Book Chapter

Chapter 2

Agriculture Crop Yield Prediction Using Deep Learning Models

S.Krithika *[1], T.A.Sangeetha †[2], Hanamant R Jakaraddi ‡[3], and N.Rajasekaran §[4]

[1]Assistant Professor, Dept. of Computer Science (PG), Kongu Arts and Science College (Autonomous), Erode
[2]Associate Professor, Dept. of Computer Applications, Kongu Arts and Science College (Autonomous), Erode
[3]Assistant Professor, Dept. of MCA, Acharya Institute of Technology, Banglore
[4]Assistant Professor, Dept. of Computer Applications, Kongu Arts and Science College (Autonomous), Erode

Abstract

Crop yield prediction is a big challenge in agricultural research. Due to the natural calamities, the farmers were not able to predict their crop yields. Hence, the prediction methodology is necessary for the researchers to identify the productivity and demand of the particular crop. Innovation in crop yield prediction models and methods can assist researchers in finding better results. The various machine learning (ML) models have been developed, and their performance has been evaluated through different research with real-time agricultural datasets. But still, the performance of the ML models is not satisfactory, and hence an improvement is needed in some factors. In this research, deep learning-based algorithms such as Deep Neural Network (DNN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short Term Memory (LSTM) were used to evaluate the

*Email: krithitup86@gmail.com Corresponding Author
†Email: tasangeethathirumalai@gmail.com
‡Email: hanamant2504@acharya.ac.in
§Email: rajasekarandpm@gmail.com

performance of the model for crop yield prediction. In this research, various experiments were performed using the DNN, CNN, RNN, and LSTM models based on the agricultural dataset. The proposed models were compared with the various features, and the Long Short Term Memory (LSTM) algorithm gave the best accuracy among the other models.

Keywords: Machine Learning. Crop Yield Prediction. Deep Learning Models. Long Short Term Memory.

1 Introduction

The agricultural industry has generated and amassed substantial volumes of data. In this market, the application of machine learning techniques has proven to be highly effective in facilitating competitive value creation through decision support through the collection, storage, and analysis of this information. It is crucial to address issues such as diminished soil fertility, inadequate irrigation infrastructure, low crop yields resulting from climate change, and subsistence-based agricultural methods. Climate, biological, and economic factors all influence agricultural output. Weather stations serve an essential function by delivering vital information regarding weather phenomena that have a direct influence on productivity. Predicting the health of crops, which has a direct impact on yield, is thus a critical issue that requires investigation. Historically, the estimation of crop and field productivity was based on the experience of the farmer (Bondre & Mahagaonkar, 2019). Anticipating the prospective crop yield represents a significant achievement for numerous stakeholders engaged in the agricultural production and trading domains. It is critical to provide producers with yield projections so they can effectively manage their resources and budget. Growers can thereby enhance the knowledge base of their economic and managerial choices, while timely identification of yield-related issues can facilitate the implementation of corrective measures for the entire crop. The ability to forecast crop yield has the potential to assist in the formulation of more informed strategies for organizing and carrying out operations. Consequently, predicting agricultural yields presents a formidable obstacle that requires resolution. The variety of seeds used, weather and soil conditions, fertilizer application, and yield level are all determinants of the plant phenotype (Thomas van Klompenburg & Catal, 2020). Despite their widespread application and criticality, the utilization of ML and DL models for agricultural yield prediction presents a number of obstacles. Training these models is a time-consuming process, particularly when they comprise a large number of layers. Furthermore, the efficacy of the models could vary depending on a multitude of factors. In addition, the performance of the most intricate models may not consistently be optimal, which complicates algorithm selection.

