International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0447
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2832 - 2841

# Prediction of Crop Yield using Deep Learning and Dimensionality Reduction Approaches

*Ramya P[1], Dr Ganesh Kumar P[2]*
*[1]Assistant Professor, Information Technology, Kamaraj College of Engineering &Technology, Virudhunagar, Tamil Nadu, India.*
*[2]Head & Professor, Information Technology, K.L.N College of Engineering, Sivagangai, Tamil Nadu, India.*
***Emails:*** *ramyapandiancsc@gmail.com[1], itz.ganeshkumar@gmail.com[2]*

## Abstract
*India has agriculture as its primary occupation. IBEF (Indian Brand Enquiry Foundation) estimates that 58% of Indians who live in rural areas depend on agriculture. Agriculture plays an important role in the life span of human being not only for the survival but for the better economic growth of the country too. Field-level crop yield prediction (CYP) is essential for quantitative and economic analysis, enabling the formulation of agricultural commodity plans, import-export strategies, and increasing farmer incomes. Crop breeding traditionally demands significant time and resources, making CYP a critical tool for forecasting higher crop production. This study introduces an effective approach combining deep learning (DL) and dimensionality reduction (DR) techniques for CYP, focusing on regional crops in India. The methodology comprises three phases: preprocessing, dimensionality reduction, and classification. Initially, agricultural data from South India is collected and preprocessed through data cleaning and normalization. Subsequently, dimensionality reduction is performed using squared exponential kernel-based principal component analysis (SEKPCA). Finally, a weight-tuned deep convolution neural network (WTDCNN) is employed to predict high crop yields profitably. Simulation results demonstrate that the proposed method achieves superior performance compared to existing techniques, with an impressive accuracy of 98.96%. The novelty of this approach lies in its integration of DL, DR, and WTDCNN, offering precise crop yield predictions tailored to Indian regional crops.*

***Keywords:*** *Crop yield prediction, Machine learning, Deep learning, Principal component analysis, Deep convolution neural network.*

## 1. Introduction
Crop Yield Prediction (CYP) is a crucial process in modern agriculture, aimed at estimating the productivity of crops based on various influencing factors. It involves the use of advanced techniques and technologies to forecast crop yields at field, regional, or national levels. Accurate CYP plays a significant role in agricultural planning, resource allocation, import-export strategies, and boosting farmer incomes, there are a number of factors that can be used to predict crop yield, including:

### 1.1. Environmental Factors
**Weather Conditions:** Temperature, rainfall, humidity, wind speed, and solar radiation. Soil Properties: Soil type, pH level, nutrient content (nitrogen, phosphorus, potassium), organic matter, and moisture levels.Topography: Elevation, slope, and aspect of the land.Climate Data: Seasonal variations, drought indices, and climate anomalies.

### 1.2.Agricultural Factors
**Crop Type:** Specific requirements and growth patterns of the crop. Sowing and Harvesting Time: Planting and harvesting dates. Irrigation Practices: Amount and timing of water application. Fertilizer Use: Type, quantity, and timing of fertilizers applied. The main objective is to reduce the workload of the farmers and increase the productivity of the farms by using technologies like I.O.T, WSNs, Remote Sensing, Drone surveillance and many more. This data can be used to develop predictive models for crop yield estimation and management. However, the large volume and complexity of agricultural data pose significant challenges for data analysis and

modeling. Therefore, there is a need for efficient and effective methods for analyzing and modeling agricultural data to improve crop yield prediction and management Agriculture is the primary food source for India's enormous population and a substantial source of economic support. Due to India's rapid population growth and critical climate changes, the food supply and demand chain must be maintained [1]. To maximize agricultural productivity, agronomic (agriculture scientist) experts have performed an important study to map, monitor, analyze, and manage yield variability. Crop production forecasts are one strategy that can help with crop management [2]. CYP for strategic plants such as rice, maize and wheat are a fascinating field of research for agro meteorologists because it is significant in national and international programming. As a result, there exist systems that estimate accuracy based on meteorological data[4]. The crop yield forecast is currently a difficult task for decision-makers at all global and local levels. Farmers can use a reliable crop production prediction model to determine what and when to sow. Crop production prediction can be accomplished through various methods [5]. Calibration crop models are more easily implemented than simulation crop models because they do not need expert knowledge or user skills, have shorter execution times, and have less storage for data limits [8]. Recently, DL has been used to develop a variety of successful computations since it is used to select the best suitable crop when several options are available [10].Previously proposed architectures for predicting crop yields are frequently hand-designed, with DL approach professionals investigating challenges. They are unable to develop ideal structures because they do not comprehend agriculture. Hence, this paper proposed a practical deep-learning approach with optimal hyper parameters tuning for CYP for Indian regional crops. The main contributions of the work are as follows:

- The pre-processing is performed based on data cleaning and normalization to remove the noise and normalize the dataset.
- The DR is performed using the SEKPCA method to reduce higher dimensional data into lower dimensional data.
- The most profitable crop yield is predicted using the WTDCNN model, and the weights of DCNN are optimally selected using the enhanced whale optimization algorithm (EWOA).

