International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

The Deep Learning Approach for Crop Selection and Yield Prediction using Bi-LSTM in Agriculture

Dr. M. Supriya*[1], V. Subitha[2], R. Sumathi[3], P. Agnes Alex Rathy[4]

Submitted: 12/01/2025 **Revised:** 22/02/2025 **Accepted:** 05/03/2025

Abstract: Accurate crop yield prediction is essential for optimizing agricultural practices, ensuring food security, and maximizing resource efficiency. Traditional methods often fail to capture the complex, sequential dependencies in agricultural data, limiting their predictive accuracy. This work focuses on improving crop yield prediction by overcoming the drawbacks of conventional methods and integrating sequential data. The presented Bi-LSTM model provides better results than other machine learning and deep learning models since it uses all dependencies of temporal data of agriculture data. The study used Agricultural Crop Yield dataset then training and testing Bi-LSTM model. The performance is compared with other methods such as Linear Regression, Random Forest and basic LSTM to determine Mean Absolute Error, Root Mean Squared Error, R^2 score and Mean Absolute Percentage Error. The Bi- LSTM model yields the best result with MAE=0.32, RMSE =0.47 and R^2 Score =0.91. It efficiently incorporates features like rainfall, usage of fertilizers, which proves its applicability in the data of crop yields data. The analysis proves Bi-LSTM to be effective in predicting crop yield and offers a sound approach for decision support in agriculture.

Keywords: *Agriculture, Bidirectional Long Short-Term Memory, Crop Selection, Deep Learning, Prediction*

1. Introduction

Agriculture remains an essential sector in the world economy contributing income, employment, food, and gross domestic product [1]. It forms the economy of many of the developing countries given that it provides employment to a significant number of population. However, agriculture has some problems it's going through like changes in climatic conditions, soil erosion, pest attacks and lack of assets [2]. Precision crops choice and yield estimation have assumed major roles in dealing with these issues, making it possible for farmers to arrive at rational choices and increase efficiency. The specific field of agriculture exhibits relationships that do not easily lend themselves to being captured by the traditional models or rule based approaches, thus resulting in relatively low accuracy of the resultant models [3]. Such models often fail to incorporate nonlinear relation, sequential dependence and large scale data into the computation of prediction.

The presence of deep learning brought new method to solve these problems; Bi-LSTM (Bidirectional LSTM) are the proactive solutions to these problems. Bi-LSTM comes out most suitable for analysis of sequential data as this model can read from past and the future which makes great sense for when handling data like in the yield prediction case [4]. In order to improve the prediction accuracy as well as to capture the non-linear patterns in climate and soil data for sustainable agriculture precise recommendations the researchers use Bi- LSTM [5]. Their implementation is a major step in enhancing the modernization of agriculture and the fight for food security on the globe [6]. This paper focuses on the improvement in crop selection strategy accuracy and predictive capability of crop yields using Bi-LSTM networks. The research thus seeks ways on yield predictions so improvements are made for actionable insights in the decision-making of farmers and policy-makers toward sustainable, resilient, and productive agricultural practices.

The remaining section are aligned as follows: Related work in Section II. In Section III, the research mechanism is explained. The experimental findings are reported and compared in Section IV. In Section V, further work is mentioned and the study is

[1, 2,3] *Department of Computer Science and Engineering Stella Mary's College of Engineering Aruthenganvilai, Tamil Nadu, India -629202*
[4]*Department of IT,St. Xavier's Catholic College of Engineering,*
Chunkankadai, Nagercoil - 629003.
** Corresponding Author Email: smily.supriya@gmail.comsupriya@stellmaryscoe.edu.in*

concluded.

2. Literature Survey

Traditional machine learning algorithms, such as SVM [7], RF [8], and Decision Trees [9], have been widely used for crop selection and yield prediction. These models are best suited for dealing with structured and often time's static data and provide prediction based on history. However, an absence of capturing the temporal dependencies is their disadvantage [10], since they do not fully employ causal characteristics of data points, considering data independently from other data points in sequence.