Deep learning is a subfield of machine learning that employs multi-layer analysis to uncover significant but concealed features within a given dataset by transforming raw

DOI:10.48001/978-81-980647-5-2-2 | ISBN: 978-81-980647-5-2

data (Hinton, 2018; Wang et al., 2010). An increased number of hidden layers in DL models can improve the accuracy of crop yield forecasts (Leong Wai Hong & Tunku Abdul Rahman, 2016). Although deep learning algorithms can provide better performance, the challenges of using deep learning techniques for crop yield prediction are lacking in the literature. They both depend on the crop type, the kind of data, the sources, and the implementation framework. We perform a systematic literature review (SLR) to get an overview of the ML and DL algorithms used for crop prediction. Machine learning, like other methodologies for yield prediction, including field surveys, crop growth models, and remote sensing, has the potential to enhance these approaches (Kale & Patil, 2019). In this study, we have investigated the power of deep learning algorithms to predict crop yields and their importance. The Deep Neural Networks (DNN) algorithm possesses the capability to effectively process non-linear data. The training of features from data is a characteristic of the Convolutional Neural Networks (CNN) algorithm. Recurrent Neural Network (RNN) shares the weight across time steps and enhances the training accuracy. The LSTM algorithm works effectively to address the issue pertaining to the long-term dependencies of RNNs.

Climate and soil conditions are factors that are considered when attempting to forecast an appropriate yield. The aim is to develop a Python-based system that intelligently employs strategies to predict the most profitable harvest under specific conditions while minimizing costs. Agarwal and Tarar's (2021) paper employs SVM as an algorithm for machine learning and LSTM and RANN as algorithms for deep learning. The optimized model, XGBoost, achieved a root mean square error (RMSE) of 0.755 and a mean absolute error (MAE) of 0.54 when trained on the original variables. By conducting a comparative analysis of different regression techniques, this paper endeavors to enhance the accuracy of yield prediction. By doing so, it provides farmers with valuable insights that can inform cultivation decisions and empower them to leverage the capabilities of predictive analytics (Khan, Mishra, & Baranidharan, 2020). Diverse deep learning-based algorithms are implemented internationally to extract useful commodities for forecasting. The integration of deep learning and data mining generates a comprehensive system for predicting crop yields, capable of establishing a connection between unprocessed data and predicted crop outputs. To estimate agricultural production, the proposed study employs a Discrete Deep Belief Network with Visual Geometry Group (VGG) Net classification method as opposed to the Tweak Chick Swarm Optimization method, with an accuracy of 97% of maintaining the baseline data distribution (Vignesh, Askarunisa, & Abirami, 2022). The results indicate that Random Forest performed the best of the employed regression techniques, including SVM, Gradient Descent, long short-term memory, and Lasso, with R2 values of 0.963, RMSE values of 0.035, and MAE values of 0.0251. Mean absolute error, root mean squared error, and R2 were utilized to validate the outcomes in conjunction with

DOI:10.48001/978-81-980647-5-2-2 | ISBN: 978-81-980647-5-2

cross-validation methods. The objective of this paper is to implement the crop selection method so that producers can resolve crop yield issues

Based on our analysis, the most frequently utilized attributes in these models are temperature, precipitation, and soil type, while the most frequently implemented algorithm is artificial neural networks. After making this observation through the examination of 50 papers utilizing machine learning, we proceeded with an additional search in electronic databases for studies employing deep learning. We ultimately located 30 papers that utilized deep learning and derived the implemented deep learning algorithms. This additional analysis indicates that convolution neural networks (CNN) are the deep learning algorithms most frequently employed in these studies, followed by long-short-term memory (LSTM) and deep neural networks (DNN) (Thomas van Klompenburg & Catal, 2020). The results of the study by Akhter and Sofi's (2021) suggest that through the integration of real-time online IoT sensor data with the analysis of a variety of agricultural data, farmers can arrive at more informed conclusions regarding the variables that influence crop growth. Ultimately, by increasing crop yields and decreasing pollution, the integration of these technologies has the potential to revolutionize contemporary agriculture. A new algorithm is presented that is augmented with a feature combination scheme. Fifteen distinct algorithms were evaluated in order to determine which ones were most suitable for irrigation. The findings indicate that the Bayes Net algorithm yields a classification accuracy of 99.59%, while the Naïve Bayes Classifier and Hoeffding Tree algorithms achieve an accuracy of 99.46% (Sharma & Kirkman, 2015). As a consequence of these findings, production rates will increase and farms will incur lower effective costs, which will contribute to the development of more resilient infrastructure and sustainable environments. Furthermore, the results acquired from this research can assist forthcoming farmers in the early detection of maladies, enhancement of crop production efficiency, and mitigation of prices during periods of global food scarcity. In addition, the researchers discuss the challenges that are associated with training the normal RNN and finds solutions to these challenges by changing the RNN into the "Vanilla LSTM network by means of a series of logical arguments.