## 2. Related Work

Martin Kuradusenge et al. [12] presented several machine-learning models for CYP. Initially, the Irish potato and maize datasets were collected, and the pre-processing operations, like removal of null values and correlation determination, were carried out to enhance the system's performance further. After that, the classification of the pre-processed data was performed using three ML models, such as random forest (RF), polynomial regression (PR), and support vector machine (SVM), for CYP. Results showed that the RF model attained bets results than the SVM and PR in predicting the crop yields of potato and maize with an RMSE of 510.8 and 129.9 on the tested datasets. Farhat Abbas et al. [13] presented a CYP system through proximal sensing and ML algorithms. Four publicly available datasets such as PE-2017, PE-2018, NB-2017, and NB-2018, were collected to perform training. The collected data were trained on the ML models such as elastic net (EN), linear regression (LR), support vector regression (SVR), and k-nearest neighbor (KNN) for predicting crop yields. The SVR achieved better results for all four tested datasets with lower RMSE than other existing schemes. Dhivya Elavarasan and P. M. Durai Raj Vincent [14] presented a hybrid approach called reinforced RF for CYP with agrarian parameters. Initially, the system collected the crop data from the agrarian dataset and the collected data was fed into the hybrid DL model, namely reinforced RF. The reinforced RF used the reinforcement learning approach in every internal node to determine the significance of the collected input data. The RF then used the most significant data determined using the reinforcement model to classify crop yield. The hybrid approach achieved better results than the existing ML models for CYP, such as SVM, LR, and KNN. Liyun Gong et al. [15] recommended hybrid DL approaches such as recurrent neural networks and temporal convolution networks for CYP. The data was collected from multiple real greenhouse sites for tomato growing. The collected data was pre-processed by performing data normalization, and the normalized data was given to the RNN to process the normalized sequence data. Finally, the output of the RNN was passed to TCN for tomato CYP. The method achieved better results than the existing related schemes for the collected datasets with lower RMSE Paudel et al. [16] proffered an ensemble of machine-learning models for large-scale CYP. Initially, the system collected the crop yield data such as crop growth simulation outputs, weather observations and yield statistics from various sources. The collected data were cleaned for classification processes. Then feature design was applied to some of the input data, and they were fed into the classifier. The ML classifiers such as SVM, KNN, ridge regression, and gradient-boosted decision trees were used for CYP. Aghila Rajagopal et al. [16] presented an optimal deep-learning model for CYP. The collected data were pre-processed, and the relevant features were extracted from the pre-processed dataset using principal component analysis. Then the selected features were further optimized using an improved chicken swarm algorithm to enhance the classifier's performance. Finally, classification was done using a discrete DBN-VGGNet classifier. The system achieved 97 % accuracy and 0.01 % MSE, which was superior to the previous state-of-the-art models. Using a remote sensing and geospatial analysis techniques Zamani et al.Sharifi and Hosseingholizadeh [19] used Sentinel-1 Synthetic Aperture Radar (SAR) data to estimate height and biomass of rice crops in Astaneh-ye Ashrafiyeh, Iran. Machine learning algorithms say Multiple Linear Regression (MLR), Support vector Regression (SVR) and relevance vector regression (RVR) were applied and RVR achieved the best results. Sharifi A develops and evaluates a specific flood detection method based on Sentinel-1 images and classifier algorithms in [20]. An algorithm [21] was proposed to extract optimized features from POLSAR images that are required for estimation. Adespeckling [22] approach based on fast independent component analysis (Fast ICA) algorithm is proposed for improving of the results when polarimetric channels are added. To improve accuracy [23] in the case of limited training samples, the researcher proposed a multiscale dual-branch residual spectral–spatial network (MDBRSSN) with attention to the hyper spectral images (HIS) classification model. The study in [24] identified suitable soil types and locations for saffron cultivation in Miyaneh City using the Weighted Linear Composition (WLC) method and remote sensing data. The previous research highlights using traditional machine-learning algorithms for CYP. Classical ML models are built with specific quantities of training data to forecast agricultural yields depending on specific criteria. However, it has several limitations. For example, features collected from data for creating traditional ML models could not be the most accurate or most representative, resulting in lower yield performance. It must be able to successfully handle data of great volume or complexity. As a result, the authors directed to suggest DL algorithms, although it still requires improvement in the model's prediction rate and computing complexity. Furthermore, previous attempts should have focused on DR, which directly predicts crop production from the dataset, which reduces the interpretation of the DL parameters and requires more storage space. As a result, the proposed system employs optimal DL and DR methodologies to estimate crop yields for Indian regional crops are tabulated as Table 1. Figure 1 shows Work flow of the Proposed system

