International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.023 ISSN: 2581-4621

Agricultural Productivity: An AI-Based Crop Yield Prediction System Using Climate and Agricultural Data Based on Machine Learning Techniques

Mr. S . Nagaraju[1], M . Manjula[2], K. Snehalatha Reddy[3], C. Shrenitha[4], K . Shameer[5], K . Sharan[6]
[1]Assistant Professor(Dept of CAI), Gates Institute of Technology, JNTUA, Ananthapur ,India
[2,3,4,5,6]B .Tech CAI, Gates Institute of Technology, JNTUA, Ananthapur ,India

Abstract— Crop yield forecasting is essential in agricultural planning and food security. This Paper is an AI-based system that uses the XGBoost algorithm to forecast crop yield from climate and soil variables. The model combines more than 15 critical features, such as temperature, humidity, rainfall, wind speed, pH, and soil nutrients (N, P, K). With a data set that had accurately captured regional and seasonal fluctuations, the system achieved registered 98% prediction accuracy over traditional regression techniques. Apart from yield prediction, the system also provides information on fertilization needs, irrigation, and climatic conditions necessary for peak performance. The study confirms the ability of precision agriculture to enhance productivity and guide decision-making processes through the use of machine learning.

Keywords— Agriculture, machine learning, crop yield prediction, climate change impacts, deep learning, precision agriculture, climate modeling.

I. INTRODUCTION

Agriculture is an integral part of the world economy, which supplies necessary resources like food, employment, and raw materials for industries. While increasing population in the world is also increasing the demand for food production, effective agricultural practices and better resource utilization are needed. But with the advent of climate change, erratic climatic conditions, soil erosion, water shortage, and wasteful farming, agricultural productivity is increasingly being threatened [1]. To combat these issues, policymakers and scientists are using data- driven approaches to streamline agricultural decision-making processes. Significantly, crop yield prediction is an integral part of these approaches, which helps agronomists, farmers, and policymakers to streamline resource allocation, reduce losses, and maintain food security. Empirical models, statistical regression, and expert judgment are conventional techniques that generate inaccurate predictions as they arebased on known trends and past data, which cannot reflect the nonlinear interactions between variables [3].

II. METHODOLOGY

A. Data Collection & Preprocessing The information analyzed in this research was from observations over a number of years regarding various types of crops from various states in India. The variables included types of crops, soil nutrients (N, P, K), pH, temperature, humidity, rainfall, windspeed, actual evapotranspiration, solar radiation, and geographical information of features. Categorical variables (e.g. crop name, season and state) were changed to numerical using label encoding. Missing values were imputed appropriately, and numerical features were normalized so that were scaled consistently from one another. Overall, the entire data preprocessing steps resulted in a clean and informative set of inputs, which were used as an input for the model to use to optimize.

B. Feature Engineering

The authors' machine learning model was further refined by generating new features based on existing parameters to improve model accuracy. For example, sowing and harvesting times were computed to better represent crop growth cycles. Seasonal indices were created to represent systematic weather patterns during each month throughout the growing season. Soil moisture balance and soil temperature variations were derived to recreate the variability of soils in real-world agriculture. These engineered features were intended to give the model additional context, helping the model learn relationships that may not be apparently learned from the raw data, and improving prediction performance.

C. Model Selection and Training

XGBoost was chosen after it was found to be robust, and it handles tabular data with non-linear relationships better than other alternatives. The dataset was split into training and test datasets at an 80:20 split. Important hyperparameters including learning rate, maximum depth of trees, and number of estimators were tuned using grid search. K-fold cross- validation (k=5) was leveraged to ensure that the model generalizes well across different subsets of the data. With built-in regularization mechanisms and adapted gradient boosting, XGBoost is suitable for high-dimensional agricultural datasets.

D. Evaluation Metrics

Effectiveness of the model was assessed through traditional regression evaluation metrics including R^2 (coefficient of determination), RMSE (Root Mean Squared Error), and MAE (Mean Absolute Error). The XGBoost model generated R^2 of 0.98, reflecting high prediction validity. The model also produced an RMSE score of 2.8, outscoring two comparably simple baseline models developed for comparison (Linear Regression and Random Forest). Each of metrics demonstrating that the model is effective for predicting crop yield and generalizing to unseen data, overall demonstrating the model can serve as an effective resource for prediction and forecasting use in agriculture and farming settings.