Yield prediction has greatly benefited from recent deep learning techniques such as LSTM networks and CNNs in enhancing the positive gains of yield prediction [11]. This is especially used while handling the time- series data because LSTM retains the information from previous sequences to make predictions on future learnings [12]. The CNNs, however, are best suited for spatial data conditioning like pattern extraction from satellite images of crop fields. Nevertheless, such models may experience difficulties in capturing bidirectional dependence in sequential data important for agricultural trends [13].Incorporating a Bi-LSTM , the forecasting capabilities of base LSTMs are strengthened since data is processed in both forward and also in the reverse [14]. This bidirectional passing facilitates the Bi-LSTM to possess a better understanding of temporal relationships, something important in agriculture applications such as seasonal change, climate fluctuation and soil movement. Bi-LSTM gives better results in the modeling of rich contextual agricultural data because it makes use of contextual information whether past or in the future. There is a list of difficulties in existing methods, several of which are mentioned below: Despite having valuable outcomes. Most of the models do not incorporate real time data to reflect environmental conditions in the model and therefore the models are not dynamic. Second, the applicability of the results is a problem since the results depend on the soil type, climate, and farming practices in the different regions and the crop type. This has created the need for better and more resilient models as are the Bi-LSTM models to help improve the prediction accuracy and versatility especially within the agriculture segment.

3. Methodology

The methodology outlines the systematic approach adopted to predict crop yields using Bi-LSTM. Preliminary steps include data cleaning and data normalization. Second, Bi-LSTM architecture is used as next step is to identify the sequential patterns and temporal dependency on the agricultural data. The training process is a complex step that uses the previous crop data to set up the model parameters and subsequently adjusts the hyper parameters to a better value. Lastly, the performance of the model is compared against different evaluation criteria to improve the accuracy of the relationship between features and crop yield. The overall workflow is shown in Fig.1.

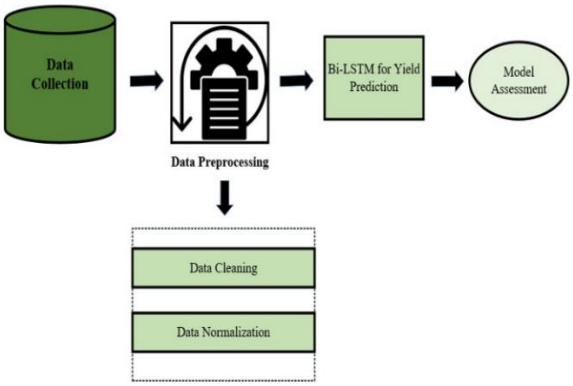

Fig. 1. Workflow of Proposed Model

3.1 Data Collection Proposed

The dataset includes consistent and extended agricultural information about more than one crop grown in different states of India for a period of 1997-2020. It contains features that offer the basis for crop yields estimation including crop type, cropping period, state, year planted, area of cultivation, production amount, annual rainfall, applied fertilizer, applied pesticide, and forecasted crop yield (quantity per area unit). This Kaggle dataset is useful in generating models in machine learning with special reference to yields, trending and recommendation of right crops from climatic characteristics, geographic location and resources. Due to its broad temporal coverage, this index provides a suitable means of assessing the impacts of shift in conditions and farming practices [15].

3.2 Data Preprocessing

Data processing relates to the preparation of data for the deep learning model for training. This includes, data pre- cleaning and transformation of feature scale. Good preprocessing leads to a faster convergence as well as to fewer overfitting, which actually leads to better performance by the model.

3.1.1 Data Cleaning

Data cleaning includes the correction of missing, inconsistent, or erroneous values to make a dataset reliable. Missing numeric features such as rainfall and fertilizer usage are imputed with statistical methods like a mean or median, whereas categorical variables like crop type or season are filled with the mode. Otherwise, if imputation is not an option, records with excessive missing data in critical fields can be dropped. Removal of duplication for redundancy and detection of outliers in features such as area, production, or pesticide usage with method like Interquartile Range (IQR).