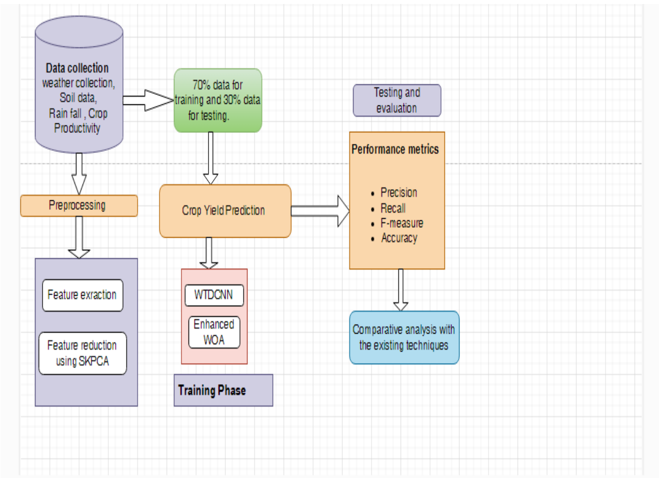


**Figure 1** **Work flow of the Proposed system**

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0447
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2832 - 2841

**Table 1 Summary of Existing Crop Yield Prediction Approaches.**

| Authors | Approach | Techniques/Classifiers Used |
|---|---|---|
| Kavitha et al. [1] | MLbased approach | Combination of DT and regression classifiers used |
| Pant et al. [3] | MLbased approach | DT, Gradient Boosting, SVM |
| Saranya et al. [4] | NeuralNetwork and Population | ANN,CNN |
| Jhajharia et al. [5] | MLbased approach | SVM,RF,LSTM, Gradient Descent, and Lasso regression |
| Gavahi et al. [6] | DLbased approach | CNN, ConvLSTM |
| Feng et al. [9] | DLbased approach | ANN and GTWR |
| Gong et al. [13] | DLbased approach | Recurrent Neural Network (RNN) and Temporal Convolutional Network (TCN), for predicting |

## 3. Proposed Methodology

This paper proposes an optimal DL model with DR approaches for CYP for Indian regional crops. Initially, the agricultural data of the south Indian region, such as rainfall, crop productivity, soil type, and weather data, are collected from publicly available data sources. Then preprocessing of the data is done by applying data cleaning and data normalization. After that, DR is made using SEKPCA, which results in lower dimensional features for CYP. Finally, CYP is made using WTDCNN. The workflow of the proposed work is shown in Figure 1. Figure 2 shows Structure of DCNN

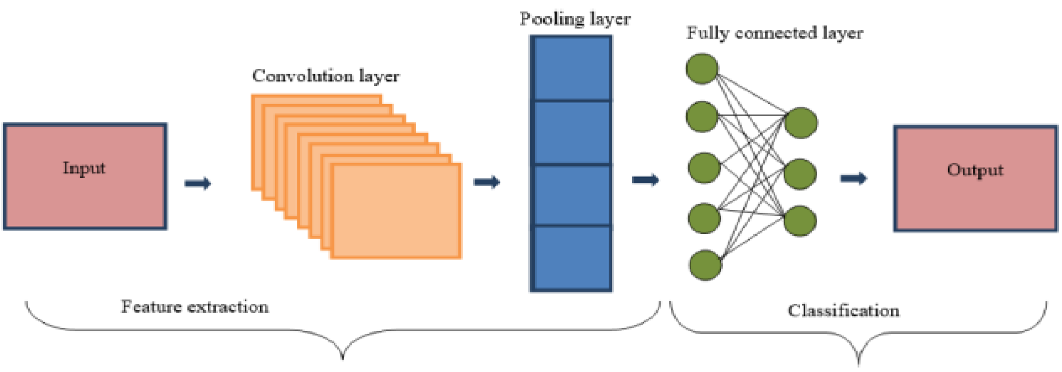


**Figure 2 Structure of DCNN**

### 3.1. Preprocessing

Accurate crop yield prediction is vital for agricultural planning and food security. Data preprocessing is a critical step that ensures data quality and enhances the performance of predictive models. This process involves cleaning raw datasets by addressing missing values, outliers, and duplicates. It also incorporates feature engineering, including scaling numerical attributes, encoding categorical variables, and deriving domain-specific features such as drought indices or growing degree days. Spatial and temporal data integration, including weather patterns, soil characteristics, and satellite imagery, enriches the dataset while ensuring consistency across time and location. The preprocessing steps are explained as

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0447
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2832 - 2841

follows.

