Indian Journal of Computer Science and Technology
Special Issue, Volume 5, Issue 1 (January-April 2026), PP: 135-139.
https://www.doi.org/10.59256/indjcst.20260501C022
International Conference on Cyber and AI Security (ICCAIS-2026)
www.indjcst.com
ISSN No: 2583-5300

IoT-Enabled Crop Yield Prediction Using Machine Learning Techniques

M. Sandhya Devi[1], Sheela Petta[2], K. Divya Harshitha[3], S. Sairam Teja[4], K. Nikhil Satya Devi Santosh[5]

[1,3,4,5] Department of Artificial Intelligence and Machine Learning, Sasi Institute of Technology and Engineering Tadepalligudem, Andhra Pradesh, India.

[2]Associate Professor, Department of Artificial Intelligence and Machine Learning, Sasi Institute of Technology and Engineering Tadepalligudem, Andhra Pradesh, India.

To Cite this Article: *M. Sandhya Devi[1], Sheela Petta[2], K. Divya Harshitha[3], S. Sairam Teja[4], K. Nikhil Satya Devi Santosh[5], "IoT-Enabled Crop Yield Prediction Using Machine Learning Techniques, Indian Journal of Computer Science and Technology Volume 05, Issue 01 (January-April 2026), PP: 135-139.*

Abstract: *Crop diseases are one of the major problems in agriculture. It has a significant impact on the productivity of the crops as well as the lives of the farmers. Detection of diseases in plants at an early stage is important in order to avoid damage to the crops and enhance the productivity of agriculture. However, the traditional method of plant disease detection is based on observation, which is a tedious process and might result in incorrect identification of diseases due to a lack of knowledge. This paper suggests an efficient method of plant disease detection using images and the EfficientNet-B0 and ViT .The EfficientNet-B0 model is a Convolutional Neural Network model that is used in the extraction of important features in images related to plant leaves, such as color changes, textures, and diseases, whereas the Vision Transformer model is used in the identification of the relationships between different areas in the images using a self-attention mechanism in order to ensure better accuracy in the classification of the images. The model is trained and validated using different datasets, including PlantVillage, PlantDoc, Vegetable images, etc. Depending upon the disease, fertilizer suggestions will also be provided with images, usage guidelines, dosage, and precautions for proper usage of the fertilizer by the farmers. Moreover, a multilingual voice assistant is also integrated into the system for providing information in English, Hindi, and Telugu, thus making it more user-friendly for farmers models.*

Key Words: *Crop Disease Detection, Deep Learning, CNN, Vision Transformer (ViT), EfficientNet-B0, Image Classification, Fertilizer Recommendation System, Multilingual Voice Assistant.*

I. INTRODUCTION

The role played by agriculture in delivering food and con- tributing to the economy cannot be overemphasized. However, there are many challenges that affect modern farms, including issues resulting from climate change, poor quality soil, absence of water, and fluctuating market prices. Making the right choices concerning what to plant, the amount of water and fertilizer used, and when to harvest is essential. Conventional methods of determining crop yield are typically dependent on past records and/or making calculations. The problem is that making calculations fails to account for the intricate connection that affects crops, hence an inaccurate result.

With the advent of machine learning and artificial intelli- gence, it has become more efficient to predict crop yield. These technologies are able to analyze large amounts of data and recognize hidden patterns within them. This helps in predicting crop yield with different conditions in the environment. The use of IoT devices has also become an added advantage in improving agriculture, as they assist in collecting various parameters such as soil moisture, temperature, and humidity. With the advent of machine learning and artificial intelli- gence, it has become more efficient to predict crop yield. These technologies are able to analyze large amounts of data and recognize hidden patterns within them. This helps in predicting crop yield with different conditions in the environment. The use of IoT devices has also become an added advantage in improving agriculture, as they assist in collecting various parameters such as soil moisture, temperature, and humidity.