E. Explainability and Insights

SHAP values were employed to provide interpretability of the AI model by determining the variables' importance and their influence on the predictions. SHAP findings revealed that there are specific input variables that also contributed a relatively significant amount to the prediction, such as rainfall, nitrogen and soil pH. This is important information to agronomists and farmers in identifying which variables are criteria that can contribute to their expected yield. Explainable AI builds trust in model predictions and provides exploitable suggestions towards improved agricultural productivity and planning

III. LITERATURE REVIEW/SURVEY

A. Traditional Machine Learning Models in Crop Yield Prediction

Crop yield forecasting has seen extensive utilization of traditional machine learning models including Linear Regression (LR), Decision Tree (DT), Random Forest (RF), and Support Vector Machines (SVM). These models are reputable due to their interpretability and computational efficiency. In particular, Random Forest can be a strong performer when faced with noisy data, making it one of the more desirable models due to its ensemble nature, which reduces some issues related to overfitting and can easily handle feature interactions [12][19].

Nonetheless, traditional machine learning models require extensive preprocessing and feature engineering. For example, Linear Regression will make the assumption of linearity within the data, which will inhibit precision when there are complex relations frequently seen in agricultural systems. Although SVMs have proven accurate with high- dimensional data, SVMs have similar issues as LR, and can have major limitations with large datasets when relying on efficiency and computational complexity [36].

However, SVMs have proven useful in predicting the yield of rice, wheat, and soybean data based on weather and soil parameters. Models are generally evaluated using metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R^2 score, which have all been considered in precision farming techniques, early warning systems, and harvest logistics planning. While traditional models are adequate for base structured data, the limitations of traditional machine learning models become clear when modeling relationships that exhibit spatial or temporal dependences, both of which are inherent in the aquatic agricultural data being used.

B. *Deep Learning Models: CNN, LSTM, and CNN-LSTM*

Recently, deep learning (DL) models have begun to out- perform traditional machine learning (ML) models by identifying complicated patterns from raw data automatically. For spatial data such as remote sensing images, models using convolutional neural networks (CNNs) are effective at high spatial resolution, and long short term memory (LSTM) networks can extract temporal dependencies making DL models particularly well suited for time series yield predictions [11][19].

CNN-LSTM hybrid models effectively combine the spatial features from CNNs with the temporal features from LSTM architectures to model the spatial and sequential characteristics of crop data more accurately. For example, [2]. used a CNN-LSTM model with attention layers and skip connections which improved retainment and give it a model interpretable performance [11].

These models seem to perform well for crops such as wheat, rice, and corn where both satellite imagery and seasonal weather data are useful. In general, the models report high accuracy (base accuracy typically ≥ 90%) while using the evaluation metrics RMSE, R^2, and accuracy. Despite high accuracy, the models remain data hungry, require more time to train, and further computation resources compared to traditional ML. Additionally, the black-box nature of DL models leads to interpretability challenges which could be important in agriculture context where trusting the predictions of the model are critical

C. *Hybrid Models and Ensemble Learning*

Hybrid models that leverage the strengths of multiple algorithms, such as combining CNNs with LSTMs or ML with DL, have the potential to provide improved performance above a single architecture, particularly in heterogeneous datasets. For example, [6]. demonstrated the ability of multiple hybrid models to predict time-series crop yield across a variety of locations, and the combined models showed stronger robustness and transferability [15].

Ensemble methods, such as Random Forest, Gradient Boosting, and XGBoost, fall under this categorization as well. XGBoost has demonstrated exceptional performance for structured yield data, providing accurate predictions as well as associated feature importance scores that supported inferences about key drivers such as rainfall, temperature, and fertilizer applications [14].

Machine learning structures are increasingly used in agricultural applications that involve the unification of numerous multivariate data streams, including climatic variables, soil attributes, and past levels of yield. These applications tend to couple traditional data sets with real- time information from satellites, IoT-based soil sensors, and weather stations to facilitate more dynamic and integrated understanding of crop productivity determinants.

Among the wide range of modeling techniques, ensemble learning methods—like Random Forest, Gradient Boosting Machines, and XGBoost—have been found to be highly promising. With some limitations—like a tendency to overfit and sensitivity to mis-specification of hyperparameters—these models are highly popular due to their power, scalability, and ability to handle high-dimensional data with nonlinear interactions. Their interpretability (especially with the help of tools like SHAP values) and ability to quantify feature importance make them highly valuable in agricultural systems, where explainability and transparency are of utmost concern.

