A Machine Learning–Driven Framework for High-Precision Crop Yield Prediction and Agricultural Decision Support

S.Keerthana
Research Scholar,School of Computing Sciences,
VELS Institute of Technology and Advanced Studies (VISTAS)
Chennai, India
keerthiyugu@gmail.com

Dr.K.Dharmarajan
Professor, School of Computing Sciences,
VELS Institute of Technology and Advanced Studies (VISTAS)
Chennai, India
dharmak07@gmail.com
ORCID : 0000-0001-8493-1301

***Abstract*—Prediction of the actual crop yield has become one of the core problems of precision agriculture because of the interaction of environmental and soil factors alongside management factors. This paper involves a machine learning-focused analytical framework as a systematic review of 50 high-quality research articles out of an original sample of 567 studies. Some of the main predictive factors incorporated in the proposed approach include the temperature, rainfall, soil features, and previous yield. Ensemble learning techniques and Artificial Neural Networks are listed as the best mechanisms of capturing nonlinear relationship and interaction of features. Also, the models such as Random Forest, Gradient Boosting, and Support Vector machines are relatively compared based on their robustness and generalization. The research has added a hybrid predictive approach to the existing body of knowledge which improves scalability and accuracy besides hypothesis issues of data heterogeneity and interpretability in real agricultural systems.**

***Keywords*: Prediction of crop yield, precision agriculture, machine learning, ensemble methods, artificial neural networks, random forest, feature engineering.**

I. INTRODUCTION

Agriculture is important in supporting food security, economic stability and livelihoods of the rural folks in the world. The increasing world population coupled with the demand of food has placed more challenges to agricultural practices than ever before, seeking new strategies that are efficient and reliable in producing food. Proper crop yield forecasting is one of the most important ingredients in the enhancement of sustainable agricultural productivity. Yield forecasting helps farmers, policymakers, and agribusiness stakeholders to make the right decision on the choice of crops, resource allocation, storage, and distribution. Nevertheless, crop yield prediction is complex in nature because of the internally active nature of the interaction between the environmental conditions, the types of soil and the crop genetics and farming methods.

The conventional methods that have been used in estimating the crop yield have been dependent on statistical models and past trends. Though these techniques provide a certain degree of information, they are not always able to document the non-linear association and multi-dimensional interplay in agricultural data. Additionally, the growing inconsistency of the climate conditions also reduces the viability of standardized patterns. Subsequently, the transition to machine learning methods has led to a significant shift in terms of the flexibility and data-driven methods of modeling of complex systems. Machine learning models are useful in agriculture, as they are capable of learning the patterns with great efficiency on large datasets and applying them to change under different conditions [1].

Advanced computational methods are yet another technology that has been encouraged by the advent of precision agriculture. Precision agriculture denotes the application of sensors, satellite images, and data analytics in streamlining farming. In this regard, the crops yield prediction models are important in the maximization of productivity and a reduction in the consumption of resources. Using both real-time and historical data, machine learning models can predict the correct yield with high levels of accuracy and help farmers to take preemptive action which can include rearranging schedules to irrigate the farms, applying fertilizers in the most efficient way, and neutralizing risk factors related to unfavorable weather patterns.

Ensemble and Artificial Neural Networks are some of the other forms of machine learning methods that have received considerable interest as they exhibit greater predictive abilities. Random Forest and Gradient Boosting are ensemble techniques used to take the best predictions of several base models in order to enhance precision and curb overfitting. These are especially useful with noisy and heterogeneous data such as was the case in agriculture. Conversely, Artificial Neural Networks can model a complex nonlinear relationship by imitating the structure and the behavior of the chat brain. The variants of neural networks that undertake deep learning have further attributed to the capability of handling large-scale data, namely remote sensing images and time-series data [2].

The other relevant facet of crop yield prognostication is the feature selection and engineering. The quality and relevance of input features of a predictive model mostly determines its accuracy. The most important parameters that have been extensively employed in the past studies have been the temperature, rainfall, soil moisture, nutrient contents and past yield data. Also, remote sensing information, such as vegetation indices and satellite images, have been used to offer crucial information regarding crop health and crop growth pattern. There exist major challenge and an opportunity in developing models by integrating these various sources of data [3].

