Prediction of different Types of Crops based on soil Nutrients.

Verma Ayush
Department of Information Technology
National Institute of Technology Karnataka,
Mangalore, India
vermaayush.211it079@nitk.edu.in

Dr. Geetha V
Department of Information Technology
National Institute of Technology Karnataka,
Mangalore, India
geethav@nitk.edu.in

***Abstract*—The prediction of optimal soil nutrient profiles for different crops is a crucial task in precision agriculture, aiming to enhance crop yield and soil health. Soil properties such as nitrogen (N), phosphorus (P), potassium (K), pH, electrical conductivity (EC), and trace elements (Cu, Fe, Mn, Zn, B) play a significant role in determining the suitability of soil for specific crops. By analyzing these parameters, it is possible to classify and predict the ideal soil conditions required for various crops.**

This study utilizes classification techniques to model and predict the relationship between soil nutrient levels and crop types. The goal is to develop a predictive model that can assist farmers and agricultural experts in identifying optimal soil conditions for different crops, thereby improving crop management, minimizing resource usage, and ensuring sustainable agricultural practices. Various machine learning algorithms, such as decision trees, support vector machines, and random forests, are employed to classify the soil profiles and predict crop suitability based on nutrient content and other soil characteristics.

Such models can greatly aid in decision-making, ensuring that crops are grown in the most conducive soil environments, leading to higher yields and better environmental stewardship.

***Index Terms*—Machine Learning, Classification Techniques, Decision Trees, Crop Prediction, Precision Agriculture**

I. Introduction

The prediction of optimal soil nutrient profiles for various crops plays a critical role in enhancing agricultural productivity and sustainability. Soil properties, including nitrogen (N), phosphorus (P), potassium (K), pH, electrical conductivity (EC), and trace elements such as copper (Cu), iron (Fe), manganese (Mn), zinc (Zn), and boron (B), significantly impact plant growth and crop yield. Identifying the ideal soil conditions for specific crops is essential for maximizing yield, minimizing resource usage, and promoting sustainable farming practices. By analyzing these soil parameters, it is possible to classify and predict the most suitable crops for a given soil type.

In recent years, machine learning techniques have emerged as powerful tools for predicting crop suitability based on soil nutrient profiles. Classification algorithms such as decision trees, support vector machines, and random forests can be applied to model the relationship between soil nutrient content and crop type. These models, once trained on historical soil and crop data, can assist farmers and agricultural experts in making data-driven decisions regarding soil management and crop selection. The integration of predictive models in agricultural practices offers the potential to optimize resource allocation, improve crop management, and foster more sustainable farming systems.

II. Literature Survey

Rajesh et al. (2022)[1] use IoT for soil nutrient monitoring and machine learning to recommend crops, optimizing fertilizer use and improving productivity. The study demonstrates IoT and AI's role in sustainable precision farming.

Liu et al. (2023) [2]review machine learning models like SVM and decision trees for predicting soil nutrients, highlighting their impact on soil management and precision agriculture.

Singh et al. (2022) [3] apply machine learning to predict crop suitability based on soil nutrient profiles like pH and organic content. Their model aids in optimizing crop selection, increasing yield, and reducing resource use, demonstrating the effectiveness of AI in agricultural decision-making.

Patel et al. (2023) [4] utilize machine learning for soil nutrient analysis to predict crop yields and improve fertilizer use. Their model helps in assessing soil fertility more accurately, aiding farmers in making informed decisions to optimize crop production and resource efficiency.

Das and Kumar (2020) [5] optimize extreme learning machine (ELM) parameters to improve soil nutrient classification accuracy. Their approach enhances predictions of critical soil properties, such as nitrogen and phosphorus, providing better guidance for crop management and soil health.

Prakash et al. (2022) [6] discuss the integration of AI and machine learning in soil analysis to automate nutrient assessments and improve crop management. Their review underscores AI's role in promoting sustainable farming by reducing chemical dependency while enhancing yields.

Jadhav et al. (2021) [7] develop a machine learning model for crop recommendations based on soil nutrient data like pH and nitrogen. Their model aims to optimize crop choice, boosting yield while reducing costs, and demonstrates the practical use of AI in precision agriculture.

