Vol.67: e24230773, 2024
https://doi.org/10.1590/1678-4324-2024230773
ISSN 1678-4324 Online Edition

BABT
Brazilian Archives of Biology and Technology

Article - Food/Feed Science and Technology

Prediction of Wheat Yield by Novel SDC-LSTM Framework

Nandini Babbar[1]*
0009-0008-6437-5326

Ashish Kumar[2]
0000-0002-6644-9754

Vivek Kumar Verma[3]
0000-0003-1998-8736

[1]Manipal University Jaipur, Department of CSE, Rajasthan, India; [2]Manipal University Jaipur, Department of CSE, Rajasthan, India; [3]Manipal University Jaipur, Department of IT, Rajasthan, India.

Editor-in-Chief: Bill Jorge Costa
Associate Editor: Fabio Alessandro Guerra

Received: 04-Aug-2023; Accepted: 24-Aug-2023

*Correspondence: nandinibabbar1@gmail.com (N.B.).

HIGHLIGHTS

- Research explores deep learning methods for predicting wheat yield.
- CNN and LSTM technology improves regional crop yield forecasting reliability.
- Deep learning-based wheat yield forecasting method for constraint resolution.
- CNN, LSTM technology improve regional agricultural forecasting reliability.

Abstract: Agriculture is the primary source of income for each country, serving as its mainstay. A promising study topic has been predicting wheat production based on environmental, soil, and water characteristics. Deep-learning-based algorithms are widely employed in crop prediction to extract significant crop traits. Wheat is linked to a variety of economic, societal, and health-related factors. Wheat yield forecasting and estimation on a regional scale, on the other hand, remains difficult. Two strategies for estimating wheat yield using deep learning (DL) models are presented in this study. To solve the limitations of regional forecasting, Convolutional Neural Networks (CNN) and Deep Learning Long Short-Term Memory (LSTM) technology are utilized to anticipate agricultural yields in a timely and reliable manner.

Keywords: Wheat-yield prediction, Machine learning (ML), Deep learning, Convolution Neural Networks (CNN), Long Short-Term Memory (LSTM).

INTRODUCTION

One of the first cereals to be domesticated in recorded agricultural history [1], wheat continues to be a significant source of nutrients and energy for both people and animals [2]. Significant socio-economic issues, as well as nutrition and health, are connected to wheat [3]. A significant wheat exporter in Eastern Europe, Ukraine, has an average yield gap of 50%; if the gap were reduced to 20% of water-limited yields, production would increase by 70 million tons. The goal is to increase the area of irrigated wheat to 400,000 hectares in 2021–2022, which is anticipated to yield 1.6 million tons. Crop switching might also be encouraged, but there would always be trade-offs on other markets, as in eastern Africa's transition from white maize to wheat

cultivation [4]. Wheat is classified into two types based on the growing season. Winter wheat and spring wheat are the two most common types of wheat. More than 70% of the wheat produced in the United States is produced from winter wheat, which is sown in the autumn and collected in the summertime. In 2019, 35 million metric tons of winter wheat were harvested from 24 million acres [5]. For local, national, and global food safety management and market strategy, accurate, timely, and geographically precise winter wheat production assessment is essential [6].

One of the three fundamental crops is wheat, and its accessibility and security are crucial. For farming decision-making and long-term prosperity, timely and trustworthy national wheat yield data are essential. Soil quality, weather information [7], agricultural practices, agricultural incentive programs, and grain market prices all have an impact on wheat output. For instance, farmers are routinely persuaded to spend more money to boost harvests [8]. A complex agroecosystem has numerous variables that are connected, such as climate, ground, and supplied fertilizers, and many of these have a nonlinear connection with yield [9, 10]. As a result, predicting wheat yields over vast geographic areas is still challenging.

Image identification, language processing, and remote sensing have all benefited significantly from ML and DL technology [11]. Support Vector Machine (SVM) and Random Forest (RF) are two well-known traditional machine learning methods that are frequently employed in the classification of satellite photos [12], parameter inverse [13], and crop production forecasting [14–16]. Supervised learning models called SVMs analyze the information for categorization and analysis. They come with teaching techniques. A supervised ML approach called random forest is used to resolve regression and classification difficulties. Meanwhile, DL techniques usually outperform traditional machine learning methods [17–19]. Convolution Neural Networks (CNN) and LSTM [20] networks, the two most popular DL models, have been used to predict and estimate agricultural productivity.

The Hochreiter and coauthors [21] presented a deep learning system for predicting agricultural productivity that was trained with a novel feature representation based on raw image histograms. The highly accurate model demonstrated strong transfer learning abilities [22,23]. This method, however, can only be applied to images with low resolution or produces forecasts on a lesser level because it needs a large number of pixels in an image to create a scatter plot. Furthermore, popular spatial aggregation representations can be employed to benefit deep neural networks like the one-dimensional CNN [24] or LSTM [25,26]. Furthermore, the bulk of these deep learning models did not assess the network's result uncertainty. Previous research improved yield prediction accuracy in both the geographical and temporal domains, but due to the difficult data processing, it was limited to partial regions [27]. Crop yield prediction on a larger scale frequently requires extensive data and extensive data processing, implying high acquisition and processing costs. As a result, the potential of deep learning for predicting regional agricultural yields has been underutilized.

According to the preceding reviews, numerous problems do not anticipate correct wheat output. Predicting wheat yield across geographical areas remains difficult. By overcoming all of these constraints, this study developed a superior novel technique known as the SDC-LSTM framework. The following notable contributions have been made by this research study:

- A Deep Convolutional Neural Network approach for recognizing dynamic features such as meteorological details and previous year's wheat yield data is introduced.
- The Mish SoftMax function in the convolution layer is presented to improve network accuracy, and a max pooling layer is implemented before the flattening layer to improve network stability.
- Furthermore, to minimize overfitting difficulties in the network, our research employs a Long-Term Short Memory to recognize static aspects of soil data.
- The Stochastic Gradient Descent algorithm is utilized to do early stopping, which improves the network's performance with training data. The static and dynamic data are combined and supplied into a fully linked layer to provide a more accurate wheat yield prediction.

Finally, our proposed network outperformed all existing techniques in predicting wheat yield. Additionally, the organization of this study report is as follows: A review of the available methods is described in Part 2. The proposed framework is further examined in Part 3. The application of the proposed approach is shown in Part 4, and the research paper is finished in Part 5.

LITERATURE SURVEY

Data analysis and process-oriented crop production concepts, which are covered in this section, are two techniques that many researchers have used in recent years to improve agricultural output prediction.

A geographically and temporally weighted neural network (GTWNN) was suggested by Feng and coauthors [28] as a way to increase forecast accuracy. Market risk management, food security, and an accurate crop production forecast are critical for agricultural trading. Crop production modeling often takes into account the global and regional climate non-stationarity that is inherent in many geographical processes, even if many multiple machine learning approaches have been created to increase prediction performance. This study used geographically weighted regression (GWR) and temporally weighted regression (TWR) to demonstrate spatial and temporal non-stationarity in winter wheat yield estimation. Then, utilizing openly accessible data sources including satellite imagery and temperature data, a geographically and temporally weighted neural net (GTWNN) model was produced by merging artificial neural networks (ANN) with geographically and temporally weighted regression (GTWR).

Support Vector Regression (SVR) combined with Sequential Forward Selection (SFS) was used by Shafiee and coauthors [29] to forecast grain yields and compare the results to those of a LASSO regressor with an internal feature selector. The ability to anticipate a variety of features, including crop output, has been demonstrated by unmanned aircraft (UAVs) and other technologies for remote sensing - based detection. To go through the vast volumes of info produced by UAV photos and extract the most relevant data, ml algorithms are currently being researched. Using derived vegetative indexes from UAV photos, this study analyses the use of two different machine learning-based regress or techniques to predict wheat crop growth.

The INSEY model was put forth by Aula and coauthors [30] to forecast grain yield potential, to determine if INSEY, well before N rate, overall precipitation, and mean ambient temperature can forecast winter wheat from September to December. For better N managerial decisions, it is essential to anticipate winter wheat crop output accurately. Although data shows that this method is effective, adding more variables to the model may improve yield forecast accuracy even further.

Aravind and coauthors [31] developed the model using weather conditions that significantly impact wheat yield production. Using meteorological data, a model with several linear neural networks and penalized regression algorithms can produce accurate, timely, and cost-effective wheat production forecasts. The study area collected wheat yield data and weather parameters over 30 years. The model was created utilizing approaches such as least absolute shrinkage and selection operator (LASSO), elastic net (ENET), artificial neural network (ANN), component analysis (PCA) combined with SMLR, and stepwise multiple linear regression (SMLR).

Srivastava and coauthors [32] presented a CNN that captures the temporal dependence of environmental variables using the 1D convolution technique. Forecasting crop productivity is affected by several interrelated variables, including genetics, weather, land, and business practices. This study shows the effectiveness of ma when compared to existing ML models for predicting winter wheat yield in Germany, the proposed CNN outperformed every other model that was put to the test. Weekly features that explicitly account for soil moisture and weather events were employed to address the seasonality.

Pang and colleagues [33], Using 2018 produce layouts from a dataset supplied by cooperating farmers, this study used a Random Forest Regression (RFR) method to construct a geographic and municipal harvest forecasting model at the low resolution for three southeastern Australian grains paddocks, both situated in Victoria (VIC), New South Wales (NSW), and South Australia (SA). The networks were validated, verified, and confirmed at the pixel level utilizing Python language for (a) geographic three-paddock composite and (b) individual paddocks. The time-series Normalized Difference Vegetation Index (NDVI) information, weather parameters, and yield data were employed. Table 1 lists the current algorithms, their methods, and their drawbacks.

