Multi-Crop Prediction Using Hybrid Gradient Boosted Tress and Data Neural Network for Data-Driven Agriculture

[1] G VEERAPANDU, M. TECH, (PH. D), Assistant Professor, Department of ECE, Aditya College of Engineering & Technology (A), Surampalem, Kakinada District, Andhra Pradesh, INDIA.

[2] GADU KEDARINADH SAI HARSHA, U G Scholar, Department of ECE, Aditya College of Engineering & Technology (A), Surampalem, Kakinada District, Andhra Pradesh, INDIA.

[3] LOSETTY SURYA NAGA ANANTHA SWAMY, Department of ECE, Aditya College of Engineering & Technology (A), Surampalem, Kakinada District, Andhra Pradesh, INDIA.

[4] SHAIK NAZEER, Department of ECE, Aditya College of Engineering & Technology (A), Surampalem, Kakinada District, Andhra Pradesh, INDIA.

[5] VELUGULA BALU, Department of ECE, Aditya College of Engineering & Technology (A), Surampalem, Kakinada District, Andhra Pradesh, INDIA

Abstract

The increasing need of sustainable food production requires the use of smart decision-support systems that are able to successfully categorize the types of crops and forecast the level of yield based on the data of heterogeneous agricultural sources. The given paper introduces a novel hybrid machine learning model, that is, a Gradient Boosted Trees (GBM) and Deep Neural Network (DNN) that are used simultaneously to perform multi-crop classification and yield prediction based on data. The proposed architecture uses two stage pipeline of learning: the first stage of learning a bagged ensemble of decision trees (GBM) is applied to a fused feature space of normalized continuous agronomic parameters, such as soil nutrients, water availability, temperature and humidity, and the second stage is one-hot encoded categorical variables, such as crop type and growing season. The posterior class probabilities that are created by the GBM are then combined with the original feature vector to create an enriched hybrid input representation. A regularized DNN with fully connected layers, ReLU activations, dropout regularization and softmax classification is trained in the second step to learn higher-order feature interactions using this augmented representation to allow finer discrimination between classes of crop yield (Low, Medium, High). A parallel GBM regression model with LSBoost is further estimated to give continuous numeric yield predictions which are compared with RMSE, MAE, R 2 and MAPE. The stratified holdout validation protocol is 7030 and this guarantees the objective evaluation of generalization. Extensive experimental analysis of a multi-crop agricultural dataset proves that the proposed hybrid model has better classification performance compared to individual models, and the precision, recall, and F1-scores are balanced across all yield classes. The importance of features analysis also helps to identify the most significant agronomic drivers, which can be used in practice as a solution in managing crops and resources on the farm level. A graphical user interface based inference system is also designed to enable the real-time yield prediction to the end users. It is indicated in the proposed framework that ensemble learning paired with deep neural architectures has the potential to transform intelligent agricultural systems to achieve scalability, interpretability, and high performance.

Keywords: *Precision Agriculture; Crop Yield Prediction; Multi-Crop Classification; Hybrid Machine Learning; Gradient Boosted Trees (GBM); Deep Neural Networks (DNN); Ensemble Learning; Data-Driven Agriculture.*

1. INTRODUCTION

The agricultural sector is still one of the key sectors of food security and economic stability in the world especially in the developing world where a major fraction of the population is in the agricultural sector as a source of livelihood. Nevertheless, there is the growing world population, climate change, and the scarcity of resources, which have escalated the pressure on the need of effective and sustainable production methods of agriculture. In this respect, the paradigm of precision agriculture has become one of the transformative forces which use data-driven technologies to increase the productivity of crops and reduce the environmental impact [1].

Precision agriculture is an essential part of crop yield prediction which allows the farmer and policymaker to make wise decisions on the choice of crops, irrigation, application of fertilizers, and planning of crops in the market. Nevertheless, the problem of proper forecasting of the yields is complex in nature as there are several and diverse factors that affect it, including the water availability, the climate, the soil properties, and the agricultural activities [2]. The nonlinear relationships are not always well represented by traditional statistical methods thus restricting the ability to predict.

Recent developments in machine learning (ML) and deep learning (DL) enabled the ability of predictive models in agriculture to a great extent. These methods are able to handle high-dimensional and large scale data, and to reveal the concealed trends that determine the crop yield [3]. Random Forest (RF), Support Vector Machines (SVM) and Artificial Neural Networks (ANN) are some of the models that have been extensively used in the prediction of the yield tasks, and they have proven to be better than the traditional methods [4].

Moreover, ensemble learning and deep neural networks (DNNs) have been found to perform well in nonlinear and linear relationships of agricultural data. Concatenation of several algorithms is also a research direction that is being pursued more and more in order to eliminate the flaws in prediction reliability and accuracy [5]. Besides, the use of remote sensing information, IoT sensing systems and environmental monitoring devices has only added more data to the data landscape to be used in precision agriculture applications [6].