II.LITERATURE REVIEW

Research carried out in recent years has also shown that there is an increasing trend in employing machine learning and deep learning-based models for crop yield prediction[4][12]. Various review papers have considered different models of crop prediction and found that models such as neural net- works, ensemble methods, and deep learning methods per- form better compared to traditional models such as statistical models[8][26]. These models perform particularly well in dealing with complexities related to crop growth and weather conditions such as climate, soil, and the like[13][23]. Factors related to weather conditions, soil types, and the like are mostly considered as predictive features in most models [13]. Hybrid models and feature selection models have also been suggested as effective models to ensure the reliability of crop yield prediction models [26].

Additional empirical research performed in different re- gions has further validated the strength of the models. In different areas of India, scientists applied different regression model types to test their strength in accurately foretelling the yields of different agricultural products such as rice, cotton, sugarcane, sorghum, etc. [7]. The ensemble model reported higher accuracy with lower prediction errors, indicating their suitability [17]. Interestingly, similar results were reported in the state of Rajasthan, whereby tree models performed better compared to linear ones, indicating their capacity to effectively interpret weather and agricultural factors [10].

Previous research undertaken in African countries has also revealed varying uses of artificial intelligence technology in agriculture [22]. Research on potato yield forecasting using deep learning techniques revealed better accuracy than con- ventional yield forecasting techniques [21]. In Rwanda, data mining techniques were used to forecast yields of maize and potato using past climate conditions, enabling better decision- making by farmers and planners [22].

More advanced methods have developed the combination of multiple methods with the aim of overcoming some of the data limitations: multitask deep learning networks, hybrid systems integrating crop simulation models, and machine learning algorithms showed significant improvements in prediction performance [26].

From the literature review, it can be summarized that ensemble and hybrid AI models have strong potentialities for efficient crop yield prediction. However, most of the related works only focus on meteorological variables and barely consider either economic or market factors. This gap suggests that more holistic prediction systems are necessary for integrating diverse data sources for practical agricultural decision-making.

III.SEARCH STRATEGY/SELECTION CRITERIA

A structured and transparent approach to the search method- ology was employed to identify relevant and high-quality research studies pertaining to the prediction of crop yield using machine learning and artificial intelligence techniques. The motive of this research is to build a strong base of research studies on crop prediction, precision agriculture, deep learning applications, intelligent farming using IoT technology, and the integration of environmental and economic factors into agricultural systems[4][12].

A. Data Sources

Material collections of research articles were sourced from reputable academic databases such as IEEE Xplore, Springer-Link, ScienceDirect, ACM Digital Library, MDPI, and Google Scholar. Since these repositories publish peer-reviewed jour- nals and conference proceedings in areas related to artificial intelligence, agricultural engineering, environmental modeling, and data analytics, they have been selected for the purpose of this review [8] [26]. Domain-specific journals related to precision agriculture and remote sensing were also reviewed in order to get research perspectives across disciplines.

B. Keyword Strategy

The search query was built using a combination of relevant keywords such as ”crop yield prediction,” ”machine learning in agriculture,” ”deep learning for crop forecasting,” ”IoT in smart farming,” and ”climate-based yield modeling,” among others, which represent commonly explored keywords in re- cent studies on AI for agriculture[12, 15]. Boolean techniques have been utilized for narrowing the search query to enhance relevance.

C. Time Frame and Screening

The major focus of the review was based on studies published between 2017 and 2024, owing to the rapid ad- vancements witnessed between these years regarding AI-based agricultural systems [27]. Where possible, older foundational studies used to establish theory were included. Following this, the review screened the titles and abstracts to exclude irrelevant or duplicate publications.

D. Inclusion and Exclusion Criteria

Included were scientific articles where machine learning, deep learning, and hybrid models were used for crop yield forecasting, and there were quantitative performance mea- sures[10][17]. Focus on environmental conditions such as climate, environmental, and soil data, along with IoT, were em- phasized[19][23]. Articles that lacked experimental validation, were focused on unrelated domains of agricultural sciences, or appeared in non-peer-reviewed journals were excluded. This filtered selection of articles ensured that only those that were technically viable were considered for final analysis.

