International Research Journal of Multidisciplinary Scope (IRJMS), 2024; 5(1):101-112

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2024.v05i01.0158

A Novel Crop Yield Prediction Using Deep Learning and Dimensionality Reduction

Kandasamy Subramaniam Leelavathi*, Marimuthu Rajasenathipathi

Department of Computer Technology, Nallamuthu Gounder Mahalingam College, Pollachi, India *Corresponding Author's Email: ksleelangmc@outlook.com

Abstract

Crop yield prediction (CYP) at the field level is crucial in quantitative and economic assessment for creating agricultural commodities plans for import-export strategies and enhancing farmer incomes. Crop breeding has always required a significant amount of time and money. CYP is developed to forecast higher crop production. This paper proposes an efficient deep learning (DL) and dimensionality reduction (DR) approaches for CYP for Indian regional crops. This paper comprised '3' phases: preprocessing, DR, and classification. Initially, the agricultural data of the south Indian region are collected from the dataset. Then preprocessing is applied to the collected dataset by performing data cleaning and normalization. After that, the DR is performed using squared exponential kernel-based principal component analysis (SEKPCA). Finally, CYP is based on a weight-tuned deep convolutional neural network (WTDCNN), which predicts the high crop yield profit. The simulation outcomes shows that the proposed method attains superior performance for CYP compared to exiting schemes with an improved accuracy of 98.96%.

Keywords: Crop Yield Prediction, Deep Convolutional Neural Network, Machine Learning, Deep Learning, Principal Component Analysis.

Introduction

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 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 is critical in food production (3). CYP for strategic plants such as rice, maize and wheat are a fascinating field of research for agrometeorologists 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).

Machine learning (ML) is one of the methods used to forecast agricultural yields, along with SVM, RF, DT, and others (6, 7). 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). Despite developing numerous ML models to increase prediction accuracy, spatial and temporal non-stationarity, inherent in many geographical phenomena, is rarely incorporated in agricultural production modelling (9). 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). It is an ML class with multiple layers of neural networks capable of learning from data (11). It seeks to produce predictions by establishing relationships between input and response variables. However, a critical difficulty with DL is its reliance on hyper-parameters, which can be avoided to improve the effectiveness of the results. 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

(Received 20th October 2023; Accepted 05th January 2024; Published 30th January 2024)

comprehend agriculture. Hence, this paper proposed a practical deep-learning approach with optimal hyperparameters 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 dimenssional 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).

The motivation is to develop an efficient deep learning (DL) and dimensionality reduction (DR) approaches for crop yield prediction (CYP) for Indian regional crops. The aim is to forecast higher crop production, which is crucial in quantitative and economic assessment for creating agricultural commodities plans for import-export strategies and enhancing farmer incomes. The paper proposes an optimal DL model with DR approaches for CYP for Indian regional crops, which attains superior performance compared to existing schemes with an improved accuracy of 98.96%.

The remaining portion of the manuscript is outlined: Section 2 gives the related work regarding CYP. Section 3 presents the proposed methodology. Section 4 explains the performance of the proposed model by comparing the results obtained, and finally, section 5 concludes the proposed work with future directions.

Farhat Abbas *et al.* (12) 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. Martin Kuradusenge *et al.* (13) 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.

Liyun Gong *et al.* (14) recommended hybrid DL approaches such as recurrent neural networks and temporal convolutional 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. The work (15) 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.

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. Dilli Paudel *et al.* (17) 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.

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.

Problem Definition

The problem addressed in this paper is crop yield prediction (CYP) for Indian regional crops. The aim is to forecast higher crop production, which is crucial in quantitative and economic assessment for creating agricultural commodities plans for import-export strategies and enhancing farmer incomes. The paper proposes an efficient deep learning (DL) and dimensionality reduction (DR) approaches for CYP for Indian regional crops.

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.