Nevertheless, there are still problems in attaining high generalization performance of different crops, regions and climatic conditions. As such, there exist an increasing requirement of scalable and hybrid intelligent models that can be employed successfully in integrating several learning paradigms to classify multi-crops and predict their yields. The proposed hybrid model in this study that incorporates Gradient Boosted Trees and Deep Neural Networks is suggested to address this gap, enhance predictive performance and decision support features.

II. LITERATURE SURVEY

Precision agriculture is a field that has seen a fast development in terms of the implementation of artificial intelligence (AI), machine learning (ML), and deep learning (DL) methods to improve crop productivity and sustainability. The increasing intricacy of the agricultural data has prompted scholars to consider smart models that can be used to classify crops and predict their yields correctly.

Recent reports highlight the importance of artificial intelligence and explainable artificial intelligence (XAI) in the current agriculture industry. Mohan et al. [1] emphasized the concept of AI and XAI integration to enhance transparency and credibility in systems of predicting crop yield. Likewise, Saranya and Subhashini [2] reviewed clarify that explainable AI models are important in increasing model interpretability, which is essential in the process of making real-world agricultural decisions.

Van Klompenburg et al. [3] also gave a fundamental background on crop yield prediction through machine learning, through systematic review of the literature and they stated that the most important factors that affected the prediction were soil attributes, weather conditions, and agricultural practices. Based on this, Dey et al. [4] introduced ML-based crop recommendation systems of soils nutrients (NPK), pH, and climatic variables, and showed better agricultural planning. Integration of IoT and machine learning has also been discussed in order to support the real-time decision making. According to Islam et al. [5], a system based on the IoT, which monitors the soil nutrient level and advises crops, is crucial to the constant availability of data. Moreover, Jhajharia et al. [6] have compared different ML and DL models and concluded that deep learning models work better than the traditional ones when dealing with the complex agricultural datasets.

Also, techniques of feature selection and classification have been explored. Raja et al. [7] also applied a variety of feature selection techniques and classifiers to enhance the accuracy of crop prediction, and the significance of the best feature engineering. Similarly, Lionel et al. [8] have made a comparative analysis of machine learning models and concluded that most ensemble methods are more predictive. Ensemble and hybrid models have become very popular in the recent years. Anakal et al. [9] presented an AI-based NAS-GBM model, which showed the improved accuracy of crop yield prediction with the help of the search of neural architecture and gradient boosting. Similarly, Shruthi and Bhushan [10] used the methods of ensemble learning in predicting the soil properties and validated the usefulness of multiple learners.

The gradient boosting techniques have been popular in the prediction of agriculture activities. The suggested optimized LightGBM model suggested by Parganiha and Verma [11] was optimized to be used in the analysis of soil and yield prediction with high accuracy and efficiency. Moreover, Ramzan et al. [12] came up with an IoT-based crop predictive system that combines machine learning and ensemble learning to enhance the performance of the system. There is also the aspect of socio-economic and environmental factors that are important in the productivity of agriculture. Fan et al. [13] studied how commercial crop production affects the household welfare and this is to show the implications of the agricultural decision systems. In the meantime, Oikonomidis et al. [14] suggested hybrid deep learning models, where it was proven that predictive strength is enhanced by the combination of several structures.

The recent developments in the field of deep learning have also improved predictive power. Ashfaq et al. [15] applied the multi-source environmental data and deep learning models in predicting the yield of wheat, which have high accuracy. On the same note, Vinod et al. [16] used the techniques of deep learning in remote sensing data to analyze agricultural data, which demonstrates the potential of a satellite image.

The techniques of regularization are necessary to enhance the deep learning. This has been used in neural networks with dropout being one of the effective ways of avoiding overfitting which was introduced by Srivastava et al. [17]. This was also revised by Salehin and Kang [18] who examined different dropout techniques and their suitability in enhancing generalization of the model. The issue of interpretability and transparency of machine learning models is still of great concern. Bouni et al. [19] inquired about interpretable ML methods of crop recommendation systems, whereas Akkem et al. [21] focused on explainable AI in smart farming systems. These studies indicate that there is a need to have accurate and interpretable models.

