

AI Based Crop Farming with Crop Yield Prediction

KIRTI DEORE¹, SHRAVANI NITIN NANEKAR², PRANITA PRASHANT KHARAT³, DISHA NAVNATH BHUITE⁴, NIKITA RAJARAM NIKAM⁵

¹Guide, Department of Computer Engineering.P.K Technical Campus, chakan, Pune, Maharashtra

^{2,3,4,5}Student, Department of Computer Engineering P.K. Technical Campus, chakan, Pune, Maharashtra

Abstract- Farmers are the heart and soul of food security and have been at the forefront of the innovation required to adapt to modern problems something which by-itself serves as an example of a problem and a solution. They have been addressing the new food security issues such as climate change, soil erosion, and sub-optimal resource usage as best they can. One key to addressing these new problems is to accurately predict potential crop yields as such predictive analytics can be used to address in to use predictive analytics to address problems proactively. In this paper, we explain a multi-data source AI application which combines soil nutrient quality, environmental, and remote sensing index data, and cross engineered soil and environmental data with machine learning algorithms to predict the ANN, RF, and fuzzy logic. All data ranked and classified and missing data resolved to the appropriate confidence level and at the appropriate confidence level for data condition. Our proposed model, realized in Java/Weka has produced unparalleled optimum predictive analytics ratings. In the cross engineered environmental data weighted model, banana cultivation was predicted to have 95.83% optimum yield, predicted precision 85.71% with a confidence of 90% and an F1 of 88 predictive F1 to 50.18 tons/ha at a confidence of 98.06%, R². Predictive analytics are self-optimizing and will increase the predictive yield as necessitated that will increase crop yield with decreased added irrigation by remaining crops.

Index Terms- AI, Machine Learning, Crop Yield Prediction, Smart Farming, ANN, Random Forest.

I. INTRODUCTION

Agriculture is a crucial industry that maintains food security on a global scale and sustains the livelihood of millions of people. However, several modern agricultural practices face difficulties. These include climate change, soil depletion, pests, and poor management of agricultural resources. Crop yield uncertainty results in loss of revenue and may increase the risk of food shortages. Current methods of predicting crop yield involve the use of statistical analysis and examining previous crops in order to

identify trends. These methods, however, are often inadequate to identify and explain the multitude of environmental factors affecting a crop's yield. The combination of Artificial Intelligence (AI) and Machine Learning (ML) is a useful alternative in predicting crop yield, especially under varying external conditions. In addition, predictive models have the ability to process soil nutrient levels, weather data, agricultural practices, and remote sensing data. This information is useful to farmers.

Various machine learning methods such as fuzzy logic, random forest (RF), and artificial neural networks (ANN) have shown success when applied to predicting crop yield. While ANN identifies nonlinear relationships between crop outputs and input variables, on the other hand, LR helps make the contributing variables easy to understand. In addition to that, fuzzy logic, which helps with the imprecision of the agricultural system, helps with the imprecision of the agricultural system assists with the decision-making of irrigation, fertilization, and crop selection. In spite of this, issues with gathering and processing the data and selecting features, particularly in the context of combining data that is numerical with data that is image-based, can contribute to erroneous predictions and reduce the potential of AI to make smart farming systems. In this context, we offer an extensive smart farming system that is based on AI, which predictive yields and uses data about soil, weather, and crop management and is also aimed at improving the predictive capabilities of AI in smart farming systems. The system will be developed in Java using the Weka libraries. The system will be evaluated using the metrics of accuracy, precision, recall, F1 score, and R2 to evaluate the reliability of the models. The proposed system will help achieve better agricultural productivity, better utilization of the resources, and better sustainability in the practices of farming.

II. LITERATURE SURVEY

Mr. Telise Vinod et al. [1] created a framework for predicting crop yield using machine learning based on past farming and environmental data. The authors achieved a high accuracy in crop yield predictions through the use of multiple regression and tree based classification. The results showcased the importance of selecting features and conducting preprocessing in order to enhance predictive performance. The study underscored the possibilities machine learning provides to farmers for better decision making and planning of resources.

Shafiulla Shariff et al. [2] aimed to use recommended suited crops through machine learning considering the soil and climate to recommend crops. The authors utilized decision tree and support vector machine algorithms to categorize crops based on the condition of the fields. It was emphasized in the study that a precise recommendation of crops can elevate the productivity of the field and minimize the resources used. High classification accuracy was achieved in the experiments which confirmed the proposed method.