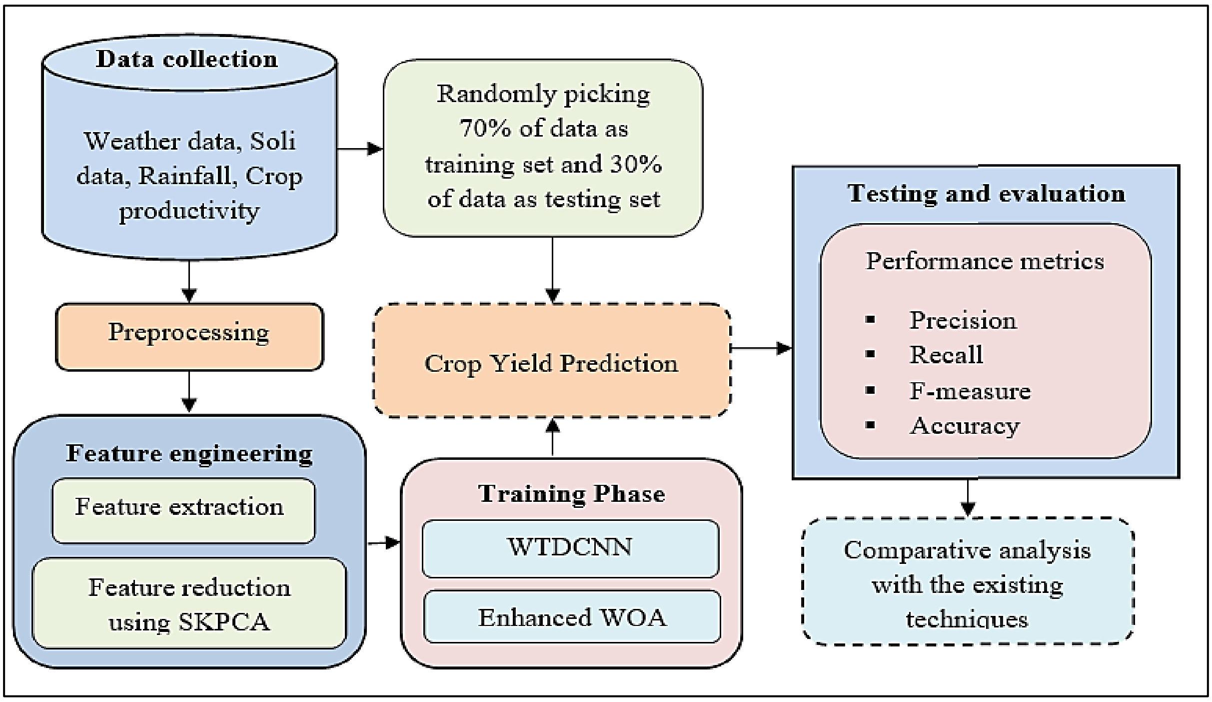

Figure 1: Workflow of the proposed system

Preprocessing

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. Following that, data preparation or preprocessing is undertaken since data is acquired from various sources. It is collected in raw format, which is not suitable for analysis. So preprocessing is most important before predicting the crop yield to improve the prediction rate. The preprocessing steps are explained as follows.

Step 1: Data cleaning

After gathering data from repositories, data cleaning is performed through missing value imputation and outliers' elimination. Missing values influence the model's accuracy in the data. As a result, the missing values are replaced with the mean or median values of the entire dataset or some other summary statistic. Outlier removal is conducted after missing value imputation to reduce noise from the dataset. The most straightforward technique to eliminate outliers from the data set is to delete them, which improves data quality.

Step 2: Normalization

After performing data cleaning, normalization of the dataset is done. Normalization aims to convert data to be dimensionless and have similar distributions. It is mathematically expressed as follows:

$$\underset{\rightarrow}{C}''_{Norm} = \frac{\underset{\rightarrow}{C}'' - \underset{\rightarrow}{C}''_{\min}}{\underset{\rightarrow}{C}''_{\max} - \underset{\rightarrow}{C}''_{\min}} \qquad [1]$$

Where, $\underset{\rightarrow}{C}''_{Norm}$ refers to the normalized data, $\underset{\rightarrow}{C}''$ indicates the original data, $\underset{\rightarrow}{C}''_{\min}$ and $\underset{\rightarrow}{C}''_{\max}$ signifies the minimum and maximum value from the data set. The dataset values are between 0 and 1 using this min-max normalization.

Step 3: Splitting the datasets

The preprocessed dataset is portioned into training and testing datasets to implement the proposed system. The proposed system randomly picks 70 % data for training and 30 % data for testing.