3.1.2 Normalization

Normalization is used in this work to bring the features into a consistent range of magnitudes, which is crucial for the Bi-LSTM models to converge. One of them is the Min-Max Scaling method, transforms features into a range of [0, 1]. It is expressed in equation (1),

$$N = \frac{N - N_{min}}{N_{max} - N_{min}} \quad (1)$$

Where, N indicates an initial value of the feature in the record. $Nmin$ and $Nmax$ are the lower and upper limit of the values for that feature. This is helpful since variables with higher numeric value such as area in hectares and rainfall in millimeters need to be normalized so the magnitude of the features does not take over the learning process.

3.3 Bi-LSTM for Crop Selection and Prediction

The Bi-LSTM framework is an advanced deep learning architecture designed specifically to capture sequential data proficiently by integrating past and future temporal dependencies within time-series datasets. This feature is particularly important for applications that require an analysis of temporal trends, including crop yield prediction, for which both historical information and prospective patterns are necessary. Each constituent element of the Bi-LSTM are given below.

3.3.1 Input Layer

The input layer of Bi-LSTM model is the first layer at which multivariate features are fed. Therefore, the multivariate features are, the total land size for cultivation (in hectares), the total yield of the crop (in metric tons), the amount of rain received in the crop producing area (in mm), fertilizer consumption (in kg and as a proportion of the total production), and pesticide consumption (in kg and as a proportion of the total production). In this work, these features are presented in a sequential manner to enable them be input in the model in a time-series form. They are yearly sequences of these features, which allows the formulated model to identify changes in them over time and learn patterns over periods of time.

3.1.2 Bi-LSTM Layer

The core component of the Bi-LSTM architecture is the Bi-LSTM layer, which enhances the traditional LSTM by incorporating two separate layers of LSTM units: while the first one filter the data from past to the present, the second one filters the data from the future to the past. This additional step gives the Bi-LSTM model better temporal insight on the input data through capturing forward in addition to backward dependencies. This interdependencies successfully learned by the Bi-LSTM layer which scans the input sequence in both directions. The forward LSTM layer learns the past streams and backward LSTM layer learns the future streams so the model learns everything about the data. This two-way phenomenon is particularly convenient for agricultural uses of the technology because it reflects both past and future data. The Bi-LSTM layer proves highly effective in capturing temporal dependencies in both directions preventing loss of valuable data needed for correct prediction of crop yields.

3.1.3 Output Layer

The last layer of the proposed Bi-LSTM model is called the output layer of the model where the computed crop yield is predicted. It often contains a single neuron with parameters that make it output a continuous value – the predicted yield. In cases of multi-output tasks, the output layer may contain multiple neurons, each representing a separate prediction (e.g., the yield corresponding to different crops). The output is then produced with an activation function best for regression model like ReLU activation function. When it comes to crop yield prediction, the linear activation function is more employed more frequently since it is effective in mapping learned characteristics from the Bi-LSTM layer directly onto a continuous target crop yield without inducing any restrictions on the result limits. This is especially so since crop yield is likely to greatly differ based on parameters including: weather conditions, application of fertilizers and the type of crops grown. The output layer effectively consolidates the information captured by the Bi-LSTM layer and translates it into a prediction that aligns with the goal

of the research and enables the accurate and easy to interpret forecast of the yield of crops. By providing an output of a continuous value, the model provides information for Input- Output analysis of resources, selection of crops and yield prediction for use in agricultural planning. The paper suggests Bi-LSTM is crucial for crop selection and yield prediction. Hence capable of handling the complexities which is crucial for understanding the trends and patterns over time, such as changes in weather, resource usage, and crop performance. The strength is the ability to model temporal dependency, which is very essential for comprehending trends and patterns through time, like weather changes, resource usage, and crop performance. Processing data both in forward and backward directions, Bi-LSTM captures those temporal relationships entirely. Additionally, Bi-LSTM also enables multivariate feature integration, allowing the model to process multiple features at one time, like rainfall, fertilizer usage, and production. This facilitates the model's learning process of intricate variable interactions, further enabling it to understand and predict how crop yield is affected by this. Relative to statistical and standard unidirectional models, Bi-LSTM offers high accuracy for a deeper analysis of sequential data; adaptability is its added advantage that allows accommodating much more time- dependent features so easily to satisfy the multiple needs of variously diverse requirements of different agricultural study datasets. This architecture is well aligned with the research objective of developing a robust and accurate crop yield prediction model, leveraging advanced deep learning techniques to address issues in agricultural forecasting.

4. Result And Discussions

The Bi-LSTM has been compared with conventional models in crop yield prediction and the result reaffirm this model is more proficient than others. Evaluation metrics also show that Bi-LSTM has better performance and prediction capacity than the other models. This model enables identifying temporal patterns and temporal relations in agricultural data better than compared methods. To make more decisions in the agricultural process, key findings such as the effects of some features such as rainfall and fertilizer usage etc. on yield are first identified.

4.1. Experimental Outcome

Crop Yield Prediction Summary provides statistical summary crop yield predictions on major crops of India with Average, Maxima, and Minima in MTs/ha. Rice has the highest average yield and is found to be best at 2.5 MTs/ha in Punjab. Wheat: The maxima is achieved at 5.0 MTs/ha attained in Uttar Pradesh. Cotton and maize have average yields that are the lowest among all the crops. The top state performers in both these crops are Gujarat and Maharashtra, respectively. Other crops have yielded varied results. Tamil Nadu stands out among the regional disparities and crop- specific yield trends. It is clearly shown in Table 1.

Table 1. Crop Yield Prediction

Crop	*Average Yield (MT/ha)*	*Max Yield (MT/ha)*	*Min Yield (MT/ha)*	*State with Highest Yield*
Rice	2.5	4.0	0.5	Punjab
Maize	2.3	3.8	0.6	Maharashtra
Wheat	2.6	5.0	1.0	Uttar Pradesh
Cotton	1.8	3.2	0.3	Gujarat
Other	1.9	3.5	0.2	Tamil Nadu

Table 2 shows the Regional Analysis of Crop Yield which compares top states with regard to crop yield, total cultivated area and production in regions of India.

Table 2. Regional Analysis of Crop Yield

State	*Average Yield (MT/ha)*	*Total Area (ha)*	*Total Production (MT)*	*Most Common Crop*
Punjab	3.0	1,200,000	3,600,000	Wheat
Maharashtra	2.8	950,000	2,660,000	Cotton
Uttar Pradesh	2.6	1,500,000	3,900,000	Rice
Tamil Nadu	2.1	800,000	1,680,000	Rice
Andhra Pradesh	2.3	850,000	1,950,000	Rice

The average yield of crop has topped at 3.0 MT/ha at Punjab, which largely cultivate wheat across 1.2 million hectares at an estimated yield of 3.6 million metric tons. Maharashtra is at the same level with an

average yield of 2.8 MT/ha, largely with cotton, while Uttar Pradesh covers much larger area, 1.5 million hectares, with lower average yield of 2.6 MT/ha and is largely used for paddy. Tamil Nadu and Andhra Pradesh have lower average yields. Rice is the prime crop in both the states, making this table very much important to show the regional variance of crop yields productivity as well as area under cultivation. It is clearly depicted in Fig.2.

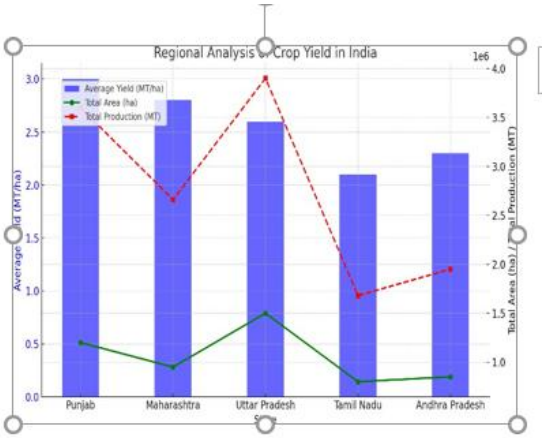

Fig. 2. Regional Analysis of Crop

The Table 3 shows how different features of agriculture determines the yield of crops. The analysis of correlation between fertilizer usage and yield reveals the strongest positive link that is 0.62 thus explaining the usage's importance. Yield is also affected by, annual rainfall slightly (0.45) and pesticide usage slightly (0.35). The area under cultivation (0.21, t= 2.93, $p < 0.05$) and crop year (0.12, t=2.03, $p < 0.05$) displayed less direct and thus have lower correlation coefficients to the yield outcomes. This analysis helps sort out items for enhancing figures concerning crop yields.

Table 3. Correlation between Features and Yield

Feature	*Correlation with Yield*
Annual Rainfall	0.45
Fertilizer Usage	0.62
Pesticide Usage	0.35

4.2. Model Comparison

The performance of each model is assessed using standard evaluation metrics it includes MAE, R^2, RMSE and MAPE. It is expressed in equation (2) to (5).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}| y_{pred}^{i} - y_{obs}^{i}| \quad (2)$$

$$R^2 = 1 - \frac{\sum n_{i=1}^{n}(y_{pred}^{i} - y_{obs}^{i})}{\sum_{obs}^{n}} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\frac{y_{pred}^{i} - y_{obs}^{i}}{n})} \quad (4)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_{pred}^{i} - y_{obs}^{i}|}{y_{obs}^{i}} \quad (5)$$

Where, n is total samples, predicted and observed data are represented as y_{pred}^{i} and y_{obs}^{i}

The Table 4. Shows Bi-LSTM has been tested for crop yield prediction using evaluation metrics with traditional methods. Bi-LSTM has been proven to be accurate in the predictions as it has a smaller MAE of 0.32 and RMSE of 0.47 than all other methods. It also has the best R^2 Score of 0.91 which points to the fact that crop yield variation can be well explained. In addition, Bi-LSTM produces the lowest result in terms of MAPE with the figure of 8.5% proving that Bi-LSTM offers the most effective way for reducing percentage errors. This comparison underlines the significance of Bi-LSTM in exploiting the chronological data for predicting the crop yield proficiently. It is clearly shown in Fig.3.

Table 4. Correlation between Features and Yields

Method	*MAE*	*RMSE*	*R^2 Score*	*MAPE (%)*
Linear Regression	0.58	0.74	0.72	18.5
Support Vector Machine (SVM)	0.51	0.67	0.78	15.8
Random Forest (RF)	0.42	0.56	0.83	12.4
Artificial Neural Network (ANN)	0.39	0.53	0.85	11.9
Long Short-Term Memory (LSTM)	0.35	0.50	0.88	9.8

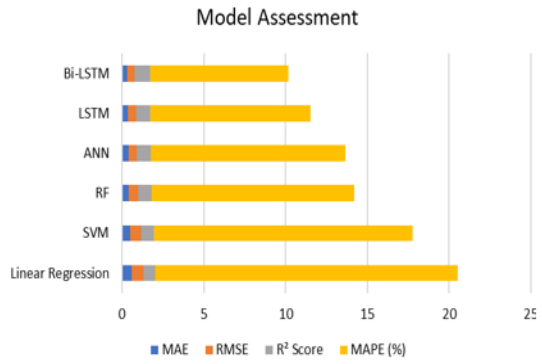

Fig. 3. Model Comparison

5. Conclusion and Future Works

Using Bi-LSTM to predict crop yields, this work demonstrates that the proposed method is superior to conventional approaches and other deep learning architectures. In fact, Bi-LSTM outperforms other models and reaches high values on the fundamental assessment criteria, based on sequential data and temporal patterns. The findings also show how factors, such as fertilizer use and rainfall, affect the crop yield and provide the framework for improving the yields through changes in practices. Future work will benefit from combining real-time data sources that are included in the IoT-based weather monitoring and satellite imagery. Furthermore, this approach could be expanded to perform multi-crop as well as regional predictions, also, fallback of developing hybrid models can scale up the problem of generalization in the field of agricultural outlooks. Other possibilities for further research could be applied to create the hybrid between Bi-LSTM and other approaches. These improvements would enable such a framework for increased scalability for large dataset, extended geographic regions, and multiple crop species, thus increasing the generality of context in the modern techniques of farming.

References

[1] E. Silamat, P. Priyono, and H. Hernawati, "Impact of the Agricultural Sector on Output, GDRP and Workforce Compensation," *Int. J. Econ. Bus. Innov. Res.*, vol. 2, no. 02, Art. no. 02, Feb. 2023.