Table 1. Comparison of existing papers

Reference Nº	Algorithms	Technique	Restrictions
[28]	Geographically and temporally weighted neural network (GTWNN)	Utilizing openly accessible data sources including satellite imagery and temperature data	R^2 values need to be improved
[29]	Support Vector Regression (SVR) combined with Sequential Forward Selection (SFS)	Using derived vegetative indexes from UAV photos	Research is still need to improve in all environments

Cont. Table 1

[30]	INSEY model	To anticipate winter wheat crop output accurately.	RMSE values need to be improved
[31]	Approaches such as least absolute shrinkage and selection operator (LASSO), elastic net (ENET), artificial neural network (ANN), component analysis (PCA) combined with SMLR, and stepwise multiple linear regression (SMLR).	Using meteorological data, a model with several linear neural networks and penalized regression algorithms	RMSE values need to be improved
[32]	A convolutional neural network (CNN) captures the temporal dependence of environmental variables using the 1-dimensional convolution technique.	Forecasting crop productivity is affected by several interrelated variables	Hybrid optimization need to improve
[33]	Random Forest Regression (RFR) method to construct geographic and municipal harvest forecasting model	The time-series Normalized Difference Vegetation Index information, weather parameters, and yield data were employed.	Accuracy need to be improved.

The difficulties seen in the aforementioned recent papers include the time-taking procedure, the network's complexity, the yield trend, and the difficulty of large-scale forecasting. To get beyond the aforementioned restrictions, one efficient solution is needed, which is covered in more detail in the next section.

DEEP LEARNING BASED ON SUPERIOR DEEP CONVOLUTIONAL LONG SHORT-TERM MEMORY

After rice, wheat (Triticum aestivum) is the 2nd most frequent primary food crop in India. The northwest of the nation is where it is most extensively grown. Wheat is a naturally thermos-sensitive crop. Changes in meteorological factors harm crop growth and development, resulting in declining yield trends. Crop yield forecasting is critical for storage, import, and export, and it assists government planners and policymakers make better product management decisions. Thus, this research proposes effective deep learning-based methods to predict wheat yield, namely, the Superior Deep Convolutional Long-Term Short-Term Memory (SDC-LSTM) Neural Network Framework. Weather data, previous year's wheat yield, and soil properties data are collected in the first level. The preprocessed data are then fed into the SDC-LSTM framework, which has different features such as dynamic and static data. Furthermore, Superior Deep Convolutional Neural Network is introduced for detecting dynamic features, in which our research proposes the Mish SoftMax function in the convolution layer for detecting weather data, wheat yield data of previous years to improve accuracy, max pooling is introduced before the flatten layer, which gives the network superior stability. In addition, Improved LSTM is proposed to detect static features (soil data [39]). The gradient descent algorithm is used early to stop overfitting in the network, improving the network with training data. Furthermore, dynamic and static features are combined. The fully associated level network receives the integrated data after which it normalizes and assesses it. Finally, the wheat yield is predicted by our SDC-LSTM framework, which is shown in Figure 1.

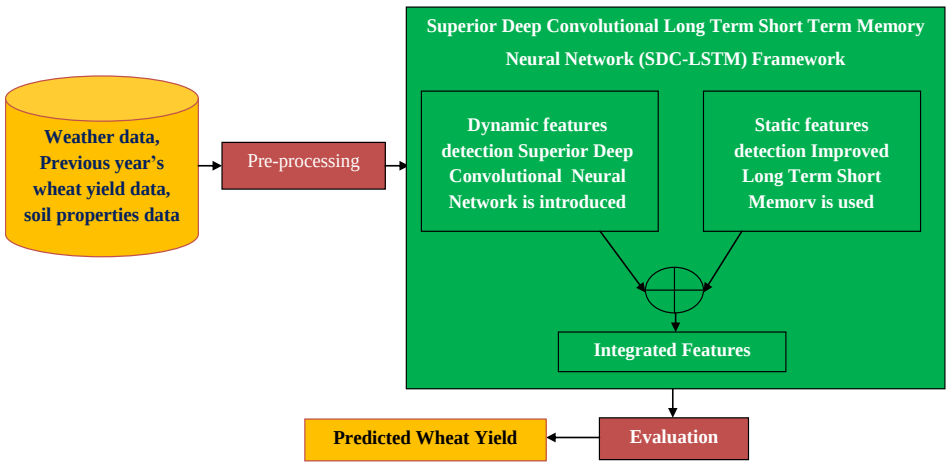

Figure 1. Structure of Proposed Deep learning-based SDC-LSTM

As a result, our proposed network improves the accuracy of wheat yield prediction, reduces complexity and processing time, and outperforms existing techniques. The following section briefly explains the process of wheat-yield prediction.

Deep Convolutional Neural Networks

We proposed a deep convolutional neural network in this paper. Deep CNN is typically composed of layers that are organized by function for detecting dynamic features which includes weather conditions of previous years. CNN has input data, an output layer, several hidden units, and billions of variables to understand difficult shapes and processes. Video and image identification, image analysis, pattern classification, computer vision, and natural language processing (NLP) are some of their uses. The CNN is composed of layers that are classified based on their functions: convolution and mish SoftMax layer, max-pooling layer, and fully-connected layer. Figure 2 illustrates the deep CNN Architecture.

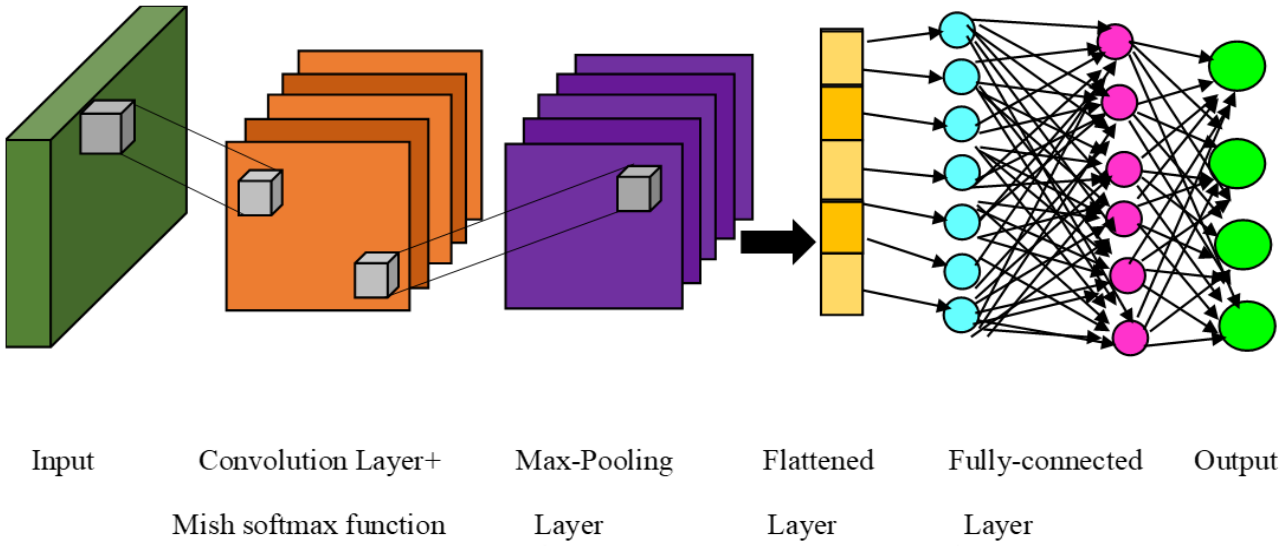

Figure 2. Deep CNN Architecture

Convolution Layer

This is the initial stage that is used to extract certain features from data images. The statistical procedure of combination between the source images and a filter of size M×M is carried out by this layer. Typical filter sizes include 3×3, 5×5, and 7×7. To increase the network's stability, we offer Mish, a brand-new self-regularized non-monotonic nonlinear activation that was inspired by the self-gating trait. Mish is arithmetically defined in equation 1:

$$f(y) = ytanh(soft\ plus\ (y)) \quad (1)$$

Max-Pooling Layer

CNNs frequently use the max-pooling layer operation after convolution layers to improve network accuracy. There are no parameters to learn in the pooling layer. The idea behind max pooling is that a large number indicates the possibility of detecting a feature. The characteristic graph's most important elements are selected using max-pooling, and these elements are included in the layer that results. It is the most commonly used method because it produces the best results. The next stage of the procedure is the flattening layer. The flattening layer combines all of the pooled feature maps' resultant 2-Dimensional arrays into a solitary lengthy continuous linear vector. The completely linked layer receives the matrix that has been flattened as input.

Fully Connected Layer

After a few convolutions and pooling layers, the CNN usually ends with several fully connected layers. These layers' tensor is transformed into a vector, and several neural network layers are added. The FC layer receives flattened images from the preceding levels. The theoretical unit procedures are often carried out

after the flattening vector flows via a few additional FC levels. The connected element's objective is to combine all the components and flatten the increased characteristics discovered by the convolution operation. Thus, the fully connected layer will provide the desired output. The following section briefly defines about Improved LSTM framework.