Dimensionality Reduction

After preprocessing, the DR of the dataset is made using squared exponential kernel-based principal component analysis (SEKPCA), which transforms the higher-dimension data into a lower dimension. PCA operates by computing the principal components and changing the basis. It solves the variable's correlation and can significantly improve crop yield detection and diagnosis of high-dimensional data in the actual production process. Even so, PCA is only effective if the variables are all highly uncorrelated. In addition, the PCA has difficulty recognizing nonlinear data models. Because 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 for each dimension using the following equation 2.

$$\underset{\rightarrow}{V}_{sv} = \frac{\underset{\rightarrow}{P}_{ds} - \mu}{\sigma} \qquad [2]$$

Where, $\underset{\rightarrow}{V}_{sv}$ refers to the scaled value, $\underset{\rightarrow}{P}_{ds}$ 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 $\underset{\rightarrow}{V}_{sv}$. It is expressed as follows:

$$\sum = SEK(v_x, v_y) \qquad [3]$$

$$SEK(v_x, v_y) = \sigma^2 \exp\left(-\frac{\left\|v_x - v_y\right\|^2}{2r^2}\right) \qquad [4]$$

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

$$\sum = \underset{\rightarrow}{M}\Lambda\left(\underset{\rightarrow}{M}\right)^T \qquad [5]$$

Where, $\underset{\rightarrow}{M}$ indicates the matrix composed of eigenvectors and Λ refers 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:

$$\underset{\rightarrow}{DR}_s = \{\underset{\rightarrow}{m}_1, \underset{\rightarrow}{m}_2, \underset{\rightarrow}{m}_3, \ldots\ldots\ldots, \underset{\rightarrow}{m}_n\} \quad [6]$$

Where, $\underset{\rightarrow}{DR}_s$ indicates a dimensionality-reduced feature set which consists of significant eigenvectors and n – refers to a total number of selected dimensions.

Crop Yield Prediction

After DR, a weight-tuned deep convolutional neural network (WTDCNN) is used for CYP. DCNN comprises several layers, each of which computes convolutional transforms before moving on to nonlinearities and pooling1 operators. In DCNN, the random weight and bias values are utilized for backpropagation training, which increases the chances of getting sup optimal results and higher loss in the prediction process. So proper tuning of weight and biases in the network is essential to enhance the detection accuracy and reduces the loss of the network. As a result, the proposed system employs an EWOA to determine the network's weights and bias values, which produces optimal results by minimizing the vanishing gradient saturation and prediction loss of the network for CYP. Figure 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 behaviour 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.