**Step 1: Data cleaning:** Data cleaning is the process of detecting, correcting, or removing errors, inconsistencies, and inaccuracies in a dataset to improve its quality and reliability for analysis or modeling. This involves handling missing values, removing duplicates, fixing outliers, standardizing formats, and ensuring data consistency and relevance. Data cleaning is to prepare a dataset that is accurate, complete, and consistent, providing a strong foundation for insightful analysis and robust predictive modeling. The dataset contains 3355 rows (input samples) and 22 columns (features), so the dimension of the matrix fed into the CNN is (3355, 1, 22), which indicates the number of data samples using a kernel size of 1 with 22 features.

**Step 2: Normalization:** Data normalization is a crucial preprocessing technique in machine learning and data analysis that transforms data into a consistent scale without distorting its relative relationships. This process is particularly essential when dealing with features measured on different scales, as it ensures that no single feature dominates the model due to its magnitude.Normalization techniques include Min-Max Normalization, which scales data to a fixed range

$$x' = (\mathbf{x - x\,min})/(\mathbf{x\,max - x\,min}) \qquad \mathbf{(1)}$$

x: Original value, xmin: Minimum value in the feature, xmax: Maximum value in the feature, x′: Normalized value.The dataset values are between 0 and 1 using this min-max normalization.

**Step 3: Splitting the Datasets:** Splitting a dataset involves dividing it into separate subsets for training and testing purposes. 70% of the data for training the model and 30% for testing its performance. The training set is used to teach the model by allowing it to learn patterns, relationships, and features within the data. The testing set, kept unseen during training, evaluates the model's generalization ability on new, unseen data. This 70-30 split ensures that the model has sufficient data to learn effectively while maintaining enough data to validate its accuracy and robustness. The split can be performed randomly or stratified to preserve the distribution of the target variable.

## 3.2. Dimensionality Reduction

After preprocessing, the DR of the dataset is made using squared exponential kernel-based principal component analysis (SEKPCA), is a nonlinear extension of traditional PCA that uses kernel functions to project data into a higher-dimensional feature space, where linear relationships can be identified. The squared exponential kernel, also known as the radial basis function (RBF) kernel, is one of the most commonly used kernels for KPCA due to its ability to capture complex, smooth, and nonlinear patterns. the relationship between different features in the preprocessed dataset is nonlinear, the proposed system incorporates squared exponential kernels (SEK) in conventional PCA, which improves the system's performance by recognizing the nonlinear data and DR of the dataset in an effective manner. Initially, consider the preprocessed dataset with dimensions and analyze the mean vector [2] for each dimension using the following Eq. (2):

$$\mathbf{V}_{\rightarrow \mathbf{sv}} = \frac{\overset{\rightarrow}{\mathbf{P}}_{\mathbf{ds}} - \mu}{\boldsymbol{\sigma}} \qquad \mathbf{(2)}$$

Where, V→sv refers to the scaled value, P→ds indicates the preprocessed dataset, μ and σ represents mean and standard deviation. Then the covariance matrix is computed using SEK, the popular kernel function for the covariance matrix estimation. Using the SEK will result in a smooth prior on functions sampled from the covariance calculation process. To summarize, the SEK function, $SEK(v_x, v_y)$models the covariance between each pair in V→sv. It is expressed as follows:

$$\sum = \mathbf{SEK(v_x, v_y)} \qquad \mathbf{(3)}$$

$$SEK(v_x, v_y) = \sigma^2 \exp\left(-\|vx - vy\|^2 / 2r^2\right) \qquad (4)$$

Where, $\sigma^2$ indicates the overall variance, r signifies the length scale. Next, the eigenvalues and eigenvectors of the covariance matrix can be computed as follows,

$$\sum = \underset{\rightarrow}{\mathbf{M}} \wedge (\underset{\rightarrow}{\mathbf{M}})^{\mathbf{T}} \qquad \mathbf{(5)}$$

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0447
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2832 - 2841

Where,M→indicates the matrix composed of eigenvectors and Λrefers to an eigenvalue diagonal matrix. These eigenvectors are unit eigenvectors whose lengths are both 1. The eigenvectors from the covariance matrix are then ordered by eigenvalue, from highest to lowest. This lists the components in descending order of importance. As a result, the less essential components must be addressed. It is mathematically expressed as follows:

$$\vec{DR}_s = \{ \vec{m1}, \vec{m2}, \vec{m3} \ldots\ldots \vec{mn} \quad (6)$$

Where, DR → s indicates a dimensionality-reduced feature set which consists of significant eigenvectors and n refers to a total number of selected dimensions. Table 2 shows Result of The Classifiers Metrics