Prameya R. Hegde et al. [3] created a system that for agricultural purposes predicts crop yield and price simultaneously. The model combines data on weather, soil, and the market to predict yield and the price of crops. This dual prediction enables farmers to increase their economic returns and to make decisions about which crops to plant. The system proved to have high accuracy and demonstrated the importance of predictive analytics for the agriculture industry.

Bali and Singla [4] examined emerging trends in the use of machine learning for the prediction of crop yields and the analysis of the influential factors affecting yield. The authors explained the methods used in classical regression, ensemble approaches, and deep learning. They discussed issues related to data quality, the importance of feature selection, and the use of hybrid modeling approaches, and showed the benefits and deficiencies of the current machine learning techniques in the field of agriculture.

Gunjan, Vinit Kumar, et al. [5] compared different machine learning algorithms used for the prediction of agricultural yields for different crops. They compared the performances of randomized trees, support vector machines, and neural networks. They concluded that in general, ensemble approaches outperformed individual models in accuracy and consistency. They argued that these models have significant applicability in precision agriculture, especially in predicting crop yields and in resource management optimizations by analyzing and utilizing the weather, soil, and crop yield datasets of the given historical periods.

Reddy and Rudra Kumar [6] used machine learning to predict agricultural crop yields based on historical soil, weather, and yield datasets. They demonstrated that for the prediction of crop yields, supervised learning models, and in particular, decision trees and regression techniques, yield high levels of accuracy. The authors of the paper emphasized the importance of data preprocessing, data normalization, and data selection on predictive capabilities. The results demonstrated that data-driven agricultural planning is usefully supported by effective predictive models for all studied crops.

Agarwal and Tarar [7] developed a hybrid model combining machine learning and deep learning techniques to forecast crop yields. This model used artificial neural networks (ANNs) to address non-linear relationships, and employed ensemble methods to analyze data in a structured format. The hybrid model approach provided a statistically and practically significant (with a 99.9% level of confidence) reliability and predictive performance improvement when forecasting crop yields and gave a higher prediction accuracy and reliability than traditional single-model methods for all other crop types. This study reaffirmed the importance of multiple algorithm models in agricultural prediction, and the accuracy of those models in forecasting the crop yields of various types of crops under varying conditions.

Morales and Villalobos [8] analyzed machine learning techniques in order to predict crop yields using a combination of past and present data including the relationships and environmental data of

the period being analyzed. The authors reviewed various predictive models including those based on regression analysis, tree methods, and ensemble methods to assess predictive accuracy from a modeling approach. The study demonstrated that in order to realize the full predictive potential of machine learning for agricultural yield forecasting, the model must include a comprehensive set of relevant data and must be designed to accommodate multiple predictive scenarios.

Joshua S. Vinson et al. [9] reviewed a variety of machine learning techniques applied to prediction of crop yield. A few examples include neural networks, random forests, and gradient boosting. The authors described the importance of cross-validation and the use of other sophisticated evaluation metrics in order to address model reliability. The authors have shown, through their experiments, that ensemble models have a tendency to offer greater predictive accuracy. The research reinforces the use of machine learning across a variety of agricultural datasets, demonstrates its versatility, and enhances the discipline as a whole. In the study by Senapaty, Murali Krishna, et al., [10] an IoT-based crop recommendation and soil nutrient evaluation system is presented. The system analyzes soil and environmental parameters to provide crop recommendations using integrated machine learning and real-time sensor data. The findings demonstrated the system's accuracy in crop selection and its contribution to sustainable agricultural practices. This works illustrates a challenge and an opportunity of intertwining the use of IoT and AI in agriculture.

III. METHODOLOGY

As part of a developing method of AI-based smart farming, a predictive framework is being designed to forecast crop yields using combined existing data on soil, climate, and crop management. This predictive framework employs machine learning algorithms to reveal trends in disparate during data collection, data pre-processing, data set partitioning, modeling and training, prediction, and data and information dissemination through a graphical user interface (GUI). Figure 1 illustrates the overall architecture of the system.