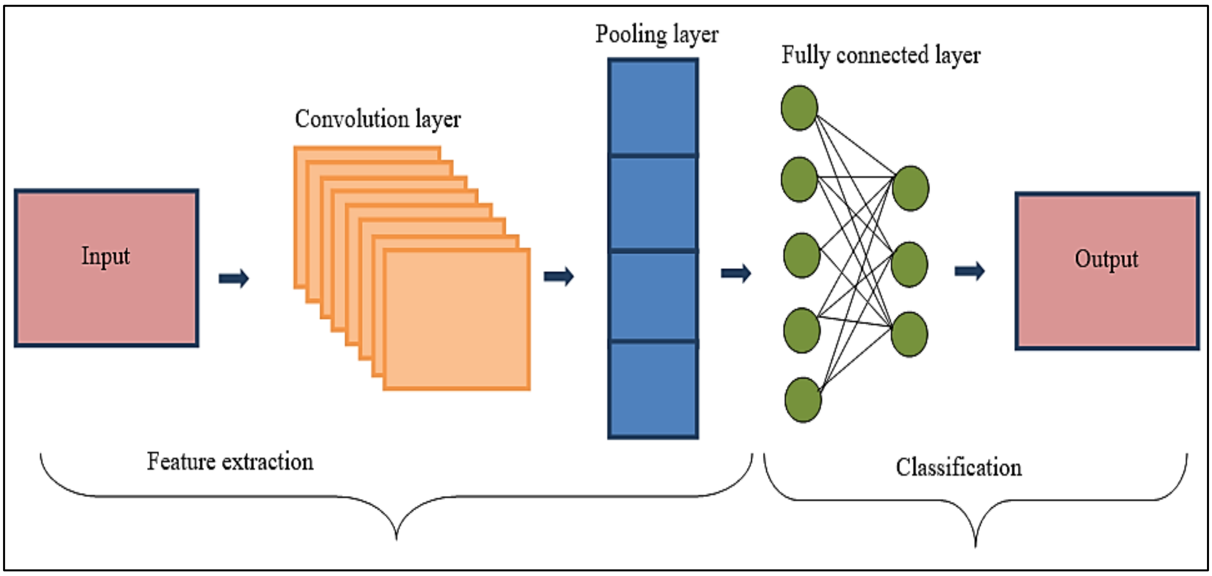

Figure 2: Structure of DCNN

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. It is written as follows:

$$\underline{Z}_{\tau+1} = \begin{cases} 2\underline{Z}_{\tau} & 0 \le \underline{Z} \le 0.5 \\ 2(1-\underline{Z}_{\tau}), & 0.5 < \underline{Z} \le 1 \end{cases} \quad [7]$$

Where, $\underline{Z}_{\tau+1}$ refers to the whales' initial population using a tent map and $\underline{Z}_{\tau}$ indicates the random population. Then the fitness $(\underline{FN}_{cal})$ of the whales in the initialized population is estimated using the classifier's mean square error (MSE). MSE is computed by taking the difference between the actual output and the predicted output of the classifier in yield prediction. It is expressed as follows:

$$\underline{FN}_{cal} = Min(MSE) \quad [8]$$

$$MSE = \frac{1}{d}\sum_{p=1}^{d}\left(q_{val} - q_{val}^{*}\right) \quad [9]$$

Where, q_{val} and q_{val}^{*} indicates the actual and predicted value of the classifier and d – refers to the number of samples in the training dataset. Then the position of the whales can be detected to surround them. The whale close to prey location is considered the best whale $\underline{Z}^{*}$ in the current population and the position of other whales is updated based on $\underline{Z}^{*}$ as follows.

$$\underline{DT} = \left|\beta \times \underline{Z}^{*}(\tau) - \underline{Z}(\tau)\right| \quad [10]$$

$$\underline{Z}(\tau+1) = \underline{Z}^{*}(\tau) - \alpha \times \underline{DT} \quad [11]$$

Where, τ indicates the current iteration and $\underline{DT}$ refers to the distance betwixt the prey $\underline{Z}^{*}(\tau)$ and the whale $\underline{Z}(\tau)$. In addition, α and β represents the coefficient vectors computed using equations (12) and (13)

$$\alpha = 2 \times l_d \times R_{num} - l_d(\tau) \quad [12]$$

$$\beta = 2 \times R_{num} \quad [13]$$

Where, R_{num} indicates a random number ranges between [0, 1] and l_d is decreased linearly from 2 to 0 over the number of iterations, Then the humpback whales' bubble-net behavior is updated using the following equation: [14]

$$\underline{Z}(\tau+1) = \begin{cases} \underline{Z}^{*}(\tau) - \alpha \times \underline{DT} & if\ p < 0.5 \\ (\underline{DT})' \times e^{h_k g_{lk}} \times \cos(2 \times \pi \times g_{lk}) + \underline{Z}^{*}(\tau) & if\ p \ge 0.5 \end{cases}$$

Whereas, p represents a random integer between [0, 1] and shows the likelihood of updating the position of whales according to the spiral updating position $(if\ p \ge 0.5)$ or shrinking encircling technique $(if\ p < 0.5)$, g_{lk} denotes an arbitrary integer between [−1, 1], and h_k defines the spiral movement shape. Then the global search process (exploration) of the whales is executed, and completed when the absolute vectorvalue is greater or equal to one. Otherwise, the algorithm implements the exploitation phase. Instead of considering the best whale $\underline{Z}^{*}$, the random value $\underline{Z}_{rand}$ is considered in the exploration phase to update whales' positions, which is expressed as follows:

$$\underline{DT} = \left|\beta \times \underline{Z}_{rand} - \underline{Z}(\tau)\right| \quad [15]$$

$$\underline{Z}(\tau+1) = \underline{Z}_{rand} - \alpha \times \underline{DT} \times \mathrm{L_f}(\xi) \quad [16]$$