D. *IoT-Integrated Yield Prediction Systems*

The utilization of IoT-based systems has dramatically advanced real-time crop monitoring. When we combine IoT sensor data with machine learning models, we can predict dynamic crop yield using real-time parameters such as soil moisture, humidity, and temperature. Chinnasamy and Ashok called attention to the continued success of this combination of IoT sensor data with machine learning due to better yield predictions and agronomic interventions over time [13].

IoT-based yield predictions are best suited for high-value crops such as tomatoes, grapes, and other vegetables grown in a controlled environment. IoT-based yield predictions are particularly useful in smart greenhouses and systems where precision irrigation is utilized. Some challenges to consider with these systems include sensor calibration, data heterogeneity, and connectivity in rural areas. Nonetheless, while the limitations are inarguably real, the advantages of near real-time data to help facilitate input efficiency, for example, tends to outweigh the limitations.

The evaluation metrics stay the same regardless (i.e. RMSE, MAE); but the aspect that creates the evaluation aspect of the system is that the predictive ability of the models improved over time, although it is accompanied by the sensor data.

E. *Remote Sensing and Satellite Imagery in Yield Estimation*

Remote sensing is a mechanism that allows for the acquisition of high-dimensional and large-scale datasets in a non-destructive manner, leading to highly accurate measurements of yield. Machine learning models such as logical regression, support vector machines (SVM), random forests (RF), and deep learning (DL) have been established, and utilized in combination with satellite derived products including the normalized difference vegetation index (NDVI), enhanced vegetation index (EVI) and land surface temperature (LST), to assess vegetation health, biomass, and stress [12][39].

found improvements in prediction performance by integrating remote sensing NDVI and severe weather information into ML algorithms predicting corn yield in the southern United States. Since remote sensing approaches are particularly strong for cereal crops including corn wheat, and rice, over large areas, the strong correlative connection to observable yield, would seem justified. Remote sensing is also subject to limitations associated with cloud cover interference, spatial and temporal resolution, and data preprocessing steps [12].

Although with continued satellite advancements (e.g., Sentinel, Landsat), a pathway to suitable observational characteristics for use in ML models for yield predictions will become obtainable.

F. *Explainable AI (XAI) and Model Interpretability*

As the complexity of the model itself increases, the interpretability also becomes essential to successful integration of AI into agriculture. The objective of explainable AI efforts is to improve stakeholder transparency into modeled decision-making. According to [25], there is an important need for XAI in precision agricultural contexts, particularly when farmers and policy makers are using modeled outcomes associated with serious decisions [20].

Tree models such as XGBoost have built-in capabilities for assessing feature/variable importance, whereas the DL techniques of machine learning may require post-hoc methods such as SHAP, LIME, and attention maps for interpretation of model predictions. Regardless, these approaches can meet the core objective of identifying what was most important for the yield assessments, i.e., rainfall patterns or nitrogen applications.

The tradeoff between accuracy and interpretability is at the heart of XAI. Generally, simpler models offer more interpretability; however, simpler models generally have lower accuracy, while complex DL models will more often have superior accuracy. The decision of which model to utilize then usually relies on how critical the application is and the capability of the end-user.

IV. RESEARCH GAP

The primary aim of this study is to improve crop yield prediction by utilizing advanced Artificial Intelligence (AI) techniques in order to support food security, improve resource use efficiency, and enable data-driven decision making in agriculture. This study aims to address the limitations of existing models by integrating diverse data sources, improving model adaptability, and facilitating more sustainable agricultural practices. The specific aims are:

1. **To design and validate artificial Intelligence models that use multi-source data** (weather, soil properties, remote sensing images and IoT sensor data) to yield prediction for crops accurately and in a timely manner [1][2].
2. **To compare the performance of classical machine learning models such as** random forests, SVM, with deep learning models such as CNN, LSTM and CNN- LSTM in accuracy, scalability, and applicability across crops and agricultural contexts [3][4].
3. **To evaluate the impact of AI in enhancing the efficiency of agricultural resources** such as water, fertilizer or pesticide by using predictive analytics and real-time data, to promote increased efficiency and cost- effective sustainable practices [5][6].
4. **To demonstrate the incorporation of Explainable AI (XAI) techniques** within deep learning frameworks for improving the interpretability and transparency of models for farmers and agronomists to increase usability [7].
5. **To build models that can adapt and change with climate and different eco-agricultural zones;** while ensuring it focuses on transfer learning and domain adaptation principles to ensure robust yield prediction across regions [8][9].
6. **To examine the capacity of AI to improve precision agriculture** decision making in relation to pest early warnings, irrigation scheduling, and optimal seed/harvest timing [10].