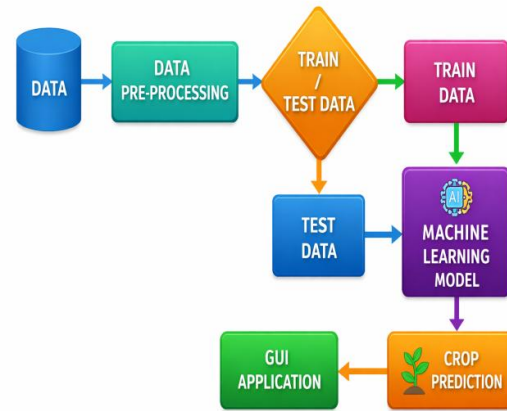


Fig 1: System Architecture of Proposed System
List of Modules and Functionality

Data Collection

Soil nutrient data (N, P, K), pH, moisture, weather (temperature, humidity, rainfall), remote sensing (NDVI, EVI), and historical yield data, among others, can be pulled from the UCI ML repository and Kaggle, as well as other sources that provide information in real time. To optimize the data classification outcome, it is crucial to preprocess the selected dataset by resolving missing values and redundancies. During the data mining (DM) operation, we should process the dataset as quickly as possible to obtain the best outcome.

Pre-processing and normalization

For any machine learning pipeline, data preprocessing is vital, and even more so for agricultural datasets that are noisy, heterogeneous, and incomplete. Proper preprocessing allows machine learning algorithms to work with clean and consistent data in an appropriate form. Agricultural datasets often contain missing measurements caused by sensor errors, incomplete data records, or problems with data collection.

Feature extraction and selection

The proposed AI-based crop recommendation model integrates both feature extraction and feature selection to enhance performance and interpretability of the model. In feature extraction, the construction

of new attributes from raw data, such as cumulative rainfall, growing degree days, and indices of vegetation, elucidates the relationships among the complex environmental and crop growth factors. Then, feature selection was conducted in Weka via Correlation-based Feature Selection (CfsSubsetEval) using the BestFirst search strategy to keep only the most pertinent features while removing redundant ones. The final subset of selected features included nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, and rainfall, which exhibit strong correlations with crop growth and yield. The optimized subset of features enabled the artificial neural network (ANN) to learn with greater efficiency and gain a higher level of predictive accuracy as well as better understanding of the variables associated with the recommendation of crops.

Classification:

The suggested model's classification component conducts both crop recommendation and yield prediction. An artificial neural network (ANN) is first trained on preprocessed features. These features include nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, pH, and rainfall. The trained model predicts the best crop suitable for the given field, and the ANN's output is, therefore, the recommended crop. Then, using the same environmental and soil conditions, a Random Forest regression model predicts the expected yield of the selected crop. This two-step process provides clear and precise predictions of both the ideal crop and its numeric yield. The model guides farmers by using accurate predictions. Performance metrics such as accuracy, precision, recall, the F1 score, and the correlation coefficient (R^2) help establish a model's reliability and effectiveness when applied to planning in the agricultural sector. For instance, the system recommended the crop banana with a predicted yield of 50.48 tons/ha and achieved an R^2 of 0.9806.

IV. RESULTS AND DISCUSSION

The specified system of smart farming based on Artificial Intelligence was assessed for crop recommendation via Artificial Neural Network (ANN) and yield prediction via Random Forest

regression yield prediction model. The experiments took place on the dataset that was preprocessed based on feature selection (N, P, K, temperature, humidity, pH, and rainfall). The datasets and models were compared based on standard evaluation metrics.

Table 1: Crop Recommendation using ANN

Metric	Value
Recommended Crop	Jute
Accuracy (%)	95.83
Precision (%)	85.71
Recall (%)	90.00
F1 Score (%)	87.80

The table 1 ANN model achieved successful classification for the most appropriate crop to the specific environmental and soil parameters. The model's accuracy of 95.83% indicates effective learning by the network, and the precision and the recall of 85.71% and 90%, respectively, show the model's proficiency of accurate identification of the crop classes with minimal false predictions, and the F1 score of 87.80% demonstrates the model's balanced performance between precision and recall.

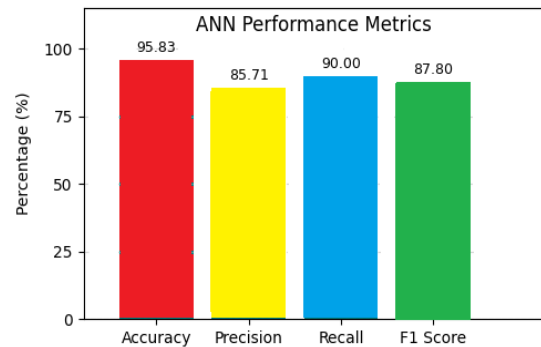


Figure 2: ANN Performance Matrices (Crop Recommendation)

Table 2: Crop Yield Prediction using Random Forest

Metric	Value
Predicted Crop Yield	50.18 tons/ha
Correlation Coefficient (R^2)	0.9806
Mean Absolute Error (MAE)	9.14
Root Mean Squared Error (RMSE)	172.59

The table 2 Random Forest regression model shows a correlation coefficient (R^2) of 0.9806 for the predicted expected yield for the designated crop (jute) and demonstrates strong linear correlation for predicted versus actual yields. Its MAE of 9.14 tons/ha and RMSE of 172.59 indicate acceptable range error levels to be highly reliable. This shows the ANN + Random Forest framework, in combination, can tell a farmer the crop to plant, and the anticipated yield.

