CROP YIELD PREDICTION USING CNN AND BI-DIRECTIONAL LSTM

Mrs.Nithya.P,M.E.,
Assistant Professor
Department of ComputerScience and Engineering
KSR Institute for Engineering and Technology
Tiruchengode
pnithyacse@ksriet.ac.in

Dr.K.Tamizharasu M.E.,Ph.d
Professor
Department of ComputerScience and Engineering
KSR Institute for Engineering and Technology
Tiruchengode
tamizharasu2013@ksriet.ac.in

Ashwin M
Department of ComputerScience and Engineering
KSR Institute for Engineering and Technology
Tiruchengode
aakashashwin2002@gmail.com

Mohanaprasath M
Department of ComputerScience and Engineering
KSR institute for Engineering and Technology
Tiruchengode
prasathmp016@gmail.com

Prasanth S
Department of ComputerScience and Engineering
KSR Institute for Engineering and Technology
Tiruchengode
prasanth.soundappan@gmail.com

Manoranjith G
Department of ComputerScience and Engineering
KSR Institute for Engineering and Technology
Tiruchengode
manoranjith636@gmail.com

ABSTRACT-Considering the support of artificial intelligence, a model that is capable of extracting features and learn from datasets has come into existence as an achievable instrument for agricultural product yield prediction. Nevertheless, by extracting essential components of crop development, sophisticated technology for agriculture assists growers succeed optimum yields from their crops. A thorough examination of the scientific literature reveals current knowledge shortages in an individual area of computational biology strategies and focuses our investigation toward how agricultural indices and variables related to the environment have an impact the production of crops. Previous investigations from the year 2012 to 2022 concerning different databases have been collected and analyzed to achieve the primary goals of this research. The research investigation examines the positive effects of using artificial intelligence for forecasting the yields of crops, discovering the most appropriate technology for remote sensing according to gathering information needs, and the different characteristics.

1. INTRODUCTION

Because of the rising concern about food security, crop yield prediction is becoming more crucial [1-3]. By calculating the quantity of nutrition that will be available for the expanding global population, advanced crop yield prediction has an important function in reducing famine [2]. Increased crop yield output is a sustainable solution for one of the most serious issues that humanity faces: hunger. In the words of the World Health Organization [1], 820 million people worldwide still lack access to enough food. By 2030, the United Nations' Sustainable Development Goals will aim to end hunger, accomplish food security, and support sustainable agriculture. According to the Food and Agricultural Organization (FAO), there would be a 60% increase in demand for food to meet the 1.3 trillion global population by 2050 [2]. Consequently, crop yield forecasting can provide essential data needed for creating a practical plan to reach the goal and end hunger [1]. Many factors affect crop productivity, making it challenging to develop a solid forecast model using conventional techniques. However, with improvements in computer technology, it is now possible to create and train an unique approach for predicting crop yield [3]. Because to its many data innovations and powerful computers, deep learning is a significant approach that is widely employed in the agriculture field [4]. A kind of machine learning known as "deep learning" uses numerous layers of neural networks to learn from unstructured, unsupervised learning in supervised, semi-supervised, or unsupervised fashion. Deep learning techniques emphasise learning abstract aspects of vast datasets, particularly pointed out by Sarker[4]. Primary understanding of the relationship between functional qualities and interacting factors is necessary to accurately forecast crop production. Both extensive datasets and highly effective algorithms are needed to examine this association, which can Deep learning can be used to achieve this. In order to anticipate crop yields using data from satellites, this article conducts an organised literature review on the issue. The purpose of doing this systematic literature review is to help us analyse how environmental indices and environmental factors affect crop development and to discover any research gaps that may exist in a particular area of deep learning technologies. By examining the benefits of employing deep learning in agricultural yield prediction, the most appropriate remote sensing technology based on the data gathering needs, and the different features that affect crop yield prediction, this systematic literature review offers a fresh perspective on research. This article does have the same management structure.

2. LITERATURE REVIEW

In [1], a machine learning method is used to predict crop yield. Science and Engineering Research Technology International Journal. This research employs the Random Forest algorithm to forecast the agricultural yield from the available data. The models were constructed using actual information from Tamil Nadu, and they were tested using samples. Crop yield forecasts can be made with precision using the Random Forest algorithm.

In [2]. Journal Plops ONE. Because of its great accuracy and precision, simplicity of usage, and utility in data analysis, RF is a useful and adaptable artificial intelligence method for agricultural output forecasts at regional and global scales. The approach that is most successful is called Random Forest, and

In [3] Crop production Ensemble Machine Learning model for prediction. International Journal of Computer Science and Software Technology (IJCSSE). AdaNaive and AdaSVM are the suggested ensemble models in this study that will be used to forecast crop production over a given time frame. AdaSVM and AdaNaive were used in the implementation. SVM and Naive Bayes algorithm efficiency is improved with AdaBoost.