Where $\underline{Z}_{rand}$ refers to the arbitrarily chosen whale from the current population, and $\mathrm{L_f}(\xi)$ indicates a levy flight mechanism, which enhances the exploration and exploitation capabilities of the algorithm. It is arandom walk where the steps are denoted regarding step lengths with a given probability distribution and is written as follows:

$$\mathrm{L_f}(\xi) \sim |\xi|^{-1-\lambda} \quad [17]$$

$$\xi = \frac{N_{ud}}{|N_{vd}|^{1/\lambda}} \quad [18]$$

Where, $\lambda(0 < \lambda \le 2)$ signifies an index, ξ indicates the step length, N_{ud} and N_{vd} represents drawn from normal distributions. After optimally chosen weights and biases, the convolution layer extracts relevant features from the dimensionality-reduced dataset. As the biases and weights accept the best values at each iteration, the likelihood of an improved model gradually rises.

$$FeatureVector = \sum\left(\underline{DR}_s + \breve{O}^{*}_{d\times d}\right) + \breve{B}^{*} \quad [19]$$

Where, $\underrightarrow{DR}_s$ refers to the dimensionality reduced dataset on which the convolutional operation is performed, $\breve{O}^*_{d\times d}$ and $\breve{B}^*$ indicates the optimal filter weights and bias selected by EWOA, and $d-$denotes the kernel size. The final feature vector obtained is then inputted into the activation layer to increase the nonlinearity in the output. ReLU is used as an activation function in the activation layer that outputs the input directly if the input value is positive. Otherwise, it will output zero. The output of the convolution layer with activation is fed into the pooling layer to minimize the input data size. The polling layers use smaller rectangular boxes of the convolution layer and produce the output by sampling the convolution's rectangular boxes. The output of the polling layers is given to the fully connected layer for performing CYP, which uses the SoftMax activation function to perform the classification. The classifier's output shows the productivity of different crops in Indian regions under various seasons that help the farmers to plant the crops according to their productivity level in future.

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.

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 (°C), maximum temperature (°C), average temperature (°C), precipitation (mm), humidity (%), pressure, dew point (°C), wind (m/s).

Performance Analysis

Here the outcomes of the proposed classification model (WTDCNN) are compared with the existing classification schemes namely with the existing DCNN, Random Forest (RF) models Deep Belief Network (DBN), and Extreme Learning Machine (ELM).The techniques are compared based on precision (PR), recall (RC), f-measure(FM) and accuracy (AC), MSE, Root Mean Square Error (RMSE), False Positive Rate (FPR), and False Negative Rate (FNR). The equations for the above metrics are given as follows

$$\mathrm{PR} = \frac{Tp}{Tp+Fp} \quad [20]$$

$$\mathrm{RC} = \frac{Tp}{Tp+Fn} \quad [21]$$

$$\mathrm{AC} = \frac{Tp}{Tp+Tn+Fp+Fn} \quad [22]$$

$$\mathrm{FM} = \frac{Tp}{Tp+1/2(Fp+Fn)} \quad [23]$$

$$\mathrm{FPR} = \frac{Fp}{Fp+Tn} \quad [24]$$

$$\mathrm{FNR} = \frac{Fn}{Fn+Tp} \quad [25]$$

$$MSE = \frac{1}{N}\sum_{i=1}^{n}(\mathrm{A}-\mathrm{B})^2 \quad [26]$$

$$RMSE = \sqrt{\mathrm{MSE}} \quad [27]$$

Where, $Tp, Tn, Fp, and\ Fn$ indicates the true positive, true negative, false positive, and false negative values of the classifier, and A and B denotes the original and predicted values of the dataset. The outcomes of the models regarding PR, RC, FM, and AC are tabulated in table 1.