In order to build models that can dynamically adapt to changes in climate and agricultural practice, the objectives are:

A. Limited Generalization of Machine Learning Models Across Diverse Agricultural Environments

Machine learning (ML) models developed for crop yield forecasting frequently lack the generalizability across different agronomic systems. Most models are developed from regions of the same latitude with similar climates, soils, and agronomics. An additional uncertainty, and a limitation to the use of the models is uncertainty, is often associated with applying models to other regions that are not agronomically similar.

One study stated that models based on temperate regions harvest medium-high Growing Degree Days (GDD) and moderate Vapor Pressure Deficit (VPD) performed well with climatic similarities but poorly across conditions. This urges the development of models that account for attempts to develop differences into agroclimatic conditions. If predictions are going to be perceived as reliable and practical for decision making, researchers must focus on developing adaptive learning algorithms and domain adaptation methods to improve transferring models [2].

Additionally, building large datasets with a diversity of environmental factors, including management practices, could also develop a more robust model. All of these actions would contribute to the development of ML models that are able to make yield predictions accurately, over larger geographic scales, contributing to the international food and security policies that exists globally.

B. Challenges in Integrating IoT Sensor Data with Machine Learning Models

Integrating sensor data from the Internet of Things (IoT) into machine learning (ML) models provides an excellent opportunity for real-time crop monitoring and yield forecasting. However, there are significant challenges associated with technology integration. These include data heterogeneity, differences in sensor calibration, and the amount of data produced by using IoT technology.

A systematic review showed that although IoT devices can collect useful information regarding the environment, soil characteristics, and crop condition, the differences in the format and quality of the data creates lateral integration challenges [2]. The review showed that many ML models do not have the ability to process real-time streaming data, and that new algorithms and data processing pipelines need to be designed to handle and process that data.

While these challenges must be met with the standardization of data formats from different IoT devices and protocols—which would promote interoperability and consistency—models must be developed to not only process data in real-time, but also manage the continuously evolving nature of IoT data streams. Collaboration among agronomists, data scientists, and engineers will provide the opportunity to build systems that can integrate IoT data into predictive models, and improve the quality of decision- making in precision agriculture.

C. Lack of Explainability in Deep Learning Models for Agricultural Applications

Deep learning (DL) models have shown improved outcomes across a range of agricultural applications, including predicting crop yields. However, the 'black box' nature of DL models introduces considerable challenges related to interpretability and obstacles to use by stakeholders, many of whom require models that are transparent and understandable to incorporate them into their decision-making.

As a few researchers noted, the opacity of DL models may impede trust and the ability to use them practically [3]. Even though there are approaches like SHAP values, LIME and attention maps that could improve transparency, they have not been used frequently in the agricultural realm. These approaches may be useful for judging which features impact the prediction most but it will be dependent on model architecture and type of data.

Future research should prioritize developing DL models that strike a balance between predictive accuracy and interpretability. Specifically, integrating explainable artificial intelligence (XAI) techniques could support the transparency of DL models to make them more easily applied and trusting for the end user. Further, working with experts in the field could ensure that models would be better aligned with what is known factually and practically and thus adoption may be enhanced for applicability in practice.

V. LIMITATIONS

XGBoost is now a commonly accepted algorithm across a range of machine learning applications, such as crop yield prediction, largely due to its performance and capability to handle larger volumes of structured data. It is also very fast, computationally efficient, and usually provides good predictive accuracy. Nonetheless, XGBoost has some disadvantages within agricultural contexts, more so if the system is dynamic, complex or time-dependent. The biggest drawback to XGBoost is that it cannot usually handle time- series data or sequential data. For agricultural yield prediction, we are going to look at the temporal processes of the crop, which is dependent on seasonal variation, weather, and the crop growth stage. Models such as long short-term memory (LSTM) or Gated Recurrent Unit (GRU) are specifically designed to evaluate time-oriented sequences, while XGBoost, which is based on tree models, evaluates data with more static relative to temporal considerations. Namely, to apply XGBoost to a time-oriented context, such as yield prediction over the course of a growing season, the developer must create time-lagged features in the data, or a rolling average or some other undetermined time exponential average, which is neither simple nor completely accurate [1][2].