2 Proposed Methodology

In the proposed work, deep learning algorithms are executed to predict the best crop yield. The proposed model conducts an experiment using a dataset of crops. In addition to climatic and soil parameters, the current atmosphere, the composition of the soil, and the area of cultivation are considered in determining which crop to cultivate and to predict the yield for the future. Deep learning models are utilized to accomplish a multitude of successful computations, such as determining the optimal crop and yield from a set of alternatives. With the aid of this method, precise crop yield forecasts

DOI:10.48001/978-81-980647-5-2-2 | ISBN: 978-81-980647-5-2

are able to be predicted accurately. This proposed model focuses on four different deep learning algorithms: Deep Neural Network (DNN), Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory (LSTM), as shown in Figure 1.

2.1 Architecture of the Proposed Model

Implementation Steps:

1. Load the crop dataset with multiple features.
2. Import the necessary libraries and packages.
3. Data preprocessing is done to enhance desired features or reduce artifacts that can bias the network.
4. The data is split into a training set and a testing set.
5. Construct the model by applying the deep learning algorithms (DNN, CNN, RNN, and LSTM) to predict a crop and its yield.
6. Use the test set to calculate the algorithm's accuracy.

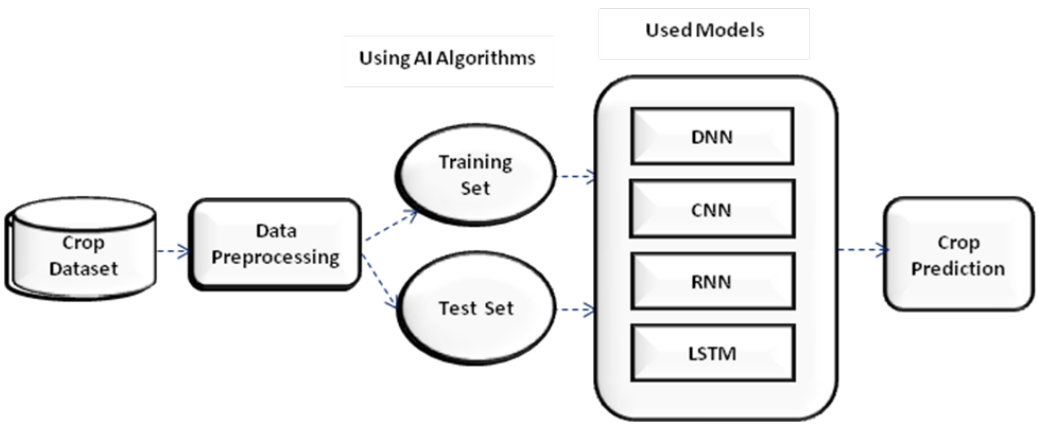

Figure 1. Proposed System

Editors: S. Pandikumar, Manish Kumar Thakur
DOI:10.48001/978-81-980647-5-2-2 | ISBN: 978-81-980647-5-2

2.2 Dataset Description

The datasets are collected from the website kaggle.com. The dataset contains multiple features such as area, crop, year, yield, rainfall, and pesticides. Table 1 displays a sample of the crop dataset. The dataset has a total of 26,297 instances, with a size of 8024 KB. The variety of crops includes types like rice, wheat, and maize. During preprocessing, the data is cleaned and normalized for all the features.

Table 1. Agricultural Data

Area	Item	Year	hg/ha_yield	average rainfall mm/year	pesticides_tonnes
Algeria	Maize	1990	16500	89.0	1828.92
Algeria	Potatoes	1990	78936	89.0	1828.92
Algeria	Rice, paddy	1990	28000	89.0	1828.92
Algeria	Sorghum	1990	16571	89.0	1828.92
Algeria	Wheat	1990	6315	89.0	1828.92
...	...	...	...	...	...
Zimbabwe	Rice, paddy	2013	22581	657.0	2550.07
Zimbabwe	Sorghum	2013	3066	657.0	2550.07
Zimbabwe	Soybeans	2013	13142	657.0	2550.07
Zimbabwe	Sweet potatoes	2013	22222	657.0	2550.07
Zimbabwe	Wheat	2013	22888	657.0	2550.07