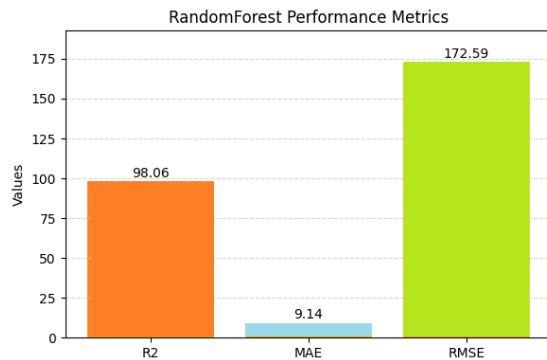


Figure 3: Random Forest Performance Metrics (Yield Prediction)

V. CONCLUSION

This paper developed a new AI based smart farming framework employing artificial neural networks (ANN) for crop recommendation and random forest regression for crop yield forecasting. Using the system farmers obtain data driven recommendations based on soil and environmental variables such as nitrogen, potassium, and phosphorus, as well as temperature, humidity, pH, and precipitation. Dimensions such as, data preprocessing, feature extraction, and attribute selection optimize the models training resulting in augmented predictive capability. An ANN model was able to reliably recommend crops demonstrated with an accuracy 95.83% and with a precision, recall and F1 scores above 85% and an associated classification performance. The Random Forest model attained a crop yield predictive capability of $R^2 = 0.9806$ with an acceptable predictive error of Mean Absolute Error (MAE = 9.14 and Root Mean Square Error (RMSE = 172.59). The framework enables actionable insights to be given to farmers for the determination of the applicable crop and anticipated

yield based on the environmental conditions and parameters. This model enhances planning, and preservation of resources, and promotes sustainable agriculture.

REFERENCES

- [1] Mr. Telise Vinod,N. Chandrasena Reddy,Rapolu Suresh Reddy,Anugu Athrey Reddy." Crop Yield Prediction using Machine Learning" International Journal of Engineering Research & Technology (IJERT) IJERTV12IS040077 Vol. 12 Issue 04, April-2023.
- [2] Shafiulla Shariff,Shwetha R B,Ramya O G,Pushpa H,Pooja K." Crop Recommendation using Machine Learning Techniques" International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 ICEI – 2022 Conference Proceedings Special Issue – 2022
- [3] Prameya R Hegde, Ashok Kumar A R . "Crop Yield and Price Prediction System for Agriculture Application" International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 IJERTV11IS070060 Vol. 11 Issue 07, July-2022
- [4] Bali, Nishu, and Anshu Singla." Emerging trends in machine learning to predict crop yield and study its influential factors: A survey." Archives of computational methods in engineering (2022): 1-18.
- [5] Gunjan, Vinit Kumar, et al." Prediction of agriculture yields using machine learning algorithms." Pro- ceedings of the 2nd International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications: ICMISC 2021. Springer Singapore, 2022.
- [6] Reddy, D. Jayanarayana, and M. Rudra Kumar. "Crop yield prediction using machine learning algo- rithm." 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2021.

- [7] Agarwal, Sonal, and Sandhya Tarar. "A hybrid approach for crop yield prediction using machine learning and deep learning algorithms." *Journal of Physics: Conference Series*. Vol. 1714. No. 1. IOP Publishing, 2021.
- [8] Morales, Alejandro, and Francisco J. Villalobos. "Using machine learning for crop yield prediction in the past or the future." *Frontiers in Plant Science* 14 (2023): 1128388.
- [9] Joshua, S. Vinson, et al. "Crop yield prediction using machine learning approaches on a wide spectrum." *Computers, Materials & Continua* 72.3 (2022).
- [10] Senapaty, Murali Krishna, Abhishek Ray, and Neelamadhab Padhy. "IoT-enabled soil nutrient analysis and crop recommendation model for precision agriculture." *Computers* 12.3 (2023): 61.