IV.PROPOSED METHODOLOGY

The system, titled ”CROP YIELD AI,” shall forecast crop yield by integrating environmental and economic data through a unified machine learning model. The intelligent system shall rely on multiple sources of data, such as weather-related parameters, including temperature, rainfall, and humidity. The parameters regarding soil health, such as nitrogen, phosphorus, potash, pH, and moisture, shall also be used for forecasting crop yield. Furthermore, the parameters regarding market conditions, such as the price of the crops, shall also be used for forecasting. The parameters regarding climate and soil health fall in line with a variety of prior study findings[13][23], while using parameters regarding IoT shall provide accuracy and precision[15][19]. Before the development of the model, the data collected is subjected to preprocessing, in which missing values are dealt with, outliers are removed, numerical data is

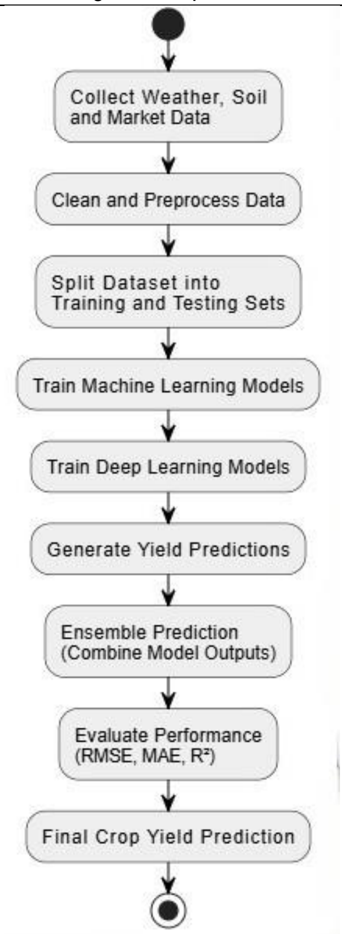

Fig. 1. Workflow of the CROP YIELD AI system

rescaled, and categorical data is encoded. Data preparation is important in order to improve the efficiency and consistency of the model [21]. A set of data prepared in this way is then divided into subsets, called training and test data. Several different predictive algorithms are supported, and these in- clude Random Forests [10], Gradient Boosting [14], CatBoost, and Deep Neural Networks [8]. These algorithms have been identified for their potential to deal with nonlinear interactions between climate, soil, and yields-related variables. To improve the robustness of the prediction models, ensemble learning is used, which combines the models for better performance [17]. The effectiveness of the model can be measured statistically through the application of various standard regression metrics, i.e., Mean Absolute Error, Root Mean Square Error, and the coefficient of determination, which are important factors in agricultural predictive research studies [10]. The output will provide information about the optimised crop estimate. The proposed solution provides a scalable and efficient solution for smart agricultural forecasting through the combination of multi-dimensional data sources and sophisticated machine learning methodologies.

V.RESULT AND DISCUSSION

The experimental analysis of the presented CROP YIELD AI framework proves the capabilities of machine learning, deep learning, and ensembling methods to achieve very accu- rate predictions regarding crop yield. The data was split into training and testing data using an 80:20 split ratio, and per- formance analysis was conducted using Mean Absolute Error, Root Mean Square Error, and Coefficient of Determination.

The Random Forest Regressor model did a good job with the agricultural data. It had an error of 0.46 when looking at the difference between predicted and actual values. The Random Forest Regressor also had an error of 0.63 when looking at the root of the average of the squared differences between predicted and actual values. The Random Forest Regressor was able to explain a lot of the variation in the data with an R^2 value of 0.86.

The Gradient Boosting Regressor model was even bet- ter than the Random Forest Regressor model. The Gradient Boosting Regressor had an error of 0.43 when looking at the difference between predicted and actual values. The Gradient Boosting Regressor had an error of 0.60 when looking at the root of the average of the squared differences between predicted and actual values. The Gradient Boosting Regressor was able to explain a lot of the variation in the data with an R^2 value of 0.88. This means the Gradient Boosting Regressor was really good, at predicting the data by keeping the errors very small. Furthermore, the CatBoost Regressor attains higher accuracy than these two predictors by recording an MAE of 0.41, RMSE of 0.57, and an R^2 of 0.90 by using efficient handling of mixed data types. Similarly, the Deep Neural Network attains the highest accuracy by recording an MAE of 0.39, RMSE of 0.55, and an R^2 of 0.91 by using predictive capabilities to learn complex nonlinear relationships between weather, soil, and market features.