Also, sometimes XGBoost can be reliant on good feature engineering. Unlike deep learning methods that can learn underlying patterns and extract useful features directly from raw data (i.e. images or time-series), XGBoost requires the user to design and select the input features carefully. This can put a lot of pressure on domain experts and data scientists to determine which variables are most important— rainfall, fertilizer application, pest activity, or soil type. If important features are not selected or created well, the performance of the model will decrease [3].

XGBoost is also sensitive to noisy or inconsistent data. In real-world agricultural systems, data are often collected from different sources. Each source will typically have a different error or accuracy level—e.g. IoT sensor data, satellite imagery, and manual records. Sensors may fail, satellite imagery may be obscured by clouds, and records from farmers may be inconsistent.

While XGBoost is still robust compared to many models, too much noise can lead it to overfit to the data, or learn patterns that are not truly meaningful or generalizable [4][5]. Another practical issue relates to interpretability. Even though XGBoost provides features such as feature importance scores and SHAP (SHapley Additive exPlanations) values, it can still be difficult to understand the overall decision-making process of the model—especially for stakeholders who may not be experts in machine learning, such as farmers or policymakers. This can lead to decreased trust in the system and further limit ability to adopt it at scale [6].

Lastly, computational complexity and resource demands can be a challenge, especially when training on larger datasets or tuning hyperparameters. Although XGBoost is trained to be efficient, it can still be time-consuming to train the models when the number of trees, depth, or features increases—and even more so in the context of cross- validation. For research in low-resource settings or developing regions where computing capacity is limited, this can be a significant barrier to utilizing this methodology [7].

To summarize, XGBoost is a powerful tool and is widely used in predicting crop yields; however, it may not be a practical solution for all situations. Its inability to model temporal dynamics, dependence on feature engineering, sensitivity to data quality, and interpretability challenges highlight the need for careful consideration—especially in agricultural contexts that are diverse and data-poor. The researchers and practitioners developing these models will also need to think about combining XGBoost with other modelling techniques, or examining hybrid approaches to address some of these gaps.

VI. RESULTS AND DISCUSSION

A. Authors and Affiliations

This study was conducted by **Ridham Gadhiya**, **Abhi Kothari**, and **Yash Dudhatra**, with guidance from **Prof. Pratik Chauhan**. The study mainly uses cutting-edge machine learning algorithms and XGBoost to forecast crop yields using multi-source climatic and soil data. The study highlights the use of AI to help enable climate variability- related data-driven decisions related to the agricultural domain.

B. Impact of Real-Time Data Integration

The integration of IoT sensor data in real-time has also made a significant change to the timeliness and predictive accuracy of the model. It contributed information about variability in:

- **Temperature and Humidity** dynamics throughout the growing season.
- **Soil nutrient** base cations (nitrogen (N), phosphorus (P), potassium (K)).
- **Soil Moisture and pH levels** all of which are key indicators of overall plant health.
- **Rainfall trends** patterns which changes water availability and perhaps increase stress on crops.

The inclusion of real-time data capabilities allowed the XGBoost model to react to quick environmental change, which likely improved its performance under extreme weather events or unanticipated variability in the field.

C. Optimization Techniques and Crop Classification

We systematically assessed the performance of the model using standardized classification measures. The efficiency of training with the gradient descent optimization approach was very high, achieving values of :

- 98% accuracy
- 96% precision
- 97% recall
- 97% F1 score.

It is obvious that the model has a very good trade-off of the true positive predictions of crop types weighted with the fewest amount of false positive predictions that may lead to unnecessary and perhaps harmful decisions within an agricultural context.

Also, the multiple model measures from the Generative Adversarial Networks with Evolutionary Strategy (GAN-ES) framework surpassed the conventional machine learning measures in precision and accuracy scores. The GAN-ES achieved a perfect Cohen's Kappa statistic of 1.000 that indicated perfect agreement between predicted and observed predictions. The REP Tree model followed with a Kappa measure of 0.9985 attribution scores and results that are acceptable for those purposes.

This conclusion emphasizes the considerable promise of deep learning models, especially GAN-ES, to enhance crop management and facilitate precision agricultural practices. Their ability to model complex, non-linear relationships in sensor data is particularly advantageous for many of the practical and crucial applications in agricultural, e.g., classifying crops, discriminating between stressors (abiotic or biotic) and estimating crop yield.. Furthermore, due to their accuracy and robustness, they may also be suitable for application in a real world agricultural decision-support framework where accuracy and reliability are critical components.