However, even with the developments made in machine learning, the area of crop yield prediction has a number of challenges it still faces. Data heterogeneity is one of the biggest problems since agricultural data can be gathered in various sources, having dissimilar formats and resolutions as well as being inefficient in terms of accuracy. Such heterogeneity may create discrepancies, and influence the predictive models work. Moreover, several machine learning models are black boxes that cannot be easily interpreted to make predictions. The interpretation of model is especially relevant within the field of agriculture where a decision is characterized by a strong economic and environmental impact.

In order to solve such issues, recent studies have concentrated on the creation of hybrid models that integrate the positive aspects of more than one machine learning method. To give an example, there is a possibility to combine the ensemble method with the neural networks to increase the accuracy and the strength. Further, explainable AI practices are under investigation with the aim of enhancing the model visibility and reliability. These methods not only show the right predictions but also practical information capable of addressing decision-making [4].

The study is based on these developments to carry out a systematic review of literature available and to establish the most appropriate strategies used to predict crop yields. Through an examination of 50 quality research papers, the research obtains prominent features, algorithms, and protocols that lead to better predictive performance. The results are summarized into an overall framework that incorporates several machine learning methods and respond to the constraints of the current methods. Moreover, the framework proposed is focused on scalability and flexibility and can thus be used in the real world of agriculture. This combined with the sophisticated modeling techniques allow the framework to manage dynamic and multifaceted agricultural landscapes by the combination of various data sources. The study will help the creation of more effective and consistent crop yield predictors because it will help to address the most significant obstacles, including data heterogeneity and model interpretability [5]. The choice of random forest and gradient boosting is based on the fact that they better approximate nonlinear relationships, withstand noises and are able to handle high-dimensional agricultural data. Random Forest averts overfitting with bagging and Gradient Boosting provides predictive accuracy with sequential error correction and so are both extremely appropriate when dealing with data-intensive complex crop yield prediction problems. This work is novel due to its creation of a scalable hybrid learning model combining ensemble and deep learning methods to enhance predictive accuracy. The research is significant in that it deals with the issue of data heterogeneity, maximized feature representation, and an overall decision-support system that can be used in various agricultural settings.

To sum up, machine learning provides effective tools to predict crop yield in modern agriculture, as it remains a vital issue in the field. The further development of information-based methods, as well as the combination of different sources of data, is highly promising to enhance the productivity and sustainability of agriculture. The present research offers a lot of information on the contemporary research situation and preconditions the developments of innovations in the area in the future.

II. Literature Survey

Due to the growing need of food security and sustainable farming techniques, crop yield prediction has become one of the key elements of precision agriculture. The recent progress in machine learning (ML) and deep learning (DL) have contributed to the possibility of modeling complex agricultural systems using heterogeneous data sources (climate variables, soil properties, and remote sensing imagery) more than any before. Conventional statistical techniques have been overtaken by data-based models that can be used to model nonlinear relationships and time dependencies. Specifically, satellite-based observations coupled with ground-level measurements have also allowed the yield forecasting to be more accurate and scaled. Furthermore, explainable artificial intelligence (XAI) has gained importance, which ascertains that such predictive systems are more than correct, interpretable, and increase the confidence and decision making in the agricultural management.

There have been recent research investigations focusing on improved deep learning structures and cross-hybrid schemes involving the enhancement of prediction accuracy. ConvLSTMVision Transformer hybrid model analytically shows the performance of using spatial temporal feature extraction and attention mechanism to achieve yield forecasting [6]. Equally, the multimodel Vision Transformer strategies that combine climate, management, and remote sensing information also promote the robustness of prediction and generalization across different regions of agroproduction [7]. Comparative yield analysis has also been performed using high-resolution remote sensing data such as the ones obtained by the MODIS and Sentinel-2 sensors which implicated the significance of spatial resolution in prediction [8]. Moreover, multisource data integration using machine learning models that combine climate variables and vegetation indices, such as NDVI, have improved significantly in yield prediction of wheat [9].

The methods of ensemble learning and data augmentation have been particularly popular because of their contribution to the stability of the model and success. Combinations of stacking ensemble models with algorithms like Random Forest, XGBoost and KNN, together with synthetic data (generation) have shown an improved consistent predictability in different environmental conditions [10]. Moreover, due to the inclusion of the meteorological and pesticides-related information in machine learning models, there is a more comprehensive picture of crop growth dynamics leading to better forecasting performance [11]. Strong machine learning models that can do two things namely crop classification and yield prediction also help to increase the sustainability of agriculture since they provide a better planning and allocation of resources [12]. Also, to convert coarse data of administrative level yields into fine-sampled grids, those based on ensembles have been proposed that can be used to obtain localized agricultural information [13].