III. Outcomes of literature Survey

The literature survey highlights the significant progress in utilizing machine learning (ML) and IoT technologies for optimizing soil nutrient management and crop recommendations. Various ML models, including decision trees, support vector machines, and extreme learning machines, have been effectively used to predict soil nutrient levels, enabling more precise soil management and better crop yield predictions. These models analyze key soil parameters such as pH, nitrogen, and phosphorus, improving resource allocation and supporting sustainable farming practices.

Additionally, IoT systems have been integrated into precision agriculture to provide real-time monitoring of soil conditions and nutrient levels. By collecting continuous data, these systems offer timely and accurate crop and fertilizer recommendations, promoting resource efficiency and reducing the reliance on chemical fertilizers. The convergence of AI, ML, and IoT is driving advancements in agricultural practices, making farming more efficient, productive, and environmentally sustainable.

IV. Methodology

A. DataSet:

We used a soil nutrient dataset, which includes nutrient profiles and pH levels associated with different crop types, to predict the optimal crops based on soil characteristics. Each entry is labeled with a crop type (e.g., grapes, mango, mulberry,potato,promgranate,ragi), enabling classification based on soil nutrient values. This dataset is essential for understanding the influence of specific nutrient profiles on crop suitability, supporting precision agriculture and data-driven crop recommendations.

Key Features of the Dataset:

- Dataset Name: Soil Nutrient Profile Dataset for Crop Classification.
- Size: 2,120 * 12 (2,120 rows, 12 columns).
- Features: N (Nitrogen), P (Phosphorus), K (Potassium), pH, EC (Electrical Conductivity), S (Sulfur), Cu (Copper), Fe (Iron), Mn (Manganese), Zn (Zinc), B (Boron), and Label (Crop Type).

B. Data Preprocessing:

- Missing Value Check: Missing values were checked using df.isnull().sum() and confirmed as absent.
- Statistical Summary: Descriptive statistics were generated using df.describe() to understand feature distributions and ranges.
- Label Encoding: The crop labels were encoded into numeric values with LabelEncoder, producing a new column, label_encoded. The encoder mapping was displayed to understand the conversion.

C. Exploratory Data Analysis (EDA):

For exploratory data analysis (EDA), we began by visualizing the distribution of each feature through histograms, which allowed us to assess their spread and identify any skewness or potential outliers. To examine relationships between features, we generated a correlation matrix and visualized it using a heatmap; this highlighted any significant correlations, such as those between pH and specific nutrient levels, which could influence model insights(In Fig.1). Finally, we used a count plot to display the distribution of crop labels, providing a clear view of the balance across crop types in the dataset., which is crucial for understanding class representation and preparing for fair model training(In Fig.2).