The ensemble model that brings together the predictions of all the models was found to be the best-performing model overall. It had the least MAE of 0.35, the least RMSE of 0.50, and the highest R^2 value of 0.93, validating the variance reduction property of ensemble learning.

Further investigation showed how vital it was to take into consideration various data. The prediction models based only on weather data presented more errors. The MAE was 0.61, and R^2 was 0.79. Additional data related to soil parameters led to better results, and the highest accuracy was obtained using the data related to the market.

Although deep learning and ensemble algorithms were asso- ciated with a higher computational time, it is worth noting that the accuracy of these algorithms justifies their use. Overall, the results have shown that the proposed CROP YIELD AI solution is accurate, scalable, reliable, and ready for use in crop yield prediction.

VI.CONCLUSION

Otherwise, integration of artificial intelligence techniques has transformed conventional crop yield prediction techniques. Machine learning techniques, such as tree-based ensemble techniques including Random Forest and Gradient Boosting Trees, show promise in handling complex crop environment datasets. In addition, deep learning techniques, including ar- tificial neural networks, convolutional neural networks, and recurrent neural networks including LSTM, improve crop yield prediction using non-linear relationships, spatial, as well as temporal features. Similarly, IoT technologies improve crop conditions, based on which crop yields can be dynamically updated.

Model	MAE	RMSE	R^2
Random Forest Regressor	0.46	0.63	0.86
Gradient Boosting Regressor	0.43	0.60	0.88
CatBoost Regressor	0.41	0.57	0.90
Deep Neural Network (DNN)	0.39	0.55	0.91

Table I
Performance Comparison of Individual Models

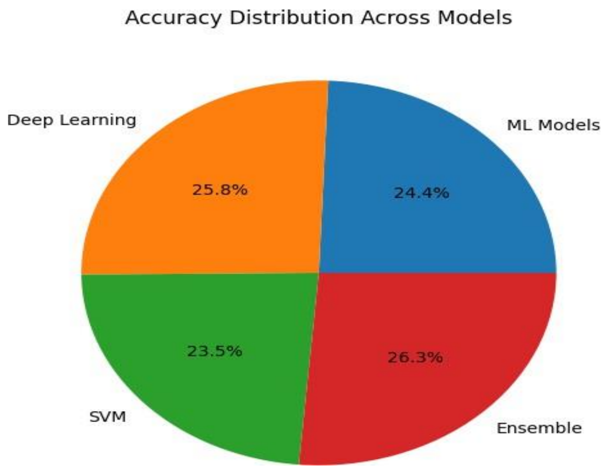

Fig. 2. Accuracy Distribution across models

Feature	Symbol	Description
Nitrogen	N	Ratio of nitrogen content in soil
Phosphorous	P	Ratio of phosphorous con- tent in soil
Potassium	K	Ratio of potassium content in soil
Temperature	–	Measured in degree Celsius (°C)
Humidity	–	Relative humidity in per- centage (%)
Soil pH	pH	Acidity or alkalinity level of soil
Rainfall	–	Rainfall measured in mil- limeters (mm)