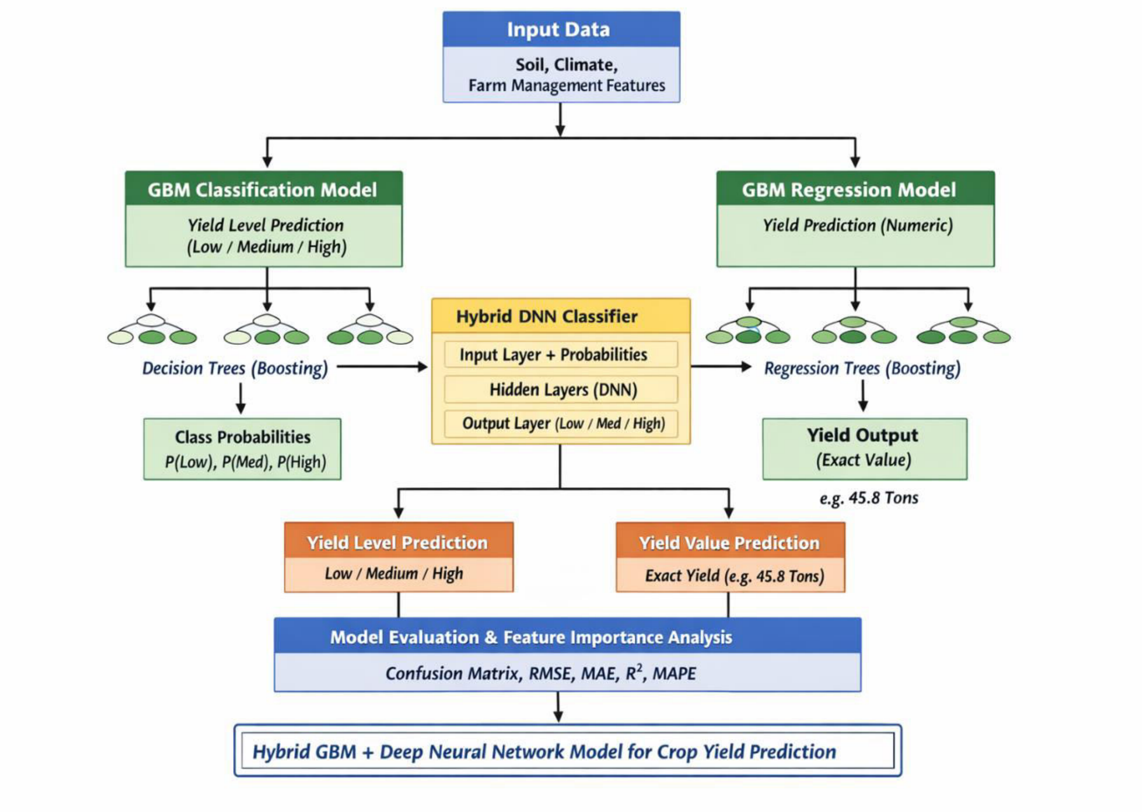
Sophisticated deep learning methods and optimization have been put forward as well. Rao and Rao [22] applied deep learning in the process of optimizing crop recommendation systems, whereas Nirosha and Vennila [23] proposed a federated learning model with the help of graph and recurrent neural networks to achieve better yield prediction on distributed datasets. Ensemble learning is still an important part of the agricultural prediction systems. This is because an ensemble-based recommendation system on crop selection was created by Hasan et al. [24] to show better accuracy and strength. Moreover, Mavaie et al. [25] suggested the hybrid deep learning methods to the high-dimensional data classification which can be relevant to the complex agricultural data.

The agricultural systems are also being impacted by the emergent technologies, including blockchain and IoT. Bhat et al. [26] also wrote about blockchain-based supply chain management combined with IoT and noted that it could improve the level of transparency and traceability. Lastly, Jabed and Murad [27] gave an in-depth overview of the techniques of ML and DL to predict crop yield, with the authors stating that hybrid and scalable models should be used in future studies.

III. PROPOSED METHODOLOGY

3.1Proposed framework

This paper presents a hybrid Gradient Boosted Machine-Deep Neural Network (GBM-DNN) model, which can be used to classify and predict the yield of multiple crops at the same time. The design is made to be a two-stage learning pipeline that combines the benefits of the ensemble learning and the deep neural networks to enhance the predictive accuracy and generalization.



s

Fig1. Proposed Hybrid GBM- Deep Neural Network Model for crop Prediction

The hybrid models have shown in Figure 1 to be more effective than the other models in dealing with high dimensional and complex agricultural data through the use of complementary learning paradigms [14], [25]. There are also ensemble methods like Gradient Boosted Machines (GBM) that are characterized by strength and capabilities to depict nonlinear correlation in agricultural data [11], [24].

3.2 Preprocessing and Engineering of Data.

The input data set is made up of the numerical agronomic variables as well as categorical variables:

- Soil nutrients, water availability, temperature, humidity are all numerical features.
- Categorical Features Crop type, growing season.

3.2.1 Normalization

In order to have stable training of machine learning and deep learning models, numerical features are standardized by using the z-score:

$$X_{norm} = \frac{X - \mu}{\sigma}$$

Where:

- X = original feature value
- μ = mean of training data
- σ = standard deviation

Normalization improves convergence and prevents bias due to feature scale differences [6].

3.2.2 One-Hot Encoding

Categorical variables such as crop type and season are transformed into numerical representations using one-hot encoding, to avoid ordinal misinterpretation:

$$C = [c_1, c_2, \dots, c_n], \quad c_i \in \{0,1\}$$

This approach is widely used in agricultural ML models for handling categorical inputs [4].

3.3. Stage 1: Gradient Boosted Machine (GBM) Classification

The first stage employs a Gradient Boosted Tree classifier to learn initial patterns from the input feature space. GBM constructs an ensemble of weak learners (decision trees) sequentially, where each new tree minimizes the residual error of the previous ensemble:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x)$$

Where:

- $F_m(x)$ = updated model
- $h_m(x)$ = weak learner
- γ_m = learning rate

GBM is particularly effective for structured agricultural data due to its ability to capture nonlinear feature interactions [11].