LSTM Framework

Crop yields have been predicted and calculated using LSTM networks, the most common methods among deep learning approaches. To locate the static data of soil information, LSTM is applied. In the ability to forecast agricultural output, for instance, you proposed an integrated architecture that was learned using a brand-new classification model based on raw picture scatter plots. The system showed good transfer learning capabilities while obtaining good precision. Furthermore, deep learning methods like the one-dimensional CNN or LSTM can use the widely used geographic grouping structure. Recurrent neural networks (RNNs) retain a feed-forward structure with a backpropagation neural loop in this type of system. By temporarily holding onto the quality of the previous output, LSTM outperforms RNN. It serves as a tiny bit of network storage that aids in feedback analysis. The vegetative indices, meteorological, and temperature information, as well as a map of all environmental parameters, were used to assess the plants. Several characteristic factors are followed throughout time, and LSTM was used to evaluate the series data for the period. Depth adaptive LSTM is a combinational network with a deep neural network that processes images and environmental factors. Then, to enhance crop prediction, the LSTM procedure was applied to time series analysis and meteorological variables. A one-dimensional convolutional network can be directly fed and process one-dimensional data. Because of the data type, the network's training time will be short, and the processing and storage space will be efficient. LSTM techniques can be derived using some equations that are following;

$$i_t = \sigma(w_i[h_{t-1}, x_t] + b_i) \tag{2}$$

$$f_t = \sigma\big(w_f[h_{t-1}, x_t] + b_f\big) \tag{3}$$

$$o_t = \sigma(w_o[h_{t-1}, x_t] + b_o) \tag{4}$$

Equation 2,3,4 is shown as a gate equation. In the equations above, i_t stands for input gates, f_t for forget gates, and o_t for output gates. $w_i, w_f \; and \; w_o$ indicate the weight for the corresponding gate neurons, while σ stands for the sigmoid function. Biases for the corresponding gates (x) are represented by b_x. At timestamp t-1, the result of the prior LSTM block is represented by h_{t-1} .

Equation 2 for the input gate can be used to decide what new info we're going to transport in the cell state (that we will see below). Equation 3 contains the forget gate, which gives the order to remove the data from the cell state. The output gate, that is used to enable the lstm frame's correct outcome at timestamp "t," is represented by the fourth equation.

The cell state (memory) at timestamp(t) is as c_t and candidate cell state $\widetilde{c_t}$ is obtained by using both f_t and i_t in the following manner.

$$\widetilde{c_t} = \tan h\,(w_c[h_{t-1}, x_t] + b_c) \tag{5}$$

$$c_t = f_t\, c_{t-1} + i_t\, \widetilde{c_t} \tag{6}$$

$$h_t = o_t\, \tan h\,(c_t) \tag{7}$$

Due to a more efficient gradient flow during backpropagation, LSTM is better at modelling longer sequences than a straightforward RNN. We can design the LSTM framework's architecture based on the aforementioned equations was shown in Figure 3.

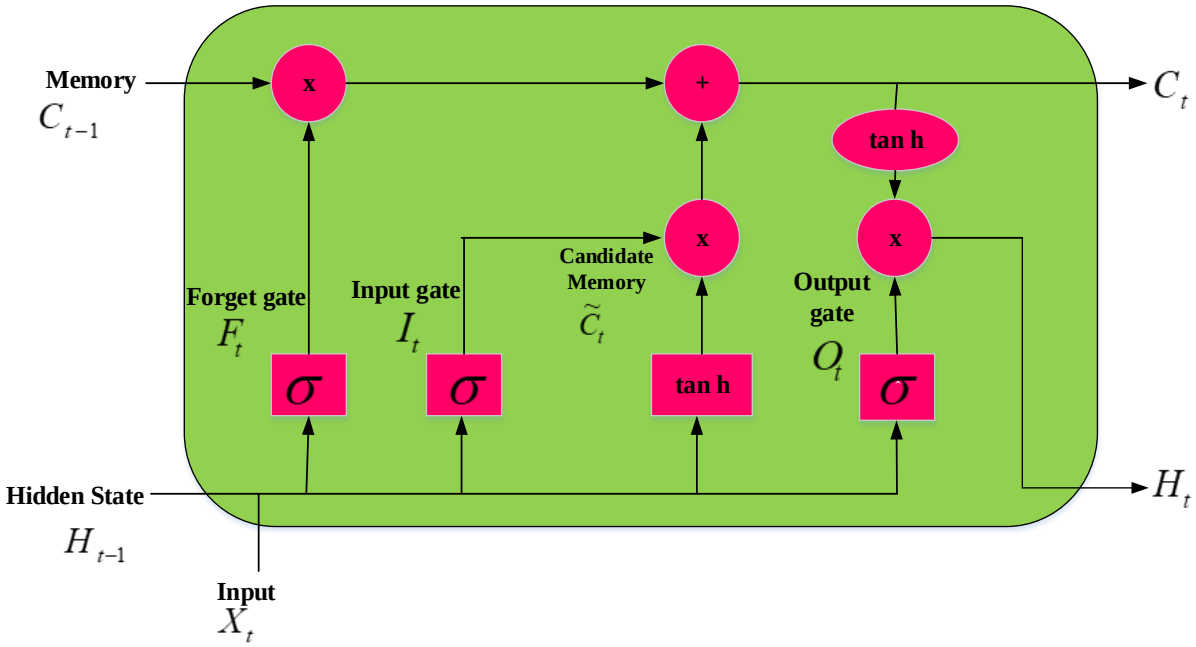

Figure 3. LSTM Architecture

Gradient Descent Algorithm

The algorithm is developed known as gradient descent is frequently used to build neural network models and ml algorithms. These models gain knowledge over time by using training examples, and the functional form in stochastic gradient especially serves as a gauge by evaluating the validity of every repetition of variable parameters.

Stochastic Gradient Descent (SGD) [34] is an innovative process for minimizing/maximizing an objective function that repeatedly tries to find the lowest or peaks. It is a stochastic approximation of gradient descent optimization. After merely examining a few training sets, SGD surpasses the existing techniques by fol. In those other terms, SGD performs very well if the amount of training datasets is enormous since it does not employ all of the training sample within every evaluation, which lowers the number of parameters and speeds up processing. Additionally, SGD can continuously modify the evaluation of the first- and second-order arrays of the elevation of each parameter in accordance with the loss function. As a result, the risk of the model converging to the local optimum is reduced. Given these advantages, we think employing SGD may offset the expense of computing and lead to result oriented.

$x_i \in R^n$ is a n-dimensional vector. $y_i \in \{1, m-1\}$ is the category of the ith training sample. Then, SGD can be detailed as follows [35].

First, give the load factor W1 a zero vector, and after that, choose a training dataset at arbitrary $(x_{it} y_{it})$ from the whole training set, where $i_t \in \{1, \ldots, m\}$ is the target of the selected training sample at the tth iteration. The objective function is

$$\min(W) = \frac{\lambda}{2} ||W||^2 + f\big(W, (x_{it} y_{it})\big). \tag{8}$$

Next, use Equation (8) to compute the gradient. The gradient may then be represented by

$$\nabla_{t=} \lambda W_t - \alpha_t y_{it} x_{it} \tag{9}$$

where $\propto_t = \begin{cases} 1, \ if \ y_{it} \langle W_t, x_{it} \rangle < 1 \\ 0, \ otherwise \end{cases}$

The updated formula of matrix W is as follows.

$$W_{t+1} = W_t - \eta_t \alpha_t \tag{10}$$

where $\eta_t = \frac{1}{(\lambda t)}$

Then an updated weight matrix W based on formulas (9) and (10) can be obtained by

$$W_{t+1} = \left(1 - \frac{1}{t}\right) W_t + y_{it} x_{it} \tag{11}$$

In practice, formula (11) finds minima or maxima by iteration. The next section follows the Integrated CNN and Improved LSTM techniques in detail.

Integrated Deep CNN and Improved LSTM

Our study proposed a new technique described in the preceding sections. We concluded from those sections that deep CNN collects all dynamic and static features such as weather data, wheat yield data and soil data from the previous year. This dynamic feature is then transferred to a SoftMax function named mish to improve network accuracy, and the acquired output is sent to a layer called the max-pooling layer to increase network stability. To detect static features in soil data, LSTM is proposed. LSTM is also used in networks to avoid overfitting issues. To detect static features in soil data, LSTM is proposed. LSTM is also used in networks to avoid overfitting issues. The gradient algorithm performs early stopping, improving the network's performance with training data. The static and dynamic features are combined before being fed into the fully connected layer network. This layer equalized and analysed our features, from which the result was derived. Consequently, our proposed Novel technique predicts wheat yield production and improves the network's stability and accuracy. Eventually, our SDC-LSTM technique outshines all other techniques. Figure 4 shows the proposed method SDC-LSTM architecture.