[4] describes a machine learning strategy for predicting crop yield based on climate factors. The paper presented at the ICCCI conference on computer communication and informatics. In the current study, a user-friendly web page called Crop Advisor was created as a software tool to anticipate the impact of meteorological conditions on crop yields.In a few Madhya Pradesh areas, the C4.5 method is utilised to determine the climatic characteristic that has the greatest impact on crop yields of particular crops. The decision tree is used to implement the paper.

In [5]. Volume 5, Issue 10, October 2016, of the International Journal of Advanced Research in Computer Science and Electronics Engineering (IJARCSEE). The interpretation of soil test findings and soil analysis are currently done on paper. This, in one way or another, has contributed to incorrect interpretation of soil test results, which has led to inaccurate recommendations of crops, soil amendments, and fertilizers to farmers, resulting in poor crop yields, deficiencies in soil micronutrients, and excessive or sparse fertilizer application. Formulas to Recommend Fertilizers and Match Crops with Soil.

In[6]. The essay that appeared in the International Journal of Engineering and Technology Research (IJRET). The main objective of this essay is to provide a user-friendly interface for farmers that offers statistics on the production of rice based on the available data. To increase crop output, several data mining techniques were used to forecast crop yields. a forecasting tool for environmental pollution, like the K-Meansalgorithm.

3. RESEARCH METHODS

Review Process: This comprehensive analysis of the literature helps in our understanding of how deep learning techniques are used to forecast crop yields using data from satellite imagery. This methodical literature study is conducted to identify the knowledge gaps in a specific area of deep learning approaches and to direct us as we examine how vegetation indices and environmental conditions affect crop development. All research papers from journals, conferences, and other online sources are not only examined for the systematic literature review, but they are also integrated and presented in accordance with the research questions indicated in our study. An excellent technique to assess a theory or evidence collection in a specific field, or to examine the veracity or validity of a particular statement, is to conduct an organized literature review theory [5]. For our comprehensive literature research, the review procedures provided by Kitchenham and Charters [6] are acceptable since they offer objectivity and transparency. The research questions are initially created based on the review criteria. According to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement [7], the review is conducted. The relevant research papers are selected using a variety of sources, including IEEE Explorer, Web of science, Scopus, Google Scholar, MDPI, and Web of Science. These providing information are evaluated and screened according to quality standards. The review is conducted using the PRISMA (prisma-statement.org) checklist in its entirety, and its findings are reported.

Research Questions: The following research questions are developed to guide the systematic review:

Q1. What methods using deep learning are used to forecast farming yields? This research aids in our discussion of both the advantages and limits of applying deep learning approaches to crop output prediction.

Q2. What approaches to remote sensing are utilized with deep learning methodologies to estimate crop yield?

With so many remote sensing technologies accessible this query assists to identify which is best suited based on the data necessary for the prediction of crop yields, such as surface area and crop type.

Q3.What are the growth metrics and external variables that are taken into account when predicting crop yield? The answer to this query teaches us more about the various characteristics that affect crop yield prediction using deep learning techniques.

Q4. What are the difficulties when forecasting agricultural yields using deep learning techniques? This query aids in our comprehension of the drawbacks and challenges inherent in the current methods.

Article Search Algorithm: A approach to looking for articles has been developed based on the stated objectives of the systematic literature review and the framed research questions. Making a successful research strategy calls for concentrating on the review's main idea rather than a broad notion. The term "deep learning" by itself will produce a large number of published articles from different application areas that are most likely unrelated to the review's objective and complicate the search process. The likelihood of straying outside the parameters of the evaluation can be decreased by redefining the search strategy as "crop yield prediction" AND "deep learning." The papers were initially retrieved from five databases, including IEEE Explorer, Science Direct, Scopus, Google Scholar, and MDPI, using these search terms. The following keywords were used to retrieve the papers from the databases in order to include any additional pertinent studies: "crop yield prediction" OR "crop yield estimation" AND "deep learning" AND "artificial intelligence" AND "smart farming". The research used articles from the past ten years (2012-2022), as deep learning approaches gained popularity after 2012 [8]. Since then, a great deal of study has been done on deep learning techniques.