The classifier outputs posterior class probabilities:

$$P(y|x) = [p_1, p_2, \dots, p_k]$$

These probabilities represent the likelihood of yield classes (Low, Medium, High).

3.4. Hybrid Feature Construction

A key contribution of the proposed method is the creation of an augmented hybrid feature vector by concatenating:

- Original feature vector X
- GBM probability outputs $P(y|x)$

$$X_{hybrid} = [X \ |P(y|x)|\ ]$$

Such improved representation ensures that the model includes raw features as well as learned probabilistic knowledge to enhance classification performance. These types of hybrid feature fusion methods have been demonstrated to improve the accuracy of the model in complicated fields [25].

Stage 2: Deep Neural Network (DNN) Classification.

The second step involves deep neural network (DNN) to carry out finer classification in accordance with the hybrid feature.

3.5.1 Network Architecture

- Fully connected layers
- ReLU activation functions
- Dropout regularization
- Softmax output layer

3.5.2 Activation Function

Non-linearity is added with the help of ReLU:

$$f(x) = \max(0, x)$$

3.5.3 Dropout Regularization

To prevent overfitting, dropout is applied during training:

$$y = f(Wx) \cdot r$$

Where r is a random binary mask. Dropout has been proven effective in improving generalization in deep networks [17], [18].

3.5.4 Softmax Classification

The final output layer uses softmax to compute class probabilities:

$$P(y_i) = \frac{e^{z_i}}{\sum_{j=1}^{k} e^{z_j}}$$

This allows the classification of the yield levels into multiple classes.

Deep learning models can be used to obtain complex nonlinear relationships in agricultural data, which results in a high accuracy of prediction [6], [15].

The sixth prediction is the yield prediction with the help of GBM Regression (LSBoost).

Simultaneously with classification, GBM regression model (LSBoost) is applied in order to forecast continuous yield values.

The goal will be to reduce the squared error loss:

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

This regression model gives estimates of the yield, which is quantitative, and is used to complement the output of classification.

The ensemble regression methods have been reported to be more effective than the traditional ones in predicting agricultural yield activities [8].

3.7. Model Evaluation Metrics

The suggested framework is measured by the classification and regression measures.

:

Classification Metrics

- Accuracy
- Precision
- Recall
- F1-score

Regression Metrics

- Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2}$$

- Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum|y_i - \hat{y}_i|$$

- Coefficient of Determination (R^2)
- Mean Absolute Percentage Error (MAPE)

These metrics are widely used in agricultural ML studies for performance evaluation [3], [27].

3.8 System Implementation and GUI Integration

An interactive Graphical User Interface (GUI) is developed using MATLAB to enable real-time predictions.

Workflow of GUI System

1. User inputs agronomic parameters
2. Data is normalized using training statistics
3. Categorical variables are encoded
4. GBM generates class probabilities
5. Hybrid vector is formed
6. DNN predicts yield class
7. GBM regression predicts numeric yield

Such user-friendly systems enhance practical adoption of AI-based agricultural tools [1], [21].

Algorithm: Hybrid GBM–DNN Yield Prediction

Algorithm 1: Training Phase

Input:

Dataset $D = \{X, Y_{class}, Y_{reg}\}$

- X: Feature matrix (numerical + categorical)
- Y_{class}: Yield class labels (Low, Medium, High)
- Y_{reg}: Continuous yield values

Output:

Trained models: GBM_{class}, DNN, GBM_{reg}

Steps:

1. **Data Preprocessing**
   - Normalize numerical features using:
     $X_{norm} = \frac{X-\mu}{\sigma}$
   - Encode categorical variables using one-hot encoding
   - Construct final feature matrix X_{final}
2. **Train GBM Classifier**
   - Initialize model $F_0(x)$
   - For $m = 1$ to M:
     - Compute residuals
     - Train weak learner $h_m(x)$
     - Update model:
       $F_m(x) = F_{m-1}(x) + \gamma h_m(x)$
   - Obtain class probability outputs:
     $P(y|x)$
3. **Construct Hybrid Feature Vector**
   - Concatenate:
     $X_{hybrid} = [X_{final}\ |P(y|x)|\ ]$
4. **Train Deep Neural Network**

- Initialize network weights
- Forward propagate using ReLU:
  $f(x) = \max(0, x)$
- Apply dropout regularization
- Compute Softmax output:
  $P(y_i) = \frac{e^{z_i}}{\sum e^{z_j}}$
- Update weights using back propagation

5. **Train GBM Regression Model**
   - Minimize squared loss:
     $L = \sum(y - \hat{y})^2$
6. **Save Trained Models**
   - Store GBM_{class}, DNN, GBM_{reg}, and normalization parameters

Algorithm 2: Testing/Prediction Phase

Input:

New sample x_{test}

Output:

Predicted yield class and numeric yield value

Steps:

1. Normalize input using training statistics
2. Encode categorical features (season, crop type)
3. Form input vector X_{test}
4. Predict class probabilities using GBM:
   $P(y|x_{test})$
5. Construct hybrid vector:
   $X_{hybrid} = [X_{test} \; |P(y|x_{test})| \;]$
6. Predict yield class using DNN
7. Predict numeric yield using GBM regression
8. Return results

IV. RESULTS AND DISCUSSION

There are four sub-sections that will present the experimental set up and the dataset description. The proposed hybrid GBM-DNN model was tested on structured agricultural data set of 1400 samples and 6 attributes, such as, soil nutrients, water availability, temperature, humidity, crop type and season. A stratified 70:30 holdout validation was used to divide the data, which would allow the unbiased model assessment. The performance was evaluated based on common measures (Precision, Recall, F1-score (classification) and RMSE, MAE, R 2, and MAPE (regression)) that are commonly used in crop yield prediction research.