**Table 2** Result of The Classifiers Metrics

| Techniques/Metrics (%) | Accuracy | Precision | Recall | F - measure |
|---|---|---|---|---|
| **Proposed WTDCNN** | 98.96 | 98.67 | 99.03 | 98.87 |
| **DCNN** | 96.98 | 96.27 | 97.03 | 96.78 |
| **DBN** | 94.43 | 94.16 | 94.66 | 94.35 |
| **ELM** | 90.34 | 90.02 | 90.46 | 90.27 |
| **RF** | 89.21 | 89.05 | 89.32 | 89.19 |

### 3.3. Crop Yield Prediction

Following DR, CYP is treated with a weight-tuned deep convolution neural network (WTDCNN). Before addressing nonlinearities, each of the layers that make up a DCNN computes convolution transforms and operators for pooling1. Back propagation training in DCNN uses random weight and bias values, increasing the likelihood of achieving suboptimal outcomes and a greater prediction loss. Therefore, to improve detection accuracy and lower network loss, the weights and biases in the network must be properly adjusted. Therefore, the suggested approach uses an EWOA to calculate the weights and bias values of the network, which minimizes the vanishing gradient saturation and prediction loss of the network for CYP and yields optimal results. Fig. 2 depicts the general structure of DCNN. The structure of DCNN comprises '4′ layers such as convolution, pooling, activation, and a fully connected layer. In DCNN, the network weight and biases are chosen randomly for backpropagation training. Instead of choosing them randomly, in the proposed system, they are selected optimally using EWOA to enhance the network's performance in yield prediction. The WOA is a new type of swarm-based optimization algorithm that mimics the humpback foraging behavior of whales. WOA employs three operators to find prey: encircling, researching, and attacking prey. The random population initialization of whales in its initial stages decreases its convergence efficiency and algorithm's quality to get optimal global solutions. In addition, in the later stages of the search process, the algorithm gets stuck into the optimal local issues, which degrades the algorithm's performance. So, the proposed system uses Tent chaotic map to initialize the population, which improves the algorithm's population diversity and convergence efficiency. In addition, the levy flight mechanism is employed for updating whales' position in the later stages of the algorithm, which prevents the system from finding locally optimal solutions. These two enhancements in conventional WOA are termed EWOA. The feature extraction method based on deep learning is implemented by VWCNN. The network consists of

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0447
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2832 - 2841

five convolution layers, five pooling layers and three fully connected layers, and each of the convolution layers is connected to a activation layer using Relu. The input image size is 200 × 200 × 3, and the output size is 6 × 6 × 128 after the fifth convolution layer and the pooling layer, that is, the size of each feature map is 6 × 6, with 128 feature maps. After the first two fully connected layers, the output is a 1024-dimensional vector. After the last fully connected layer, a 5- dimensional vector is output and input to the Softmax classifier for classification. The algorithm starts by initializing the population of the individuals in the search space using a chaotic tent map. Tent chaotic maps have improved population distribution uniformity and search speed, reducing the influence of the initial population distribution. Table 3 shows Analysis of Classification Error

**Table 3** Analysis of Classification Error

| Classifiers/Metrics | MSE | RMSE | FPR | FNR | FRR |
|---|---|---|---|---|---|
| Proposed WTDCNN | 0.034 | 0.219 | 0.029 | 0.065 | 0.061 |
| DCNN | 0.095 | 0.298 | 0.089 | 0.194 | 0.187 |
| DBN | 0.124 | 0.367 | 0.121 | 0.258 | 0.223 |
| ELM | 0.345 | 0.412 | 0.334 | 0.322 | 0.305 |
| RF | 0.398 | 0.483 | 0.379 | 0.423 | 0.402 |