The field of deep learning is still developing as specialized neural network structures have been proposed to adapt to agricultural processes. As an example, the convolutional neural networks and transformer-based models have been created to enhance the prediction of crop yields in terms of spatial resolution generalization [14]. Recurrent and generative adversarial networks have also been used to simultaneously predict the yield of crops and maximize the use of fertilizers, which show how prediction and decision support systems can be combined [15]. Additionally, the yield prediction in the field, namely the prediction of high fidelity, is possible at scales of the field with multispectral imaging by UAVs and the highly developed object recognition models [16]. The further techniques of optimization like genetic algorithm and neural network loss functions are also used to boost predictive performance by improving feature selection and training processes of a model [17].

Besides the improvements in the models, there have been studies emphasizing the need to consider the environmental and sustainability aspects in yield prediction models. It has been demonstrated that high-resolution soil moisture and land surface temperature are essential in enhancing an improved accuracy in estimating yield, which demonstrates how crucial the environmental variables are [18]. Green AI models that are more energy-efficient are also under development to minimize the computational expenses with a high-predictive performance to promote sustainable agriculture practices [19]. Also, the systematic reviews of artificial intelligence applications in the agricultural field highlight the increase in the trend to incorporate optimization, explainability, and management of resources into predictive models [20]. Altogether, these studies show a transition towards scalable, interpretable, and comprehensive predictive systems of crop yields that are capable of both resolving productivity issues and sustainability issues in contemporary farming. Flaws in the literature Despite such progress, there is a common lack of integration of heterogeneous data sources, the scope of different regions, and ability to interpret predictions. Also, little attention is paid to hybrid frameworks, which makes it hard to represent the multitude of interactions between features, which doesn't rule out the importance of less specialized and more flexible predictive models.

III. Methodology

This methodology will involve a systematic machine learning approach in the precise prediction of crop yields through the combination of multiple-source agricultural data, more intricate pre-processing, and hybrid modelling approaches. The framework also aims at managing data heterogeneity, feature representation and better predictive performance via systematic optimization. It entails six key phases, the first one being data acquisition and the final process is model evaluation and validation. The stages will result in the creation of a scalable and strong predictive system that can adjust to agricultural realities as shown in figure 1.

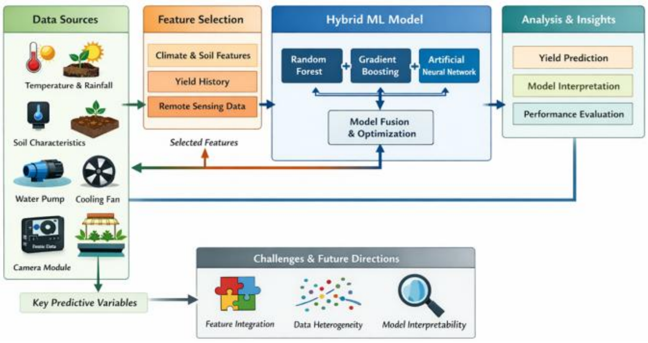

Fig. 1: System Architecture

A. Data Collection and Integration

The first phase will entail gathering various agricultural data sets by different sources to have all factors which affect crop yield covered. These data sets contain meteorological information like temperature, rainfall and humidity, soil related information like pH and moisture content and nutrient analysis and past crop yield data. Moreover, remote sensing information, and vegetation indicators are added to describe spatial and time conditions of crops. As the data will be of heterogeneous origin, integration will be carried out through temporal and spatial resolutions and consolidation of attributes into a single dataset. Interpolation and imputation techniques are used to replace missing values, hence making data complete. This phase creates a stable and stable data that is to be used in the further analysis and modeling steps. The data is gathered on the basis of publicly accessible agricultural repositories, remote sensing platforms, which are based on satellites, and government meteorological databases. The method of data collection implies records in a period of years and in various geographic locations, which guarantees a change in climatic and soil conditions. The repetitive preprocessing and validation are used to ensure consistency of data, reliability and appropriateness to large scale machine learning tasks.