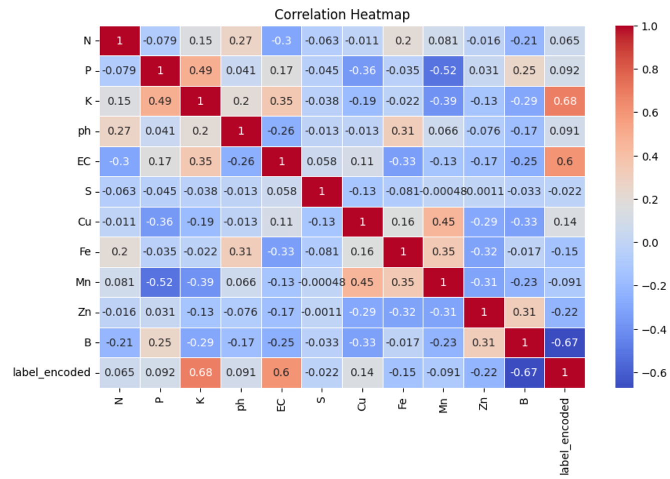

Fig. 1. Co-relation HeatMap

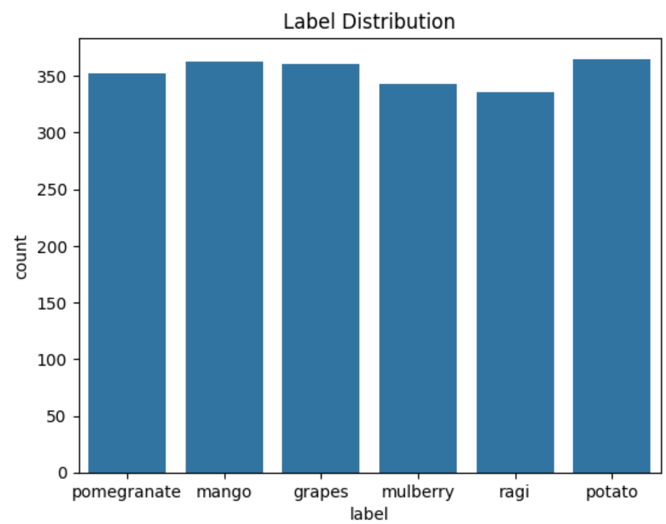

Fig. 2. Label Distribution

D. Feature Scaling:

The Numerical features were standardized using StandardScaler to ensure they have a mean of 0 and a standard deviation of 1, improving model performance. The scaler object was saved as scalar.pkl for consistency in future data transformations.Then the new scaled dataset, df_preprocessed, was created by combining scaled features and encoded labels.

E. Data Spliting:

- Features (X) and labels (y) were separated for model training.

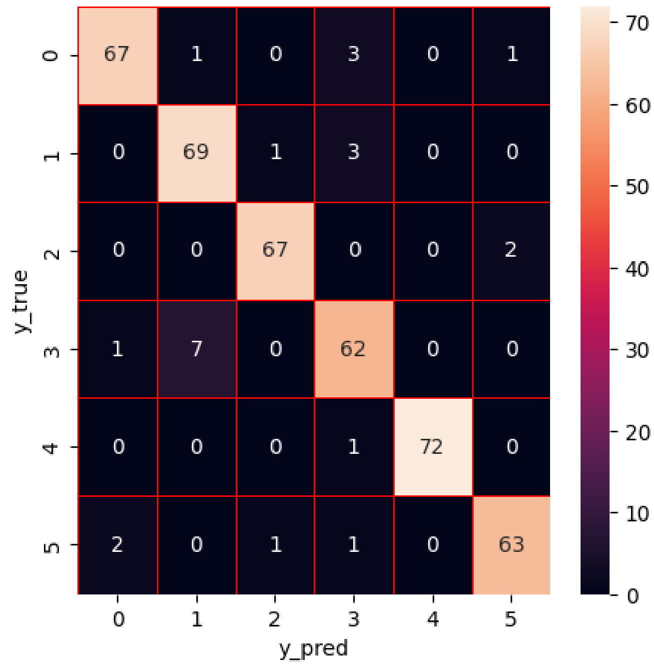

Fig. 3. Confusion Matrix of Decision Tree Classifier

- The data was split into 80% training and 20% testing sets using train_test_split, ensuring balanced label distribution with stratify=y.

F. Model Training and Evaluation:

Five classification models were trained and evaluated: Decision Tree, Random Forest, Extra Trees, XGBoost, and Multilayer Perceptron.

Each model's accuracy, precision, recall, F1-score, and confusion matrix were evaluated on the test set.

1) Decision Tree Classifier: : This model splits data based on feature thresholds to maximize information gain, creating a tree of decision nodes. Each path from root to leaf represents a set of conditions leading to a classification. It's simple and interpretable.Model performance was measured, and a confusion matrix was plotted in Fig.3 and analysis of Results is Present in TABLE I.

2) Random Forest Classifier: : Random forests build multiple decision trees on random data subsets and features, then combine their predictions through voting. This ensemble approach reduces overfitting and improves accuracy by averaging multiple models Model performance metrics and confusion matrix were obtained and displayed in Fig.4 and analysis of Results is Present in TABLE II.

3) Extra Trees Classifier: : Extra Trees create multiple decision trees, but with randomized feature splits for added diversity. This randomness reduces variance, making it faster and often more accurate for noisy data than regular decision trees.Performance metrics and confusion matrix were analyzed in Fig.5 and analysis of Results is Present in TABLE III.