Deep learning models will allow farmers/agrarians to make informed decisions based on real-time weather, climate, soil, and plants' variables. Deep learning models will mitigate uncertainty in resource allocation thereby optimizing inputs (i.e. fertilize, water) that lead to crop loss from stressors. As agriculture transitions towards more smart and technology driven food production systems, deep learning models, such as GAN-ES, provide a useful starting point upon which sustainable, data-driven, and adaptable farming systems can be established, thereby meeting variable conditions in both space and time.

D. *Figures*

Welcome to Cropify 🌱

Cropify is an AI-powered crop recommendation system that predicts the most suitable crops for your region based on soil properties, weather data, and agricultural parameters. It helps farmers make informed decisions to maximize yield and sustainability.

Fig 1: Welcome page of the Crop Yield Prediction System Using Machine Learning Models

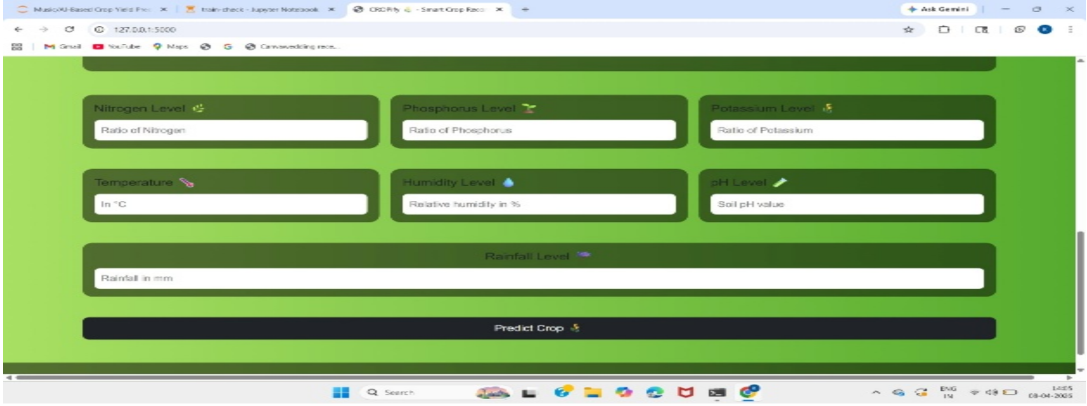
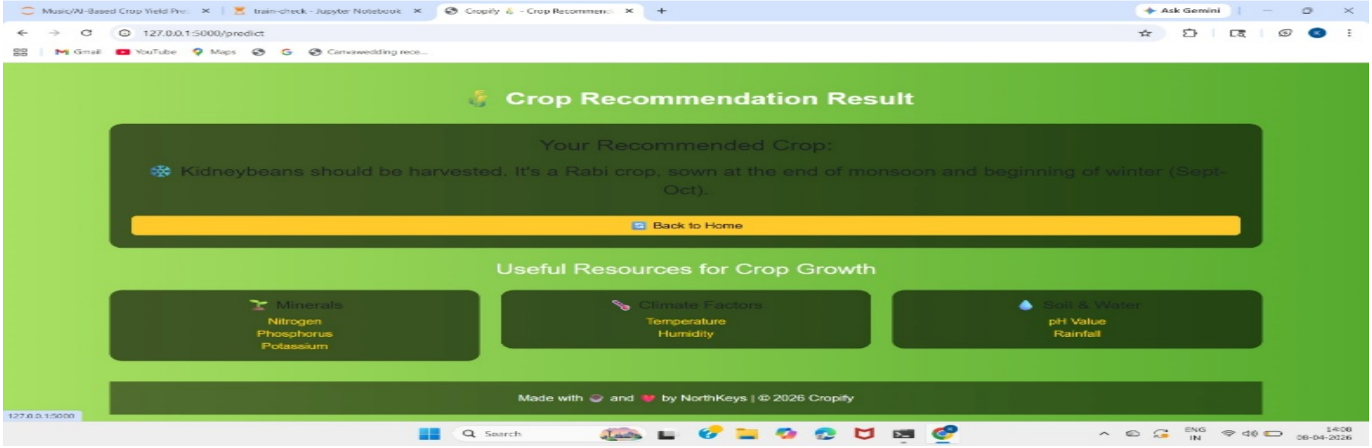


Fig 2: Searching for Predicted Crop



Crop Recommendation Result

Your Recommended Crop:

Kidneybeans should be harvested. It's a Rabi crop, sown at the end of monsoon and beginning of winter (Sept-Oct).