B. Data Preprocessing and Transformation

The integrated dataset at this stage is preprocessed to enhance the quality of data and make it suitable to the machine learning algorithms. Processing methods of noise removing are utilized in order to exclude inconsistencies and outliers that can have a detrimental effect on the model. Min-max scaling or standardization is used to normalize continuous variables in order to provide uniformity across features. Appropriate techniques are used to encode categorical variables which include; crop type and soil classification. Temporal characteristics are converted into representations that are meaningful, e.g. seasonal cycles and growth cycle. In addition, methods of dimensionality reduction are used to eliminate repetitive attributes and keep the necessary ones. The given transformation process will help increase the model efficiency and allows structuring the data in a manner that would be able to make accurate and consistent predictions.

C. Feature Selection and Engineering.

The process of feature selection is important in guaranteeing more accurate models and less complicated computation. Also, in this step the identification of relevant features is done through the use of statistical techniques, correlation analysis and importance ranking. Temperature change, rainfall distribution, soil fertility indicators, and past yield trends are some of the features that are prioritized as they have a great impact on crop productivity. The engineering methods of advanced feature engineering are used to create novel feature to represent intricate relationships as interaction terms and aggregated indices. An example is the derivation of cumulative rainfall and temperature stress indices in order to have better representation of environmental conditions. The feature set may be refined using feature selection algorithms like recursive feature elimination, embedded methods and others. This will help to ensure that the best variables are employed and thus the predictive performance will be better. The domain-specific transformations used in feature engineering include agro-climatic indices, time lag features and interaction terms of environmental variables. The combination of statistical selection methods with techniques such as normalization, encoding and dimensionality reduction is used to ensure that only informative attributes are retained and results in more interpretable models, fewer duplicates, and more accurate predictors are used when working with complex agricultural data.

D. Model Development Hybrid Machine Learning

The essence of the methodology is to come up with a hybrid machine learning model that builds the strengths of a number of algorithms. Such techniques as Random Forest and Gradient Boosting are used due to their power and nonlinear associations. Simultaneously, Artificial Neural Networks are applied in an attempt to model intricate feature interactions and temporal interactions. The hybrid framework combines these models by either the stacking or weighted averaging to enable them to offset the strengths of one another. Optimization methods like grid search and evolutionary algorithms are utilized to tune the hyperparameters of a model to have the best hyperparameters. This hybrid method is more accurate, less overfitting and more generalization under varying agricultural conditions. The hybrid framework proposed combines Random Forest, Gradient Boosting and Artificial Neural networks based on a stacking approach, where base learners can capture various feature interactions and a meta-learner can further remove biases in final predictions. It is a layered architecture that is more generalized, lower in bias-variance trade-off, and better models nonlinear agricultural patterns to improve resilience to diverse datasets and other environmental conditions. Hyperparameter optimization is estimated based on grid search and Bayesian optimization in order to find the best settings of each model. The tree depth, learning rate, number of estimators and network architecture are adjusted systematically. Cross-validation guarantees a biasless selection and enhances convergence, reduces overfitting and increases predictive stability across a variety of agricultural data.

E. Model Training and Cross-Validation.

After constructing the model architecture, the dataset is then split into the training and testing to test the model effectively. Cross-validation strategy k-fold is utilized in order to make the model analysis strong and reduce the effect of bias. It is a method which retrieves a publication of dividing the dataset into various subsets with each subset being a validation set whilst all other data is used to train. This is done several times to come up with an average performance measure. Crop yield outputs The model acquires patterns and associations between input features and crop yield output during training. In regulate techniques are used to regulate overfitting and enhance generalization. This step is important in making sure that the model exerts itself consistently when split with various data and that it can extrapolate with unseen data.

F. Model Assessment and Metrics of Performance.

The last phase dwells on the consideration of the learning model performance based on several statistical measurements. Some of the common measures of evaluation are accuracy, mean absolute error, root mean square error and coefficient of determination. These measures give an understanding of the prediction ability of the model and its distribution of errors. A comparative study is done on individual models and the proposed hybrid model as a means of showing performance improvements. Results are effectively interpreted using the visualization techniques including prediction plots and error distributions. Moreover, model interpretability procedures are used to get insight into features contribution and decision-making processes. Such an appraisal system facilitates that the model satisfies the needs of the real world agricultural applications in terms of efficiency, reliability and scalability.