2.3 Experimental Environment

The proposed deep learning algorithms are implemented in Python, using the following libraries:

- tensorflow: to import the dataset,
- keras: to use ANN functionalities,
- numpy and matplotlib: to check the prediction and plot the image,
- sklearn.model: to split the data into training and test sets,
- DNN: Dense, Flatten, Dropout layers,
- CNN: Conv1D, MaxPooling1D layers,
- RNN: LSTM.

DOI:10.48001/978-81-980647-5-2-2 | ISBN: 978-81-980647-5-2

2.4 Proposed Algorithms

2.4.1 Deep Neural Network (DNN)

Feed-forward networks (FDNs) are deep neural networks (DNNs), where data flows unidirectionally from the input layer to the output layer. This indicates that the procedure continues without further interaction with the node. Some studies identify the use of deep learning approaches to aid in the prediction of diseases affecting crops, aiming to support agriculture.

Steps for DNN Implementation:

1. Import the required libraries.
2. Load the dataset using Keras and preprocess it.
3. Build the neural network model using keras.sequential.
4. Create an input layer with the keras.layers.Flatten method.
5. Set the hidden layer, defining the number of neurons and using the relu activation function.
6. Define the output layer with the keras.layers.Dense method and assign the softmax activation function.
7. Compile the model using the compile method, setting arguments like optimizer="adam", loss="sparse_categorical", and metrics like accuracy, MAE, MSE, and R2.
8. Pass the input data and target labels through the network, adjusting weights and biases.
9. Assess the test dataset using model.evaluate.
10. Predict on testing data with model.predict.

2.4.2 Convolutional Neural Network (CNN)

The Convolutional Neural Network procedure begins with the convolution operation, where a filter scans the input data to identify specific features. The outcome is a feature map that highlights detected features in the data. Using this feature map as input for the next layer, CNN constructs a hierarchical data representation. The CNN component considers internal temporal dependencies in meteorological data and spatial dependencies in soil data obtained at various depths.

Steps for CNN Implementation:

DOI:10.48001/978-81-980647-5-2-2 | ISBN: 978-81-980647-5-2

1. Import necessary libraries and packages.
2. Load the dataset and set labels.
3. Reshape input data to have a single channel.
4. Build the model and add layers (input, hidden, and output) using Conv1D, MaxPooling1D, and Flatten.
5. Apply the sigmoid method for activation.
6. Adjust the output layer.
7. Compile the model, setting optimizer, loss, and metrics.
8. Train the model to predict accuracy.

2.4.3 Recurrent Neural Network (RNN)

The fundamental processing unit of an RNN is the recurrent neuron, which sustains a hidden state, allowing the network to capture sequential dependencies by retaining prior inputs. Through a hidden layer, the RNN resolves issues involving sequential data.

Steps for RNN Implementation:

1. Define the network using Keras in layers.
2. Provide the network with a single time-step of input.
3. Compute the current state from the previous state and current input.
4. The current state h_t becomes h_{t-1} for the next time step.
5. Iterate over time steps, retaining the information from the entire previous state.
6. After completing all time steps, utilize the final current state to compute the output.
7. If weights are updated, errors are back-propagated through the network.

2.4.4 Long-Short Term Memory (LSTM)

LSTM techniques are used to make accurate predictions for the future yield of various agricultural products, providing the network with short-term memory across many phases.

Steps for LSTM Implementation:

1. Import Keras and LSTM.
2. Set input, hidden, and output layers.

DOI:10.48001/978-81-980647-5-2-2 | ISBN: 978-81-980647-5-2

3. Specify the activation function.
4. Compile the network with specified parameters.
5. Fit the model to the training data.
6. Check both input patterns matrix X and output patterns array y.
7. Evaluate the network on training data.
8. Predict the model accuracy.