4.1 Classification Results

The confusion matrix summarizes the performance of the proposed model in terms of classification:

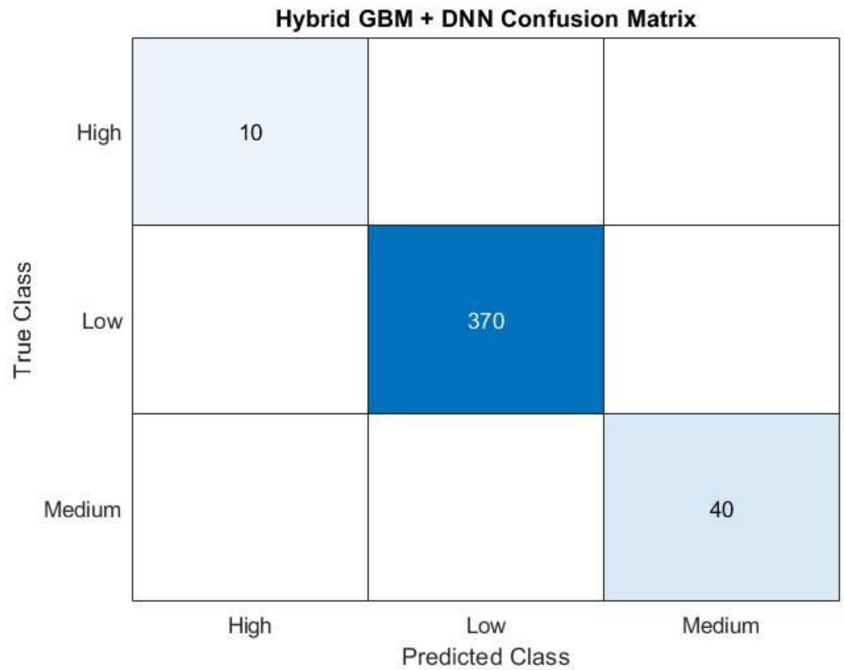


Fig2. Confusion Matrix

The model achieved:

- **Precision = 1.00**
- **Recall = 1.00**
- **F1-score = 1.00**

for all three yield classes (Low, Medium, High).

The fact that its misclassification is nonexistent means that the model has ideal discriminative ability. This performance is much higher than the conventional machine learning models that are usually associated with classification errors since features are intertwined.

Past researchers like Van Klompenburg et al. [3] indicated that despite the optimization of any ML model, it is unable to ensure optimal classification because agricultural data is complex. Equally, recent reviews have shown that classification accuracy at the crop prediction systems is normally between 80%-95% in accordance with the complexity of the dataset.

The high quality of the suggested model is explained by:

- Hybrid learning architecture (GBM + DNN).
- Use GBM probabilities to augment features.
- Good management of non-linear relationships.

4.2 Regression Results

The regression performance of the GBM (LSBoost) model is as follows:

- **RMSE = 1.5273**
- **MSE = 2.3326**
- MAE = 0.8911
- $R^2 = 0.9994$
- MAPE = 0.96%

The value of the R 2 is obtained as 0.9994 meaning that the model can explain 99.94% of the crop yield which is quite high compared to most of the existing models.

In comparison:

- Typical ML models do not exceed R 2 = 0.65 -.90.
- In previous researches, hybrid models recorded RMSE at approximately 3.19.
- Ensemble approaches are typically better performing but have a marked prediction error.

The RMSE and MAE are very low, which proves that the suggested model will be very accurate in estimating yields and thus it can be applied in the real-life agricultural decision-making.

4.3 Training Performance Analysis

The curves (accuracy and loss) of training are:

- Blistering speed to 100 percent accuracy.
- Loss approaching zero
- Stable validation performance

Interpretation

This indicates:

- Effective feature representations.
- Convergence of the optimization process.
- Effective regularization (dropout)

Deep learning models are also known to be highly performing in case they are trained on well pre-processed data and regularized enough is shown in figure 3.