Table 1: Results of the classifiers regarding detection 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

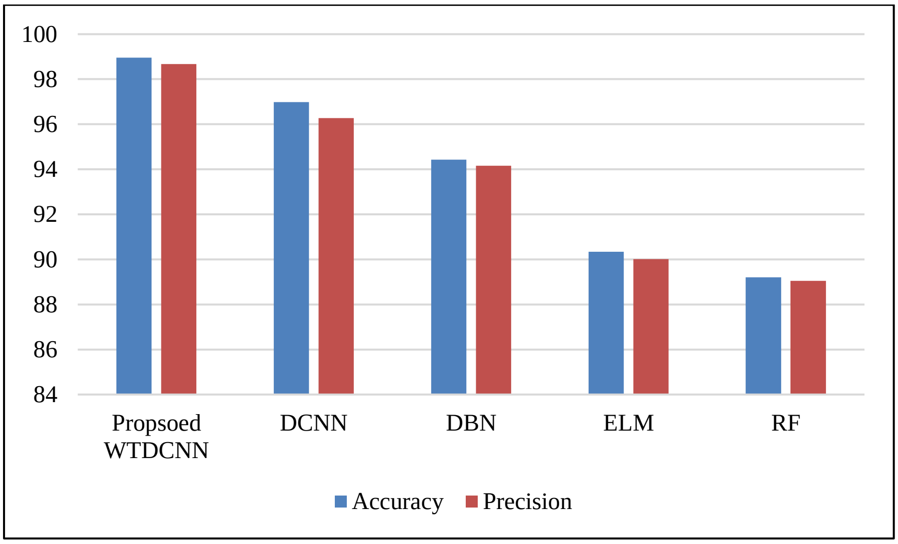

Figure 3: Analysis of PR and AC

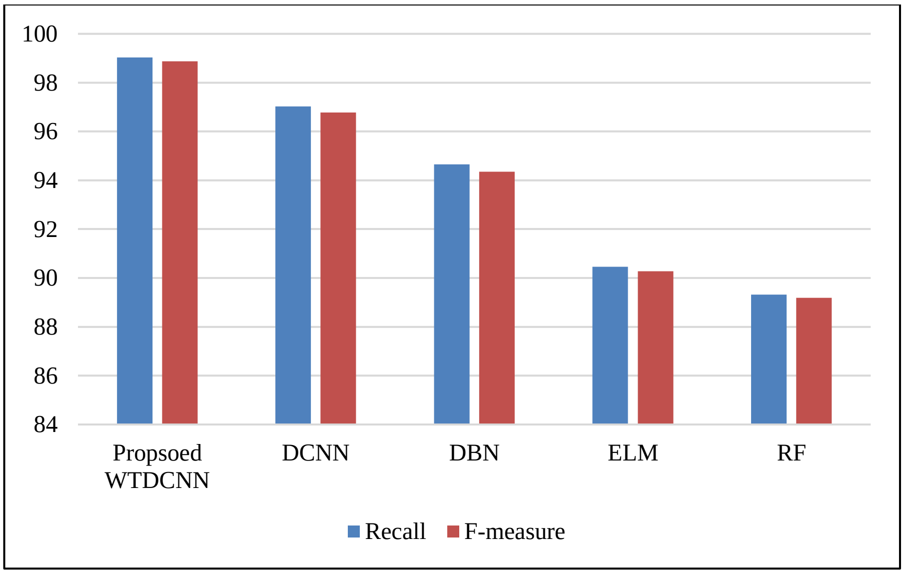

Figure 4: RC and FM analysis

Table 1 has demonstrated that DL can play an important role in CYP, and our results confirmed the same. Despite being based on essential performance criteria, the outcomes are compared with other state-of-the-art methodologies. The suggested WTDCNN produces better results than the existing ones. For example, the existing DCNN achieves accuracy, precision, recall, and f-measure of 96.98%, 96.27%, 97.03%, and 96.78%, respectively. Also, the existing RF attains minimal 89.21% accuracy, 89.05% precision, 89.32% recall, and 89.19% f-measure, which is lower than the proposed one, because the proposed one achieves maximum accuracy of 98.96% along with 98.67% precision, 99.03% recall, and 98.87%f-measure. Similarly, considering other existing methods (DBN and ELM), the proposed one achieves more excellent performance. Thus, the outcomes proved that the proposed one outperformed the conventional methodology. The diagrammatic representation of the table 1 is shown in figure 3.

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 2. Table 2 demonstrates the outcomes of the proposed one is investigated 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 and 6.