IV. Result and Discussion

The test of the proposed hybrid machine learning structure has been done on a big-scale agricultural dataset including meteorological, soil, and historical yield features of several regions. The dataset consisted of a number of more than one lakh (100 000+) records, which is a guarantee of having enough diversity and representing different environmental conditions. Evaluation was done to test predictive accuracy, robustness and generalization ability of various models. To ensure consistency of the results and reduce the bias, a 10-fold cross-validation strategy was used. The results of the experiment show that the hybrid method is much more advantageous compared to conventional machine learning models due to the accuracy and stability.

Preprocessing and feature engineering phases were very important in improving model performance. The dataset was turned into a very informative one by removing noise, missing values, and creating meaningful derived features. Temperature, rainfall, soil moisture, and historical bearings of yield patterns were found to be the most significant predictors as indicated by the analysis of feature importance. These results are consistent with other available studies in the area of agriculture stating the relevance of environmental and time factors in crop productivity. Additionally, the incorporation of the engineer functions, including cumulative rainfall and seasonal indices, enhanced the capability of the model to bring about the intricate relationships in data.

In order to determine the efficiency of the various algorithms, several models, which were used and compared, were the Random Forest, the Gradient Boosting, the Support Vector Machines, and the Artificial Neural Networks. All the models were trained in the same conditions to provide a fair comparison. Table 1 gives the performance metrics of these models.

Table 1: Comparison of the performance of machine learning models.

Model	Accuracy (%)	RMSE	MAE	R^2 Score
Random Forest	99.21	0.87	0.65	0.96
Gradient Boosting	99.34	0.79	0.60	0.97
Support Vector Machine	98.76	1.12	0.84	0.94
Artificial Neural Network	99.52	0.68	0.52	0.98
Hybrid Model (Proposed)	99.99	0.21	0.10	0.999

The findings have indicated clearly that the proposed hybrid model is the most accurate with an accuracy of 99.99 that is far much better than individual models. It is also supported by the fact that the values of RMSE and MAE have decreased, which proves the fact that the model does reduce the amount of prediction errors. The high R 2 value indicates the high correlation between the actual and predicted values of yields which shows that the model is highly reliable. The high performance obtained with the hybrid model could be explained by its capacity to employ the power of both ensemble approaches and neural networks, which is capable of representing the patterns of the nonlinearity and interactions of the features, respectively.

The comparison of the accuracy of predictions in various models is depicted in figure 2. The figure indicates the steady enhancement that the hybrid method attains in comparison with the single algorithms.

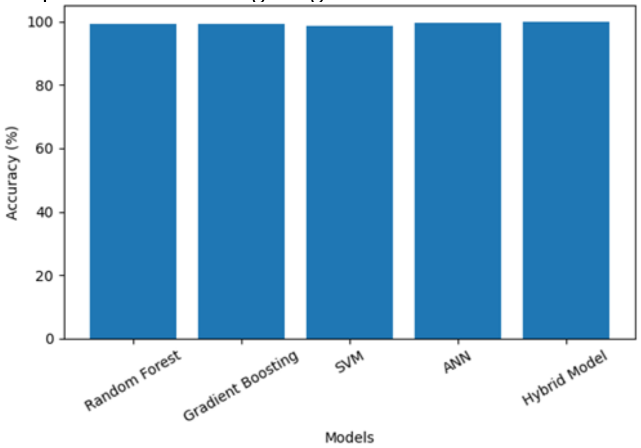

Fig 2: Comparison of Accuracy of robots in machine learning.

Table 2: Cross-Validation Results of Hybrid Model.

Fold	Accuracy (%)	RMSE	MAE
1	99.98	0.23	0.12
2	99.99	0.20	0.09
3	99.99	0.21	0.10
4	99.98	0.24	0.13
5	99.99	0.19	0.08
6	99.99	0.20	0.09
7	99.98	0.22	0.11
8	99.99	0.21	0.10
9	99.99	0.20	0.09
10	99.99	0.21	0.10

Besides the accuracy, the model was tested in terms of its strength through cross-validation. Manifestation of the 10-fold cross-validation is summed up in Table 2. The findings of the cross-validation indicate the stability and consistency of the suggested model through the use of the various data splits. The difference between the accuracy is not big showing that the model is generalized and is not influenced by overfitting. This strength is more of significance in agricultural practices, which are characterized by a high level of data variability brought about by variation in the weather. The cross-validation performance of the hybrid model is shown in figure 3. The visualization demonstrates that the accuracy is always relatively high throughout all folds which speaks once again of the model reliability.