Article Evaluation Standards: The retrieved articles are initially chosen based on criteria like the standard of the journal, the kind of deep learning techniques used, and any remote sensing technology applied to the research. Understanding the keywords and article selection are made easier by analysing the abstracts of papers. Based on the following factors, irrelevant articles were eliminated: • Publications that use machine learning techniques for crop yield prediction; • Publications without open access; •

Literature searches for articles published before 2012; • Articles in languages other than English. Articles that are related to agriculture but do not fall under crop yield prediction. 51 articles total are left after all exclusion factors have been applied.

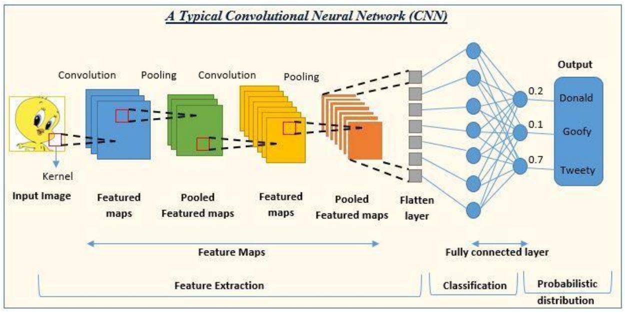
4. OVERVIEW OF THE EXISTING APPROACHES

Deep Learning: Due to their limited applicability and uncertainty, traditional approaches like the static logistic approach and the mechanistic approach make it difficult to develop a crop yield prediction model that is accurate [9,10]. For the prediction of crop output, many studies have used machine learning techniques like regression trees, random forests, multivariate regression, association rule mining, and artificial neural networks [9, 11]. The output, crop yield, gets treated by machine learning models as an implicit consequence of the input variables, such as weather factors and soil conditions, both of which may be very complicated [11]. Additionally, supervised learning techniques used in machine learning miss the nonlinear connection between the variables used as inputs as well as outcomes [12]. However, new technological developments have made it possible to create a sophisticated model for predicting crop yields using deep learning. Deep learning is a subset of machine learning that uses hierarchical structures to connect to the other layers. It differs from other conventional machine learning techniques in that it can analyse both unlabeled and unstructured data [13].

Since deep learning can analyse enormous datasets, discover relationships between different factors, and employ nonlinear functions, it is widely used in the agricultural sector. In an unsupervised setting, these methods can extract features for sizable datasets. Deep learning techniques outperform conventional machine learning methods in feature extraction [13].Deep learning has a powerful ability to extract features from the available data because an accurate crop yield prediction depends on the variables affecting crop growth. Deep neural networks have a number of nonlinear layers that at each level extract information from the untested incoming data. The capacity to extract features from the material at hand. The nonlinear layers in deep neural networks transform the unproven input data into a specially extracted shape at each layer [14]. Finding the irregular correlation between the input and response factors requires the application of deep neural networks with a variety of hidden layers [14]. However, they are challenging to train and require modern technology and methods for optimization [15]. Therefore, adding more hidden layers can be useful but comes with some limitations that can be overcome by using particular techniques.

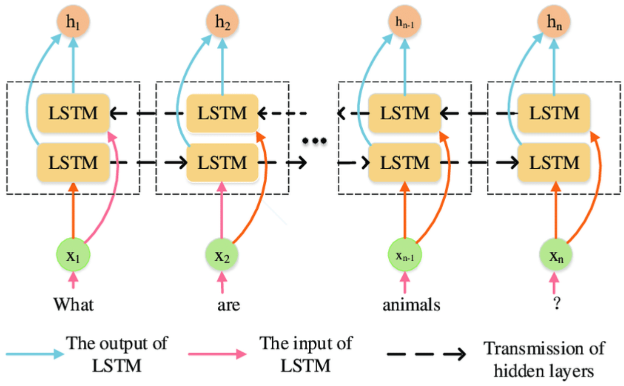
Deep neural networks (DNN): A DNN is a specific type of feedback neural network that has several entirely interconnected hidden layers. The placed layers typically use activation functions like Relu and loss functions like L2 normalization and mean squared error.

Convolution Neural Network (CNN): A CNN includes layers like convolution layers, pooling layers, and fully connected layers in comparison to conventional neural network approaches, which aids in quickly identifying salient features within the data. Feature extraction is carried out by the activation function and convolution algorithm that make up the convolution layer [19]. A filter and feature map are involved in the convolution process. A feature map is the output of a filter, which is a set of weights given to the input.