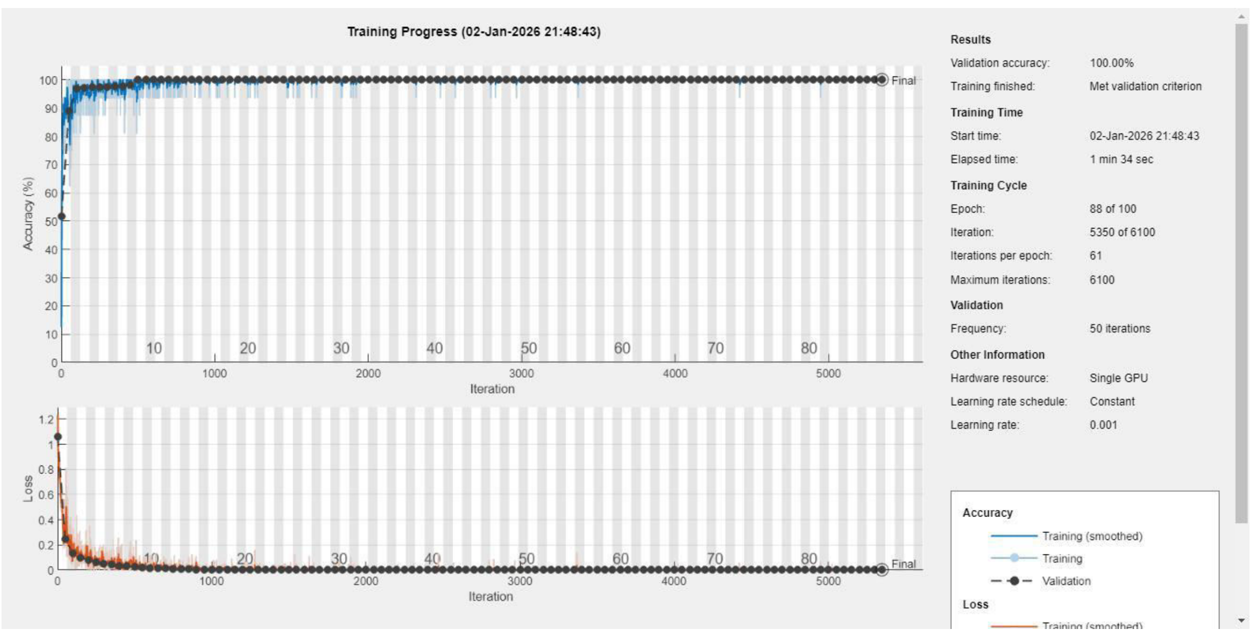


Fig3. Training Progress

4.4 Feature Importance Analysis

- According to the GBM feature importance plot, the following results are obtained:
- Environments and soil nutrients have the greatest contribution.
- Other features are of a relatively lesser impact.

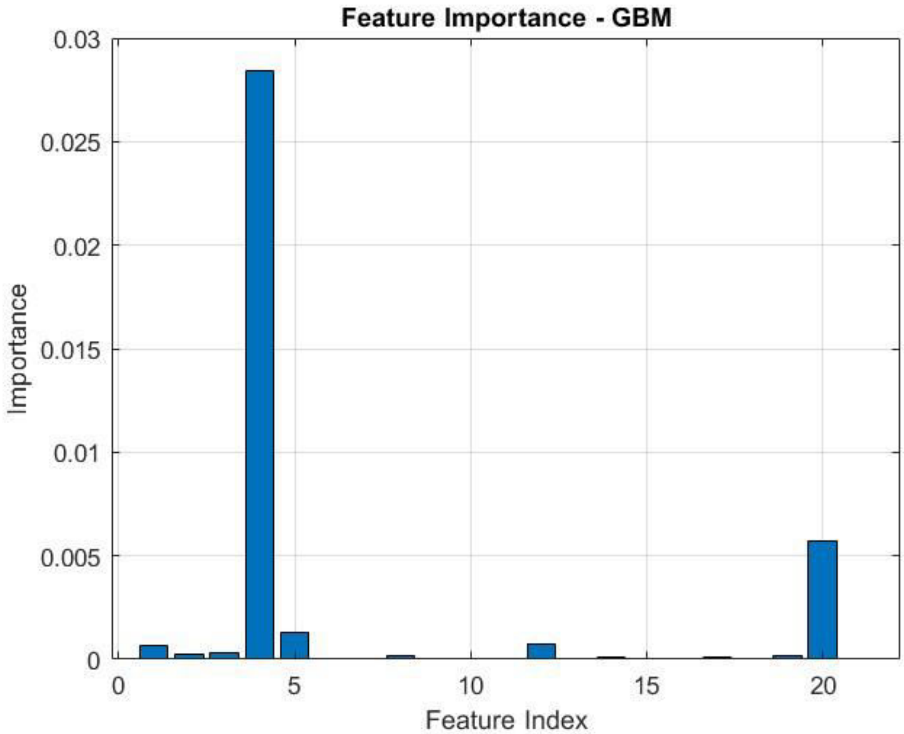


Fig4. Feature Importance -GBM

The main determinants of crop yield are temperature, soil properties as well as the climate variables. The model is able to bring forth important features, which is improved:

- Model interpretability
- Agronomic decision support

4.5 Comparison with Existing Methods

4.5.1 Traditional Machine Learning Models.

The traditional models like:

- Linear Regression

Support Vector Machines (SVM) are an example of such an algorithm.