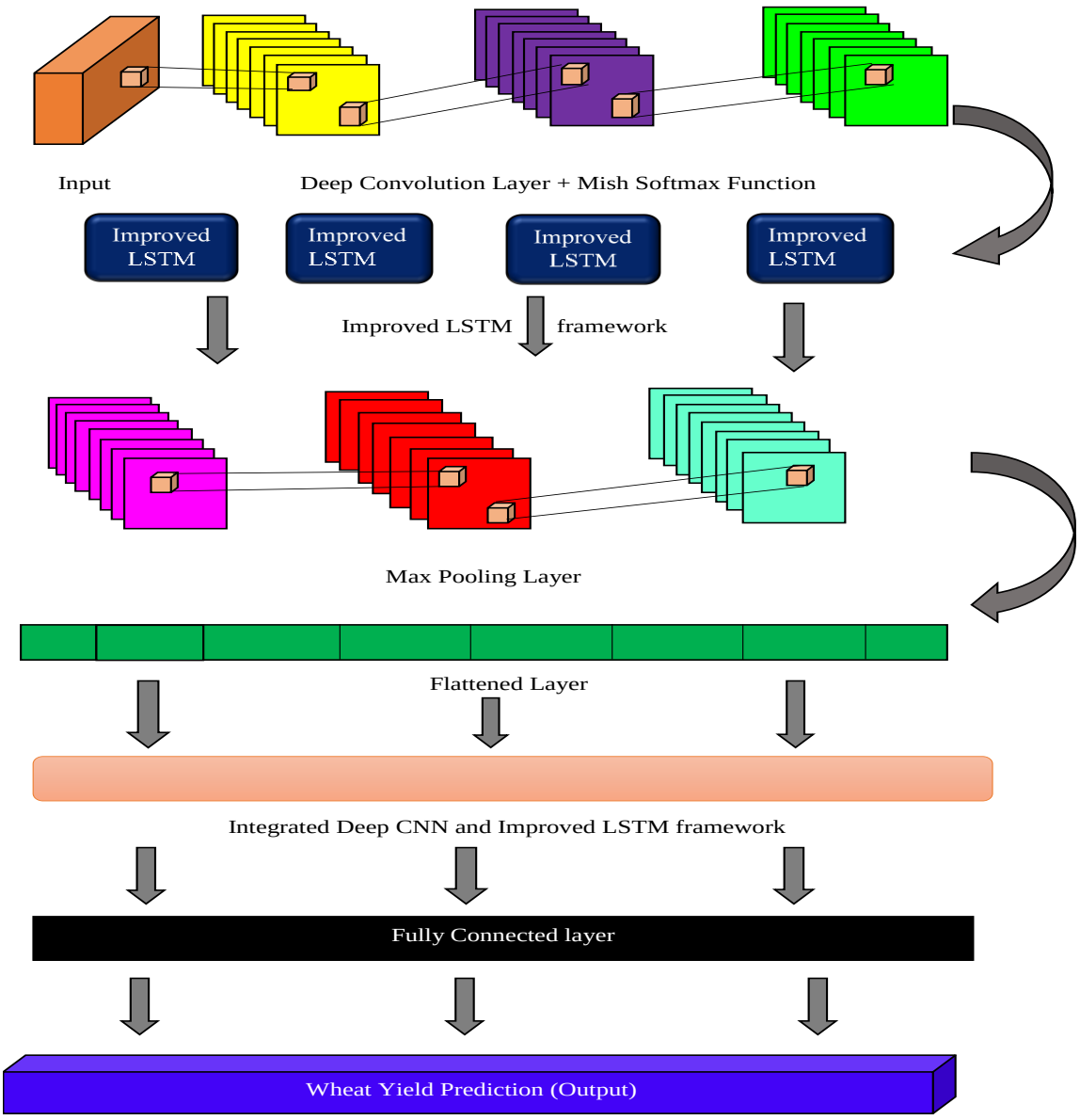

Figure 4. Proposed framework Architecture

Algorithm: Wheat-a yield prediction using Deep CNN and LSTM

Start

Step1: Deep Convolution layer receives input message.

Step2: The input message is convolved using the Mish-SoftMax and Deep Convolution algorithms.

Step3: Therefore, a convoluted message is sent to the upgraded LSTM framework.

Step4: The data are now transferred by the LSTM framework to the Max-pooling layer.

Step5: The following layer, which produces superior results, is flattened.

Step6: For better performance, the Flattened layer transmits the enhanced messages to the Integrated Deep CNN and LSTM layer.

Step7: The fully connected layer receives the integrated data as its input data.

Step8: Finally, our proposed method succeeds with great precision.

End

RESULTS AND DISCUSSION

This section outlines the implementation outcomes and the performance of our current proposal. Also signifies the comparison results of the existing works.

Tool : Python
OS : Windows 7 (64 bit)
Processor : Intel Premium
RAM : 8GB RAM

Dataset Description

1. The datasets that have been gathered such as rainfall data from https://data.world/rajanand/rainfall-in-india, Crop production DataSet from https://data.world/thatzprem/agriculture-india. Temperature dataset from https://data.gov.in/catalog/all-india-seasonal-and-annual-minmax-temperature-series data's from the Uttar Pradesh, India. This dataset only contains data from year 1901 and 2015, so we are taking and processing those data. The collected rainfall data contains the maximum rain, the minimum rainfall in the month of January to December. The humidity is evaluated based on the rainfall. Following that, temperature data have the months from January to December temperature and the yearly temperature in Uttar Pradesh. The wind is evaluated based on the temperature, if the temperature is high means it considered to as no wind. Finally, the crop data contains different types of crops, maximum yield, minimum yield, and temperature data. Then, collected data are properly arranged in one file which is shown in the following table 2. From the data, this research splits training and testing data as 70:30.

Table 2. Data Collection Sample

Crop	State	Temperature		Rainfall		Humidity	Cost of Cultivation	Cost of Cultivation	Cost of Production	Yield	Support price
		Max. Temp	Min. Temp	Max. Rain fall	Min. Rainfall						
Wheat	Uttar Pradesh	34	25	3722.8	2365.8	54	18979.38	31902.74	769.84	34.99	1975

Performance Analysis

Figure 5 depicts India's monthly rainfall pattern. The month is shown by the X-axis, while the millimeters per hour of rainfall are shown by the Y-axis. When compared to September, August has a humid climate. In this, the month of November has the lowest rainfall, followed by April.

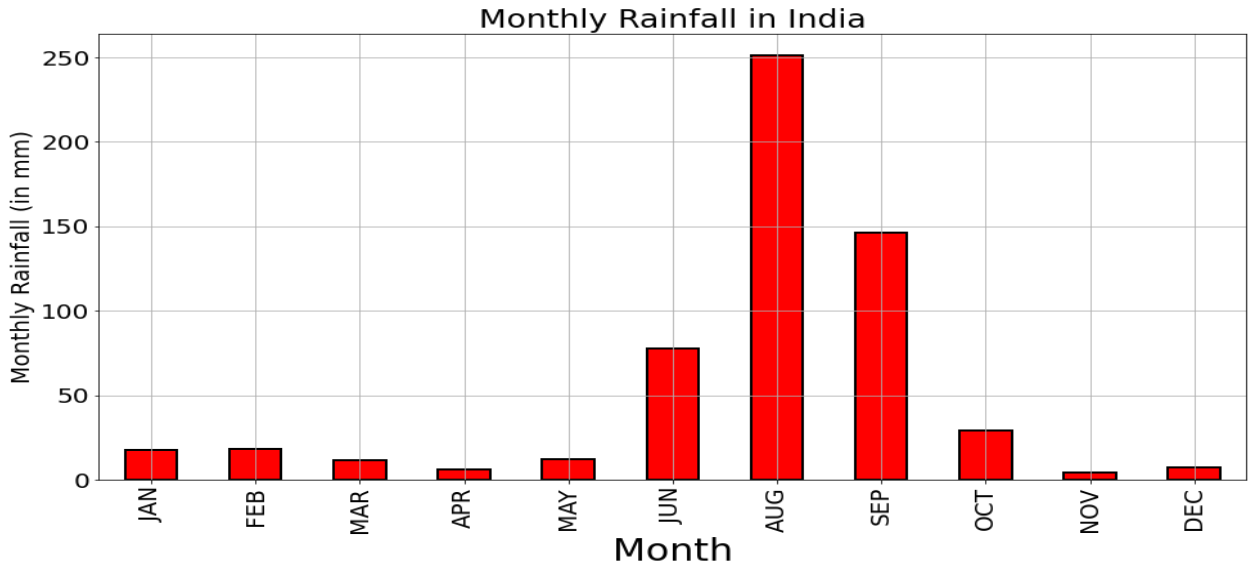

Figure 5. Monthly rainfall

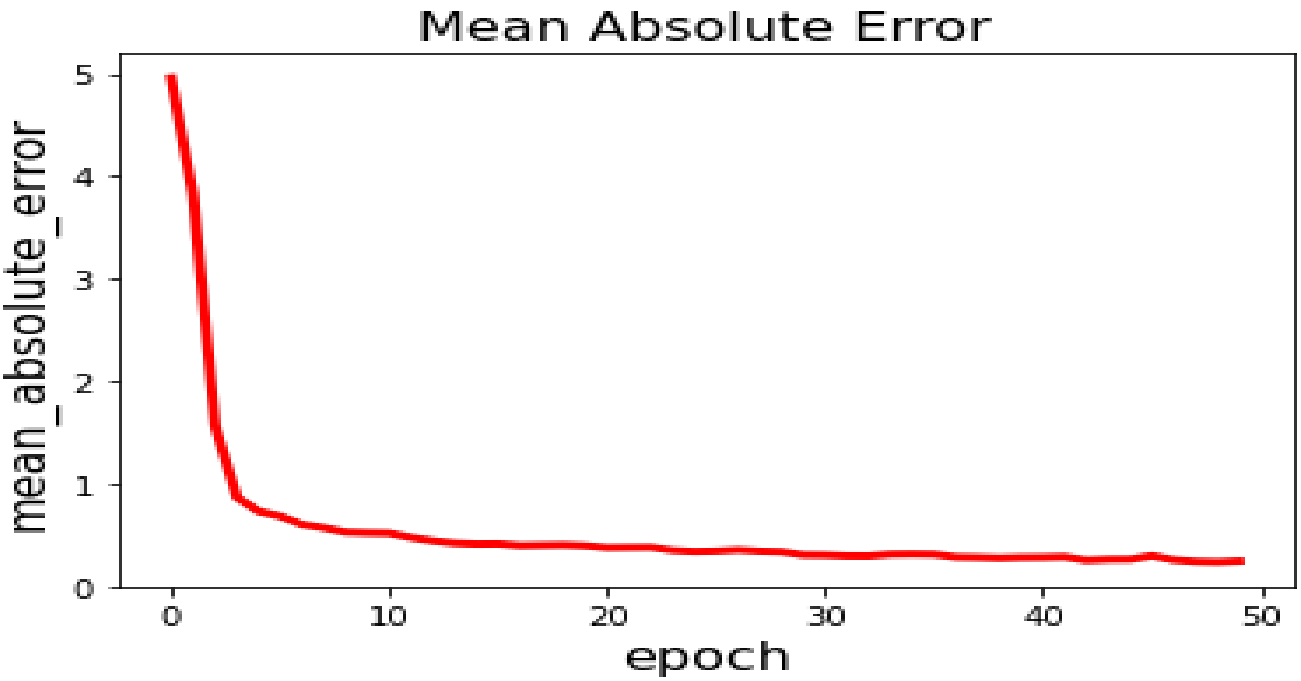

Figure 6. Mean Absolute Error

A metric known as the Mean Absolute Error (MAE) calculates the average size of faults in a series of forecasts without taking orders into account. By dividing the total cost of actual mistakes by the sample size, the MAE is calculated.

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|y_j - x_j|$$

$$MAE = \frac{1}{n}\sum_{j=1}^{n}|e_i| \quad (12)$$

The MAE is calculated using Equation 12. The MAE is shown in Figure 6 to find the epoch's average errors. Because of our proposed technique, the MAE error has decreased since epoch 10.

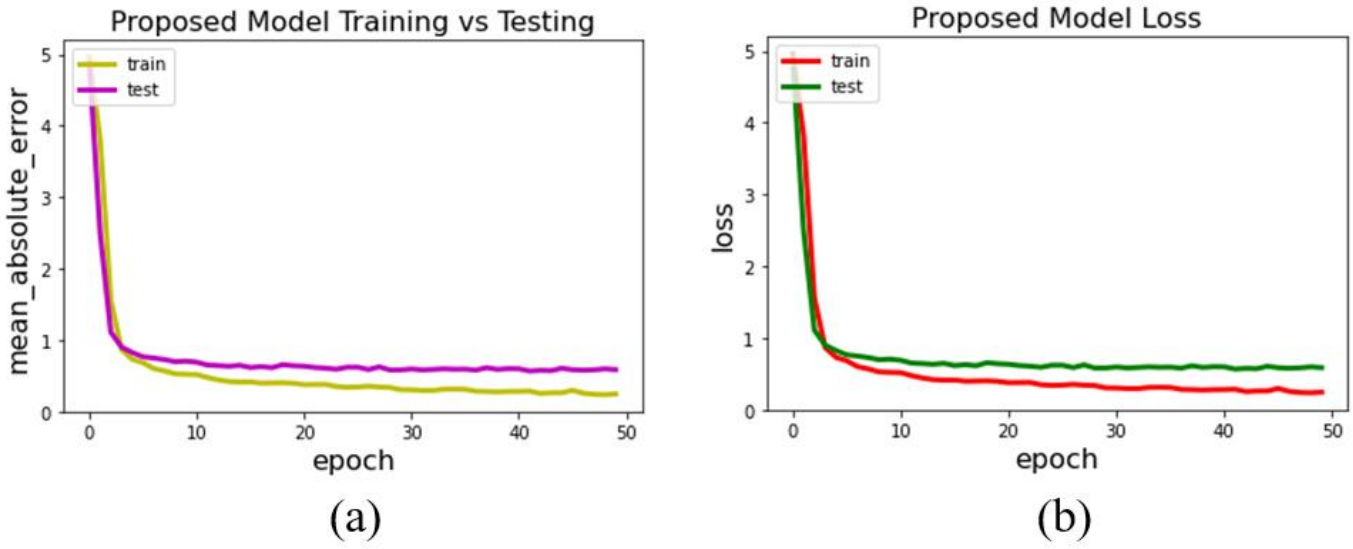

Figure 7. (a) Training vs Testing, (b) Proposed model loss

The training and testing process for our proposed scheme is shown in Figure 7a. In this, the training proposed method is very less when compared to the testing proposed model. Figure 7b shows the proposed model loss where the training model loss is maximum compared to the testing model loss.

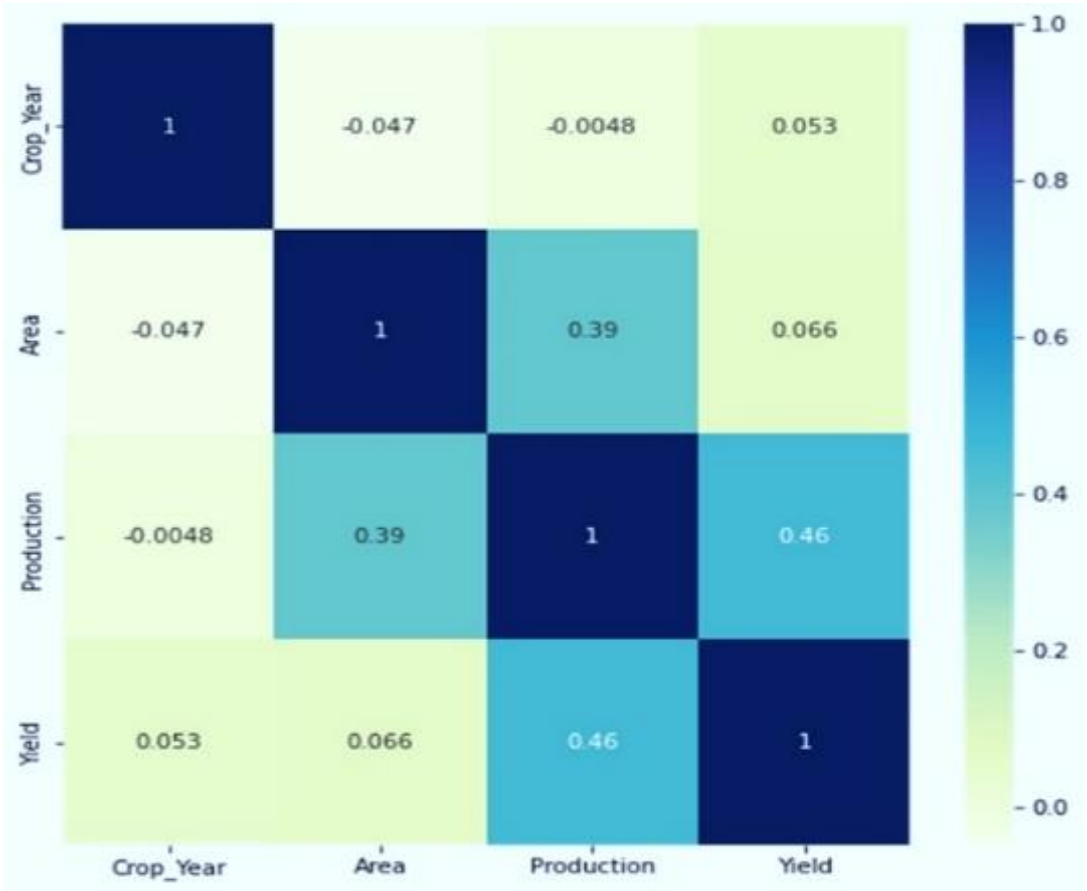

Figure 8. Confusion Matrix for wheat yield prediction

Figure 8 illustrates the Confusion Matrix for prediction of wheat yield. The diagonal matrix reflects the amount of units for which the predicted value is close to the actual value, but off-diagonal components are incorrectly labelled by the classification. Wheat-yield prediction is gained better because of our proposed novel technique.

Error Analysis

The following Table 3 shows the error analysis for calculating errors

Table 3 Methods and Percentage of Error

Errors	Accuracy %
MAE	0
RMSE	1
MSE	2
MAPE	13.42
R2	0.99

Mean Absolute Error

MAE is a metric for mistakes among number of pairs reporting the same phenomena. Examples of Y vs X include evaluations of predicted with observable, succeeding time against project commencement, and one evaluating method against another. By determining the proportion of absolute mistakes by the sample size, the MAE is calculated.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - x| \tag{13}$$

Using formula 13, the MAE value is 0%, indicating that our proposed models have less error than the other current models.

Root Mean Square Error

One common approach to assessing a model's error in forecasting quantitative results is to use the RMSE. The following is its official definition

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\widehat{y_i} - y_i)^2} \tag{14}$$

Where n is the number of observations, $\widehat{y_1}, \widehat{y_2}, \widehat{y_3} \ldots \ldots \widehat{y_n}$ are predicted values and $y_1, y_2, y_{3,\ldots\ldots} y_n$ is the observed values. The proposed RMSE value in percentage is 1 and gained a very less prediction error compared to other models.

Mean Square Error

The MSE of a regression line shows how close it is to a series of parameters. By doubling the lengths between both the endpoints and the coefficient of determination, it can do this (these distances are the "errors"). Any unfavorable indications must be removed using squaring. Substantial variations are also valued more highly. It's called the mean squared error since you're measuring the mean of many faults. The MSE decreases as the prediction gets better.

$$MSE = \frac{1}{n}\sum(y_i - x_i)^2 \tag{15}$$

Where, n denotes number of items, Σ is summation notation, y_i is the real level and x_i is the prediction rate. In our proposed method, gained 2% of error using equation 15. The error is very low and gained a better forecast.

Mean Absolute Percentage Error

The MAPE, which is a statistic, is used to gauge how well a prediction model makes a prediction. The formula's ratio, which represents the precision, is as follows:

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{16}$$

Where A_t is the actual value and F_t is the forecast value. By the true value, they split their variance A_t. The proposed method accuracy is 13.42% using equation 16. Thus, we outperform all the other existing techniques.

Co-efficient of Determination

The coefficient of determination, abbreviated R2, is a widely used regression performance indicator. It is a measure of how much of a dependent variable's variation is explained by a regression model and is defined by:

$$R^2 = \frac{1-\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})} \quad (17)$$

By using formula 17, the obtained co-efficient of determination is 0.99%. In contrast to all the other categories, R2 value is very great. Thus, compared to all the other existing methods proposed method outshines uniquely.

Equating all the above formulas, we plot a graph using these metrics.

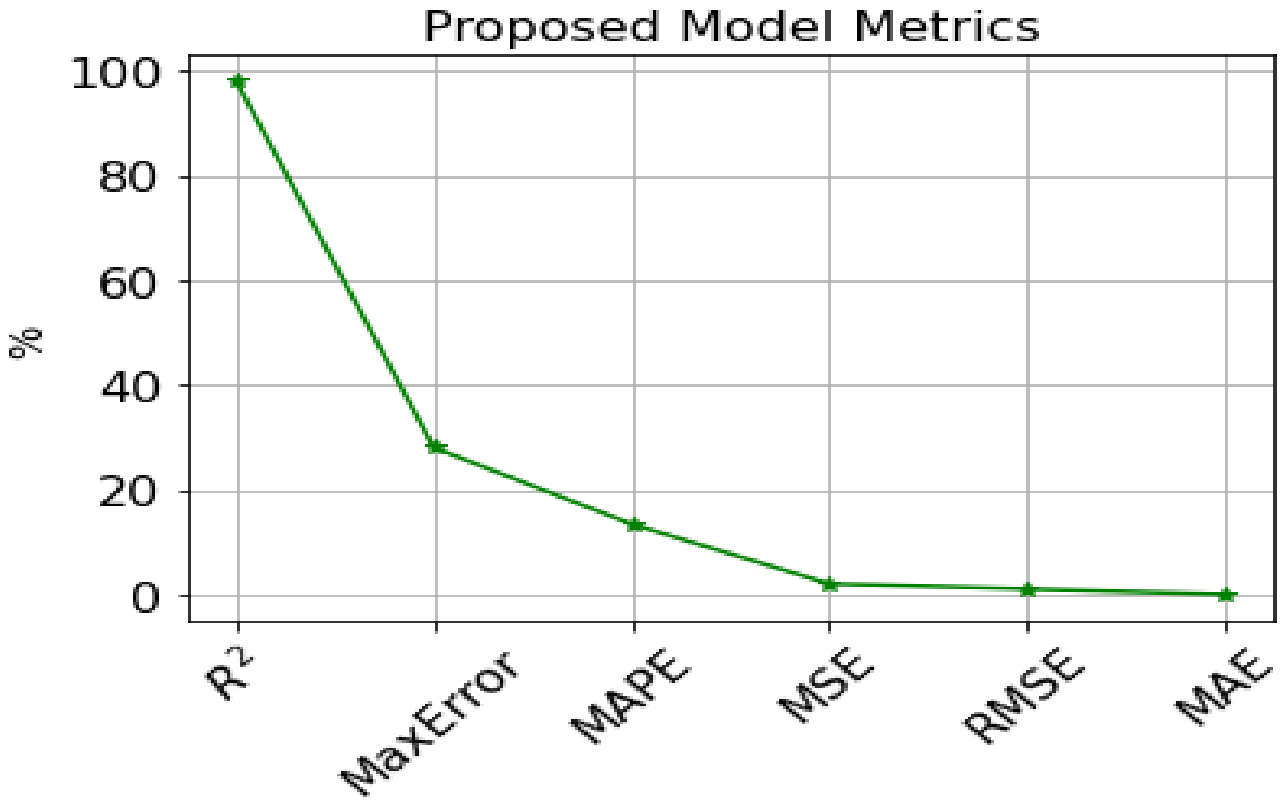

Figure 9. Proposed Model Metrics

Figure 9 illustrates the proposed model metrics. This picture shows the percentage of models in which the accuracy of R2 is high while the Mean Absolute Error is very low.

Comparison Analysis

This section discusses the suggested method's comparison results, in which our unique technique is compared to a reference line approach such as MAE, RMSE, MSE, MAPE, R^2, and maximum error.

MAE comparative analysis

The MAE comparative analysis is shown in Figure 10. Multiple Linear Regression (MLR) approaches have the highest percentage of error in this diagram, followed by other techniques. However, the error rate of proposed innovative strategy is extremely low. As a result, our model outperforms all other models.

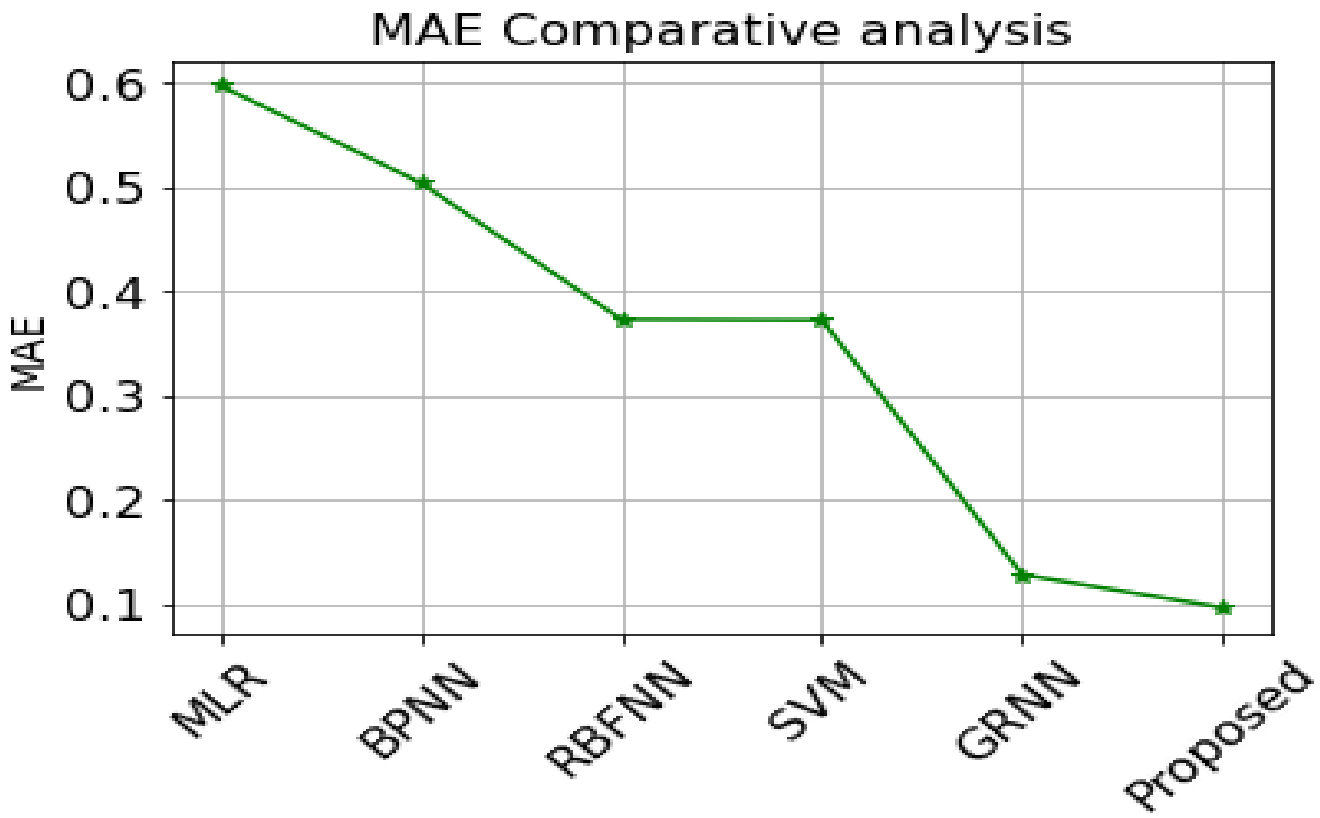

Figure 10. MAE comparative analysis

The following table 4 shows the method and its errors for MAE. The Mean Absolute Error is reduced by employing the proposed SD CNN-LSTM approach. This research achieves less error when compared to the baseline as MLR [36], Back Propagation Neural Network (BPNN) [36], Radial Basis Function Neural Network (RBFNN) [36], SVM [36], and General Regression Neural Network's (GRNN) [36], such as 0.6, 0.52, 0.37, 0.37, and 0.13. As a result, compared to existing approaches, the novel, distinctive technology has an error as 0.

Table 4. Methods and errors for MAE

Method	**MAE**
MLR [36]	0.6
BPNN [36]	0.52
RBFNN [36]	0.37
SVM [36]	0.37
GRNN [36]	0.13
Proposed	0.0

MSE comparative analysis

The MSE comparison is shown in Figure 11. This illustration demonstrates that all other techniques have a substantial occurrence of errors. Compared to previous strategies, the MSE error of proposed model technique is relatively low. As a result, the proposed framework surpasses all previous approaches.