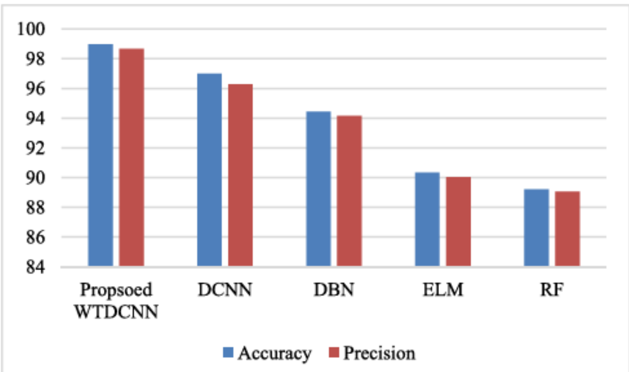


**Figure 3** Analysis of PR and AC

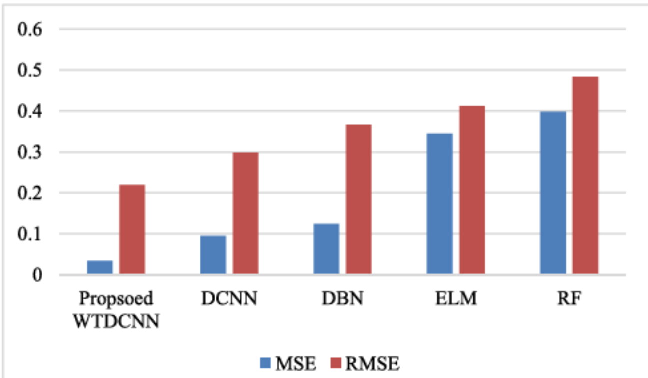


**Figure 5** MSE and RMSE Analysis

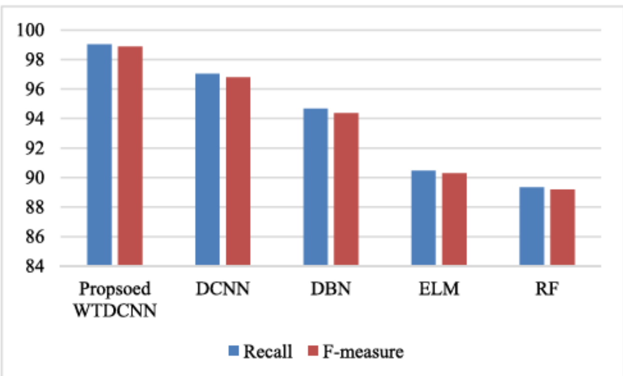


**Figure 4** RC and FM Analysis

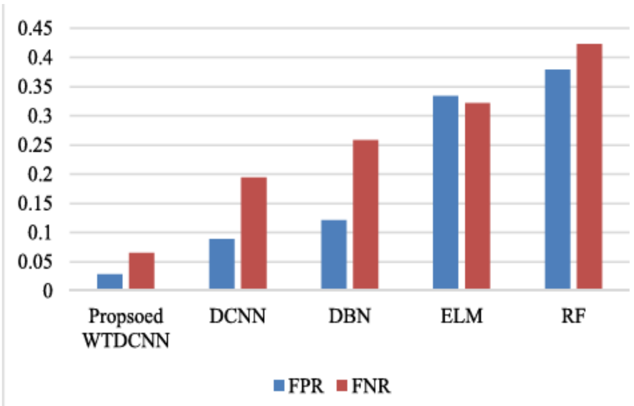


**Figure 6** FPR and FNR Analysis

## 4. Results and Discussion

This section looks at the experimental findings of the suggested yield prediction for Indian regional crops utilizing efficient DL and DR methodologies. The proposed methodology is compared with existing schemes for CYP regarding classification metrics. The predictions were made in Python with an Intel Core i7–8550 CPU, an NVIDIA GEFORCE MX130 graphics card, and 8.0 GB of RAM.

### 4.1. Dataset Descriptions

The proposed system collected the crop production data from the publicly available data source using https://data.world/thatzprem/ agriculture-india, which consists of State Name, District Name, Crop, Year, Season, Crop, Crop class, Area, and Production Yield. Also, the weather data are collected from the Indian website, which consists of minimum temperature ( ◦ temperature ( ◦ C), maximum temperature ( ◦ C), average C), precipitation (mm), humidity (%), pressure, dew point ( ◦ C), wind (m/s).