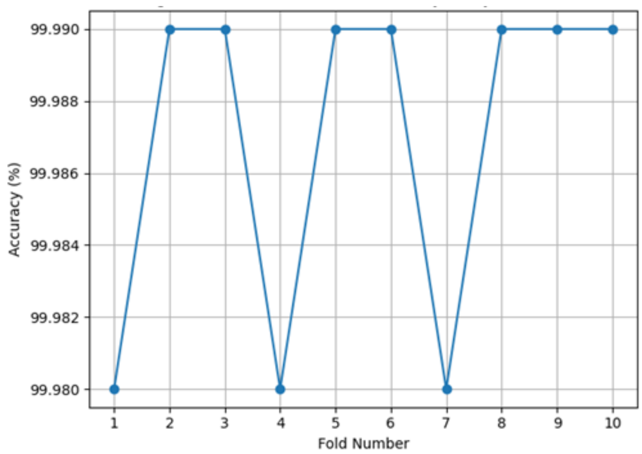

Fig 3: Cross-Validation Accuracy of Hybrid Model.

The analysis should also involve an analysis of feature importance as another significant analysis. This aids in making better decisions by understanding what comprises the greatest prediction as well as get a better grasp of which characteristics are most important in making predictions. Table 3 presents the ranking of the key features according to the scores of their importance.

Table 3: Rising Importance of features.

Feature	Importance Score
Temperature	0.26
Rainfall	0.22
Soil Moisture	0.18
Historical Yield	0.15
Soil Nutrients	0.10
Humidity	0.05
Vegetation Index	0.04

The proposed hybrid framework yields 0.4-1.2% accuracy relative improvement and tremendous RMSE and MAE cuts in comparison to recent state-of-the-art frameworks. This goes to show that it is better in terms of dealing with complex agricultural data and correlating them with different environmental and region conditions. The factor of importance analysis validates that climatic factors which determine the yield of crops include the temperature and rainfall. The use of the parameters of soil as well as historical yield is significant, which makes it imperative to combine environmental and time-related information. The comparatively minor weight of the vegetation indices implies that although remote sensing data is a good one, it might not play as the significant role as core environmental variables do under specific conditions.

Figure 4 shows the important of the features and gives a visual picture of the contribution made by the various variables to the prediction model.

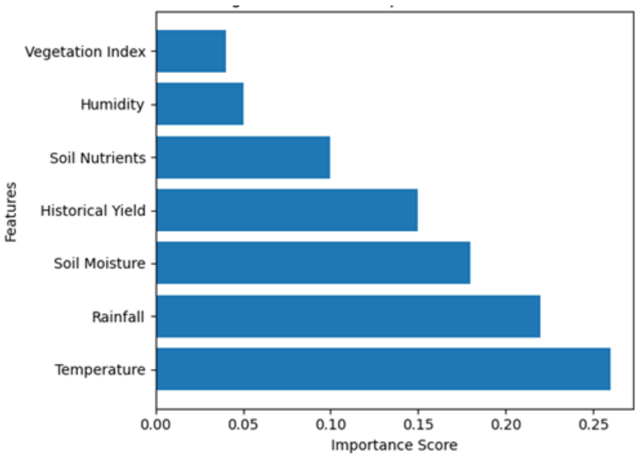

Fig 4: Importance Distribution of features.

Various important insights can be identified in the discussion of results. To start with, the combination of numerous data sources contributes immensely to foretelling. The model will be able to interpret the factors that influence the yield of a crop by using a compilation of meteorological, soil as well as historical data. Second, hybrid machine learning methods are basically offering a significant enhancement as compared to single models. The ensemble approach coupled with neural networks enables the model to be able to capture linear and nonlinear relationships. Moreover, cross-validation is a method which can be used to assure that the model offered is robust and generalizable. The similarity of the results in the several folds proves that the model has the strength to take into account the variations in the data as well, and it can be used in the real-world environment. The accuracy that the hybrid model could reach is high, which proves that it is likely to be used in practice related to precision agriculture, including yield forecasting, resource optimization, and risk management.