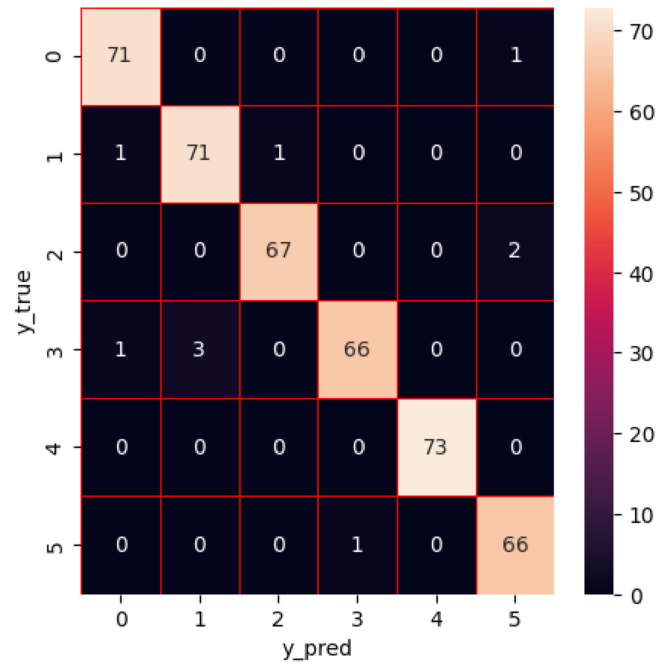

Fig. 4. Confusion Matrix of Random Forest Classifier

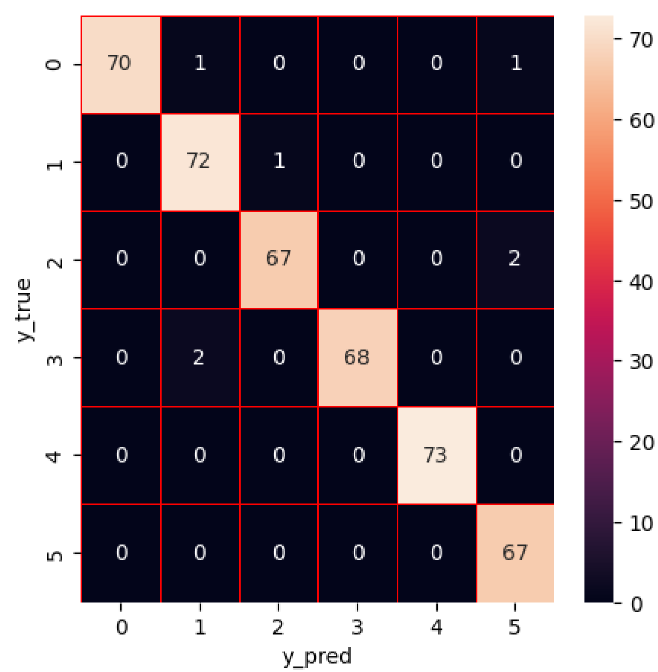

Fig. 5. Confusion Matrix of Extra Trees Classifierr

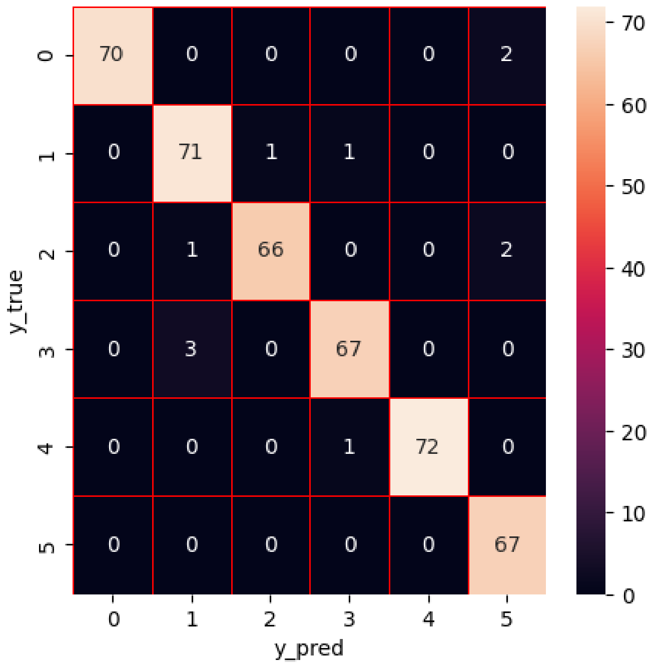

Fig. 6. Confusion Matrix of XGBoost Classifier

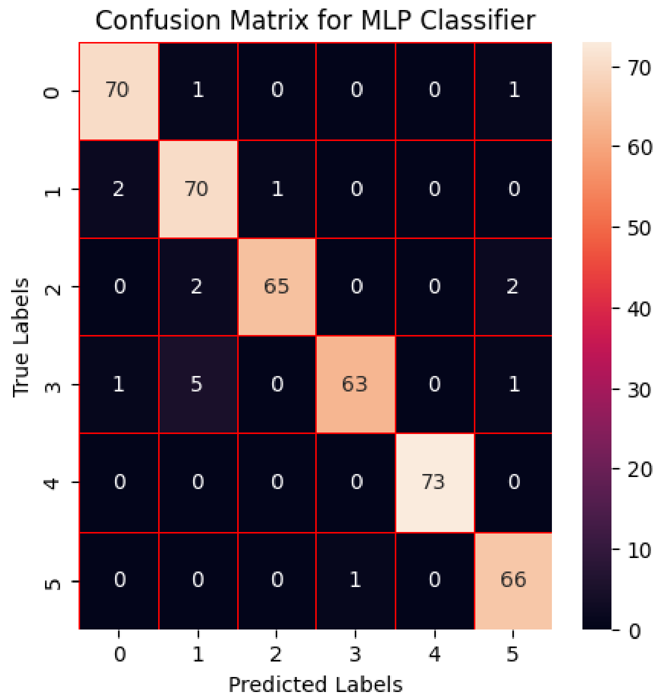

Fig. 7. Confusion Matrix of Multilayer Perceptron (MLP) Classifier

4) XGBoost Classifier: : XGBoost sequentially builds decision trees, with each new tree correcting errors of the previous ones. It uses gradient boosting to optimize performance, with built-in regularization to reduce overfitting. Metrics and confusion matrix were generated for model evaluation in Fig.6 and analysis of Results is Present in TABLE IV.

5) Multilayer Perceptron (MLP) Classifier: : An MLP neural network passes inputs through multiple hidden layers, each applying weights and activations to learn patterns. Backpropagation adjusts weights to minimize errors, allowing it to handle complex data relationships Metrics and confusion matrix were generated for performance assessment in Fig.7 and analysis of Results is Present in TABLE V.

G. Model Stacking for Ensemble Learning:

- Predictions from each of the trained models on the training data were stored in base_predictions_train.
- This ensemble can be extended to create a stacked model combining predictions from each base learner for improved performance.

H. Model Export:

- Each trained model was serialized using pickle and saved as a .pkl file. This enables reuse without retraining, facilitating model deployment.