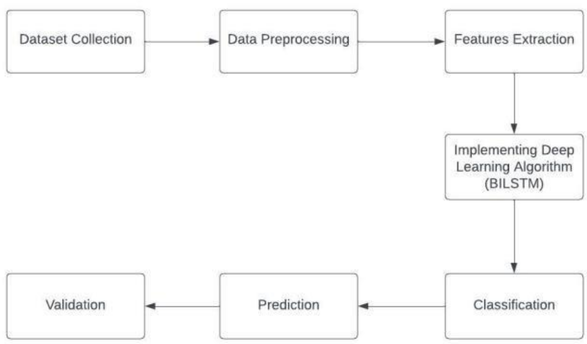


Short-Term Long-Term Memory (LSTM): The LSTM is a specific kind of recurrent neural network (RNN) that, when used with the right gradient-based algorithm, may gather time- dependent information. An input layer, one or more LSTM layers, and the output layer are the first three levels in the chain structure which make up the LSTM. The input gate, forget gate, and output gate are the three gates which the LSTM utilizes to control cell state and output. These gates are more apt to function as layers in artificial neural networks that can regulate the transfer of information [14]. In

LSTM layers, each cell has three gates: the input gate selects the information to be kept, the forget gate selects the amount of prior information to be forgotten, and the reserve gate chooses the amount of current input that is kept back a bit.



4.2. SEARCH STRATEGY



Estimating crop yield is a crucial job in agriculture that aids farmers in making choices about their crops.

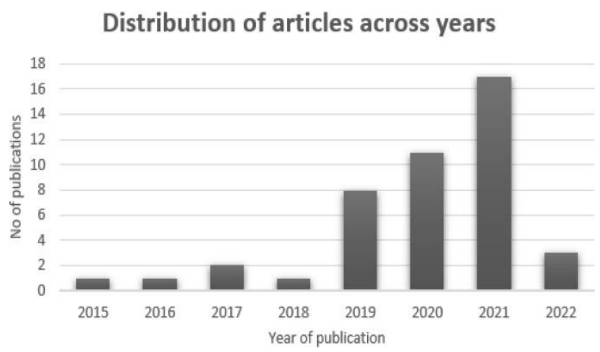
1. DATA COLLECTION: Gather past crop information as well as environmental data from various sources, such as temperature, rainfall.
2. DATA PREPROCESSING: Sort through the information and remove any unusual patterns or missing data. Normalize the data to make sure that each feature is the same size.
3. FEATURE EXTRACTION: Irrigation Both the quantity and timing of irrigation have a significant impact on crop development. This data can be gathered using irrigation logs and soil moisture sensors.
4. IMPLEMENTING DEEP LEARNING BILSTM: A deep learning algorithm trains neural networks to learn from data and make predictions by autonomously extracting high-level features from the data.

5. PROJECTION: To predict the quantity of crops that will be harvested in a specific season or location, crop yield prediction uses statistical and machine learning techniques.

6. VALIDATION: In the field of agriculture, validation entails assessing the precision and dependability of crop yield prediction models using different datasets or testing techniques.

5. RESULTS AND DISCUSSION

The reviewed documents are evaluated, summarized, and subjected to analysis. Demonstrates the amount of articles that were published among 2012 and 2022. The papers that we examined for the years 2012 to 2014 did not meet the condition of using deep learning to predict crop yield. It is obvious that crop production prediction research is essential.

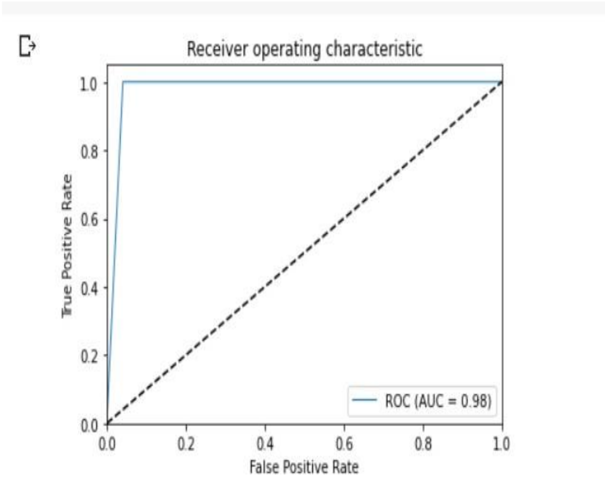



```
[ ] score, acc = model.evaluate(test_X, test_Y, verbose=0)
    print('Test accuracy:', acc)

    Test accuracy: 0.8973334635257721
```

```
from sklearn.metrics import confusion_matrix
cm = confusion_matrix(test_Y.argmax(axis=1), y_pred.argmax(axis=1))
print(cm)

[[4052]]
```



	precision	recall	f1-score	support
1	1.00	0.37	0.54	4052
2	0.00	0.00	0.00	0
4	0.00	0.00	0.00	0
accuracy			0.37	4052
macro avg	0.33	0.12	0.18	4052
weighted avg	1.00	0.37	0.54	4052