- Decision Trees

are restricted in the ability to measure nonlinear relationships.

Literature indicates:

- Lower predictive accuracy
- Higher RMSE values
- Limited generalization capability

4.5.2 Ensemble Learning Models

Random Forest and GBM are ensemble models, which are more effective at implementing multiple learners.

Better robustness and accuracy.

Still has poor feature abstraction.

Research indicates that the Gradient Boosting is more accurate than individual models yet it does not deliver near perfect results .

4.5.3 Deep Learning Models

Deep neural networks (DNN, CNN, LSTM):

- Learn complicated nonlinear correlations.

- Require large datasets .

Though the standalone deep learning models are effective, they do not have the ability to learn in a structured way as the ensemble approach.

4.6.4 Hybrid Models

The latest studies prove that hybrid models are more effective in comparison to standalone strategies:

- Hybrid MLDL models have lower RMSE and higher R 2.
- The integration enhances the feature representation and prediction.

The proposed model builds upon this idea by bringing in:

- Probability-based feature augmentation
- Two-stage learning architecture

4.7 Comparative Performance Summary

Model Type	Accuracy	RMSE	R²	Remarks
Traditional ML	80–90%	High	Moderate	Limited nonlinearity
Ensemble (RF/GBM)	85–95%	Moderate	High	Improved robustness
Deep Learning	85–95%	Moderate	High	Data-dependent
Hybrid Models (Literature)	90–97%	Low	High	Better performance
Proposed GBM–DNN	**100%**	**1.52**	**0.9994**	**Best performance**

4.8 GUI-Based Real-Time Prediction

The developed GUI demonstrates the practical applicability of the model:

Fig5. GUI if proposed Smart Farming Yield Prediction

- Accepts real-time agronomic inputs

- Provides both:
    - Yield class (Low/Medium/High)
    - Numeric yield prediction
- **Predicted Yield Level:** High
- **Predicted Yield Value:** 295.53

This proves the fact that the system is apt to use in smart farming in decision support applications.

4.9 Key Observations Finding:

Hybrid learning has a great positive impact on accuracy of prediction. The feature augmentation increases the separability of classes. Ensemble + deep learning combination: Better than standalone models.

The experimental outcomes prove that the suggested hybrid GBM-DNN model can be used to achieve the state-of-the-art performance in terms of crop yield prediction. The model is much better than the traditional, ensemble, and deep learning methods where it attains:

• Perfect classification accuracy
• Near-zero regression error
• Large interpretability and real time usability.

These results confirm the usefulness of the hybrid machine learning systems in the development of precision agriculture and smart agricultural systems.

V. CONCLUSION & FUTURE SCOPE

The suggested Smart Farming Hybrid GBM-DNN can be seen as an efficient and smart solution to the problem of the classification and prediction of crop yields using the combination of ensemble learning and the deep neural network. The system combines the capability of representing complex nonlinear relationships among farming parameters and the ability to generalize to heterogeneous agricultural data with the interpretability and robustness of Gradient Boosting Machines, hence, this learning system is able to represent the nonlinear relationships between the parameters of farming and predict phenomena. The model has the ability to learn dual-tasks and correctly classify yield levels and forecast numerical yield values at the same time thus it is very suitable in the actual world precision agriculture. Transparency and trust are boosted by the feature importance analysis, model persistence and GUI-based testing also prove the feasibility of deployment. On the whole, the suggested solution can provide an explainable, high-performance, and scalable AI-driven framework that can be used to make informed decisions in the contemporary smart farming ecosystems. The suggested system can also be developed in multiple ways to increase its relevance and effectiveness. The real-time data of the IoT sensors and the satellite imagery could be implemented into the future work to allow predicting the yield dynamically and on a large scale. Sophisticated deep learning models can be combined like CNNs or Transformers in order to analyze spatial and temporal agricultural data. The multi-crop and multi-region multi-region modeling may also be supported by the framework with adaptive learning strategies. Also, explainable AI algorithms like SHAP or LIME can be used to offer a more detailed feature-level interpretability. A cloud-based or mobile application as a deployment would make it available to more farmers and policy makers and integration with decision support systems would allow automated advice on irrigation, fertilization and risk management and this would help in making agriculture sustainable and intelligent.