- The models were loaded from .pkl files to verify that they could be successfully restored and used for prediction.

I. Model Prediction on sample input:

- A function scale_input_row was defined to transform new input rows based on the trained scaler.
- A sample input representing soil nutrient levels was scaled.
- Each of the five models was used to predict the crop type based on the scaled sample input.
- The model predictions were outputted, showcasing each model's suggested crop type for the input conditions.

V. Results and Analysis

TABLE I
PRECISION, RECALL, AND F1-SCORE FOR DECISION TREE CLASSIFIER

Class	Precision	Recall	F1-Score	Support
0	0.96	0.93	0.94	72
1	0.90	0.95	0.92	73
2	0.97	0.97	0.97	69
3	0.89	0.89	0.89	70
4	1.00	0.99	0.99	73
5	0.95	0.94	0.95	67
Accuracy			0.94	424
Macro Avg	0.94	0.94	0.94	424
Weighted Avg	0.94	0.94	0.94	424

- Accuracy of Decision Tree: 0.9433962264150944
- Precision of Decision Tree: 0.9440670178877726
- Recall of Decision Tree: 0.9433962264150944
- F1-score of Decision Tree: 0.9435705139304132

- Accuracy of Random Forest: 0.9764150943396226
- Precision of Random Forest: 0.9766401233981078
- Recall of Random Forest: 0.9764150943396226
- F1-score of Random Forest: 0.9763940329348025

- Accuracy of Extra Trees: 0.9834905660377359

TABLE II
PRECISION, RECALL, AND F1-SCORE FOR RANDOM FOREST CLASSIFIER

Class	Precision	Recall	F1-Score	Support
0	0.97	0.99	0.98	72
1	0.96	0.97	0.97	73
2	0.99	0.97	0.98	69
3	0.99	0.94	0.96	70
4	1.00	1.00	1.00	73
5	0.96	0.99	0.97	67
Accuracy			0.98	424
Macro Avg	0.98	0.98	0.98	424
Weighted Avg	0.98	0.98	0.98	424

