INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

ORIGINAL ARTICLE

Check for updates

OPEN ACCESS

Received: 19/11/2025

Accepted: 14/12/2025

Published: 31/12/2025

Citation: Gayathri P, Jayanthi C (2025) IoT Aware Piecewise Optimized Convolutional Deep Belief Network for Crop Yield Prediction in Cloud Environment. Indian Journal of Science and Technology 18(48): 3878-3892. https://doi.org/10.17485/IJST/v18i48.1860

* **Corresponding authors**.

msgayathriphd@gmail.com

jayanthibabu117@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

IoT Aware Piecewise Optimized Convolutional Deep Belief Network for Crop Yield Prediction in Cloud Environment

P Gayathri[1]*, C Jayanthi[1]*

1 PG and Research Department of Computer Science, Government Arts College (Autonomous) Karur-05, Affiliated to Bharathidasan University, Tiruchirappalli, Tamil Nadu, India

Abstract

Objectives: To effectively enhance crop yield prediction accuracy and reduce the prediction time. To achieve accurate yield prediction results and minimize the error rate. **Methods:** A novel model called Piecewise Bayesian Optimized Convolutional Deep Belief Network (PIBO-CDBN) is proposed to deliver highly accurate predictions. It includes data harvesting, data augmentation, preprocessing, feature selection and classification. Performance of the PIBO-CDBN model is evaluated in python using Smart Farming Sensor Data for Yield Prediction dataset based on several key metrics with number of input samples. **Findings:** Results demonstrate that the PIBO-CDBN achieves superior accuracy by 96.44%, precision by 97.17% and recall 97.74% and F1-score by 97.45%. Also, the PIBO-CDBN reduces RMSE by 0.079, and MAPE by 0.0021 than traditional techniques. This optimized fine-tuning process improves the deep neural network's performance and overall learning accuracy in crop yield prediction tasks. **Novelty:** PIBO-CDBN utilizes Improved Convolutional deep belief network (CDBN) with several layers. Data preprocessing addressing missing values using imputation and eliminating outliers. Evaporative Cooling Relief mechanism identifies relevant features. Multivariate Bayesian linear regression analyzes the features and generates final yield predictions. During fine-tuning, error back-propagation algorithms adjust hyperparameters optimally using Adaptive Moment optimizer method.

Keywords: Crop Yield Prediction in Cloud Environment; IoT; Piecewise interpolative robust Autoencoder model; Improved Convolutional deep belief network; Evaporative Cooling Relief mechanism

1 Introduction

Agricultural yield prediction is the process of estimating the amount of crops that harvested from a specific area of land. This estimation plays a vital role in farming practices and food supply planning. Conventionally, yield prediction has relied on analyzing historical agricultural data, weather patterns, and soil conditions. However, these conventional methods often lack the precision for modern large-scale farming. To address this, advanced machine learning and deep learning techniques have been increased for more accurate and timelier predictions[1]. Attention mechanism combined a one-dimensional convolutional neural network and long short-term memory (AM-CNN-LSTM) was developed to perform crop yield estimation[1]. However, it did not minimize time. A Deep Neural Network (DNN) model was designed to forecast cotton yield[2]. Multi-crop and multi-regional analyses were not performed. A novel DNN framework was developed to predict crop yields[3]. Machine learning (ML) and Deep learning (DL) model were developed[4], [5]. Deep integrated models were developed to identify seasonal crop productivity[6]. DL was integrated to enhance rice yield predictions[7]. LSTM with Attention Mechanism was developed based on classification[8]. ConvLSTM was developed for soybean yield prediction[9]. New deep ensemble model was developed by various data sources[10]. Hybrid feature selection approach and optimized ML were developed[11]. Innovative DL was designed with higher accuracy[12]. Boosting enabled efficient ML was developed for accurate prediction of crop yield[13]. Advance DL[14] and LSTM[15] were introduced to detect crop types. Meta-knowledge-guided framework was introduced in[16]. A deep transfer learning[17] and CNN-Transformer[18] were developed for soybean and wheat yield estimation. To enhance the accuracy, Deep Observation-Centric Simple Online and Real-Time Tracking (OC-SORT) model was proposed[19]. Multi-attribute weighted tree-based support vector machine (MAWT-SVM)[20] and integration of IoT and ML[21] were developed for predicting crop yields. CNN-Bidirectional Gated Recurrent Unit (BiGRU) deep learning model was developed[22]. Innovative crop recommendation system using artificial intelligence was designed[23] for supporting sustainable agricultural practices. Predicting crop yields through the use of machine learning was developed in[24]. Experiments are carried out with variety of performance metrics, to demonstrate the improvement of the PIBO-CDBN over conventional deep learning methods. Structure of this paper is organized by. Section 2 outlines the proposed PIBO-CDBN. Section 3 presents the results and discussion. Section 4 concludes the paper with key findings.

Research gap: Agricultural yield prediction is the process of estimating the amount of crops that harvested from a particular area of land. A conventional Deep Neural Network provides insufficient accuracy. Convolutional neural network, long short-term memory gives deficient latency. Optimization technique and deep transfer learning do not achieve higher efficiency. Every existing method unable to focus on computation cost minimization and time minimization. Concerning these issues, a new technique called PIBO-CDBN is developed in cloud computing for improving crop yield prediction with minimum time consumption.

1.1 Novelty of this study

The novelty of PIBO-CDBN model is listed as below.

- To improve IoT aware crop yield prediction, a novel PIBO-CDBN has been developed by introducing the Improved Convolutional deep belief network which includes different processes namely preprocessing, feature selection and prediction.
- The novelty of preprocessing model is performed to eliminate outlier data, thereby reducing the overall size of the dataset by using an improved Convolutional Deep Belief Network (CDBN) method in the PIBO-CDBN model for reducing crop yield prediction time. Additionally, an Evaporative Cooling Relief mechanism is employed to identify and retain significant features while discarding less relevant ones.
- Novel method of CDBN method multivariate Bayesian linear regression technique is employed to improve crop yield prediction accuracy and minimize the error. In addition, the Adaptive Moment optimizer is applied to provide accurate prediction results by reducing the error rate.
- Finally, a comprehensive experimental assessment is carried out, incorporating a variety of performance metrics, to demonstrate the improvement of the PIBO-CDBN over conventional deep learning methods.

2 Methodology

Agriculture is a keystone of the national economy and plays a crucial role in ensuring food security. Novel PIBO-CDBN is introduced for crop yield prediction in cloud computing.

Figure 1 presents the architectural overview of PIBO-CDBN. This system utilizes IoT-based sensors and smart devices deployed across agricultural fields to gather real-time data on soil characteristics, weather patterns and other environmental factors. The collected data is securely transmitted and stored on a cloud-based platform for further analysis. It is structured namely data acquisition, preprocessing, feature selection, classification and hyperparameter tuning.