3 Experimental Results and Discussion

The implementation is done in Python by importing libraries for prediction. The crop dataset, saved in CSV format, is loaded and examined for features. Data preprocessing techniques remove empty and duplicate data from the dataset. Summary statistics of the numerical and categorical variables are computed, and EDA is applied to finalize the data. Figures 2 and 3 illustrate yield per year and pesticide usage per crop. These analyses help train the model to make predictions with the test dataset.

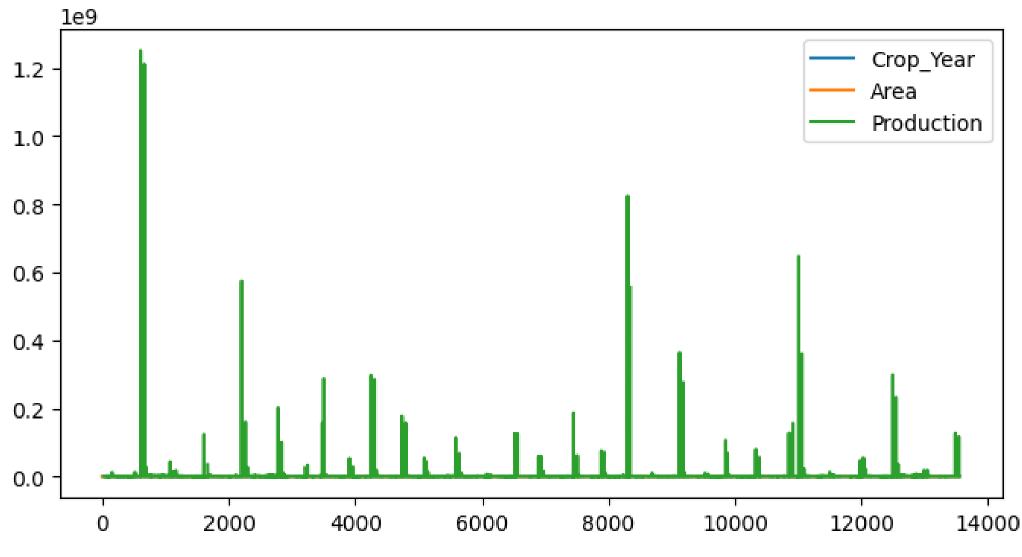

Figure 2. Crop production based on year and area wise

The initiation of the research involves with the collection of dataset pertaining to agriculture. It then continued with the execution of data preprocessing after importing the

Editors: S. Pandikumar, Manish Kumar Thakur
DOI:10.48001/978-81-980647-5-2-2 | ISBN: 978-81-980647-5-2