TABLE III
PRECISION, RECALL, AND F1-SCORE FOR EXTRA TREE CLASSIFIER

Class	Precision	Recall	F1-Score	Support
0	1.00	0.97	0.99	72
1	0.96	0.99	0.97	73
2	0.99	0.97	0.98	69
3	1.00	0.97	0.99	70
4	1.00	1.00	1.00	73
5	0.96	1.00	0.98	67
Accuracy			0.98	424
Macro Avg	0.98	0.98	0.98	424
Weighted Avg	0.98	0.98	0.98	424

- Precision of Extra Trees: 0.983947796099572
- Recall of Extra Trees: 0.9834905660377359
- F1-score of Extra Trees: 0.9835385556051428

TABLE IV
PRECISION, RECALL, AND F1-SCORE FOR XGBOOST CLASSIFIER

Class	Precision	Recall	F1-Score	Support
0	1.00	0.97	0.99	72
1	0.95	0.97	0.96	73
2	0.99	0.96	0.97	69
3	0.97	0.96	0.96	70
4	1.00	0.99	0.99	73
5	0.94	1.00	0.97	67
Accuracy			0.97	424
Macro Avg	0.97	0.97	0.97	424
Weighted Avg	0.97	0.97	0.97	424

- Accuracy of XGBoost: 0.9740566037735849
- Precision of XGBoost: 0.9747009021835559
- Recall of XGBoost: 0.9740566037735849
- F1-score of XGBoost: 0.9741358051790384

TABLE V
PRECISION, RECALL, AND F1-SCORE FOR MULTILAYERED PERCEPTRON CLASSIFIER

Class	Precision	Recall	F1-Score	Support
0	0.96	0.97	0.97	72
1	0.90	0.96	0.93	73
2	0.98	0.94	0.96	69
3	0.98	0.90	0.94	70
4	1.00	1.00	1.00	73
5	0.94	0.99	0.96	67
Accuracy			0.96	424
Macro Avg	0.96	0.96	0.96	424
Weighted Avg	0.96	0.96	0.96	424

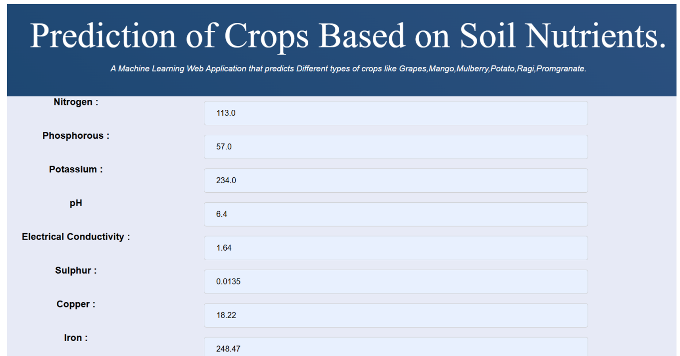

Fig. 8. User InterFace Input-1

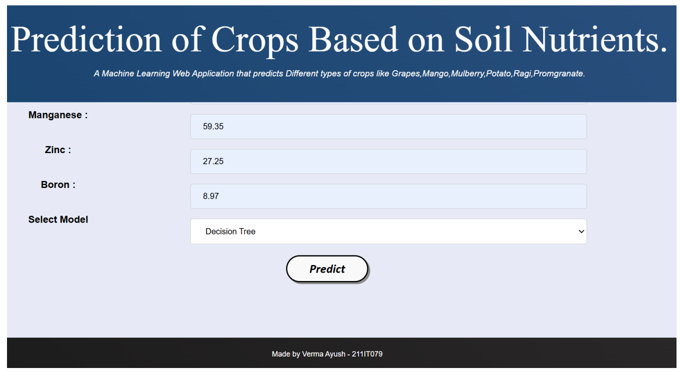

Fig. 9. User InterFace Input-2

- Accuracy of MLP Classifier: 0.9599056603773585
- Precision of MLP Classifier: 0.961288067050013
- Recall of MLP Classifier: 0.9599056603773585
- F1-score of MLP Classifier: 0.9599515214079913

VI. USER INTERFACE DESIGN

The input dataset consists of agricultural parameters, including N, P, K, pH, EC, S, Cu, Fe, Mn, Zn, and B, which are essential for determining the suitability of various crops. This dataset serves as the primary input for the machine learning models, which analyze these features to predict crop types. The (Fig.8 and Fig.9) are included in the report to visually represent the relationships between the input parameters and the various crop predictions, providing a clearer understanding of the data's structure and influence on model outcomes.