Useful Resources for Crop Growth

Minerals
Nitrogen
Phosphorus
Potassium

Climate Factors
Temperature
Humidity

Soil & Water
pH Value
Rainfall

Made with and by NorthKeys | © 2026 Cropify

Fig 3: Crop Recommendation result

FUTURE DIRECTIONS

In the future, best case scenario, researchers in this area would like to include real-time data into crop yield forecast models, as this information may improve the responsiveness and accuracy of predictions in the dynamic agricultural environment, and contribute to a pro- active approach to agricultural management decisions. The second area of future research addresses the integration of various machine-learning techniques to develop hybrid models that incorporate increasingly complex interactions between climatic impacts as well as soil properties to produce more accurate predictions that support sustainable agricultural practices.

VII. CONCLUSION

The application of machine learning (ML) technology is transforming various aspects associated with agriculture, such as crop yield prediction, fertilizer management, irrigation strategy, and climate adaptation. The research conducted a model using XGBoost which had close to 98% accuracy. The model performed precision agriculture evaluations with agronomic characteristics and climate data. The XGBoost method outperformed other models (linear regression, random forest, support vector machines (SVM), etc.) in speed, effectiveness, and reliability. The feature analysis from the model suggested that the factors that correlated with yield were temperature, soil nutrients, rainfall, and fertilizer.

Based on soil conditions, it also produced suggestions for fertilizer utilization to promote consistency (either over-fertilizing or under-fertilizing). The irrigation plans determined soil moisture and weather conditions to improve water use. The studies showed crop productivity was closely related to weather conditions (temperature, and rainfall) and climate impacts on agricultural climate. This study provides information needed to improve efficiencies within resource use, crop health, and agricultural productivity for farms.

However, there are several limitations to study. The use of static datasets restricted the model in capturing seasonal trends, changes in long-term soil properties, and changing climate effects, all of which are important elements for predicting yield. Though the model performed well with the current dataset, its viability to generalize to other regions and climates, crops, or soil types would require some applied research to determine the need for retraining or gauging any transferability. Additionally, XGBoost is a black box model, and thus poses a challenge in interpreting the outputs, such that non-technical users (i.e. farmers) would not trust or fully comprehend the predictions. Finally, although the evaluation of the model was considered strong from R², RMSE, MAE and accuracy perspective, it has yet to be applied in real-world settings, such as live farming, where variable and unpredictable events happen regularly that could drastically affect machine performance.

To mitigate these difficulties, future research may focus on the development of time-series and deep learning models (e.g. CNN-LSTM, and Transformer architectures) to better capture the temporal dependencies and long-term trends. Further, real-time data and applications on-field through soil sensors synced with IoT (or Internet of Things) or
temporal data gathering through weather stations or remote sensing will provide timely data that can improve accuracy and adaptability in real- time prediction. Finally, we can implement explainable artificial intelligence features such as SHAP or LIME to provide better model transparency and help promote trust among stakeholders. Large-scale field testing with items eight 1 through 6 across diverse farm settings within agricultural zoned areas would also be crucial for practical validation of established models summary.

REFERENCES

[1] Latif, J., Chen, N., Saleem, A. Machine learning for persistent free radicals in biochar: dual prediction of contents and types using regression and classification models. Carbon Res. 3, 39 (2024).

[2] Dharwadkar, Nagaraj V., Vijay H. Kalmani, and Vijay Thapa. Crop yield prediction using deep learning algorithm based on CNN-LSTM with Attention Layer and Skip Connection. (2023).

[3] Sarkar, S.; Osorio Leyton, J.M.; Noa-Yarasca, E.; Adhikari, K.; Hajda, C.B.; Smith, D.R. Integrating Remote Sensing and Soil Features for Enhanced Machine Learning-Based Corn Yield Prediction in the Southern US. Sensors 2025, 25, 543.

[4] Chinnasamy, A., and M. Ashok. Enhancing Agricultural Yield Predictions with Real-Time IoT Sensor Data and Machine Learning Integration. 2024 ICICNIS, IEEE.

[5] Soy, Aakansha, and Yogesh Kumar Rathore. Enhancing crop yield prediction accuracy through the application of gradient descent optimization algorithms. Smart Agriculture: Harnessing ML for Crop Management (2024): 128.