### 4.2. Performance Analysis

Figure 3 shows the outcomes of the models regarding PR, RC, FM, and AC. From the figure it was clear that the proposed model attains better results than the existing schemes. The proposed WTDCNN attains the PR of 98.67, which is higher than DCNN (96.27), DBN (94.16), ELM (90.02), and RF (89.21). Likewise, the proposed method attains highest accuracy than other existing schemes i.e., the WTDCNN attains an AC of 98.96, whereas the existing schemes such as DCNN, DBN, ELM, and RF attains an AC of 96.98, 94.43, 90.34, and 89.21, which are lower than the proposed scheme. Figure 4 shows the outcomes of the models regarding RC and FM. From the figure it was clear that the proposed model attains better results than the existing schemes. The proposed WTDCNN attains the FM of 98.87, which is higher than DCNN (96.78), DBN (94.35), ELM (90.27), and RF (89.19). Likewise, the proposed method attains highest RC than other existing schemes i.e., the WTDCNN attains an RC of 99.03, whereas the existing schemes such as DCNN, DBN, ELM, and RF attains an RC of 97.03, 94.66, 90.46, and 89.32, which are lower than the proposed scheme. Next, the outcomes of the proposed one are investigated based on error metrics, namely, MSE, RMSE, FPR, FNR, and FRR metrics. This could be given in Table 3. Table 2 demonstrates the outcomes of the proposed one is investi gated against the existing DCNN, DBN, ELM, and RF methods in terms of MSE, RMSE, FPR, FNR, and FRR. The results showed that the proposed method obtains better performance than the existing models by achieving lower error values in classification. The proposed has MSE, RMSE, FPR, FNR, and FRR of 0.034, 0.219, 0.029, 0.065, and 0.061, respectively, which showed more excellent performance than the existing methods because the existing method has higher error values. However, the system is considered a sound system if the system has lower error values. Henceforth, it proved that the proposed system achieved superior performance than the previous existing schemes for accurate CYP. The diagrammatic representation of the Table 2 is given in Figure 5. Figure 6 shows the MSE and RMSE of the proposed and existing clas sifiers. It was clear that the proposed model attains better results than the existing schemes. The proposed WTDCNN attains the MSE of 0.034, which is lower than DCNN (0.095), DBN (0.124), ELM (0.345), and RF (0.398). Likewise, the proposed method attains lowest RMSE than other existing schemes i.e., the WTDCNN attains an RMSE of 0.219, whereas the existing schemes such as DCNN, DBN, ELM, and RF attains an RMSE of 0.298, 0.367, 0.412, and 0.483 which are lower than the proposed scheme (Algorithm 1). Figure 6 shows the FPR and FNR of the proposed and existing classifiers. It was clear that the proposed model attains better results than the existing schemes. The proposed WTDCNN attains the FPR of 0.029, which is lower than DCNN (0.089), DBN (0.121), ELM (0.334), and RF (0.379). Likewise, the proposed method attains lowest FNR than other existing schemes i.e., the WTDCNN attains an FNR of 0.065, whereas the existing schemes such as DCNN, DBN, ELM, and RF attains an RMSE of 0.194, 0.258, 0.322, and 0.423 which are lower than the proposed scheme. The results of our outperforming WTDCNN model demonstrate the novelty of the proposed work by using the proper data preprocessing methods, architecture, and

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0447
e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2832 - 2841

hyperparameters values. Clearly, the pro posed one first preprocesses the dataset before prediction and efficiently utilizes the DR method. So, these approaches are more efficient in making predictions.

## Conclusion And Future Works

This paper suggests efficient DL and DR approaches for CYP for In dian regional crops. The proposed system comprises '3 ′ main phases: preprocessing, DR, and classification. The results of the proposed work are weighted against the conventional DCNN, DBN, ELM, and RF con cerning the accuracy, precision, recall, f-measure, MSE, RMSE, FPR, FNR, and FRR. The outcomes of the proposed one have significant performance because it achieves maximum accuracy of 98.96 % along with 98.67 % precision, 99.03 % recall, and 98.87 % f-measure. In addition, the proposed one attains lower error values of 0.034 MSE, 0.219 RMSE, 0.029 FPR, 0.065 and FNR. The outcomes concluded that the proposed optimal DL approach with a practical DR approach ach ieves superior results than the existing state-of-the-art schemes for CYP. The authors suggest that the proposed method can be extended to other regions and crops to improve the generalizability of the model. They also suggest that the proposed method can be further improved by incor porating additional data sources, such as satellite imagery and remote sensing data. Finally, the authors suggest that the proposed method can be used to develop decision support systems for farmers to help them make informed decisions about crop management and production. In future, the proposed methodology can be integrated with precision agriculture technologies to provide real-time crop yield prediction and decision-making support to farmers. The proposed methodology can be used to develop a web-based application that can be accessed by farmers and policymakers to make informed decisions about crop production and import-export strategies.

## References

[1]. M. Kavita, P. Mathur, Crop yield estimation in India using machine learning, in: Proceedings of the IEEE 5th International Conference on Computing Communication and Automation (ICCCA), IEEE, 2020, pp. 220–224.

[2]. H. Burdett, C. Wellen, Statistical and machine learning methods for crop yield prediction in the context of precision agriculture, Precision Agriculture 23 (5) (2022) 1553–1574.

[3]. J. Pant, R.P. Pant, M.K. Singh, D.P. Singh, H. Pant, Analysis of agricultural crop yield prediction using statistical techniques of machine learning, Mater. Today Proc. 46 (2021) 10922–10926.

[4]. C.P. Saranya, N. Nagarajan, Efficient agricultural yield prediction using metaheuristic optimized artificial neural network using Hadoop framework, Soft Comput. 24 (2020) 12659–12669.

[5]. M. Annamalai, P. Muthiah, An early prediction of tumor in heart by cardiac masses classification in echocardiogram images using robust back propagation neural network classifier, Braz. Arch. Biol. Technol. 65 (2022), https://doi.org/10.1590/ 1678-4324-2022210316.

