RESEARCH ARTICLE

DOI: 10.34256/ijcci2311

International Journal of Computer Communication and Informatics

Survival and Comparative study on Different Artificial Intelligence Techniques for Crop Yield Prediction

S. Saritha [a,*], G. Abel Thangaraja [a]

[a] Department of Computer Science, Shri Nehru MahaVidyalaya College of Arts and Science, Coimbatore-641050, Tamil Nadu, India.

[*] Corresponding Author: sarithajayabrabu@gmail.com

Received: 04-02-2023, Revised: 18-04-2023, Accepted: 02-05-2023, Published: 30-05-2023

Abstract: Agriculture is an essential, important sector in the wide-reaching context. Farming helps to satisfy the basic need of food for every living being. Agriculture is considered the broadest economic sector. The crop yield is a significant part of food security and improves the drastic manner by human population. The quality and quantity of the yield touch the high rate of production. Farmers require timely advice to predict crop productivity. The strategic analysis also helps to increase crop production to meet the growing food demand. The forecasting of crop yield is a process of forecasting crop yield by using historical data. Machine learning provides a revolution in the agricultural field by changing the income scenario and growing an optimum crop. Many researchers carried out their research to deal with forecasting crop yield. In this way, accurate prediction of crop yield was improved. But, failed to reduce the crop yield prediction time and the accuracy level was not enhanced by existing methods.

Keywords: Agriculture, Economic Sector, Strategic Analysis, Crop Yield Prediction, Machine Learning

1. Introduction

Important producers of crops are using the food cycle. The aim of farmers is to achieve better crop yield at minimum investment. In order to, owing to several reasons the majority of farmers are not getting the expected crop yield. Climate change, rainfall, soil health, water scarcity, and crop disease are various conditions that influence cultivation. Tradition methods are not enough to handle the huge demands of satisfying shifting consumer preferences. The evolution of machine learning and deep learning in agricultural sectors helps farmers to increase efficiency and accuracy in the decision-making of crop yield. To address the range of crop production, management, and precision agricultural challenges are used for accurate crop yield prediction.

The Aim of this paper mainly includes two parts: First, an inclusive evaluation of significant kinds of literature based on crop yield prediction was improve the conceptual framework and explores the propositions; next, refers to the limitation of the existing literature, and the way of future work is proposed.

The major objective of the research work is contributed as follows:

(1) The quantitative analysis of crop yield prediction was used for the proposed method. Compared with previous studies, the recommended work reflects the maximum accuracy in the prediction of crop yield.

(2) We list six sections, which help readers understand and view the literature from different dimensions by using literature classification.

(3) The designed graphical illustration is used to obtain the various prediction models.

(4) Keywords of the current research.

The rest of the paper is outlined as follows: Section 1, the writing background of the article is the preamble. Section 2 is a review of the research method and downsides of existing crop yield prediction methods. Section 3 describes the study and analysis of conventional crop yield prediction methods, including the development trend. Section 4, identifies the possible comparison between different crop prediction algorithms. In Section 5, discusses the limitations of existing forecasting methods and provides recommendations for future research. Section 6, the conclusion of the paper.

2. Literature Review

In [1], CubeSat imagery with the APSIM crop model was introduced to merge the daily high-resolution to predict the yield. The designed approach used APSIM to instruct linear regression that combined the yield within the simulated leaf area index (LAI). A significant effort has focused on acquiring spatially distributed information on crop status and yield using satellite-based remote sensing platforms. Though, prediction accuracy was not enhanced by the yield prediction approach. The DLMLP neural network was introduced [2] with noteworthy success to address the forecast-related issues of crop yield. With robust machine learning, crop yield prediction was carried out in an accurate manner on soil health parameters. To develop the estimated wheatcrop yield by using estimated soil health parameters like SAR backscatter and optical remote sensing satellite data parameters. But prediction time was not reduced by DLMLP neural network.

The fuzzy Enumeration Crop Prediction Algorithm (FECPA) was introduced in [3] to enhance crop yield prediction accuracy. The fuzzy strategy minimized the plan to emphasize self-organizing contours. However, the error rate was not reduced by FECPA. A mathematical model was introduced in [4] to predict the pest incidence level in banana crops. An alternative solution was obtained with 4.0 Industry technologies and a system of IoT sensor networks in banana plantations. This model helped the producers to program dates more effectually in order to control and improve the management of pests. However, prediction accuracy was improved; a mathematical model did not reduce computational complexity.

In [5], a performer-based deep learning framework was proposed for crop yield prediction. The designed framework was employed with neural network design for predicting barley yields. This model was specifically made for genomic selection datasets. However, it failed to minimize the prediction time by the farmer-based deep learning framework. An accurate crop yield prediction model was introduced [6] to predict genotype response with parameters of weekly weather. A temporal attention mechanism was designed with the LSTM model for interpretability. In this proposed model are geospatial data without field-scale farming management data and a variety of information is indiscernible. Heat stress during the flowering stage was not considered in this research even though less impact on seed yielding. Whereas, the computational complexity was not reduced.

The long trend was discussed in [7] for reproduction evolution and yield aging effects, yield-related characteristics and disease resistance features for cereal crops with a spectrum of genotypes. The crop performance and disease resistance of varieties are obtained with aging effects, minimum yield, and maximal disease susceptibility. However, the computational cost was not reduced. A multi-sensor method was designed in [8] for drought-induced agricultural impact prediction. Lasso regression with satellite data was determined for relative yield anomalies. The designed method was robust to extreme drought events. The computational complexity was not minimized by the multi-sensor method.

In [9], a machine learning method was performed to forecast yields cultivated all over India. The machine learning method employed a trained model to find the patterns among data for crop prediction.

However, the prediction time was not reduced by the machine learning method. The CNN-based approach was designed [10] for IoT technology for gathering information. A decision Tree was employed to forecast the crop condition with a suitable solution. Whereas, the error rate was not minimized by CNN based approach.

In DSS LANDS were performed [11] to predict potato disease in Sardinia. The disease severity was predicted through Feed-forward Neural Network and SVM Classification depending on meteorological parameters. However, the prediction accuracy was not reduced. A multi-layered perceptron model was introduced in [12-16] for disease classification to detect plant diseases and increase production. But the error rate was not reduced by the multi-layered perceptron model.

3. Crop Yield Prediction

Predicting crop yield is an important role in decision-making at global, regional, and field levels. In order to address the essential rising challenges in food security, particularly in an era of global climate change. To extract the main crop features for prediction Decision support models are broadly used. Accurate yield predictions not only help farmers make informed economic and management decisions but also support famine prevention efforts. Machine learning is an important decision support tool for crop yield prediction, including supporting decisions on what crops to grow and what to do during the growing season of the crops. Different machine learning algorithms are applied to support crop yield prediction research.

3.1 Early season prediction of within-field crop yield variability by assimilating CubeSat data into a crop model

A method of crop prediction to blend the APSIM crop model with daily high-resolution CubeSat data. The temporal growth of main crop processes over the course of the growing season, which eliminates regional heterogeneity in yield forecast, has been described using the APSIM. The APSIM is used in this method that trains a linear regression to correlate the simulated yield and simulated LAI. The ideal regression date for the LAI to offer the highest yield prediction is then found using the relationship. Weeks before the ideal regression date, CubeSat-based LAI into APSIM will generate end-of-season high-resolution (3 m) crop maps. An appropriately chosen set of pictures and in-situ data were available for evaluation, and the planned approach was shown on a rainfed maize field. To adapt the regression model in a single assimilation phase, the approach did not require field data. Before the ideal regression date, it was probably possible to generate yield predictions with high accuracy. With a significant connection to measurements that were separately obtained, the yield seasonal variation was replicated. Early-season spatially explicit crop productivity has a major potential to improve digital agriculture goals and developed end-of-season output predictions.

3.2 DLMLP and remote sensing approach for soil health-based crop yield estimation

A DLMLP model was determined for crop yield estimation. Based on the accuracy level; the designed model estimated the soil health parameters like soil moisture, EC, and SOC. The appraised soil health parameters are discussed with satellite data matrics for wheat crops and tracking soil health and their yield. DLMLP model was employed for the estimation of soil moisture, soil salinity, and SOC. The crop yield estimation was carried out using the per-pixel soil health metrics values and satellite data parameters. DLMLP model was employed with hidden, dropout, and activation layers for the evaluation of soil health parameters. The DLMLP model with soil health-based parameters obtains the accurate yield estimation and validates even the in absence of soil health parameter values. The designed model determines wheat crop growth for the early stages with the help of soil health parameters. The exposed soil parameters are used for crop yield prediction.

3.3 Superior fuzzy enumeration crop prediction algorithm for big data agriculture applications

The big data cloud function is a main data storage and data analysis environment. Big data weather and crop forecasting is the new research area in the agricultural structure and large data evaluation. FECPA was proposed to enhance crop yield prediction accuracy by using agricultural big data and developing effective programs for data classification. Preprocessing was employed to examine the collected data and identified the absent values for attributes. Non-existent values were filled with corresponding values by checking previous and past outcomes. The conventional algorithm gives minimal feature set value compare to Naive Bayes, CNN. The minimized dimension information was determined to predict climate as a reasonable outcome.

3.4 Thrips incidence prediction in organic banana crop with machine learning

A mathematical model was performed to forecast the pest incidence level in banana crops through SVM. The designed model facilitated decision-making for farmers in their region of Piura, Peru. The proposed model included data collection with a data recording system, binary classification description for pest incidence, and experiment management, input variable selection, SVM prediction design. The existence of pests like red spot thrips, Black Sigatoka was the issues faced by the producers and the final product quality. With the help of industry technology, 4.0 and the installation of IoT sensor networks in banana plantations provides an alternate solution. The designed model determined the classification of pest incidence levels through Machine learning methods through atmospheric variables measured with an IoT sensor network. A support Vector Machine was employed to improve the accuracy level. The producer's better management of pest control by strategically scheduling dates for spraying through the implementation of the designed model. As well as, to provide the optimum product quality at minimum cost.

3.5 Multimodal performers for genomic selection and crop yield prediction

For the purpose of predicting crop yield forecasting by using single nucleotide polymorphisms and weather data, a performer-based deep learning architecture has been developed. A Multimodal Performer network's self-attention maps indicated that the model makes significant associations between genotype and weather data. To update breeding decisions and shorten breeding cycles, the breeder used the proposed scheme. The performer-based paradigm was used for genomic selection in animal husbandry contexts, such as salmon breeding for improved Omega-3 fatty acid synthesis. The neural networks process is genotype sequence and weather data by using a separate network.

The genotype sequence was produced by CNN architecture and the weather data was processed through MLP. To study particular filters in each layer of its pipeline in CNN. By partitioning data into two pipelines, a modality-specific network was employed to attain optimal performance. Performer-based deep learning framework minimized the RMSE.

3.6 Crop yield prediction integrating genotype and weather variables using deep learning

An accurate crop yield prediction model was performed to improve farming reproduction for monitoring across different climatic conditions. The forecast was made in order to prepare for the impact of climate threats on crop production. In order to build the LSTM and RNN, the performance data from the North American Uniform Soybean Tests (UST) was used. The genotype response in various environments was predicted using pedigree-relatedness metrics and weekly meteorological parameters. The designed prediction model included least absolute shrinkage and selection operator (LASSO) regression, SVR-RBF, and data-driven USDA model for yield prediction. For LSTM models, a temporal attention mechanism was used to provide the interpretability of time frames during the growing season. Plant breeders benefited greatly from the insights that the designed models 'outputs supplied. The proposed LSTM for soybean crop development against weather makes predictions relatively accurate.

4 Experimental Settings

The different crop yield prediction methods were discussed with several numbers of data points to conduct the experiments. Different crop yield prediction parameters are analyzed for enhancing crop yield prediction performance. In conducting, the experiments, the rice data points from SMART FASAL Dataset are taken asinput. The dataset URL is given as http://smartfasal.in/ftp-dataset-portal/. The dataset has 13 attributes and 42666 instances. The attributes are soil moisture 1, soil moisture 2, soil moisture 3, soil pressure, soil temperature, soil humidity, etc., The experimental analysis is carried out with six methods, namely yield prediction approach, DLMLP model, FECPA, mathematical model, performer-based deep learning framework, and accurate crop yield prediction model. The quantitative analysis of crop yield prediction is compared with different metrics like,

- Prediction accuracy
- Prediction time and
- Error rate

4.1 Analysis on Prediction Accuracy

Prediction accuracy is ratio of number of data points to correctly predict to total number of data points. Prediction accuracy is calculated as

$$Prediction\ Accuracy = \frac{Number\ of\ Data\ Points\ correctly\ predicted}{Total\ number\ of\ data\ points} * 100 \quad (1)$$

From (1), the prediction accuracy is calculated. Prediction accuracy is measured in percentage (%). When the prediction accuracy is higher, the method is said to be more efficient.

Table 1. Tabulation of Prediction Accuracy

Number of Data points	Prediction Accuracy (%)					
	Yield prediction approach	DLMLP model	FECPA	Mathematical model	Performer based deep learning framework	Accurate crop yield prediction model
100	75	81	64	70	68	60
200	77	83	66	72	69	62
300	79	85	68	75	72	64
400	76	82	65	73	70	63
500	78	84	67	76	71	65
600	80	87	69	78	74	68
700	82	89	72	79	76	70
800	84	91	75	81	78	72
900	86	93	77	83	80	74
1000	88	95	80	85	82	75

Table 1, describes the prediction accuracy with respect to a number of data points ranging from 100 to 1000. Prediction accuracy comparison takes place on the existing Fuzzy rule-based method, new wildfire detection solution, FFRZ Map, Parallel SVM method, spatial framework, and New End-to-End framework. Let us consider the number of forest fire data as 300, the prediction accuracy of the yield prediction approach, DLMLP model, FECPA, mathematical model, performer-based deep learning framework, and accurate crop yield prediction model is 79%, 85%, 68%, 75%,72%, and 64% respectively. Figure 1 is a graphical representation of prediction accuracy.

In figure 1, prediction accuracy for the different numbers of data points. From the figure, the forecasting accuracy using the DLMLP model is higher when compared to the yield prediction approach, Fuzzy Enumeration Crop Prediction Algorithm (FECPA), mathematical model, performer-based deep learning framework, and accurate crop yield prediction model. This is because of applying the Rectified Linear Unit (RELU) activation function for soil health parameter estimation. The crop yield estimation was due to the larger data size in the crop yield estimation stage. Consequently, the prediction accuracy of the proposed DLMLP model is improved by 8%, 24%, 13%, 18%, and 29% when compared to the existing methods respectively.