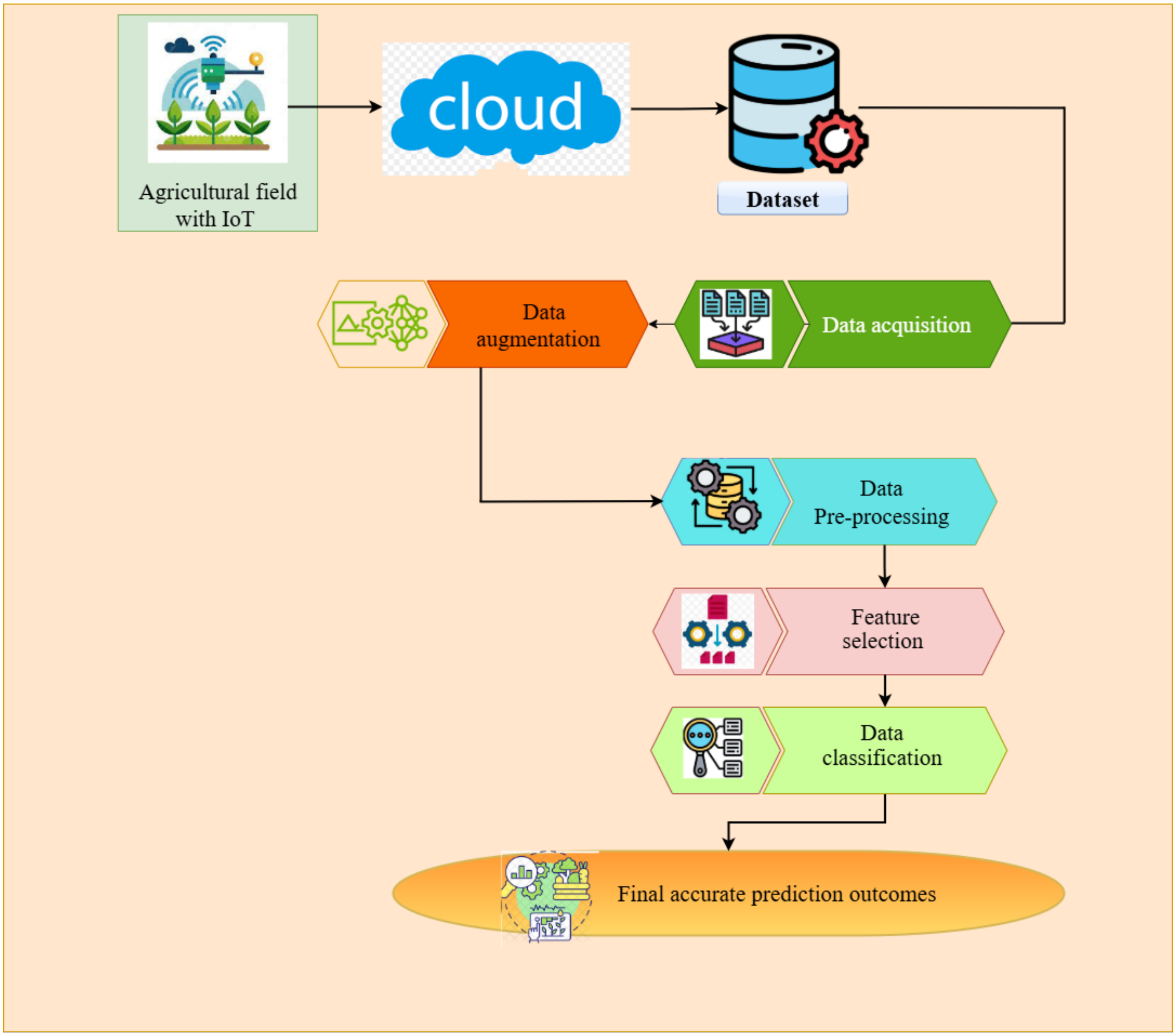

Fig 1. Architecture of Proposed PIBO-CDBN Model

2.1 Data Acquisition

Data acquisition refers to process of collecting the input raw data for early prediction using Smart Farming Sensor Data and Yield Prediction dataset[25] https://www.kaggle.com/datasets/atharvasoundankar/smart-farming-sensor-data-for-yield-prediction. It comprises of information collected from 500 farms located across varied geographic regions, including India, the United States and parts of Africa. The proposed method includes 22 distinct attributes and 500 data instances. These attributes are illustrated in Table 1.

2.2 Data Augmentation

Data augmentation refers to the process of artificially increasing size of the dataset by creating new, modified versions of the existing data. For instance, in this dataset of 500 farm records with 22 attributes, data augmentation techniques simulate new environmental or operational scenarios. After data collection, data augmentation is performed using Piecewise interpolative robust Autoencoder model for generating synthetic data and to increase the records from 50 to 5000. The additional samples are

Table 1. Attribute Description

S. No	Attributes or Column name	Description	S. No	Attributes or Column name	Description
1	farm_id	Unique ID for each smart farm	12	pesticide_usage_ml	Daily pesticide usage in milliliters
2	region	Geographic region (e.g., North India, South USA)	13	sowing_date	Date when crop was sown
3	crop_type	Types of crop grown such as Wheat, Rice, Maize, Cotton, Soybean	14	harvest_date	Date when crop was harvested
4	soil_moisture_%	Soil moisture content in percentage	15	total_days	Crop growth duration (harvest - sowing)
5	soil_pH	Soil pH level (5.5–7.5 typical range)	16	yield_kg_per_hectare	Target variable: Crop yield in kilograms per hectare
6	temperature_C	Average temperature during crop cycle (in °C)	17	sensor_id	ID of the IoT sensor reporting the data
7	rainfall_mm	Total rainfall received in mm	18	Timestamp	Random in-cycle timestamp when the data snapshot was recorded
8	humidity_%	Average humidity level in percentage	19	Latitude	Farm location latitude (10.0 - 35.0 range)
9	sunlight_hours	Average sunlight hours received per day	20	Longitude	Farm location longitude (70.0 - 90.0 range)
10	irrigation_type	Type of irrigation: Drip, Sprinkler, Manual, None	21	NDVI_index	Normalized Difference Vegetation Index (0.3 - 0.9)
11	fertilizer_type	Fertilizer used: Organic, Inorganic, Mixed	22	crop_disease_status	Crop disease status: None, Mild, Moderate, Severe

generated in the form of synthetic data which is a more advanced approach than traditional data augmentation. This process uses the Piecewise Interpolative Robust Autoencoder model to learn the original data sample (500) and generate new data samples (5,000 records). The objective is to address data scarcity by creating a large, diverse, and representative dataset. To ensure the quality and utility of the generated synthetic data, 10-cross validation process is employed.