Table 2: 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

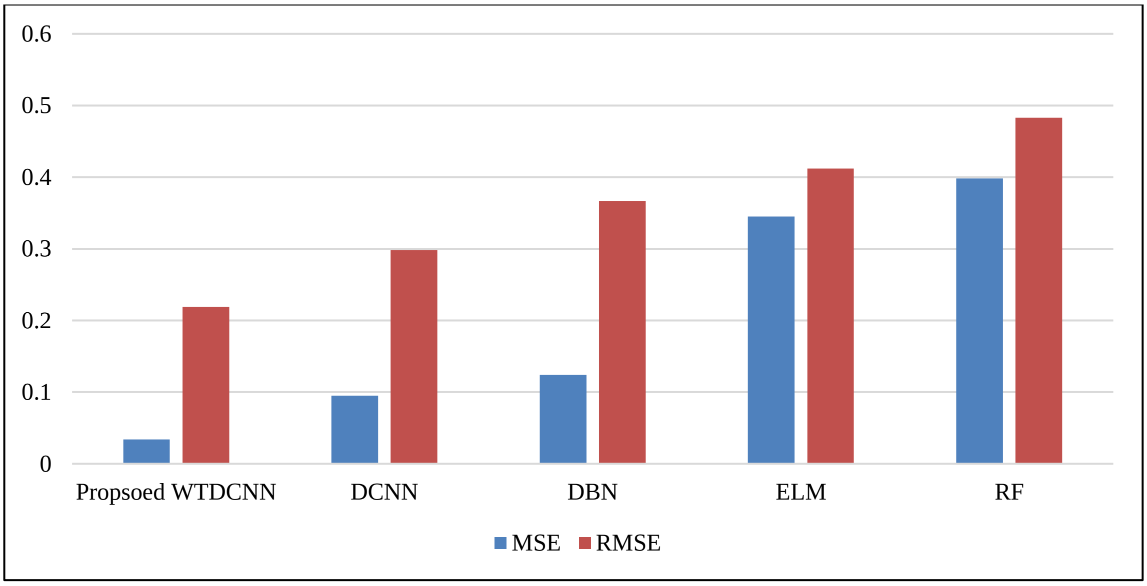

Figure 5: MSE and RMSE analysis

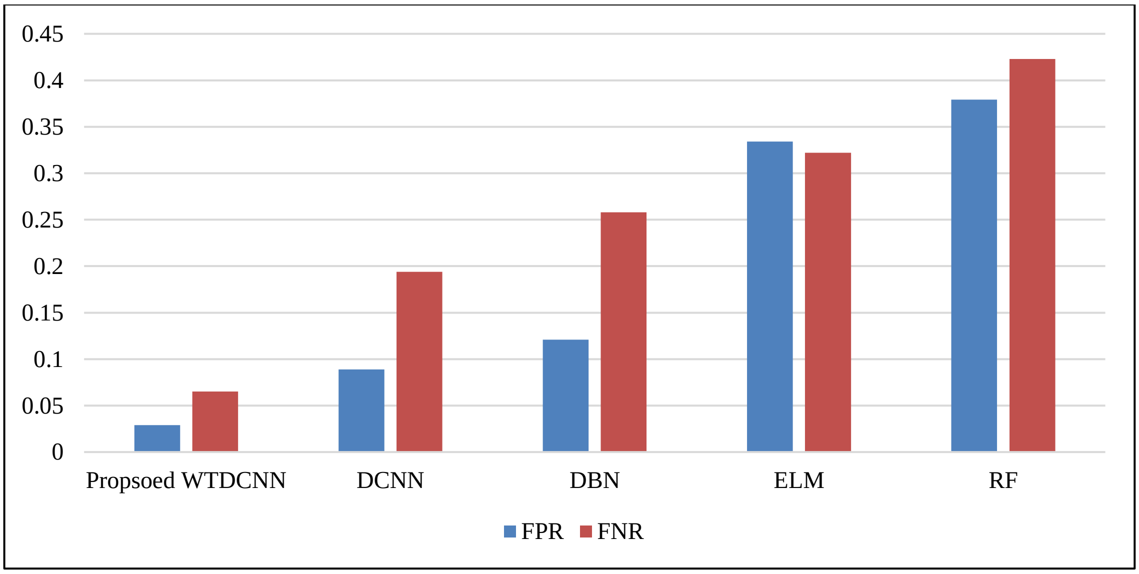

Figure 6: FPR and FNR analysis

Figure 5 shows the MSE and RMSE 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 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.
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 hyperparameters values. Clearly, the proposed one first preprocesses the dataset before prediction and efficiently utilizes the DR method. So, these approaches are more efficient in making predictions.