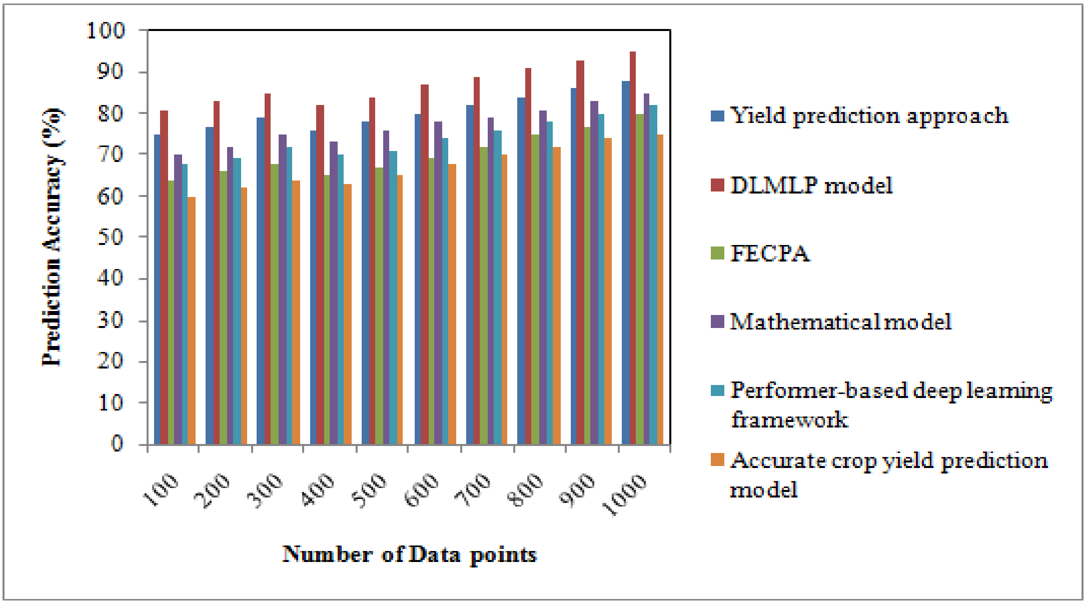

Figure 1. Measurement of Prediction Accuracy

4.2 Error Rate

Error Rate is the ratio of the number of data points that are incorrectly predicted to the total number of data points. It is calculated as,

$$Error\ Rate = \frac{number\ of\ data\ points\ incorrectly\ predicted}{Total\ Number\ of\ Data\ Points} * 100 \quad (2)$$

From (2) is measured in terms of percentage (%). When the error rate is lesser, the method is said to be more efficient.

Table 2, exemplifies the error rate with respect to the number of data points ranging from 100 to 1000. Error rate comparison takes place on the existing Fuzzy rule-based method, new wildfire detection solution, FFRZ Map, Parallel SVM method, Spatial framework, and New End-to-End framework. Let us consider the number of forest fire data as 500, the error rate of yield prediction approach, DLMLP model, FECPA, mathematical model, performer-based deep learning framework, and accurate crop yield forecasting model is 33%, 26%, 32%, 37%, 21%, and 22% respectively.

Figure 2, describes the error rate for the different numbers of data points. From the figure, the error rate using a performer-based deep learning framework is lesser when compared to the yield prediction approach, DLMLP model, FECPA, mathematical model, and accurate crop yield prediction model respectively.

Table 2. Tabulation of Error Rate

Number of Data points	Error Rate (%)					
	Yield prediction approach	DLMLP model	FECPA	Mathematical model	Performer based deep learning framework	Accurate crop yield prediction model
100	28	22	27	31	15	19
200	30	24	29	33	17	21
300	32	27	31	36	20	23
400	35	29	34	39	22	25
500	33	26	32	37	21	22
600	31	24	30	35	19	20
700	29	22	27	33	16	18
800	32	25	29	36	18	21
900	34	27	31	39	20	23
1000	36	29	33	42	23	25