[2] P. Khatri, P. Kumar, K. S. Shakya, M. C. Kirlas, and K. K. Tiwari, "Understanding the intertwined nature of rising multiple risks in modern agriculture and food system," *Environ. Dev. Sustain.*, vol. 26, no. 9, pp. 24107–24150, Sep. 2024, doi: 10.1007/s10668-023-03638-

[3] A. Gupta and P. Nahar, "Classification and yield prediction in smart agriculture system using IoT," *J. Ambient Intell. Humaniz. Comput.*,vol. 14, no. 8, pp. 10235–10244, Aug. 2023, doi: 10.1007/s12652-021- 03685-w.

[4] M. Fathi, R. Shah-Hosseini, and A. Moghimi, "3D-ResNet-BiLSTM Model: A Deep Learning Model for County-Level Soybean Yield Prediction with Time-Series Sentinel-1, Sentinel-2 Imagery, and Daymet Data," *Remote Sens.*, vol. 15, no. 23, Art. no. 23, Jan. 2023, doi: 10.3390/rs15235551.

[5] B. Chen *et al.*, "A joint learning Im-BiLSTM model for incomplete time-series Sentinel-2A data imputation and crop classification," *Int.J. Appl. Earth Obs. Geoinformation*, vol. 108, p. 102762, Apr. 2022, doi: 10.1016/j.jag.2022.102762.

[6] C. Sun, H. Zhang, L. Xu, C. Wang, and L. Li, "Rice Mapping Using a BiLSTM-Attention Model from Multitemporal Sentinel-1 Data," *Agriculture*, vol. 11, no. 10, Art. no. 10, Oct. 2021, doi: 10.3390/agriculture11100977.

[7] M. Rajakumaran, "Crop yield prediction using multi-attribute weighted tree-based support vector machine - ScienceDirect." Accessed: Nov. 28, 2024. [Online]Available: https://www.sciencedirect.com/science/article/pii /S26659174230033 80#sec3

[8] K.P.Uvarajan and K.Usha, "Implement A System For Crop Selection And Yield Prediction Using Random Forest Algorithm," *Int. J. Commun. Comput. Technol.*, vol. 12, no. 1, Art. no. 1, Mar. 2024.

[9] M. K. Senapaty, A. Ray, and N. Padhy, "A Decision Support System for Crop Recommendation Using Machine Learning Classification Algorithms," *Agriculture*, vol. 14, no. 8, Art. no. 8, Aug. 2024, doi: 10.3390/agriculture14081256.

[10] D. R. I. M. Setiadi, A. Susanto, K. Nugroho, A. R. Muslikh, A. A. Ojugo, and H.-S. Gan, "Rice Yield Forecasting Using Hybrid Quantum Deep Learning Model," *Computers*, vol. 13, no. 8, Art. no. 8, Aug. 2024, doi:

10.3390/computers13080191.

[11] Y. Wang *et al.*, "Progress in Research on Deep Learning-Based Crop Yield Prediction," *Agronomy*, vol. 14, no. 10, Art. no. 10, Oct. 2024, doi: 10.3390/agronomy14102264.

[12] Y. Mahale *et al.*, "Crop recommendation and forecasting system for Maharashtra using machine learning with LSTM: a novel expectation- maximization technique," *Discov. Sustain.*, vol. 5, no. 1, p. 134, Jun. 2024, doi: 10.1007/s43621-024-00292-5.

[13] J. Dong, "Estimating reference crop evapotranspiration using improved convolutional bidirectional long short-term memory network by multi-head attention mechanism in the four climatic zones of China - ScienceDirect." Accessed: Nov. 28, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S03783774230053 09

[14] M. V. Krishna, K. Swaroopa, G. SwarnaLatha, and V. Yasaswani, "Crop yield prediction in India based on mayfly optimization empowered attention-bi-directional long short-term memory (LSTM)," *Multimed. Tools Appl.*, vol. 83, no. 10, pp. 29841–29858, Mar. 2024, doi: 10.1007/s11042-023-16807-7.

[15] A. Gupta, "Agricultural Crop Yield in Indian States Dataset." Accessed:Nov. 28, 2024. [Online].Available: https://www.kaggle.com/datasets/akshatgupta7/crop-yield-in-indian- states-dataset