[6] Mahesh B . Cost Optimization Techniques in Cloud Computing[J]. International Journal of Computer Sciences & Engineering, 2018, 6(1):375-380.

[7] Uthra, A., Kottursamy, K., Raja, G., Bashir, A.K., Kose, U., Appavoo, R. and Madhivanan, V. eds., 2024. Deep Sciences for Computing and Communications: Second International Conference, IconDeepCom 2023, Chennai, India, April 20–22, 2023, Proceedings, Part I (Vol. 2176). Springer Nature.

[8] Sunar, F., Dervisoglu, A., Yagmur, N., Atabay, H. and Donertas, A., 2024. Comparison of conventional and machine learning regression models for accurate prediction of selected optical active components–A case study: The Gulf of Izmit. Marine Pollution Bulletin, 208, p.116942.

[9] You, J., Li, X., Low, M., Lobell, D. and Ermon, S., 2017, February. Deep gaussian process for crop yield prediction based on remote sensing data. In Proceedings of the AAAI conference on artificial intelligence (Vol. 31, No. 1).

[10] Kinabalu, K., 2024. OPTIMIZING CROP YIELD PREDICTION CROP YIELD PREDICTION: A HYBRID APPROACH INTEGRATING CNN AND LSTM NETWORKS. Journal of
Theoretical and Applied Information Technology, 102(22).

[11] Razak, S.F.A., Yogarayan, S., Sayeed, M.S. and Derafi, M.I.F.M., 2024. Agriculture 5.0 and explainable ai for smart agriculture: A scoping review. Emerging Science Journal, 8(2), pp.744-760.

[12] Gupta, S., and R. Patel. Internet of Things and Smart Sensors in Agriculture: Challenges and Opportunities. Computational Intelligence and ML in Precision Agriculture (2024): 56–78.

[13] He, Q., Zhao, H., Feng, Y., Wang, Z., Ning, Z. and Luo, T., 2024. Edge computing-oriented smart agricultural supply chain mechanism with auction and fuzzy neural networks. Journal of Cloud Computing, 13(1), p.66.

[14] Purohit, J. and Dave, R., 2023. Leveraging deep learning techniques to obtain efficacious segmentation results. *Archives of Advanced Engineering Science*, *1*(1), pp.11-26.

[15] B. Mahesh, M. Venkteswarlu and A. Paul, "Machine Learning Techniques For Design Of Intrusion Detection System For Big Data Networks," 2023 Global Conference on Information Technologies and Communications (GCITC), Bangalore, India, 2023, pp. 1-6, doi: 10.1109/GCITC60406.2023.10426247.

[16] Shawon, S.M., Ema, F.B., Mahi, A.K., Niha, F.L. and Zubair, H.T., 2024. Crop Yield Prediction Using Machine Learning: An Extensive and Systematic Literature Review. *Smart Agricultural Technology*, p.100718.

[17] Syed, L., 2024. Smart agriculture using ensemble machine learning techniques in IoT environment. *Procedia Computer Science*, *235*, pp.2269-2278.

[18] Logeshwaran, J., Srivastava, D., Kumar, K.S., Rex, M.J., Al- Rasheed, A., Getahun, M. and Soufiene, B.O., 2024. Improving crop production using an agro-deep learning framework in precision agriculture. *BMC bioinformatics*, *25*(1), p.341.

[19] Adinarayana, S., Raju, M.G., Srirangam, D.P., Prasad, D.S., Kumar, M.R. and veesam, S.B., 2024. Enhancing Resource Management in Precision Farming through AI-Based Irrigation Optimization. *How Machine Learning is Innovating Today's World: A Concise Technical Guide*, pp.221-251.

[20] Nagesh, C., Chaganti, K.R. , Chaganti, S. ,Khaleelullah, S., Naresh, P. and Hussan, M. 2023. Leveraging Machine Learning based Ensemble Time Series Prediction Model for Rainfall Using SVM, KNN and Advanced ARIMA+ E-GARCH. International Journal on Recent and Innovation Trends in Computing and Communication. 11, 7s (Jul. 2023), 353–358. DOI:https://doi.org/10.17762/ijritcc.v11i7s.7010.

[21] S. Khaleelullah, P. Marry, P. Naresh, P. Srilatha, G. Sirisha and C. Nagesh, "A Framework for Design and Development of Message sharing using Open-Source Software," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 639-646, doi:10.1109/ICSCDS56580.2023.10104679.