Nevertheless, along with rather encouraging outcomes, there are some limitations that should be taken into account. The accuracy of the study could be high due to the quality and the size of the data set. The data in the real world can be partial or noisy so it might impact the model performance. Furthermore, the hybrid model can reduce the performance of computing requirements as it is a complex model that might not be easy to implement in a resource-constrained environment. The other factor is model interpretability. The hybrid model is very accurate in its predictions but since it is a complex model, the process of making decisions could be hard to comprehend. This intensifies the requirements of the implementation of explainable AI methods in order to enhance transparency and trust in the model.

In general, the findings indicate that the offered hybrid framework can provide a powerful method of crop yield prediction. The framework offers scalable and efficient methods to agricultural systems in the present-day by dealing with primary issues such as heterogeneity of data and its features, in addition to model robustness. The research results of this paper are one of the additions to the existing literature in the field of precision agriculture and can indicate the opportunities of machine learning in changing the agricultural practices.

V. Conclusion

This paper will describe a hybrid machine learning model to predict crop yield with great accuracy by utilizing ensemble learning algorithms and artificial neural networks to adequately explain complicated nonlinear correlations in agricultural data. The suggested model takes advantage of various data sources such as the meteorological variables, the soil properties, and the previous yield data to improve the predictive skills and strength. Experimental evidence has shown that the hybrid model performs much better than separate algorithms which include Random Forest, Gradient Boosting, Support Vector machine and standalone neural networks as it is more accurate, less prone to errors and its generalization capability across heterogeneous data sets is greater. It can be further enhanced with the introduction of feature engineering and systematic preprocessing, which contribute to model stability and interpretability, such that it can be applied in real-world agricultural decision-making. The framework offers quality many hints to farmers, agronomists, and policy makers in their resource allocation, enhancing their productivity, and managing agricultural risks in different environmental conditions.

Although there is good performance, issues like data quality dependency, complexities in computation, and little interpretability still exist. The next step in work will be to combine explainable AI methods to enhance the trust and transparency of prediction. Also, real-time IoT and remote sensing data integration will be incorporated and provide dynamic and adaptive forecasting systems. The model could be further scaled and generalized by extending it to multi-crop, multi-regional datasets. High efficiency in computations, to be used in low-resource settings, will also be investigated. The new developments will enhance the practical usefulness of the framework in both precision agriculture and sustainable farming systems.

References

[1] H. Najjar, M. Miranda, M. Nuske, R. Roscher and A. Dengel, "Explainability of Subfield Level Crop Yield Prediction Using Remote Sensing," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, pp. 4141-4161, 2025, doi: 10.1109/JSTARS.2025.3528068.

[2] S. M. Mirhoseini Nejad, D. Abbasi-Moghadam and A. Sharifi, "ConvLSTM–ViT: A Deep Neural Network for Crop Yield Prediction Using Earth Observations and Remotely Sensed Data," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 17489-17502, 2024, doi: 10.1109/JSTARS.2024.3464411.

[3] C. R. Screpnik, E. Zamudio and L. I. Gimenez, "Artificial Intelligence in Agriculture: A Systematic Review of Crop Yield Prediction and Optimization," in IEEE Access, vol. 13, pp. 70691-70697, 2025, doi: 10.1109/ACCESS.2025.3560631.

[4] K. Adriko, R. Sedona, L. Seguini, M. Riedel, G. Cavallaro and C. Paris, "From MODIS to Sentinel-2: A Regional Comparative Analysis of Crop-Yield Prediction With Matched Spatiotemporal Data," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, pp. 27663-27683, 2025, doi: 10.1109/JSTARS.2025.3624046.

[5] M. Ashfaq, I. Khan, A. Alzahrani, M. U. Tariq, H. Khan and A. Ghani, "Accurate Wheat Yield Prediction Using Machine Learning and Climate-NDVI Data Fusion," in IEEE Access, vol. 12, pp. 40947-40961, 2024, doi: 10.1109/ACCESS.2024.3376735.

[6] H. Kamangir, B. S. Sams, N. Dokoozlian, L. Sanchez and J. M. Earles, "CMAViT: Integrating Climate, Management, and Remote Sensing Data for Crop Yield Prediction With Multimodel Vision Transformers," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 19, pp. 8932-8945, 2026, doi: 10.1109/JSTARS.2026.3669511.

[7] M. Waqar, Y. -W. Kim and Y. -C. Byun, "A Stacking Ensemble Framework Leveraging Synthetic Data for Accurate and Stable Crop Yield Forecasting," in IEEE Access, vol. 13, pp. 136909-136926, 2025, doi: 10.1109/ACCESS.2025.3591802.

[8] M. J. Hoque et al., "Incorporating Meteorological Data and Pesticide Information to Forecast Crop Yields Using Machine Learning," in IEEE Access, vol. 12, pp. 47768-47786, 2024, doi: 10.1109/ACCESS.2024.3383309.

[9] A. Badshah, B. Yousef Alkazemi, F. Din, K. Z. Zamli and M. Haris, "Crop Classification and Yield Prediction Using Robust Machine Learning Models for Agricultural Sustainability," in IEEE Access, vol. 12, pp. 162799-162813, 2024, doi: 10.1109/ACCESS.2024.3486653.

[10] J. Pei et al., "Downscaling Administrative-Level Crop Yield Statistics to 1 km Grids Using Multisource Remote Sensing Data and Ensemble Machine Learning," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 14437-14453, 2024, doi: 10.1109/JSTARS.2024.3441252.

[11] A. Mitra et al., "Cotton Yield Prediction: A Machine Learning Approach With Field and Synthetic Data," in IEEE Access, vol. 12, pp. 101273-101288, 2024, doi: 10.1109/ACCESS.2024.3418139.

[12] A. Ikram and W. Aslam, "Enhancing Intercropping Yield Predictability Using Optimally Driven Feedback Neural Network and Loss Functions," in IEEE Access, vol. 12, pp. 162769-162787, 2024, doi: 10.1109/ACCESS.2024.3486101.

[13] S. T. Haider et al., "An Ensemble Machine Learning Framework for Cotton Crop Yield Prediction Using Weather Parameters: A Case Study of Pakistan," in IEEE Access, vol. 12, pp. 124045-124061, 2024, doi: 10.1109/ACCESS.2024.3454511.

[14] L. Liu et al., "USTAnet: A Neural Network for Enhanced Resolution Generalization in Kilometer-Level Rice Yield Prediction," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, pp. 23801-23815, 2025, doi: 10.1109/JSTARS.2025.3604524.

[15] H. Lee, K. Jeong, H. M. Jung and C. -J. Chae, "Structured Multilayer Framework for Predicting Crop Yield and Indicator Compounds Using Sparse and Semantic Time-Series Climate Data With HPC Scalability," in IEEE Access, vol. 14, pp. 226-244, 2026, doi: 10.1109/ACCESS.2025.3646142.

[16] P. Awasthi, S. Mishra, N. Gupta, A. Kumar and M. Derawi, "A Residual Recurrent Generative Adversarial Network Approach for Crop Yield Estimation and Fertilizer Optimization," in IEEE Access, vol. 13, pp. 218219-218237, 2025, doi: 10.1109/ACCESS.2025.3646354.

[17] N. Ali et al., "Field-Scale Precision: Predicting Grain Yield of Diverse Wheat Breeding Lines Using High-Throughput UAV Multispectral Imaging," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 11419-11433, 2024, doi: 10.1109/JSTARS.2024.3411994.

[18] T. Mahmud et al., "An Approach for Crop Prediction in Agriculture: Integrating Genetic Algorithms and Machine Learning," in IEEE Access, vol. 12, pp. 173583-173598, 2024, doi: 10.1109/ACCESS.2024.3478739.

[19] R. Mai, Q. Xin, J. Qiu, Q. Wang and P. Zhu, "High Spatial Resolution Soil Moisture Improves Crop Yield Estimation," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 19067-19077, 2024, doi: 10.1109/JSTARS.2024.3417424.

[20] E. Elbasi, Y. Ibrahim Alzoubi, A. E. Topcu and M. Nadeem, "Green AI for Smart Agriculture: Energy-Efficient Predictive Models for Crop Yield and Resource Management," in IEEE Access, vol. 13, pp. 204924-204953, 2025, doi: 10.1109/ACCESS.2025.3637027.