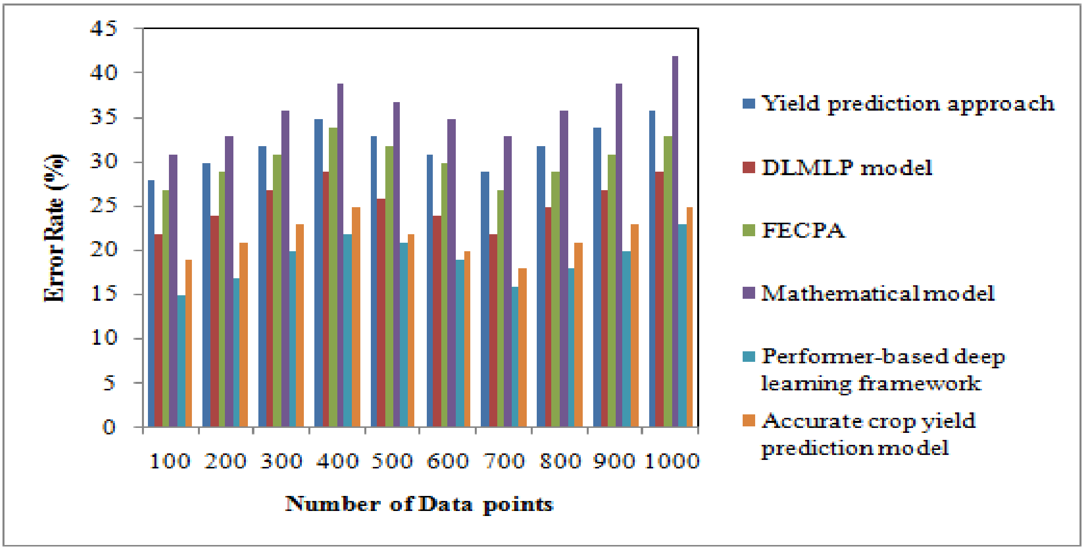

Figure 2. Measurement of Error Rate

4.3 Performance on Prediction Time

Prediction time is product of the total number of data points and time consumed to predict one data point. It is formulated as,

$$Prediction\ Time = Number\ of\ data\ points * Time\ consumed\ to\ predict\ one\ data \tag{3}$$

Table 3. Tabulation of Prediction Time

Number of Data points	Prediction Time (ms)					
	Yield prediction approach	DLMLP model	FECPA	Mathematical model	Performer based deep learning framework	Accurate crop yield prediction model
100	12	16	19	24	26	30
200	15	18	21	26	28	32
300	17	20	23	29	31	35
400	19	22	25	32	33	37
500	21	26	28	34	35	39
600	23	29	30	36	37	40
700	25	32	33	38	39	42
800	27	34	36	40	43	45
900	30	37	39	42	46	48
1000	32	40	42	44	49	51

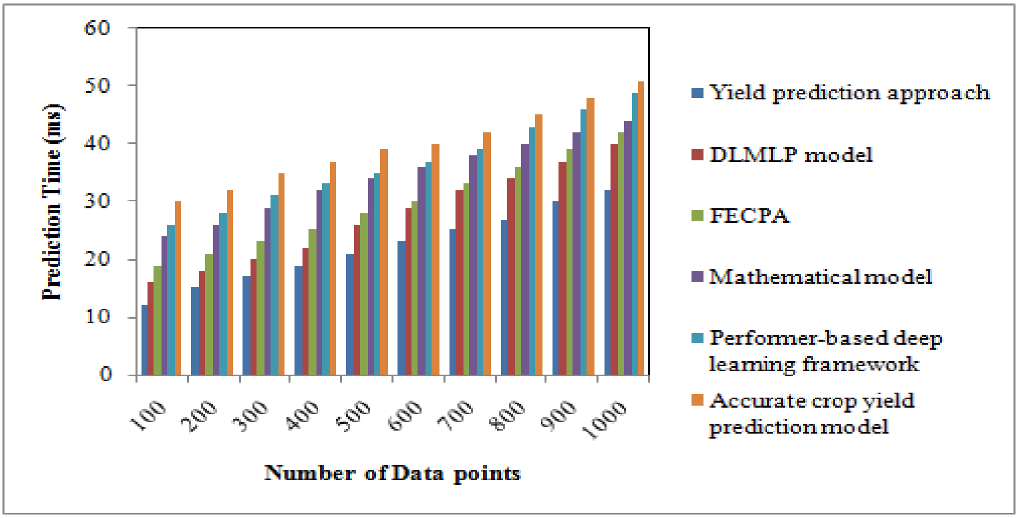

Figure 3. Measurement of Prediction Time (ms)

From (3), the prediction time is determined. Prediction time is measured in milliseconds (ms). When the prediction time is lesser, the method is said to be more efficient.