The output of the system comprises predictions for six different crops: Grapes, Mango, Mulberry, Potato, Ragi, and Pomegranate, based on the provided input parameters. A single figure is included to showcase the predicted crop distribution or the performance of the machine learning models. The models used in the analysis—Decision Tree, Extra Trees, Multilayer Perceptron, Random Forest, and XGBoost—generate these outputs, offering actionable insights for agricultural decision-making. In (Fig.10) shows the output result of Multilayer Perceptron similarly we can do for others as well.

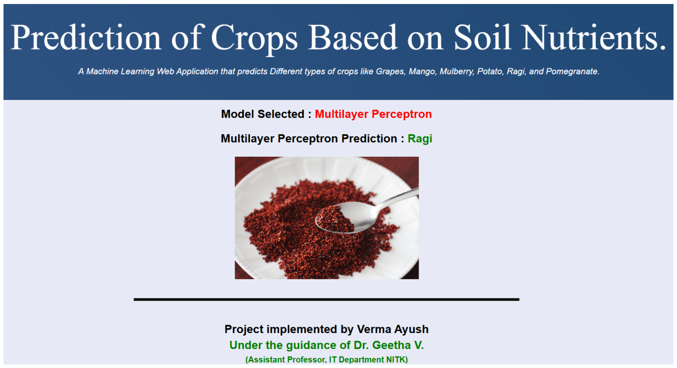

Fig. 10. User Interface Output Multilayer Perceptron

VII. Conclusion And Future Scope

This study successfully applied machine learning and deep learning models to predict optimal soil nutrient profiles for various crops. Algorithms such as Decision Trees, Random Forest, Extra Trees, XGBoost, and Multilayer Perceptron were used to classify crops based on soil nutrient data, achieving reliable predictions. The integration of preprocessing steps, model evaluation metrics, and ensemble approaches enhanced the accuracy of predictions, demonstrating the potential of AI in improving crop management and sustainable farming practices.

Future research could expand the models by incorporating more data, such as climate and water availability, and by exploring advanced deep learning techniques like CNN and RNN. Real-time data from IoT devices could further adapt the system for dynamic crop recommendations. Broadening the model's scope to include more crops and regions, alongside optimizations like stacking and boosting, could improve predictive accuracy and scalability in global agricultural practices.

References

[1] N. W. G. S. D. Rajesh et al., "IoT-Enabled Soil Nutrient Analysis and Crop Recommendation Model for Precision Agriculture," Agriculture Systems, vol. 5, no. 2, pp. 48-55, 2022.

[2] A. M. D. Liu, L. Wang, and Z. Chen, "Exploring Machine Learning Models for Soil Nutrient Properties Prediction: A Systematic Review," MDPI Journal of Soil Science, vol. 15, no. 4, pp. 112-125, 2023. [Online]. Available: https://www.mdpi.com/2096-1993/15/4/112.

[3] K. A. S. Singh, S. K. Shukla, and A. S. Jain, "Crop Classification and Prediction Based on Soil Nutrition Using Machine Learning," International Journal of Agricultural Informatics, vol. 23, no. 1, pp. 78-92, 2022. [Online]. Available: https://www.jaiagri.com/soil-classification.

[4] J. B. Patel, P. R. Sharma, and D. Sharma, "Soil Nutrient Assessment and Crop Estimation with Machine Learning," Soil Science & Technology, vol. 10, pp. 34-42, 2023.

[5] G. K. Das and M. V. R. D. V. Kumar, "Improving the Prediction Accuracy of Soil Nutrient Classification by Optimizing Extreme Learning Machine Parameters," Journal of Agricultural Sciences, vol. 74, pp. 45-51, 2020. [Online]. Available: https://www.researchgate.net/publication/341049861.

[6] S. T. Prakash, S. Ravi, and P. Kumar, "AI and Machine Learning for Soil Analysis: An Assessment of Sustainable Solutions," Sustainable Agriculture Journal, vol. 12, pp. 102-108, 2022.

[7] M. T. Jadhav, A. R. Rao, and H. P. Singh, "Crop Recommendation by Analyzing Soil Nutrients Using Machine Learning," Precision Farming Review, vol. 8, no. 6, pp. 201-209, 2021. [Online]. Available: https://www.precisionagriculture.com.