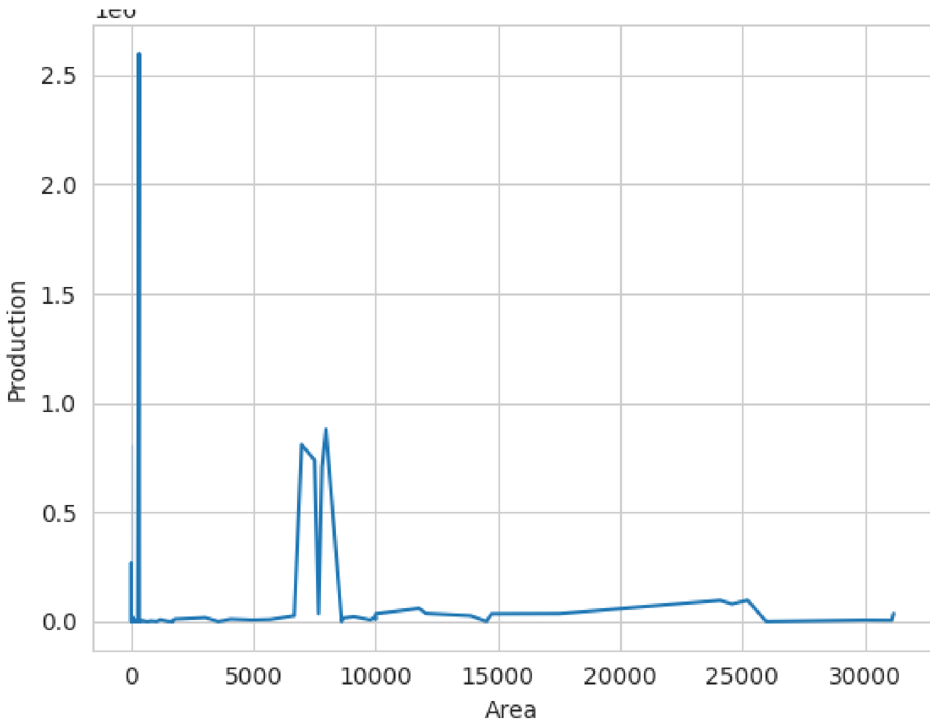

Figure 3. Crop Production Growth Based on the Area

required libraries and packages. At next step, Separated from the data are the trained and test sets. Ultimately, a model is developed by incorporating the necessary DL algorithms, which determine the optimal crop and yield to be cultivated on a specific plot of land.Based on the results shown in Table 2 shows the performance analysis of the proposed algorithm and its accuracy for crop yield prediction was measured. The proposed Deep Learning algorithms such as Deep Neural Network(DNN) produce an accuracy of 80.92%, Convolutional Neural Network(CNN) produced 90.20%, Recurrent Neural Network(RNN) produced 87.18, Long Short Term Memory(LSTM produced 96.5%. In figure 4 the accuracy were analyzed. Among the proposed four algorithms the LSTM give 96.5% of accuracy in agricultural data. It proven that the deep Learning algorithms provide more efficient way to predict the crop and its yield in a better manner.

DOI:10.48001/978-81-980647-5-2-2 | ISBN: 978-81-980647-5-2

Table 2. Performance Analysis

Model	Features	Crops	Accuracy
DNN	Area, crop, year, yield, rainfall, and pesticides	Wheat, Rice, Maize, Millets, Pea, Potatoes, Green Gram, Soybeans, Sugarcane.	80.92
CNN	Area, crop, year, yield, rainfall, and pesticides	Wheat, Rice, Maize, Millets, Pea, Potatoes, Green Gram, Soybeans, Sugarcane.	90.20
RNN	Area, crop, year, yield, rainfall, and pesticides	Wheat, Rice, Maize, Millets, Pea, Potatoes, Green Gram, Soybeans, Sugarcane.	87.18
LSTM	Area, crop, year, yield, rainfall, and pesticides	Wheat, Rice, Maize, Millets, Pea, Potatoes, Green Gram, Soybeans, Sugarcane.	96.5