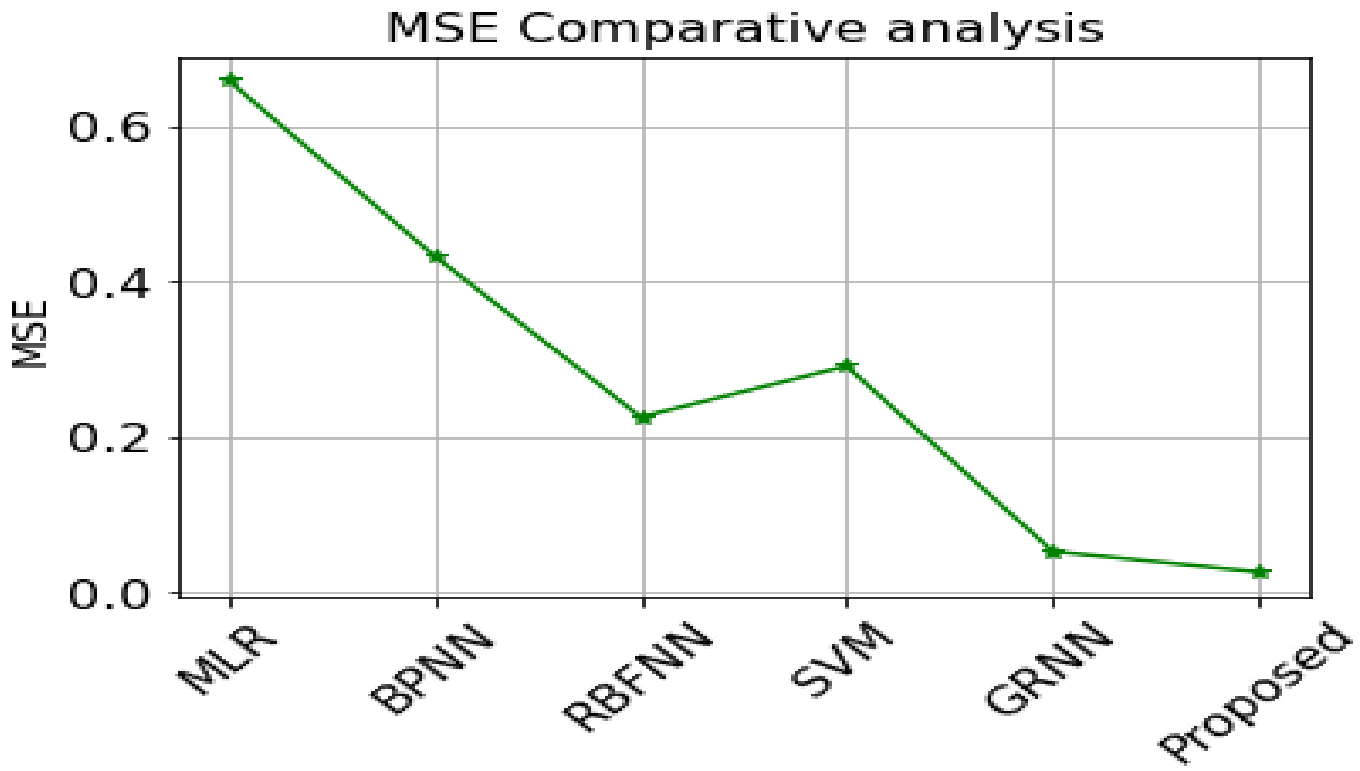

Figure 11. MSE comparative analysis

The following table 5 shows the method and its errors for MSE method. The Mean Square Error is reduced by employing the proposed SDCNN-LSTM approach. This research achieves less mean square error when compared to the baseline as MLR [36], BPNN [36], RBFNN [36], SVM [36], and GRNN [36], such as 0.7, 0.42, 0.24, 0.31, and 0.03. As a result, compared to existing approaches, the novel, distinctive technology has an error as 0. 01.

Table 5. Method and errors for MSE

Method	**MSE**
MLR [36]	0.7
BPNN [36]	0.42
RBFNN [3]	0.24
SVM [36]	0.31
GRNN [36]	0.03
Proposed	0.01

RMSE Comparative analysis

Figure 12 depicts the RMSE comparative analysis. IN BPNN, RMSE is quite high, while GRNN is very low. However, compared to other methodologies, our model achieves minimal RMSE errors ie) 0.01. As a result, we once again outperform all other strategies.

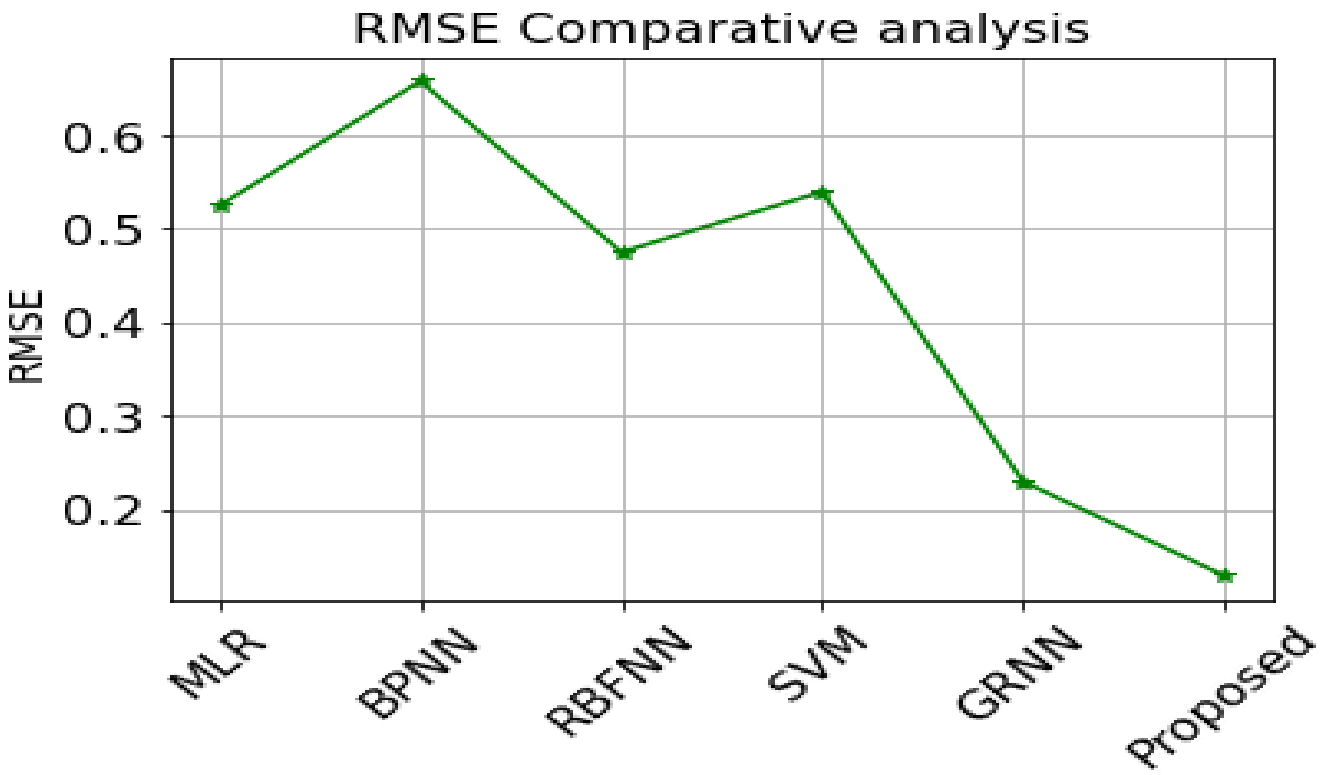

Figure 12. RMSE comparative analysis

The following table 6 shows the comparative analysis of methods for RMSE. The Root Mean Square Error is reduced by employing the proposed SDCNN-LSTM approach. This research achieves less root mean square error when compared to the baseline as MLR [36], BPNN [36], RBFNN [36], SVM [36], and GRNN [36], such as 0.52, 0.65, 0.48, 0.53, and 0.22. As a result, compared to existing approaches, the novel, distinctive technology has an error as 0. 01.

Table 6. Comparative analysis for RMSE

Method	RMSE
MLR [36]	0.52
BPNN [36]	0.65
RBFNN [36]	0.48
SVM [36]	0.53
GRNN [36]	0.22
Proposed	0.01

R^2 Comparative Analysis

The comparative analysis of R2 is shown in Figure 13. The coefficient of determination (R2) value of all the other approaches in this figure is the lowest, while our proposed model has the greatest Coefficient of determination. The proposed strategy, however, outperforms all previous ways.