Table II Input Features

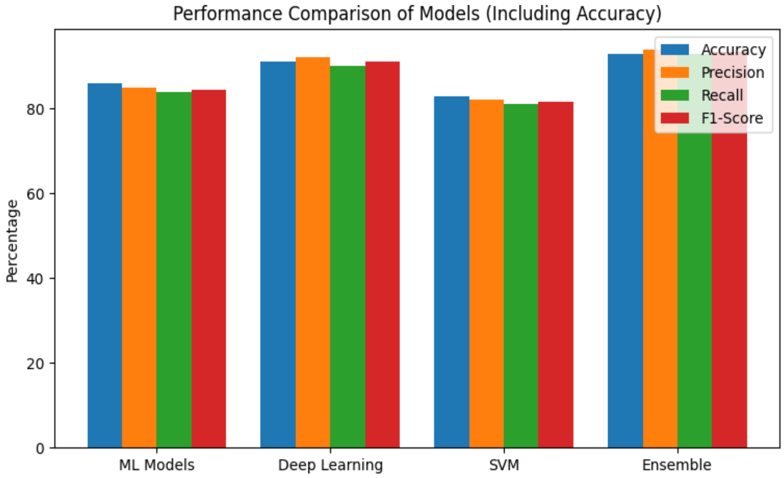

Fig. 3. Perfomance Comparison across models

However, studies have shown that current research primarily focuses on climatic and soil aspects, ignoring other aspects such as economic factors. This shows that the application of prediction systems for efficient agricultural decision-making is highly limited. Besides, other challenges such as poor quality of data, region-based models, computational complexities, and lack of transparency limit their application.

In Conclusion the research presented in the proposed CROP YIELD AI framework addresses these concerns through the incorporation of environmental, actual-time sensor data, and economic data in the hybrid ensemble. The aim is that this combined model will promote the development of highly effective, scalable, and economic crop yield predictability for rational agricultural decision-making. The research contributes to closing existing research gaps to promote more effective and sustainable development in intelligent agricultural systems. Improved adaptive models, AI, as well as the integration of data sources, will enhance the reliability of future crop prediction models.

REFERENCES

[1] J. Karthikeyan and A. Murugan, "Enhancing crop yield prediction using machine and deep learning models," *SEEJPH*, vol. 24, pp.

[2] A. Islam et al., "Crop yield prediction using IoT and ensemble machine learning," in *Proc. ICCIT*, 2023, pp. 1–6.

[3] P. Delfani et al., "Integrative approaches in agriculture using IoT and AI," *Precision Agriculture*, vol. 25, no. 4, pp. 2589

[4] V. Ramesh and P. Kumaresan, "Machine learning and deep learning for crop yield prediction: A review," *Nat. Env. Poll. Tech.*, vol. 23, no. 4, pp. 2071

[5] M. K. Siam, Sushil J. Jangirala, and S. Senthil Kumar, "IoT-based smart device for crop yield prediction,"

[6] F. M. Talaat, "Crop yield prediction algorithm using IoT and active learn- ing," *Neural Computing and Applications*, vol. 35, pp. 17281–17292

[7] J. Gupta *et al.*, "Machine learning approach for crop yield prediction and recommendation," *MIT IJCSIT*, vol. 7, no. 1, pp.

[8] A. Kamilaris and F. Prenafeta-Boldu´, "Deep learning in agriculture," *Computers and Electronics in Agriculture*, vol. 147, pp

[9] S. Shinde and S. Shah, "Review of machine learning in crop yield prediction," *Int. J. Eng. Res.*, vol. 9, pp. 110–115, 2020.

[10] R. Feng et al., "Crop yield prediction using Random Forest," *IEEE Access*, vol. 7, pp. 180326–180335, 2019.

[11] A. Kumar et al., "Machine learning-based crop prediction model," *Int. J. Comput. Sci.*, vol. 8, no. 2, pp. 34–40, 2019.

[12] L. Chlingaryan et al., "Machine learning in agriculture: A review," *Sensors*, vol. 18, no. 12, pp. 1–29, 2018.

[13] S. Sarker et al., "Predicting crop yield using climate data," *Agricultural Systems*, vol. 184, p. 102903, 2020.

[14] P. Lal et al., "Weather-based yield prediction using ML," *Environmental Modelling & Software*, vol. 115, pp. 102–112, 2019.

[15] R. Ravi et al., "Smart agriculture using IoT and ML," *Procedia Com- puter Science*, vol. 171, pp. 1789–1796, 2020.

[16] K. Patil and M. Shirkhedkar, "Agricultural prediction using ML tech- niques," *Int. J. Sci. Tech.*, vol. 6, pp. 22–28, 2018.

[17] A. Ali and M. Chaudhary, "Crop yield forecasting using ensemble learning," *IEEE Access*, vol. 8, pp. 191287–191297, 2020.

[18] S. Zhou et al., "Evaluation of ML models for agricultural prediction,"*Computers in Agriculture*, vol. 3, pp. 45–56, 2017.

[19] M. Misra et al., "IoT-based smart farming systems," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2758–2768, 2020.