Table 3, describes the prediction time with respect to a number of data points ranging from 100 to 1000. Error rate comparison takes place on the existing Fuzzy rule-based method, new wildfire detection solution, FFRZ Map, Parallel SVM method, Spatial framework, and New End-to-End framework. Let us consider the number of forest fire data as 800, the prediction time of yield prediction approach, DLMLP model, FECPA, mathematical model, performer-based deep learning framework, and accurate crop yield prediction model is 27ms, 34ms, 36ms, 40ms, 43ms, and 45ms respectively.

Figure 3, discusses the prediction time for different numbers of data points. In above the figure, prediction time using the yield prediction approach is lesser when compared to the DLMLP model, FECPA, mathematical model, performer-based deep learning framework, and accurate crop yield prediction model. This is because of using a particle filter that combined CubeSat-based LAI into APSIM to provide an end-of-season high-resolution yield map weeks before the optimal regression date. As a result, the prediction time of the yield prediction approach is minimized by 19%, 26%, 37%, 41%, and 46% when compared to the DLMLP model, FECPA, mathematical model, performer-based deep learning framework, and accurate crop yield prediction model respectively.

5. Discussion and Limitation on Crop Yield Prediction Methods

To yield prediction approach was used to train linear regression that transmitted the simulated yield to simulated LAI. The relationship was employed to identify the optimal regression date where LAI presented the best yield. The procedures not required in-field data to regulate the regression model with a single assimilation step. But prediction accuracy was not improved by the yield prediction approach. DLMLP neural network was performed for crop yield prediction based on soil health parameters and tracking soil health. For anticipating the early growth stages of wheat crops, the developed network evaluated soil health characteristics. But, the prediction time was not reduced by DLMLP neural network.

FECPA was performed for crop yield prediction. FECPA improved the crop yield prediction accuracy from agricultural big data for data classification. The fuzzy strategy minimized the highlights of self-organizing contours. The reduced dimension information was employed to predict climate as a reasonable outcome. However, FECA did not lower the mistake rate. To categorize the amount of pest prevalence based on machine learning methods, a mathematical model was proposed. The IoT sensor network-measured atmospheric variables were employed in the proposed methodology. Though the prediction accuracy improved; a mathematical model did not reduce computational complexity.

For the purpose of predicting crop yield forecasting using single nucleotide polymorphisms and weather information, a performer-based deep learning architecture has been developed. To help with breeding decisions and speed up the breeding cycle, the Multimodal Performer network made significant connections between genotype and meteorological data.

Performer-based deep learning framework reduced the RMSE. But prediction time was not reduced by the performer-based deep learning framework. Accurate crop yield prediction model leveraged pedigree relatedness measures to predict the genotype response in multiple environments The interpretability of the time frames during the growing season was made possible by a temporal attention mechanism. However; the computational complexity was not reduced.

6. Future work

In future work, the survival study is to perform efficient crop yield prediction with minimal time consumption and higher accuracy by machine learning and deep learning techniques.

7. Conclusion

A comparison of different crop yield forecasting methods is studied. In this study, it is clear that the prediction time was not reduced by the performer-based deep learning framework. In addition, the computational complexity was not reduced. While the prediction accuracy improved, it failed to minimize computational complexity. In addition, the prediction time was not reduced by DLMLP neural network. The wide range of experiments on existing methods obtains the performance of crop yield prediction with its limitations. To conclude the result of this research work can be performed by using machine learning techniques for developing the performance of crop yield prediction with better accuracy and minimum time consumption.