Conclusion

This paper suggests efficient DL and DR approaches for CYP for Indian 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 concerning 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 achieves superior results than the existing state-of-the-art schemes for CYP.

Abbreviations

Nil

Acknowledgments

Nil

Author's contribution

Leelavathi KS: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation and Writing-original paper draft. Rajasenathipathi M: Validation, Supervision and Project Administration.

Conflicts of interest

The authors declare that they have no competing interests.

Ethical statement

The work is not being considered for publication or released as a preprint elsewhere.

Funding

No funding received by any government or private concern.

References

1. Kavita M, Mathur P. Crop Yield Estimation in India Using Machine Learning. In: 2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA), 30-31 October 2020, Greater Noida, India. pp. 220-224.
2. Burdett H, Wellen C. Statistical and machine learning methods for crop yield prediction in the context of precision agriculture. Precision Agriculture. 2022;23(5):1553-1574.
3. Pant J, Pant RP, Singh MK, Singh DP, Pant H. Analysis of agricultural crop yield prediction using statistical techniques of machine learning. Materials Today: Proceedings. 2021;46:10922-10926.
4. Saranya CP, Nagarajan N. Efficient agricultural yield prediction using metaheuristic optimized artificial neural network using Hadoop framework. Soft Computing. 2020;24:12659-12669.
5. Jhajharia K, Mathur P, Jain S, Nijhawan S. Crop Yield Prediction using Machine Learning and Deep Learning Techniques. Procedia Computer Science. 2023;218:406-417Gavahi, K., Abbaszadeh, P., &Moradkhani, H. (2021). DeepYield: A combined convolutional neural network with long short-term memory for crop yield forecasting. *Expert Systems with Applications*, *184*, 115511.
6. Nishant PS, Venkat PS, Avinash BL, Jabber B. Crop yield prediction based on Indian agriculture using machine learning. In: 2020 International Conference for Emerging Technology (INCET); 2020 June; pp. 1-4. IEEE.

7. Shahhosseini M, Hu G, Huber I, Archontoulis SV. Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. Scientific Reports. 2021;11(1):1-15.
8. Feng L, Wang Y, Zhang Z, Du Q. Geographically and temporally weighted neural network for winter wheat yield prediction. Remote Sensing of Environment. 2021;262:112514.
9. Feng L, Wang Y, Zhang Z, Du Q. Geographically and temporally weighted neural network for winter wheat yield prediction. Remote Sensing of Environment. 2021;262:112514.
10. Muruganantham P, Wibowo S, Grandhi S, Samrat NH, Islam N. A systematic literature review on crop yield prediction with deep learning and remote sensing. Remote Sensing. 2022;14(9):1990Abbas, F., Afzaal, H., Farooque, A. A., & Tang, S. (2020). Crop yield prediction through proximal sensing and machine learning algorithms. *Agronomy*, *10*(7), 1046.
11. Kuradusenge M, Hitimana E, Hanyurwimfura D, Rukundo P, Mtonga K, Mukasine A, et al. Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize. Agriculture. 2023;13(1):225.
12. Gong L, Yu M, Jiang S, Cutsuridis V, Pearson S. Deep learning based prediction on greenhouse crop yield combined TCN and RNN. Sensors. 2021;21(13):4537.
13. Elavarasan D, Vincent PDR. A reinforced random forest model for enhanced crop yield prediction by integrating agrarian parameters. Journal of Ambient Intelligence and Humanized Computing. 2021:1-14.
14. Rajagopal A, Jha S, Khari M, Ahmad S, Alouffi B, Alharbi A. A novel approach in prediction of crop production using recurrent cuckoo search optimization neural networks. Applied Sciences. 2021;11(21):9816.
15. Paudel D, Boogaard H, de Wit A, Janssen S, Osinga S, Pylianidis C, et al. Machine learning for large-scale crop yield forecasting. Agricultural Systems. 2021;187:103016.