DOI:10.48001/978-81-980647-5-2-2 | ISBN: 978-81-980647-5-2

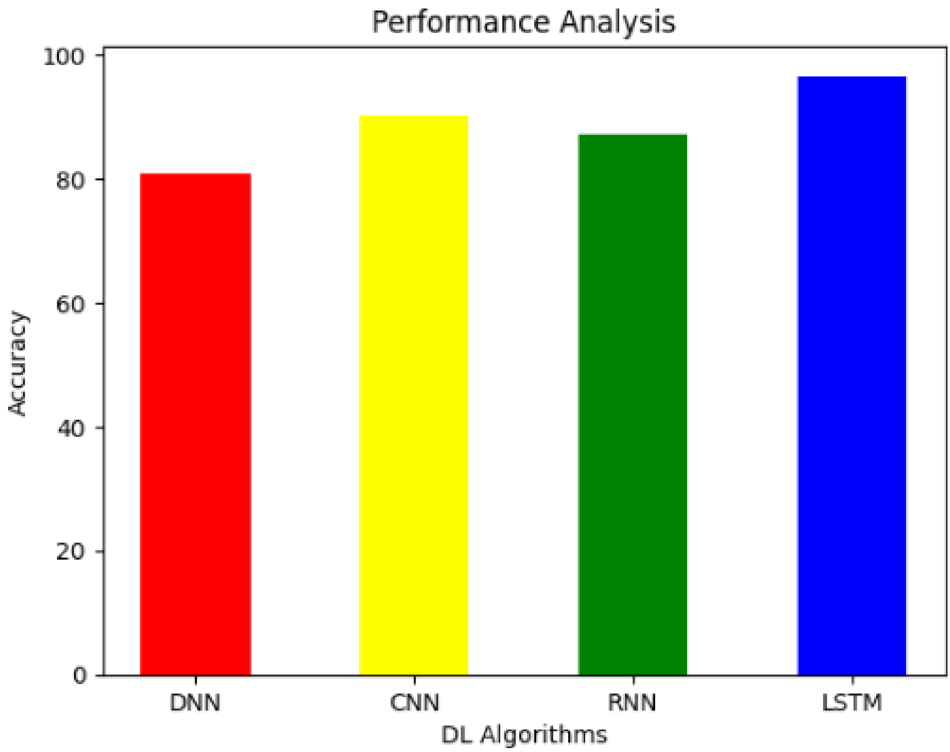

Figure 4. Performance Analysis

4 Conclusion

The proposed methodology utilizes various deep learning algorithms for predicting agricultural data. The agricultural data set that is used for this research contains various features like cultivation year, area, soil, crop yield, rainfall, and fertilizers used. This research takes these data's as input to the proposed model, and the model was analyzed using the data's and given the best result of predicting the yield. Deep learning techniques such as DNN, CNN, RNN, and LSTM were used for the prediction. Among these four techniques, the proposed methodology will suggest the most cost-effective and productive techniques to help the farmers cultivate the appropriate crop and get a better yield. Hence, the proposed research stated that there is an improvement in the accuracy of the LSTM model. The accuracy of LSTM is found to be 96.5% by applying the agricultural dataset. Comparing the other three algorithms, the LSTM gives better results for predicting agricultural data. It really helps the agricultural researchers to suggest ways to increase yield in cultivation.

Editors: S. Pandikumar, Manish Kumar Thakur
DOI:10.48001/978-81-980647-5-2-2 | ISBN: 978-81-980647-5-2

References

Agarwal, S., & Tarar, S. (2021). A hybrid approach for crop yield prediction using machine learning and deep learning algorithms. Journal of Physics: Conference Series. https://doi.org/10.1088/1742-6596/1714/1/012012

Akhter, R., & Sofi, S. A. (2021). Precision agriculture using iot data analytics and machine learning. Journal of King Saud University - Computer and Information Sciences, 34(5), 5603–5612. https://doi.org/10.1016/j.jksuci.2021.05.013

Bondre, D. A., & Mahagaonkar, S. (2019). Prediction of crop yield and fertilizer recommendation using machine learning algorithms. International Journal of Engineering Applied Sciences and Technology. https://doi.org/:10.33564/IJEAST.2019.v04i05.055

Hinton, G. (2018). Deep learning-a technology with the potential to transform health care. JAMA - Journal of the American Medical Association, 320(11), 1101–1102. https://doi.org/10.1001/jama.2018.11100

Kale, S. S., & Patil, P. S. (2019). A machine learning approach to predict crop yield and success rate. Pune Section International Conference (PuneCon). https://doi.org/10.1109/PuneCon46936.2019.9105741

Khan, R., Mishra, P., & Baranidharan, B. (2020). Crop yield prediction using gradient boosting regression. International Journal of Innovative Technology and Exploring Engineering (IJITEE), 9(3), January. https://doi.org/10.35940/ijitee.C8879.019320

Leong Wai Hong, B., & Tunku Abdul Rahman, U. (2016). Food Ordering System Using Mobile Phone.

Sharma, P. N., & Kirkman, B. L. (2015). Leveraging Leaders: A Literature Review and Future Lines of Inquiry for Empowering Leadership Research. Group and Organization Management, 40(2), 193–237. https://doi.org/10.1177/1059601115574906

Thomas van Klompenburg, A. K., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. Agronomy, 10(11). https://doi.org/10.1016/j.compag.2020.105709

Vignesh, K., Askarunisa, A., & Abirami, A. M. (2022). Optimized deep learning methods for crop yield prediction. Computer Systems Science Engineering. https://doi.org/10.32604/csse.2023.024475

Wang, S., Guidice, R. M., Tansky, J. W., & Wang, Z. M. (2010). When RD spending is not enough: The critical role of culture when you really want to innovate. Human Resource Management, 49(4), 767–792. https://doi.org/10.1002/hrm.20365

Editors: S. Pandikumar, Manish Kumar Thakur
DOI:10.48001/978-81-980647-5-2-2 | ISBN: 978-81-980647-5-2