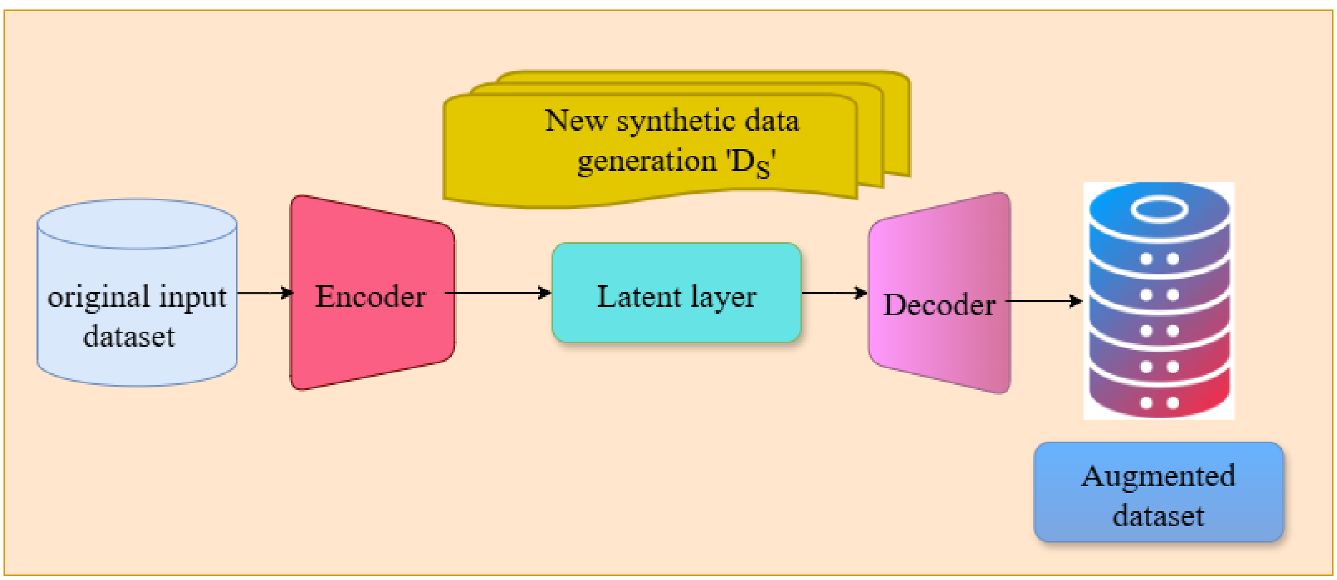

Fig 2. Structure of Piecewise interpolative robust Autoencoder model

Figure 2 illustrates the schematic structure of a Piecewise interpolative robust Autoencoder model which comprises encoder and decoder. Consider dataset 'DS' and data samples 'D' as well as features $\{F_1, F_2, \ldots, F_m\}$ are arranged in form of matrix. Input matrix 'IM' is formulated as,

$$IM = \begin{bmatrix} F_1 & F_2 & \ldots & F_m \\ D_{11} & D_{12} & \ldots & D_{1n} \\ D_{21} & D_{22} & \ldots & D_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ D_{m1} & D_{m2} & \ldots & D_{mn} \end{bmatrix} \tag{1}$$

Where, each column is number of features$F = \{F_1, F_2, \ldots, F_m\}$, each row is number of data samples or instances or records '$D = \{D_1, D_2, \ldots, D_k\}$'. Piecewise interpolative is employed for measuring the weighted average as follows,

$$D_S = \sum_{i=1}^{k} \frac{w_i * D_i}{w_i} \tag{2}$$

$$w_i = \exp\left(-\frac{|D_i - \mu|^2}{\sigma}\right) \tag{3}$$

Where, $D_S$is synthetic data value, $w_i$is weight assigned to neighboring data points 'D_i', $k$is number of neighboring data points used in the weighted average calculation. Aim of decoder is to reduce the reconstruction error in the low-dimensional data.

$$\arg min\, |IM - Z| \tag{4}$$

Where, IM is to a standard deep autoencoder to learn the low-dimensional representations, $Z$is reconstructed output of decoder, $\arg\min$is argument of minimum function. These generated samples are then passed to the generative AI model for accurate crop yield prediction.

2.3 Improved Convolutional Deep Belief Network Based Crop Yield Prediction

Proposed model uses Improved Convolutional deep belief network (CDBN)to provide efficient and accurate data classification while reducing computational time. By combining Deep Belief Networks (DBNs) with the structure of Convolutional Neural Networks (CNNs), the CDBN enhances both feature selection and classification tasks. The architecture consists of multiple layers of Convolutional Restricted Boltzmann Machines (CRBMs) stacked together to simplify learning and minimize errors during training. Figure 3 shows the detailed structure of the CDBN.

Figure 3 presents the architecture of a CDBN developed for accurate crop yield prediction. CDBN is divided into layer-wise pre-training and fine-tuning. During the layer-wise pre-training phase, each layer of the CDBN independently learns to transform its input through weighted computations and transfers the output to the subsequent layer. In the fine-tuning phase, backpropagation is applied across the entire network to refine the model parameters. The main aim is to reduce the prediction error and improve overall prediction performance. CDBNs utilizes the RBMs for layer-wise training. These are stochastic neural networks composed of two layers namely a hidden layer and a pooling layer. CDBN consists of three layers such as an initial visible layer, followed by alternating hidden and pooling layers and output layer. The network processes input augmented data samples. Each layer is made up of neurons that transmit the input data to the next layer. The visible layer directly receives the augmented data. The output generated by one RBM serves as the input to the visible layer of the next RBM, facilitating a layer-by-layer training process in Figure 3. CDBNs consist of training set $\{D_i, Y_k\}$ where $D_i$is training augmented data samples$D = \{D_1, D_2, \ldots, D_n\}$' collected from dataset and a label or output 'Y_k' is prediction output. Input samples are connected to a weight '$\beta_1, \beta_2, \ldots, \beta_n$' and included with bias 'b'. Therefore, the neuron activation probability in the visible layer 'P_V' is given below,

$$P_V = A\left(\sum_{i=1}^{n} D_i * \beta_i\right) + b_v \tag{5}$$

Where, P_V is neuron activation probability in visible layer, A is sigmoid activation function, 'D_i' is input data samples, β_i is weights in visible layer, '$*$' is convolution operator, $b_v$is bias of visible layer. When neuron activation probability $P_V = 1$, input data samples are transferred into the hidden layer of first RBM.

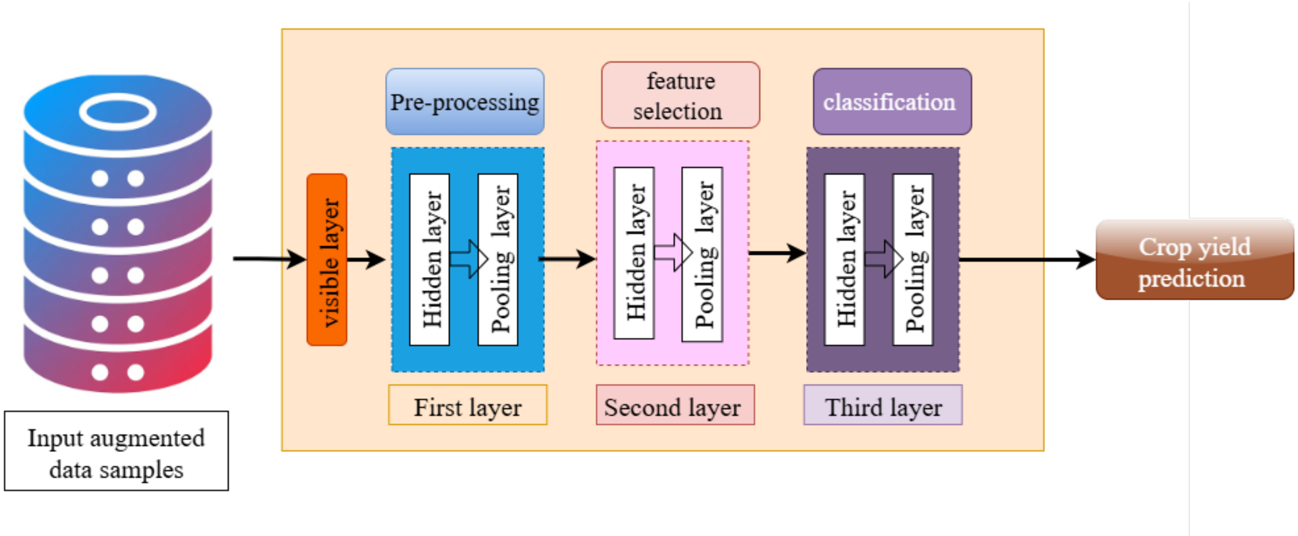

Fig 3. Architecture of CDBN

2.3.1. Data Preprocessing

Data preprocessing step aimed at identifying and correcting inconsistencies within the dataset to enhance its overall quality. This procedure involves addressing missing values, eliminating outliers. Proper data preprocessing significantly improves the reliability and accuracy of crop yield prediction models. Proposed model executes preprocessing by handling both missing data and outlier's removal within the input data matrix. Missing data refers to the absence of values in particular cells for certain features within a dataset. Sibson method is employed for estimating the missing values by taking a weighted average of nearby known data points, where weights are influenced by proximity and distribution.

$$D_{miss} = \sum_{i=1}^{k} D_i \delta_i \tag{6}$$

Where, $D_{miss}$is missing data output, $\delta_i$is weight assigned to neighboring data points 'D_i', $k$is number of neighboring data points. Higher weight is assigned to neighboring data points, while lower weight is given to other data points. By this way, the missing data are filled in the input database. After data handling process, Rosner's generalized statistical test is used in proposed PIBO-CDBN for outlier detection. Outlier detection is a vital task in data analysis for ensuring the quality. Rosner's generalized statistical test is used in proposed PIBO-CDBN to compute the deviation. Rosner's generalized statistical test is employed to identify outlier's data.

$$ST = \frac{\sum_{i=1}^{n} |D_i - \mu_D|}{\sigma} \tag{7}$$

Where ST isstatistical test, D_i is known data samples in input matrix, $\mu_D$is mean of the all the data samples, $\sigma$is deviation between the mean and data samples. If the statistical test (ST) value is high, the corresponding data point is considered an outlier. Otherwise, the data is said to be normal. Identified outliers are removed for further processing.

2.3.2. Feature Selection

Proposed PIBO-CDBN model employs an Evaporative Cooling Relief mechanism within the second hidden layer selecting the more relevant features. Evaporative Cooling Relief approach is an advanced feature selection technique designed for incorporating an iterative 'evaporative' mechanism that gradually removes the least informative features. Main aim is inspired by evaporative cooling, where the system becomes increasingly refined as less useful features are eliminated. In this approach, features are evaluated using similarity scores, which estimate their ability to distinguish between different classes based on nearest-neighbor analysis. These scores are then combined with mutual information, providing a more robust measure of feature relevance by capturing both linear and non-linear dependencies with the target variable. For each iteration, features with the lowest similarity scores are probabilistically removed, allowing the model to focus more on the most meaningful features from the dataset. At first, temperature is pre-defined as 'T'. This temperature controls the probability of feature removal. For each feature F in training and holdout (test) set, compute its importance score using Marczewski-Steinhaus similarity with respect to the target.

$$MSC = \frac{F_j \cap TG_k}{\sum F_j + \sum TG_k - F_j \cap TG_k} \tag{8}$$

Where, $MSC$is Marczewski-Steinhaussimilarity coefficient (i.e. importance score of feature), F_j is feature vector, $TG_k$is target (i.e. crop yield prediction), the intersection symbol '$\bigcap$' is mutual independence between the features and target which are statistically dependent, $\sum F_j$is the sum of $F_j$score, $\sum TG_k$is sum of TG_kscore. 'MSC' returns output values from 0 to $1[0 \leq MSC \leq 1]$. Absolute difference between the two scores is measured as follows,

$$\Delta d_F = |MSC_F(t) - MSC_F(h)| \tag{9}$$

Where, Δd_F is absolute difference between iimportance score of features in training set '$MSC_F(t)$' and similarity score of features in holdout set '$MSC_F(h)$'. For each feature, calculate the probability of evaporation (i.e., being removed) as follows,

$$P(F) = \exp\left(-0.5 * \frac{MSC_F(t)}{T.\Delta d_F}\right) \tag{10}$$

Where, $P(F)$ is probability of evaporation, $MSC_F(t)$ is importance score of features in training data samples, $T$is initial temperature, Δd_F is difference betweenimportance score of features in training set and holdout set. High probability '$P(F)$' means that the feature is said to be removed while features with low probability are retained as significant features.

2.3.3. Classification

Proposed approach applies Multivariate Bayesian Linear Regression by considering multiple output variables simultaneously. Bayesian framework provides robust predictions with large data samples. Likelihood function for the Bayesian linear regression model is formulated,

$$\rho(Y|D, \vartheta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left[-0.5\,\frac{1}{\sigma^2}(Y - D\,\vartheta)^T(Y - D\,\vartheta)\right] \tag{11}$$

Where,$\rho(Y|D, \vartheta, \sigma^2)$ islikelihood function,Yisobserved output (crop yields), D is input data samples with selected features, ϑ is regression coefficients, σ is variance, $n$is number of data samples. Based on high likelihood function, the algorithm predicts the crop yield with collecting relevant features that influence crop growth, such as weather conditions, soil quality and farming practices. This likelihood function measures the relationships between multiple interrelated crop yield outcomes and the selected features samples through the estimating regression coefficients. In this way, crop yield prediction is performed through the regression model. During the fine-tuning phase of the proposed model, the prediction error 'PE' is measured for each sample.

$$PE = [Z_{Ac} - Z_{Ob}]^2 \tag{12}$$

Where, Z_{Ac} is actual output, Z_{Ob} is observed prediction output. To reduce the error, the proposed approach employs the stochastic gradient descent algorithm to iteratively update the weights between the layers of deep learning model. Adaptive Moment optimizer method is employed to enhance the weight adjustment process for faster and more efficient convergence.

$$\beta_{t+1} = \beta_t - \eta\,\frac{M_t}{\sqrt{V_t} + \varepsilon} \tag{13}$$

$$\text{Where, } M_t = \gamma_1 M_{t-1} + (1 - \gamma_1)\left[\frac{\partial PR}{\partial \beta_t}\right] \tag{14}$$

$$V_t = \gamma_2 V_{t-1} + (1 - \gamma_2)\left[\frac{\partial PR}{\partial \beta_t}\right]^2 \tag{15}$$

Where, β_{t+1} is updated weight, β_t is current weight, η is learning rate($\eta < 1$), '$\frac{\partial PR}{\partial \beta_t}$' is partial derivative of prediction error 'PR' regarding current weight 'β_t', γ denotes a Momentum parameter (0.9), M_t is first moment (i.e. mean), V_t is second moment (i.e. variance). Update procedure continues iteratively until the error gets minimized. This fine-tuning process repeats until model attains lowest achievable prediction error. Once optimal weights are determined, final prediction output is generated using the softmax activation function in the output layer.

$$Y = A_{softmax}\left[h_t\right] \tag{16}$$

Where, Y is final output of crop yield prediction, $A_{softmax}$ is softmax activation function, which outputs a probability distribution over the multi-class prediction results based on the regression output from third hidden layer 'h_t'. As a result, algorithm provides better prediction accuracy in crop yield while minimizing the error.

Algorithm 1: Improved Convolutional Deep Belief Network
Input: **Dataset 'DS', features** $F_1, F_2, \ldots, F_m$ and augmented data samples $D_1, D_2, \ldots, D_n$
Output: Improve the crop yield prediction accuracy
Begin
Step 1: Collect number of augmented data samples $D_1, D_2, \ldots, D_n$ from the dataset
Step 2: **Input features** $F_1, F_2, \ldots, F_m$ and data samples $D_1, D_2, \ldots, D_n$ given to visible layer
Step 3: **For each** data samples D_i **do**
Step 4: Compute the neuron activation probability using (5)
Step 5: End For
Step 6: **For each** data samples D_i **do**
Step 7: Perform missing data imputation using (6)
Step 8: **Measure the** generalized statistical test using **(7)**
Step 9: **if** ($ST > T$) **then**
Step 10: Data samples called as outlier
Step 11: **else**
Step 12: Data samples called as normal
Step 13: **End if**
Step 14: **End for**
Step 15: Preprocessed data samples
Step 16: **For each preprocessed** data samples in training set
Step 17: **For each preprocessed** data samples in holdout set
Step 18: Measure the feature importance score using (8)
Step 19: Compute the absolute difference using (9)
Step 20: Calculate the probability of evaporation '$P(F)$' using (10)
Step 21: **if** ($\arg\max P(F)$) **then**
Step 22: Remove the features
Step 23: **else**
Step 24: Select the relevant the features
Step 25: **End if**
Step 26: **End for**
Step 27: **End for**
Step 28: For each selected feature with data samples
Step 29: Measure the likelihood using Bayesian linear regression using (11)
Step 30: **if** $(\rho\left(Y \middle| D, \vartheta, \sigma^2\right) = 1)$ **then**
Step 31: predicts the yield for particular crop

Step 32: End if
Step 33: For each prediction results
Step 35: Compute the prediction error 'PE' using (12)
Step 36: Update the weight using (13)
Step 37:If$arg\ min\ PE$
Step 38: Obtain the final prediction outcomes using (16)**at the output layer**
Step 39: End if
Step 40: End for
End

3 Results and Discussion

Experimental assessment of the proposed PIBO-CDBN model and existing methods namely AM-CNN-LSTM [1] and DNN [2] are implemented using python high level language. In order to conduct the experiment, Smart Farming Sensor Data for Yield Prediction dataset is considered. For the experimental analysis, the crop yield prediction dataset is employed to estimate yield outcomes in kilograms per hectare. The dataset comprises CSV files collected from 500 farms located across various regions, including India, the United States, and parts of Africa. It contains 22 features and 500 records, supporting accurate crop yield forecasting. Simulations are analyzed using performance parameters namely prediction accuracy, precision, recall, F1-score, root mean square error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE). Table 2 lists the hyper-parameters and their description employed in our proposed method.

Table 2. Hyper-parameters and Description

S. No	Hyperparameters	Description
1	Number of layers	Five layers employed (one input layer, three hidden layers and one output layer)
2	Activation function used in first hidden layer	Sibson method and Rosner's generalized statistical test are applied
3	Activation function used in second hidden layer	Marczewski-Steinhaus similarity employed
4	Activation function used in third hidden layer	Bayesian linear regression model utilized
5	Activation function used in the output layer	Softmax activation function used
6	Learning rate	Value of learning rate is 0.01used
7	Batch size	Batch size is 64 used
8	Number of epochs	Number of epochs is 10 employed

3.1 Result of Crop yield prediction accuracy (CPA)

It is computed by the ratio of correctly predicted yield values.

$$CPA = \left(\frac{TP+TN}{TP+TN+FP+FN}\right) * 100 \tag{17}$$

Where, $TP$and TN are true positive and negative, $FP$and FN are false positive and negative.

Table 3 displays a performance comparison of crop yield prediction accuracy using three methods namely PIBO-CDBN model, AM-CNN-LSTM (1) and DNN (2). The horizontal axis denotes the number of input data samples, while the vertical axis reflects the prediction accuracy achieved. The results indicate that the proposed PIBO-CDBN model consistently increases the performance of existing methods in (1) and (2). Let us consider 500 data samples, PIBO-CDBN model increases the accuracy of 96.4%, significantly higher than the 93% and 91% observed by methods (1) and (2), respectively. The evaluation was repeated across several data samples to evaluate accuracy. After achieving ten results, performance of PIBO-CDBN model is compared to existing methods. The average results indicate that the PIBO-CDBN model demonstrated improved accuracy results approximately 4% and 6% compared to (1) and (2), respectively. The higher performance of accuracy is achieved in CDBN by using Bayesian linear regression which efficiently analyzes the input data samples. Based on these regression outputs,

Table 3. Crop yield prediction accuracy versus Data samples

Number of samples	Crop yield prediction accuracy (%)		
	Proposed PIBO-CDBN	**AM-CNN-LSTM (1)**	**DNN (2)**
500	96.4	93	91
1000	96.33	92.98	90.86
1500	96.21	92.56	90.82
2000	96.15	92.32	91.33
2500	96.75	93.02	91.36
3000	96.89	93.22	91.74
3500	96.52	93.54	91.88
4000	96.99	93.63	91.32
4500	96.12	93.88	91.88
5000	96.05	93.06	91.22

the model accurately predicts the crop yield. Further refinement is achieved through adaptive moment optimizer, which fine-tunes model parameters, reduces prediction errors, and boosts the overall accuracy of the system.

3.2 Result of Precision

Precision (**Prec**) is measured by computing the proportion of correctly identified positive cases (true positives) in crop yield out of all instances predicted as positive.

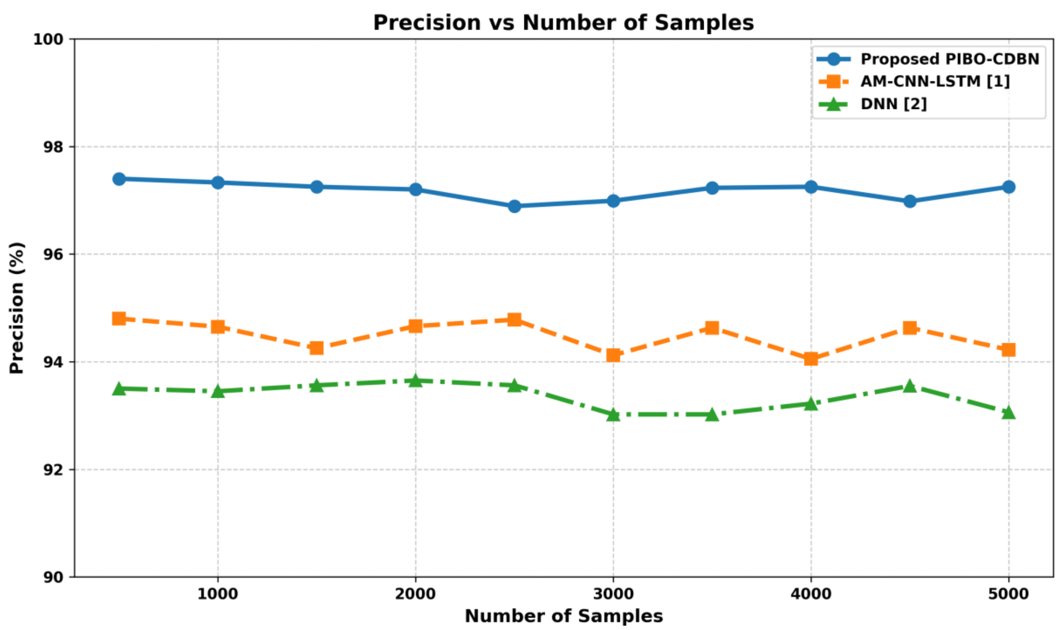

Fig 4. Graphical analysis of precision

Figure 4 demonstrates the precision. However, considering 500 samples, the proposed PIBO-CDBN model achieved a precision of 97.40%, whereas (1) and (2) observed the precisions of 94.80% and 93.50%. PIBO-CDBN enhances precision by 3% and 4% over (1), (2). This enhanced performance results from the utilization of a deep learning classifier. This approach effectively increases the true positive rate while reducing false positives, significantly boosting the overall precision in crop yield prediction.

3.3 Result of Recall or Sensitivity

It measures the model's ability to accurately detect predicts the crop yield.

$$Recall= \left(\frac{TP}{TP+FN}\right) * 100 \tag{19}$$

Table 4. Recall versus Data samples

Number of samples	Recall (%)		
	Proposed PIBO-CDBN	**AM-CNN-LSTM (1)**	**DNN (2)**
500	97.91	96.05	94.73
1000	97.88	95.89	94.62
1500	97.75	95.88	94.55
2000	97.65	95.75	94.36
2500	97.55	95.65	94.21
3000	97.68	95.74	94.05
3500	97.86	95.36	94.12
4000	97.66	95.62	94.05
4500	97.89	95.55	94.08
5000	97.66	95.45	94.33

Table 4 presents the recall. Contrary to existing methods, PIBO-CDBN of recall is enhanced. PIBO-CDBN demonstrates a recall improvement of approximately 2% and 4% over (1), (2). The Bayesian Regression implemented into the CDBN by providing a probabilistic framework for predicting continuous crop yield values. This combination enhances recall by ensuring that the model effectively identifies all relevant yield cases, reducing missed predictions and improving overall accuracy in crop yield forecasting.

3.4 Result of F1 score or F-measure

F1 score represents the harmonic average of precision and recall.

$$F1-Score = 2 * \left(\frac{\mathbf{Prec} * Recall}{\mathbf{Prec} * Recall}\right) \tag{20}$$

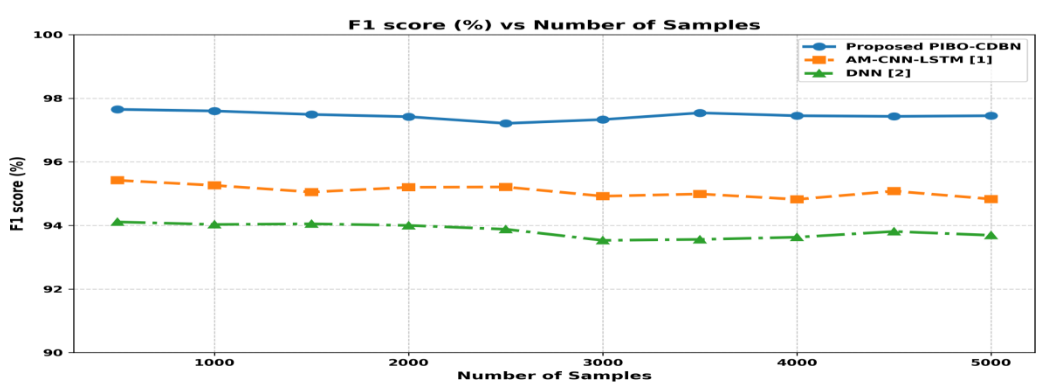

Fig 5. Graphical analyses of F1 score

Figure 5 illustrates the relationship between the F1-score. PIBO-CDBN of F1-score is improved by 3% and 4% over (1), (2). The precision and recall performance is enhanced to the PIBO-CDBN for multiclass prediction scenarios.

3.5 Result of Root Mean Square Error (RMSE)

RMSE is employed to assess the accuracy of crop yield predictions by quantifying the difference between predicted and actual values.

$$\mathbf{RMSE} = \left[\sqrt{\frac{\left(\mathbf{Y}_{\mathbf{Actual}} - \mathbf{Y}_{\mathbf{predicted}}\right)^2}{\mathbf{n}}}\right] \tag{21}$$

Where, Y_{Actual} and $Y_{predicted}$ are actual and predicted crop yield results, n is data samples.

Table 5. RMSE versus Data samples

Number of samples	RMSE		
	Proposed PIBO-CDBN	**AM-CNN-LSTM (1)**	**DNN (2)**
500	0.161	0.313	0.402
1000	0.116	0.221	0.289
1500	0.097	0.192	0.237
2000	0.086	0.171	0.193
2500	0.065	0.139	0.172
3000	0.056	0.123	0.150
3500	0.058	0.109	0.137
4000	0.047	0.100	0.137
4500	0.057	0.091	0.121
5000	0.055	0.098	0.124

Table 5 displays RMSE. PIBO-CDBN reduces RMSE by 48% and 59% compared to (1), (2). The reason for lesser RMSE is to apply Adaptive Moment optimizer method to adjust the weights. By efficiently determining the optimal set of weights, the algorithm improves the model for crop yield prediction.

3.6 Result of Mean square Error (MAE)

MAE is used to evaluate the performance of a crop yield prediction model by measuring the squared difference between predicted and actual values.

$$MSE = \left[\frac{\left(\mathbf{Y}_{\mathbf{Actual}} - \mathbf{Y}_{\mathbf{predicted}}\right)^2}{\mathbf{n}}\right] * 100 \tag{22}$$

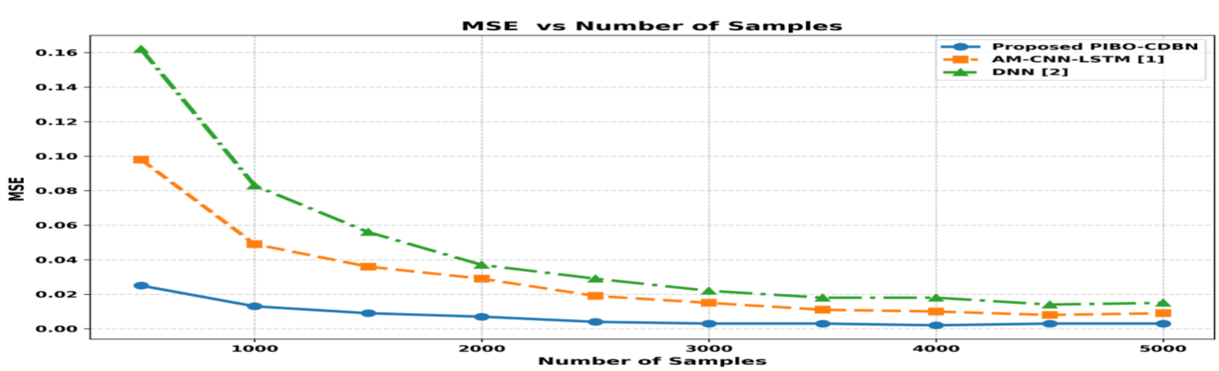

Fig 6. Graphical analyses of MSE

Figure 6 provides MSE. Proposed method records an MSE significantly is lesser than the existing methods. This reduction in MSE highlights the robustness of the PIBO-CDBN model in handling increasing data volumes while maintaining high prediction reliability. Results of PIBO-CDBN minimize MSE by 74% and 84% than (1), (2).

3.7 Result of Mean Absolute Percentage Error (MAPE)

MAPE is used to evaluate the accuracy of a model in predicting crop yield by expressing the prediction error as a percentage of the actual values.

$$MAPE = \frac{1}{n}\left|\frac{\mathbf{Y}_{\mathbf{Actual}} - \mathbf{Y}_{\mathbf{predicted}}}{\mathbf{Y}_{\mathbf{Actual}}}\right| * 100 \tag{23}$$

Table 6. MAPE versus Data samples

Number of samples	MAPE		
	Proposed PIBO-CDBN	**AM-CNN-LSTM (1)**	**DNN (2)**
500	0.0074	0.0150	0.0197
1000	0.0038	0.0075	0.0100
1500	0.0026	0.0053	0.0067
2000	0.0020	0.0041	0.0047
2500	0.0013	0.0030	0.0037
3000	0.0010	0.0024	0.0030
3500	0.0010	0.0019	0.0025
4000	0.0007	0.0017	0.0023
4500	0.0008	0.0014	0.0019
5000	0.0008	0.0014	0.0019

Table 6 presents MAPE. PIBO-CDBN of MAPE reduced 74% and 84% over [2]. Contrary to conventional methods, PIBO-CDBN achieves a significant reduction in MAPE. The MAPE reduced is due to the integration of the Adaptive Moment optimizer to fine tuning algorithm in crop yield prediction.

3.8 Result of Crop yield prediction time (CPT)

It denotes the total time taken by the algorithm to carry out the crop yield prediction task.

$$CPT = \sum_{i=1}^{n} D_i * time(CP) \tag{24}$$

Where, 'D_i' is data and '$time(CP)$' is actual time consumed in crop yield prediction.

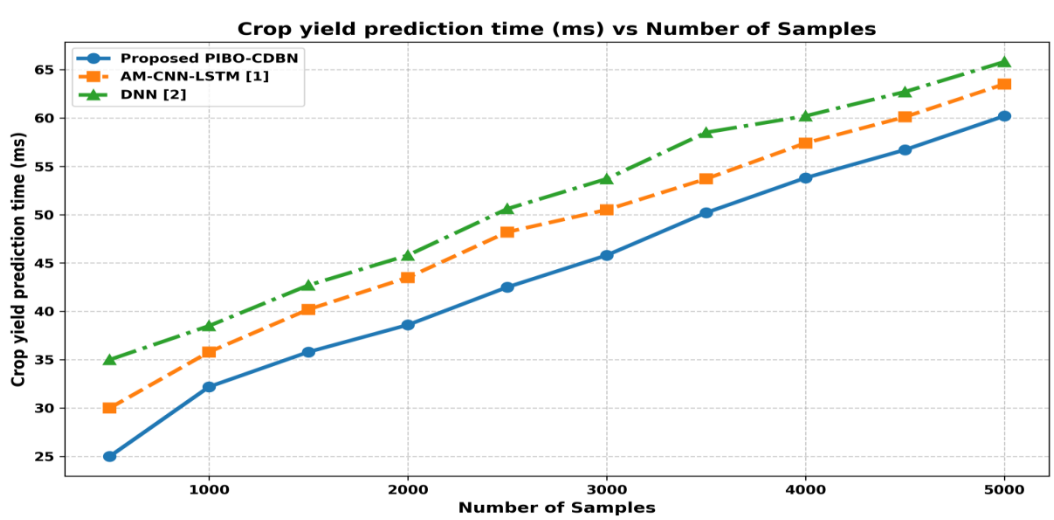

Fig 7. Graphical analyses of Crop yield prediction time

Figure 7 showcases the performance evaluation of crop yields prediction time. However, the PIBO-CDBN demonstrated reduced prediction time than other two models. However, in experiments involving 500 samples, the PIBO-CDBN, existing (1) and (2) achieved prediction time of 25ms, 30ms and 35ms. PIBO-CDBN achieved a reduction in prediction time by 9% and 17% than (1), (2). This is because of Marczewski-Steinhaus similarity for identifying only the most relevant features that minimizes computational time, leading to faster and more efficient crop yield prediction without compromising accuracy.

4 Discussion

The most influential features for crop yield are soil quality and nutrient availability, water management, climate conditions, crop genetics, and management practices. Soil health (including fertility, pH, organic matter content, and structure) is the foundation of productive agriculture. Water is essential for plant survival and growth processes. The timing and amount of water are more critical than the total rainfall received, making efficient irrigation systems in many regions. Temperature, sunlight, and rainfall patterns directly affect photosynthesis, growth rates, and the length of the growing season.

Improving these features in the real world is critical for ensuring food security, increasing farmer income, and promoting environmental sustainability. In this work, PIBO-CDBN model is applied for achieving the accurate crop yield prediction by including data collection, augmentation, preprocessing, feature selection, and classification. Higher yields translate to increased income and economic stability for farmers, especially smallholders in developing countries.

5 Conclusion

Accurate crop yield forecasting, especially across extensive agricultural regions, plays a vital role in handling global food security issues. This study proposes a novel PIBO-CDBN designed to deliver high-precision crop yield predictions while minimizing prediction time. Data preprocessing handle missing values and outlier detection. Evaporative cooling relief approach selects relevant features with lower time. Bayesian linear regression is employed for analyzing the features and generates final yield predictions. Results of PIBO-CDBN enhanced prediction accuracy while reducing both mean square error rate by 0.007 and processing time by 44.08ms than traditional deep learning techniques. The proposed method fails to consider the memory consumption and specificity parameter. In future work, the novel deep learning method will be employed for crop yield prediction.

References

1) Li M, Wang P, Tansey K, Zhang Y, Guo F, Liu J, et al. An interpretable wheat yield estimation model using an attention mechanism-based deep learning framework with multiple remotely sensed variables. *International Journal of Applied Earth Observation and Geoinformation*. 2025;140:1–12. 10.1016/j.jag.2025.104579.
2) Quddus AM, Chowdhury S, Jasper W, Das L. Enhancing cotton yield prediction with robust deep neural network-based framework. *European Journal of Agronomy*. 2025;170:1–12. 10.1016/j.eja.2025.127750.
3) Demirhan H. A deep learning framework for prediction of crop yield in Australia under the impact of climate change. *Information Processing in Agriculture*. 2025;12(1):125–138. 10.1016/j.inpa.2024.04.004.
4) El-Kenawy ESM, Alhussan AA, Khodadadi N, Mirjalili S, Eid MM. Predicting Potato Crop Yield with Machine Learning and Deep Learning for Sustainable Agriculture. *Potato Research*. 2025;68:759–792. 10.1007/s11540-024-09753-w.
5) Subramaniam LK, Marimuthu R. Crop yield prediction using effective deep learning and dimensionality reduction approaches for Indian regional crops. *e-Prime - Advances in Electrical Engineering, Electronics and Energy*. 2024;8:1–9. 10.1016/j.prime.2024.100611.
6) Mustafa G, Moazzam MA, Nawaz A, Ali T, Alsekait DM, Alattas AS, et al. ECP-IEM: Enhancing seasonal crop productivity with deep integrated models. *PLOS ONE*. 2025;20(5):1–23. 10.1371/journal.pone.0316682.
7) Jeong S, Ko J, o Ban J, Shin T, m Yeom J. Deep learning-enhanced remote sensing-integrated crop modeling for rice yield prediction. *Ecological Informatics*. 2024;84:1–11. 10.1016/j.ecoinf.2024.102886.
8) Renukaradya NG, Rao KG, Jayachandra AB. Classification and Estimation of Crop Yield Prediction in Karnataka using LSTM with Attention Mechanism. *International Journal of Intelligent Systems and Applications in Engineering*. 2024;12(3):1–15. Available from: https://ijisae.org/index.php/IJISAE/article/view/5226.
9) Nejad SMM, Abbasi-Moghadam D, Sharifi A. ConvLSTM–ViT: A Deep Neural Network for Crop Yield Prediction Using Earth Observations and Remotely Sensed Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2024;17:17489–17502. 10.1109/JSTARS.2024.3464411.
10) Yewle AD, Mirzayeva L, Karakuş O. Multi-modal data fusion and deep ensemble learning for accurate crop yield prediction. *Remote Sensing Applications: Society and Environment*. 2025;38:1–20. 10.1016/j.rsase.2025.101613.
11) Abdel-salam M, Kumar N, Mahajan S. A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning. *Neural Computing and Applications*. 2024;36:20723–20750. 10.1007/s00521-024-10226-x.
12) Zijin MO, Yanxiong WU. Precision crop yield prediction model based on deep learning and multimodal data fusion. *Turkish Journal of Agriculture and Forestry*. 2025;49(3):580–597. 10.55730/1300-011X.3288.
13) Nagesh OS, Budaraju RR, Kulkarni SS, Vinay M, Ajibade SSM, Chopra M, et al. Boosting enabled efficient machine learning technique for accurate prediction of crop yield towards precision agriculture. *Discover Sustainability*. 2024;5:1–9. 10.1007/s43621-024-00254-x.

14) Sambhe NU, Deshmukh M, Kumar LA, Chavan S, Ranjan NP. MCYP-DeepNet: Nutrition and Temperature Based Season Wise Multi Crop Yield Prediction Using DeepNet 230 Classifier. *Optical Memory and Neural Networks*. 2024;33:236–253. 10.3103/S1060992X24700115.
15) Ashfaq M, Khan I, Shah D, Ali S, Tahir M. Predicting wheat yield using deep learning and multi-source environmental data. *Scientific Reports*. 2025;15:1–20. 10.1038/s41598-025-11780-7.
16) Tunio MH, Li JP, Zeng X, Akhtar F, Shah AS, Ahmed A, et al. Meta-knowledge guided Bayesian optimization framework for robust crop yield estimation. *Journal of King Saud University - Computer and Information Sciences*. 2024;36(1):1–14. 10.1016/j.jksuci.2023.101895.
17) Huber F, Inderka A, Steinhage V. Leveraging Remote Sensing Data for Yield Prediction with Deep Transfer Learning. *Sensors*. 2024;24(3):1–18. 10.3390/s24030770.
18) Du J, Zhang Y, Wang P, Tansey K, Liu J, Zhang S. Enhancing Winter Wheat Yield Estimation With a CNN-Transformer Hybrid Framework Utilizing Multiple Remotely Sensed Parameters. *IEEE Transactions on Geoscience and Remote Sensing*. 2025;63:1–14. 10.1109/TGRS.2025.3544434.
19) Tan S, Kuang Z, Jin B. AppleYOLO: Apple yield estimation method using improved YOLOv8 based on Deep OC-SORT. *Expert Systems with Applications*. 2025;272:1–21. 10.1016/j.eswa.2025.126764.
20) Rajakumaran M, Arulselvan G, Subashree S, Sindhuja R. Crop yield prediction using multi-attribute weighted tree-based support vector machine. *Measurement: Sensors*. 2024;31:1–8. 10.1016/j.measen.2023.101002.
21) Kuradusenge M, Hitimana E, Mtonga K, Gatera A, Habiyaremye J, Ngabonziza J, et al. SMART-CYPS: an intelligent internet of things and machine learning powered crop yield prediction system for food security. *Discover Internet of Things*. 2024;4:1–20. 10.1007/s43926-024-00079-0.
22) Lu J, Fu H, Tang X, Liu Z, Huang J, Zou W, et al. GOA-optimized deep learning for soybean yield estimation using multi-source remote sensing data. *Scientific Reports*. 2024;14:1–19. 10.1038/s41598-024-57278-6.
23) Aarthi S, Manimegalai S, Sakthivel R. AI-based Smart Crop Recommendation System for Sustainable Agricultural Production: A Data-driven Approach to Minimize Resource Use and Maximize Yield. *Madras Agricultural Journal*. 2025. 10.29321/MAJ.10.701221.
24) Balasundaram A, Suraboina VS, Maka B, Thalupula A. Smart agriculture yield prediction using machine learning. *AIP Conference Proceedings*. 2025 Feb;3237(1):020024. 10.1063/5.0248222.
25) Soundankar A. Smart Farming Sensor Data for Yield Prediction. Precision Agriculture Sensor Logs, Environmental Factors & Crop Yield Data. 2025 Feb. Available from: https://www.kaggle.com/datasets/atharvasoundankar/smart-farming-sensor-data-for-yield-prediction.