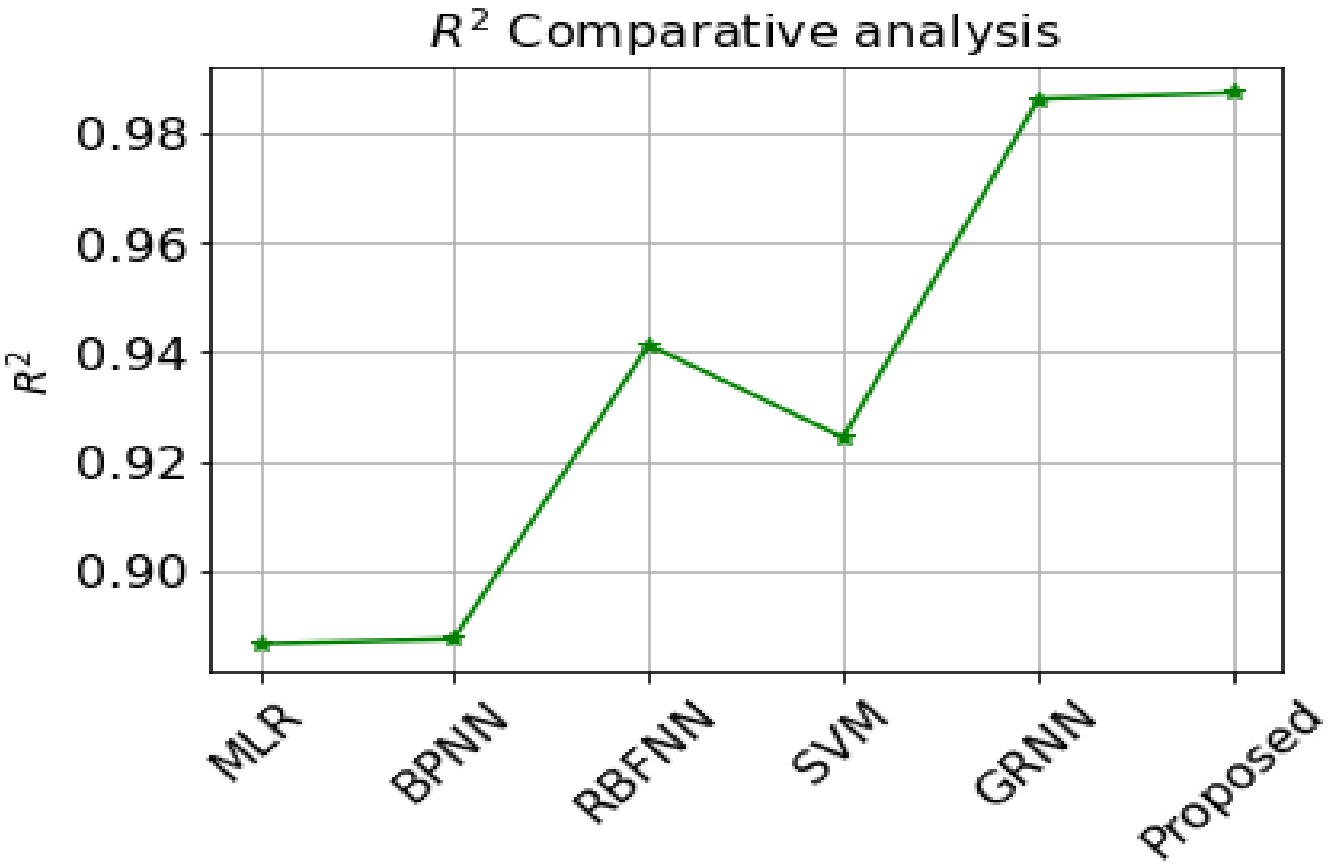

Figure 13. R2 comparative analysis

The following Table 7 shows the comparative Analysis of various methods for R2. The R2 is increased by employing the proposed SDCNN-LSTM approach. This research achieves high R2 error when compared to the baseline as MLR [36], BPNN [36], RBFNN [36], SVM, and GRNN [36], such as 0.30, 0.32, 0.94, 0.924, and 0.985. As a result, compared to existing approaches, our novel, distinctive technology has an error as 0. 99.

Table 7. Comparative analysis for R2

Method	**R^2**
MLR [36]	0.30
BPNN [36]	0.32
RBFNN [36]	0.94
SVM [36]	0.924
GRNN [36]	0.985
Proposed	0.99

Therefore, the proposed approach obtains less error when compared to the existing research that illustrates the high accuracy. This research offers a useful deep learning-based method for forecasting wheat yield. In this study, we test accuracy, evaluate its performance against other current methodologies, and achieve 99% accuracy. The accuracy of the MSE, RMSE, R^2, and MAE values are shown in our result section, and our error value is low. The graph shown in this approach demonstrates how well our innovative strategy performs. We compare each technique to the proposed approach and get superior outcomes.

CONCLUSION

The wheat yield prediction approach was studied using deep learning techniques in this work. The proposed method predicts wheat production by collecting soil and weather data, which is then pre-processed. The pre-processed data is sent to a framework with better deep convolutional long short-term memory. For

higher accuracy, we are utilizing an enhanced Long Short-term approach and gradient descent algorithm to eliminate overfitting and early stopping difficulties in the network. Data from the integrated layer is supplied into the fully linked layer to be normalized. Finally, the work accurately predicts wheat yield and surpasses all existing strategies. For improved outcomes in the future, researchers are urged to adopt a hybrid deep learning-based crop yield or wheat yield prediction, as well as enriched machine learning techniques. Furthermore, we can include a fertilizer dataset to preserve crops, and then we can anticipate wheat yield and production improvements, which will outperform all other techniques

Funding: This Research received no external funding
Acknowledgement: The authors would like to thank the Deanship of GITAM (Deemed to be University) for supporting this work.
Conflict of Interest: The Author Declare no conflict of Interest

REFERENCE

1. Bell GD. The history of wheat cultivation. In Wheat breeding: its scientific basis 1987;31-49.
2. Shewry PR, Hey SJ. The contribution of wheat to human diet and health. FES 2015; 4(3): 178-202.
3. Curtis T, Halford NG. Food security: the challenge of increasing wheat yield and the importance of not compromising food safety. AAB. 2014; 164(3): 354-72.
4. USDA, 2020a. United States Department of Agriculture Foreign Agricultural Service. Available online: https://ipad.fas.usda.gov/cropexplorer/cropview/. 2020.
5. USDA, 2020b. United States Department of Agriculture National Agricultural Statistics Service. Available online: https://www.nass.usda.gov/ 2020.
6. Cai Y, Guan K, Lobell D, Potgieter AB, Wang S, Peng J, et.al. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. AFM. 2019; 274: 144-59.
7. JP A. MapReduce and optimized deep network for rainfall prediction in agriculture. CJ. 2020; 63(6): 900-12.
8. Lobell DB, Cassman KG, Field CB. Crop yield gaps: their importance, magnitudes, and causes. ARER. 2009; 34: 179-204.
9. Basso B, Cammarano D, Carfagna E. Review of crop yield forecasting methods and early warning systems. InProceedings of the first meeting of the scientific advisory committee of the global strategy to improve agricultural and rural statistics, FAO Headquarters, Rome, Italy 2013;41:1-56.
10. Schlenker W, Roberts MJ. Nonlinear effects of weather on corn yields. RAE. 2006;28(3):391-8.
11. Patrício DI, Rieder R. Computer vision and artificial intelligence in precision agriculture for grain crops: A systematic review. CEA. 2018;153:69-81.
12. Feng Q, Liu J, Gong J. UAV remote sensing for urban vegetation mapping using random forest and texture analysis. RS. 2015;7(1):1074-94.
13. Feng Q, Liu J, Gong J. Urban flood mapping based on unmanned aerial vehicle remote sensing and random forest classifier—A case of Yuyao, China. Water. 2015;7(4):1437-55.
14. Ali I, Greifeneder F, Stamenkovic J, Neumann M, Notarnicola C. Review of machine learning approaches for biomass and soil moisture retrievals from remote sensing data. RS. 2015;7(12):16398-421.
15. Ip RH, Ang LM, Seng KP, Broster JC, Pratley JE. Big data and machine learning for crop protection. Comp EA. 2018;151:376-83.
16. Jeong JH, Resop JP, Mueller ND, Fleisher DH, Yun K, Butler EE, et.al. Random forests for global and regional crop yield predictions. PloS one. 2016;11(6):e0156571.
17. Han J, Zhang Z, Cao J, Luo Y, Zhang L, Li Z, et.al. Prediction of winter wheat yield based on multi-source data and machine learning in China. Remote Sensing. 2020;12(2):236.
18. Kuwata K, Shibasaki R. Estimating corn yield in the united states with modis evi and machine learning methods. ISPRS APRSSIS. 2016;3:131-6.
19. Kim N, Lee YW. Machine learning approaches to corn yield estimation using satellite images and climate data: a case of Iowa State. KSSGPC. 2016;34(4):383-90.
20. Feng Q, Yang J, Zhu D, Liu J, Guo H, Bayartungalag B, et.al. Integrating multitemporal Sentinel-1/2 data for coastal land cover classification using a multibranch convolutional neural network: A case of the Yellow River Delta. RS. 2019;11(9):1006.
21. Hochreiter S, Schmidhuber J. Long short-term memory. NC. 1997;9(8):1735-80.
22. You J, Li X, Low M, Lobell D, Ermon S. Deep gaussian process for crop yield prediction based on remote sensing data. InProceedings of the AAAI conference on artificial intelligence 2017;31(1).
23. Wang AX, Tran C, Desai N, Lobell D, Ermon S. Deep transfer learning for crop yield prediction with remote sensing data. InProceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies. 2018;1-5.
24. Wolanin A, Mateo-García G, Camps-Valls G, Gómez-Chova L, Meroni M, Duveiller G, et.al. Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt. ERL. 2020;15(2):024019.
25. Jiang Z, Liu C, Hendricks NP, Ganapathysubramanian B, Hayes DJ, Sarkar S. Predicting county level corn yields using deep long short term memory models. arXiv preprint arXiv:1805.12044. 2018 May 30.

26. Cunha RL, Silva B, Netto MA. A scalable machine learning system for pre-season agriculture yield forecast. In2018 IEEE 14th International Conference on e-Science (e-Science). 2018;423-30.
27. Aghighi H, Azadbakht M, Ashourloo D, Shahrabi HS, Radiom S. Machine learning regression techniques for the silage maize yield prediction using time-series images of Landsat 8 OLI. IEEE JSTAEO RS. 2018; 11(12): 4563-77.
28. Feng L, Wang Y, Zhang Z, Du Q. Geographically and temporally weighted neural network for winter wheat yield prediction. RSE. 2021;262:112514.
29. Shafiee S, Lied LM, Burud I, Dieseth JA, Alsheikh M, Lillemo M. Sequential forward selection and support vector regression in comparison to LASSO regression for spring wheat yield prediction based on UAV imagery. Computers and Electronics in Agriculture. 2021;183:106036.
30. Aula L, Omara P, Nambi E, Oyebiyi FB, Dhillon J, Eickhoff E, et.al. Active optical sensor measurements and weather variables for predicting winter wheat yield. AJ. 2021; 113(3): 2742-51.
31. Aravind KS, Vashisth A, Krishanan P, Das B. Wheat yield prediction based on weather parameters using multiple linear, neural network and penalised regression models. Journal of Agrometeorology. 2022;24(1):18-25.
32. Srivastava AK, Safaei N, Khaki S, Lopez G, Zeng W, Ewert F, et.al. Winter wheat yield prediction using convolutional neural networks from environmental and phenological data. Scientific reports. 2022;12(1):3215.
33. Pang A, Chang MW, Chen Y. Evaluation of random forests (RF) for regional and local-scale wheat yield prediction in southeast Australia. Sensors. 2022;22(3):717.
34. Zinkevich M, Weimer M, Li L, Smola A. Parallelized stochastic gradient descent. Advances in neural information processing systems. 2010;23.
35. Hardt M, Recht B, Singer Y. Train faster, generalize better: Stability of stochastic gradient descent. In International conference on machine learning 2016;1225-1234.
36. Joshua V, Priyadharson SM, Kannadasan R. Exploration of machine learning approaches for paddy yield prediction in eastern part of Tamilnadu. Agronomy. 2021;11(10):2068.