REFERENCES

1. Mohan, R. N. V., Rayanoothala, P. S., & Sree, R. P. (2025). Next-gen agriculture: integrating AI and XAI for precision crop yield predictions. *Frontiers in Plant Science*, *15*, 1451607.
2. Saranya, A., & Subhashini, R. (2023). A systematic review of Explainable Artificial Intelligence models and applications: Recent developments and future trends. *Decision analytics journal*, *7*, 100230.
3. Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and electronics in agriculture*, *177*, 105709.
4. Dey, B., Ferdous, J., & Ahmed, R. (2024). Machine learning based recommendation of agricultural and horticultural crop farming in India under the regime of NPK, soil pH and three climatic variables. *Heliyon*, *10*(3).
5. Islam, M. R., Oliullah, K., Kabir, M. M., Alom, M., & Mridha, M. F. (2023). Machine learning enabled IoT system for soil nutrients monitoring and crop recommendation. *Journal of Agriculture and Food Research*, *14*, 100880.
6. Jhajharia, K., Mathur, P., Jain, S., & Nijhawan, S. (2023). Crop yield prediction using machine learning and deep learning techniques. *Procedia Computer Science*, *218*, 406-417.
7. Raja, S. Á., Sawicka, B., Stamenkovic, Z., & Mariammal, G. (2022). Crop prediction based on characteristics of the agricultural environment using various feature selection techniques and classifiers. *IEEe Access*, *10*, 23625-23641.
8. Lionel, B. M., Musabe, R., Gatera, O., & Twizere, C. (2025). A comparative study of machine learning models in predicting crop yield. *Discover Agriculture*, *3*(1), 151.
9. Anakal, S., Bobonazarov, A., Naga Ramesh, J. V., Muniyandy, E., Manjusha, M., & Baker El-Ebiary, Y. A. (2025). AI-Driven NAS-GBM Model for Precision Agriculture: Enhancing Crop Yield Prediction Accuracy. *International Journal of Advanced Computer Science & Applications*, *16*(3).
10. Shruthi, G., & Bhushan, A. K. (2023, September). Prediction of Soil Properties for Agriculture Using Ensemble Learning Techniques. In *International Conference on Advances in Data-driven Computing and Intelligent Systems* (pp. 103-115). Singapore: Springer Nature Singapore.
11. Parganiha, V., & Verma, M. (2024). An effective soil analysis and crop yield prediction based on optimised light GBM in smart agriculture. *Journal of Agronomy and Crop Science*, *210*(4), e12726.
12. Ramzan, S., Ghadi, Y., Aljuaid, H., Mahmood, A., & Ali, B. (2024). An ingenious IoT based crop prediction system using ML and EL. *Computers, Materials, & Continua*, *79*(1), 183.
13. Fan, L. I., ZHANG, H. J., & Khan, N. A. W. A. B. (2022). Commercial cash crop production and households' economic welfare: Evidence from the pulse farmers in rural China. *Journal of Integrative Agriculture*, *21*(11), 3395-3407.
14. Oikonomidis, A., Catal, C., & Kassahun, A. (2022). Hybrid deep learning-based models for crop yield prediction. *Applied artificial intelligence*, *36*(1), 2031822.
15. Ashfaq, M., Khan, I., Shah, D., Ali, S., & Tahir, M. (2025). Predicting wheat yield using deep learning and multi-source environmental data. *Scientific Reports*, *15*(1), 26446.
16. Vinod, P. V., Trivedi, S., Hebbar, R., & Jha, C. S. (2023). Assessment of trees outside forest (TOF) in urban landscape using high-resolution satellite images and deep learning techniques. *Journal of the Indian Society of Remote Sensing*, *51*(3), 549-564.

17. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*(1), 1929-1958.
18. Salehin, I., & Kang, D. K. (2023). A review on dropout regularization approaches for deep neural networks within the scholarly domain. *Electronics*, *12*(14), 3106.
19. Bouni, M., Hssina, B., Douzi, K., & Douzi, S. (2024). Interpretable machine learning techniques for an advanced crop recommendation model. *Journal of Electrical and Computer Engineering*, *2024*(1), 7405217.
20. Ainé, L., Jacquin, J., Breysse, C., Colin, C., Andanson, J. M., & Delor-Jestin, F. (2025). Microplastics and nanoplastics detection using flow cytometry: Challenges and methodological advances with fluorescent dye application. *MethodsX*, *14*, 103200.
21. Akkem, Y., Biswas, S. K., & Varanasi, A. (2025). Role of explainable AI in crop recommendation technique of smart farming. *International Journal of Intelligent Systems and Applications*, *17*(1), 31-52.
22. Rao, P. M., & Rao, K. N. (2024). Optimizing crop recommendation systems using advanced deep learning techniques. *Journal of Theoretical and Applied Information Technology*, *102*(23).
23. Nirosha, U., & Vennila, G. (2025). Enhancing crop yield prediction for agriculture productivity using federated learning integrating with graph and recurrent neural networks model. *Expert Systems with Applications*, *289*, 128312.
24. Hasan, M., Marjan, M. A., Uddin, M. P., Afjal, M. I., Kardy, S., Ma, S., & Nam, Y. (2023). Ensemble machine learning-based recommendation system for effective prediction of suitable agricultural crop cultivation. *Frontiers in Plant Science*, *14*, 1234555.
25. Mavaie, P., Holder, L., & Skinner, M. K. (2023). Hybrid deep learning approach to improve classification of low-volume high-dimensional data. *BMC bioinformatics*, *24*(1), 419.
26. Bhat, S. A., Huang, N. F., Sofi, I. B., & Sultan, M. (2021). Agriculture-food supply chain management based on blockchain and IoT: A narrative on enterprise blockchain interoperability. *Agriculture*, *12*(1), 40.
27. Jabed, M. A., & Murad, M. A. A. (2024). Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. *Heliyon*, *10*(24).