References

[1] N. Arvanitopoulos, R. Achanta and S. Süsstrunk, Single Image Reflection Suppression, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, (2017) 1752-1760. https://doi.org/10.1109/CVPR.2017.190

[2] N. Arvanitopoulos, R. Achanta and S. Süsstrunk, (2017) Single Image Reflection Suppression, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, 1752-1760. https://doi.org/10.1109/CVPR.2017.190

[3] Q. Fan, J. Yang, G. Hua, B. Chen, D. Wipf, A generic deep architecture for single image reflection removal and image smoothing. In Proceedings of the IEEE International Conference on Computer Vision, (2017) 3238-3247.

[4] K. Gai, Z. Shi, C. Zhang, Blind separation of superimposed moving images using image statistics, IEEE transactions on pattern analysis and machine intelligence, 34(1) (2011) 19-32. https://doi.org/10.1109/TPAMI.2011.87

[5] Q. Wen, Y. Tan, J. Qin, W. Liu, G. Han, S. He, Single image reflection removal beyond linearity. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, (2019) 3771-3779. https://doi.org/10.1109/CVPR.2019.00389

[6] X.Guo, X. Cao, Y. Ma, Robust separation of reflection from multiple images. In Proceedings of the IEEE conference on computer vision and pattern recognition, (2014) 2187-2194.

[7] K. He, S. Soundarajan, X. Cao, J. Hopcroft, M. Huang, (2015) Revealing multiple layers of hidden community structure in networks, arXiv preprint arXiv:1501.05700.

[8] X. Shi, Z. Chen, H. Wang, D.Y. Yeung, W.K. Wong, W.C. Woo, Convolutional LSTM network: A machine learning approach for precipitation nowcasting, Advances in neural information processing systems, (2015) 802-810.

[9] J. Yang, D. Gong, L. Liu, Q. Shi, Seeing deeply and bidirectionally: A deep learning approach for single image reflection removal. In Proceedings of the european conference on computer vision (ECCV), (2018) 654-669. https://doi.org/10.1007/978-3-030-01219-9_40

[10] D. Lee, M.H. Yang, & S. Oh, (2018) Generative single image reflection separation. arXiv preprint arXiv:1801.04102.

[11] A. Levin, A. Zomet, Y. Weiss, Learning to perceive transparency from the statistics of natural scenes. Advances in Neural Information Processing Systems, (2002) 1271-1278.

[12] M. Jin, S. Susstrunk, P. Favaro, Learning to see through reflections. In 2018 IEEE International Conference on Computational Photography (ICCP), (2018) 1-12.

[13] X. Zhang, R. Ng, Q. Chen, Single image reflection separation with perceptual losses. In Proceedings of the IEEE conference on computer vision and pattern recognition, (2018) 4786-4794. https://doi.org/10.1109/CVPR.2018.00503

[14] C. Sun, S. Liu, T. Yang, B. Zeng, Z. Wang, & G. Liu, Automatic reflection removal using gradient intensity and motion cues. In Proceedings of the 24th ACM international conference on Multimedia, (2016) 466-470. https://doi.org/10.1145/2964284.2967264

[15] R. Szeliski, S. Avidan, P. Anandan, Layer extraction from multiple images containing reflections and transparency. In Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No. PR00662) 1 (2000) 246-253.

[16] M.A.C. Tuncer, A.C. Gurbuz, Ground reflection removal in compressive sensing ground penetrating radars, IEEE Geoscience and remote sensing letters, 9(1) (2011) 23-27. https://doi.org/10.1109/LGRS.2011.2158981

Funding

No funding was received for conducting this study.

Conflict of interest

The Authors have no conflicts of interest to declare that they are relevant to the content of this article.

About The License