6. CONCLUSIONS

The study structured review of the literature on the use of algorithms based on deep learning for crop yield prediction is given in this article. The goal of this comprehensive review of the literature was to determine the research gaps that still needed to be filled in a specific area of deep learning methodologies and to provide useful information on how vegetation indices and environmental variables affect crop yield prediction. This review has given a variety of deep learning features, approaches, and factors utilized for predicting crop yield. All of the studies focused on different crops, geological locations, and features. In general, when compared to conventional machine learning methods, the success rate and precision associated with the deep learning strategy for crop yield prediction are better. All deep learning techniques are equally successful at predicting crop yield. Depending on the model's factors and parameters. However, CNN- and LSTM-based deep learning techniques are the most effective in predicting crop output. CNN has the ability to identify significant features that may affect the prediction of the yield of crops. Furthermore, LSTM not only recognizes the data fluctuation pattern but also the dependent connection that exists between the time-series data.

7. FUTURE SCOPE

A key component of contemporary agriculture is crop yield prediction, which entails estimating the volume of crop that will be grown in a specific field or area. At the moment, agricultural yield predictions are made using a variety of data sources, such as environmental information, historical climate information, and crop growth models. However, there are a lot of other sources of data that might be included, including real-time sensor data from fields, drone data, and satellite imagery. It might be feasible to increase the precision of crop yield predictions by adding more data. Predictions based on patterns and trends discovered from previous data can be made using machine learning algorithms, possibly producing predictions which are more correct. Additionally, precision agriculture practices, such as optimizing methods of farming with data and technology, can help increase crop yields and produce more precise forecasts. In order to optimize profits and avoid waste, planting strategies might gain from the integration of market information. We can anticipate seeing more precise and effective crop yield forecasts in the future thanks to ongoing advances in technology and the combination of different sources of information, working to create a more sustainable and effective agriculture sector.

8. REFERENCES

1. Quanyin Zhu and J.iajunZong, "Apply Grey Prediction in the Agriculture Production Price", Fourth International Conference on Multimedia Information Networking and Security, 2012.

2. Michael Gurstein, "A Decision Support System to Assist the Rural Poor in Bangladesh", IEEE Technology and Society Magazines, September 2013.

3. "Prediction of Crop Yield Using Data Mining Approach" in Computational Intelligence and Communication Networks (CICN), International Conference, 12–14 December 2015.

4. Tng Zhang, "Solving large-scale linear prediction algorithm", Proceedings of the 21st International Conference on Machine Learning

5. D. Ramesh, B. Vishnu Vardhan, “Analysis Of Crop Yield Prediction Using Data Mining Techniques”, 2015 [7] A.A. Raorane, R. V. Kulkarni, “Data Mining: An effective tool for yield estimation in the agricultural sector”< IJETTCS, vol.1, no. 2, pp. 75-79, 2012.

6. E. I. Papageorgiou, A. T. Markinos, and T. A. Gemtos, "Fuzzy cognitive map based approach for predicting yield on cotton crop production as a basis for decision support system in predtiction agriculture application", Elsevier, vol. 0, no. 11, pp. 3643-3657, 2011.

7. James V. Zidek and Luke Bornn, "Efficient stabilisation of crop yield prediction in the Canadian Prairies", Elsevier, vol. 0, no. 0, 2012, pp. 223-232.

8. R.Sujatha and P. Isakki, "Crop Yield estimation using classification techniques", Computational Intelligence and Communication Network (CICN), International Conference 7-9 Jan. 2016.Rakesh Kumar, M.P. Singh, Prabhat Kumar And J.P. Singh, “Crop Selection Method to Maximize Crop Yield Rate Using Machine Learning Technique” International Conference 6-8 May. 2015.

9. [8] S. Djodiltachoumy and E. Manjula published "A Model for Prediction of Crop Yield" in the International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4, March 2017.

10. "Agriculture Yield Prediction Using Prediction Analytic Techniques" International Conference on Contemporary Computing and Informatics 2016, S. Nagini, Dr. T. V. Rajini Kanth, and B.V. Kiranmayee.

11. Yung-Hsingpeng, Chin-Shun Hsu, and Po-Chuang Huang, "Developing Crop Price Forecasting Service Using Open Data from Taiwan Markets" Nov. 20-22, 2015.

12. S.Veenadhari, Bharat Misra, and CD Singh, "Machine Learning approach for forecasting crop yield based on climatic parameters" 2014 International Computer Conference Fig.7. Time Accuracy Complexity VOLUME 8, ISSUE 1, JANUARY 2020, INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH