of crop prediction models using data analytics and ML techniques: a review." *Multimedia Tools and Applications* 83.13 (2024): 37813-37838.

[13] Attri, Ishana, Lalit Kumar Awasthi, and Teek Parval Sharma. "Machine learning in agriculture: a review of crop management applications." *Multimedia Tools and Applications* 83.5 (2024): 12875-12915.

[14] Wei, Yang, et al. "The dawn of intelligent technologies in tea industry." *Trends in Food Science & Technology* (2024): 104337.

[15] Dey, Biplob, Jannatul Ferdous, and Romel Ahmed. "Machine learning based recommendation of agricultural and horticultural crop farming in India under the regime of NPK, soil pH and three climatic variables." *Heliyon* 10.3 (2024).

[16] Zheng, Pengfei, et al. "Online System for Monitoring the Degree of Fermentation of Oolong Tea Using Integrated Visible–Near-Infrared Spectroscopy and Image-Processing Technologies." *Foods* 13.11 (2024): 1708.

[17] Demir, Sinan, Mert Dedeoğlu, and Levent Başayiğit. "Yield prediction models of organic oil rose farming with agricultural unmanned aerial vehicles (UAVs) images and machine learnaing algorithms." *Remote Sensing Applications: Society and Environment* 33 (2024): 101131.

[18] Kouters, J. M. *Uncovering smallholder patterns in agricultural-water-soil characteristics: an unsupervised machine learning method*. MS thesis. University of Twente, 2024.

[19] Sajitha, P., et al. "A Review on Machine Learning and Deep Learning Image-based Plant Disease Classification for Industrial Farming Systems." *Journal of Industrial Information Integration* (2024): 100572.

[20] Kakade, Subodh, et al. "A Study of Weed Management in Agricultural Plants using IoT and Machine Learning." *2024 International Conference on Emerging Smart Computing and Informatics (ESCI)*. IEEE, 2024.

[21] Lu, Huayu, et al. "Multiple Soil Health Assessment Methods for Evaluating Effects of Organic Fertilization in Farmland Soil of Agro-Pastoral Ecotone." *Agriculture* 14.4 (2024): 572.

[22] Mohyuddin, Ghulam, et al. "Evaluation of Machine Learning approaches for precision Farming in Smart Agriculture System-A comprehensive Review." *IEEE Access* (2024).

[23] Shi, Chang, Qingwen Zhang, and Bowei Yu. "Higher improvement in soil health by animal-sourced than plant-sourced organic materials through optimized substitution." *Agriculture, Ecosystems & Environment* 363 (2024): 108875.

[24] Chen, Yifan, et al. "Deep learning and targeted metabolomics-based monitoring of chewing insects in tea plants and screening defense compounds." *Plant, Cell & Environment* 47.2 (2024): 698-713.

[25] Usama, Mohammad. "Application of Digital Technologies & Remote Sensing in Precision Agriculture for Sustainable Crop Production." *Geospatial Technology to Support Communities and Policy: Pathways to Resiliency*. Cham: Springer Nature Switzerland, 2024. 203-223.

[26] Prova, Nuzhat, et al. "Enhancing Crop Management: Ensemble Machine Learning for Real-Time Crop Recommendation System from Sensor Data." *Available at SSRN 4817360* (2024).

[27] Huang, Zhenxiong, et al. "Advanced deep learning algorithm for instant discriminating of tea leave stress symptoms by smartphone-based detection." *Plant Physiology and Biochemistry* (2024): 108769.

[28] Tamilselvi, R., K. S. Mohanasathiya, and S. Prasath. "Developed a Smooth Support Vector Machine to Predict the Crop Production in Alluvial Soil and Red Soil Regions of Tamil Nadu, India." *NATURALISTA CAMPANO* 28.1 (2024): 279-297.

[29] Mahinda, TG Nimal, et al. "Drivers of Adoption: Mechanical Harvesting in the Corporate Sector Tea Plantations in Sri Lanka." *Millennial Asia* (2024): 09763996241239414.

[30] Li, Yicong, et al. "The classification, detection and 'SMART'control of the nine sins of tea fraud." *Trends in Food Science & Technology* (2024): 104565.

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title: Soil Condition Monitoring For Organic Tea Plantations Using Machine Learning

Student ID: 201-15-13794
201-15-14013

CO Description for FYDP

CO	CO Descriptions	PO
Phase -I		
CO1	Integrate recently gained and previously acquired knowledge to Soil Condition Monitoring For Organic Tea Plantations for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish these goals for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
Phase -II		
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**
CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**

CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (With Knowledge Profile)	References
1.	EP1: Depth of Knowledge required	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	The project demonstrates **specialist knowledge (K4)** by conducting machine learning with GaussianNB, Random Forest, Decision Tree, SVC, MLP enhancing Soil Condition Monitoring For Organic Tea Plantations correctly.	**Page no:** [12-17] **Section:** [1.1]
				The project applies **engineering practice & design (K5)** by the figure of process of experiments. The project addresses **engineering practice & technology (K6)** by employing Machine Learning Model,	**Page no:** [18] **Section:** [3.2] **Page no:** [19-23] **Section:** [3.4]

				This project ensures **to K8 (Research Literature)** by synthesizing insights from recent studies, Soil Condition Monitoring For Organic Tea Plantations Using Machine Learning, showcasing a comprehensive understanding of current methodologies.	**Page no:** [8-10] **Section:** [2.2, 2.3]
2.	EP2: Range of Conflicting Requirements	Yes	CO2, and CO7	This project addresses **EP-2** by recognizing the hurdles in Soil Condition Monitoring For Organic Tea Plantations, including limitations of traditional methods and the complexities of integrating using machine learning. Through comparative analysis, it confronts challenges in understanding spatial distributions, offering insights for refining diagnostic methodologies.	**Page no:** [11-12] **Section:** [2.4, 2.5]
3.	EP3: Depth of analysis required	Yes	CO2, and CO6	This project addresses **EP-3** by meticulously comparing experimental outcomes, highlighting using deep and transfer learning as the chosen significant solution for enhancing Soil Condition Monitoring For Organic Tea Plantations Approaches.	**Page no:** [18-19] **Section:** [4.2]
4.	EP4: Familiarity of Issues	Yes	CO8	This project's interdisciplinary approach extends beyond computer science and engineering, impacting Soil Condition Monitoring For Organic Tea Plantations Detection **EP-4.**	**Page no:** [23-31] **Section:** [2.2, 2.4]
5.	EP5: Extends of application codes	No	CO5	N/A	N/A

6.	EP6: Extends of stakeholders involved and conflicting requirements	No	CO8	N/A	N/A
7.	EP7: Interdependence	Yes	CO5	This project's comprehensive approach addresses high-level problems by integrating various components across data collection, statistical analysis, and proposed methodology, ensuring a holistic solution to complex challenges in data collection from local nursery, Interiors decorated with tea and garden which ensures **EP-7.**	**Page no:** [26-34] **Section:** [3.2, 3.3, 3.4]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	EA1: Range of resources	Yes	CO11	Our project utilizes diverse resources such as high-performance computing infrastructure, GPUs, and ethical considerations to ensure systematic research and contribute to advancements Soil Condition Monitoring Detection.	**Page no:** [26-27] **Section:** [3.5]
2.	EA2: Level of interaction	No		N/A	N/A
3.	EA3: Innovation	No		N/A	N/A
4.	EA4: Consequences for society and the environment	Yes		This project contributes to society by improving Soil Condition Monitoring Detection methods, while also promoting environmental sustainability by employing efficient computational resources and adhering to ethical guidelines for data privacy.	**Page no:** [27-28] **Section:** [5.1, 5.2, 5.4]

5.	EA-5: Familiarity	Yes		This project expands upon existing research by examining a novel approach in Soil Condition Monitoring through deep and deep learning model demonstrated through preliminary terminologies and a comprehensive comparative analysis, offering new insights into the field.	**Page no:** [23-29] **Section:** [2.1, 2.3]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project addresses **CO4** by integrating effective project management and financial oversight, ensuring meticulous planning, resource allocation, and budget estimation for optimal resource utilization throughout the research lifecycle.	**Page no:** [20] **Section:** [1.6]
2	CO9	Yes	The project demonstrates adherence to ethical principles by prioritizing privacy, obtaining informed consent, and transparently documenting the research process, ensuring responsible knowledge dissemination and societal well-being through the ethical application of advanced classification technologies which comply with **CO9.**	**Page no:** [28] **Section:** [5.3]
3	CO10	No	N/A	N/A
4	CO12	Yes	The project's dedication to continuous learning **(CO12)** and adaptation within the dynamic technological landscape is reflected in its comprehensive data collection, rigorous statistical analysis, meticulous methodology development, and thorough experimental results and analysis, showcasing a commitment to staying updated and refining techniques to address modern challenges.	**Page no:** [27-28, 31-32] **Section:** [3.2, 3.3, 3.4, 3.5, 4.2]

Imrun2024

ORIGINALITY REPORT

7%	4%	4%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	1%
2	Submitted to Liverpool John Moores University Student Paper	1%
3	www.eurchembull.com Internet Source	1%
4	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
5	www.mdpi.com Internet Source	<1%
6	bc.pollub.pl Internet Source	<1%
7	Ananya Sarker, Must. Asma Yasmin, Md. Atikur Rahman, Md. Harun Or Rashid, Bristi Rani Roy. "Credit Card Fraud Detection Using Machine Learning Techniques", Journal of Computer and Communications, 2024 Publication	<1%