[6]. K. Gavahi, P. Abbaszadeh, H. Moradkhani, DeepYield: a combined convolutional neural network with long short-term memory for crop yield forecasting, Expert Syst. Appl. 184 (2021) 115511.

[7]. P.S. Nishant, P.S. Venkat, B.L. Avinash, B. Jabber, Crop yield prediction based on Indian agriculture using machine learning, in: Proceedings of the International Conference for Emerging Technology (INCET), IEEE, 2020, pp. 1–4.

[8]. S. Pokhariyal, N.R. Patel, A. Govind, Machine learning-driven remote sensing applications for agriculture in India–a systematic review, Agronomy 13 (9) (2023) 2302.

[9]. M. Shahhosseini, G. Hu, I. Huber, S.V. Archontoulis, Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt, Sci. Rep. 11 (1) (2021) 1–15.

[10]. R. Ali, A. Manikandan, J. Xu, A novel framework of adaptive fuzzy-GLCM segmentation and fuzzy with capsules network (F-CapsNet) classification, Neural

e ISSN: 2584-2854
Volume: 03
Issue: 09 September 2025
Page No: 2832 - 2841

Comput. Appl. (2023), https://doi.org/10.1007/s00521-023-08666-y.
[11]. A. Manikandan, M. Ponni Bala, Intracardiac mass detection and classification using double convolutional neural network classifier, J. Eng. Res. 11 (2A) (2023) 272–280, https://doi.org/10.36909/jer.12237.
[12]. P. Muruganantham, S. Wibowo, S. Grandhi, N.H. Samrat, N. Islam, A systematic literature review on crop yield prediction with deep learning and remote sensing, Remote Sens. 14 (9) (2022) 1990.
[13]. A. Joshi, B. Pradhan, S. Gite, S. Chakraborty, Remote-sensing data and deep- learning techniques in crop mapping and yield prediction: a systematic review, Remote Sens. 15 (8) (2023) 2014.
[14]. F. Abbas, H. Afzaal, A.A. Farooque, S. Tang, Crop yield prediction through proximal sensing and machine learning algorithms, Agronomy 10 (7) (2020) 1046.
[15]. M.M. Islam, M.A.A. Adil, M.A. Talukder, M.K.U. Ahamed, M.A. Uddin, M.K. Hasan, S.K. Debnath, S. Sharmin, M. Rahman, DeepCrop: deep learning-based crop disease prediction with web application, J. Agric. Food Res. 14 (2023) 100764.
[16]. M. Kuradusenge, E. Hitimana, D. Hanyurwimfura, P. Rukundo, K. Mtonga, A. Mukasine, C. Uwitonze, J Ngabonziza, A. Uwamahoro, Crop yield prediction using machine learning models: case of Irish Potato and Maize, Agriculture 13 (1) (2023) 225.
[17]. Palaniappan, M. & Annamalai, M.. (2019). Advances in signal and image processing in biomedical applications. 10.5772/intechopen.88759.
[18]. D. Elavarasan, P.D.R. Vincent, A reinforced random forest model for enhanced crop yield prediction by integrating agrarian parameters, J. Ambient Intell. Humaniz. Comput. 12 (2021) 10009–10022.
[19]. A. Rajagopal, S. Jha, M. Khari, S. Ahmad, B. Alouffi, A. Alharbi, A novel approach in prediction of crop production using recurrent cuckoo search optimization neural networks, Appl. Sci. 11 (21) (2021) 9816
[20]. Kolli, S. & V., Praveen & John, A. & Manikandan, A.. (2023). Internet of things for pervasive and personalized healthcare: architecture, technologies, components, applications, and prototype development. 10.4018/978-1-6684-8913-0.ch008
[21]. A.R. Venmathi, S. David, E. Govinda, K. Ganapriya, R. Dhanapal, A. Manikandan, An automatic brain tumors detection and classification using deep convolutional neural network with VGG-19, in: Proceedings of the 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, 2023, pp. 1–5, https://doi.org/ 10.1109/ ICAECA 56562. 2023.10200949.
[22]. A. Sharifi, M. Hosseingholizadeh, Application of Sentinel-1 data to estimate height and biomass of rice crop in Astaneh-ye Ashrafiyeh, Iran, J. Indian Soc. Remote Sen. 48 (2020) 11–19
[23]. A. Sharifi, Development of a method for flood detection based on Sentinel-1 images and classifier algorithms, Water Environ. J. 35 (3) (2021) 924–929.
[24]. A. Kosari, et al., Remote sensing satellite's attitude control system: rapid performance sizing for passive scan imaging mode, Aircr. Eng. Aerosp. Technol. 92 (7) (2020) 1073–1083, https://doi.org/10.1108/aeat-02-2020-0030.