Integrated Knowledge Guided Graph Attention Network with Mixture Density Network for Crop Yield Prediction

Priyanka. K[1*], *Mohan* Kumar. A V[2], *Abdul* Lateef Haroon. P S[3], *Sanjay* Madaan[4], and *Achyutha* Prasad. N[5]

[1]Department of Information Science and Engineering, Nitte Meenakshi Institute of Technology, Nitte (Deemed to be University), Bengaluru, India
[2]Department of Engineering and Technology (CSE), Vyasa Deemed to be University, Sattva Global City, Bengaluru, India
[3]Department of Electronics and Communication Engineering, Ballari Institute of Technology and Management, Ballari, India
[4]Department of Computer Engineering and Applications, GLA University, Mathura, India
[5]Department of Computer Science and Engineering, East West Institute of Technology, Bengaluru, India

Abstract. In recent years, the accurately and timely predict the crop yield has become hard to achieve because crop growth has complex and changing patterns across space and time. However, the existing model struggle to capture dynamic spatial temporal, nonlinear and external environmental factors such as temperature, soil conditions and precipitation, due to its graph performance. To overcome this problem, this paper proposed a Knowledge guided Graph Attention network with a Mixture Density Network (KGAT-MDN) to predict crop yield with more flexibility. The proposed KGAT-MDN first uses a 3D Convolutional Neural Network (CNN) to extracts important features. Next, the graph attention network used to model, the proposed method as a node-level regression problem on a graph built from the distribution system. After, mixture density networks are used to predict multiple possible outcomes at once, along with their importance, which are represented as Gaussian distributions. At last, location-aware Spatial Attention Graph Network is used, which uses geospatial knowledge to combine the features of nearby areas used for the final result. The experiments results obtained 0.50 for RMSE, 0.95 for R^2 and 10.11 for MAPE on Crop yield prediction dataset by outperforming the existing KSTAGE in predicting crop yield in agriculture.

1 Introduction

Crop yield prediction is the process of predicting the number of crops grown in a specific area. It helps governments, businesses and farmers make better decisions regarding agricultural production [1]. The need for accurate pre-harvest yield prediction has increased among the different stakeholders in the agricultural food chain [2]. In the past ten years Artificial Intelligence (AI) has been a more effective method for predicting crop production under different farming conditions [3]. The reliability of these methods depends on how clearly their predictions are explained in terms of human-understandable terms [4]. Modern farming is closely linked to precision agriculture, in which advanced technologies are used to achieve higher yields and improve resource use of resources [5]. Accurate prediction of crop yield is difficult because it depends on many factors such as climate, soil, location, and seed variety [6].

The inconsistency in data availability and quality, especially in rural areas, reduces the effectiveness of these models. However, some models can accurately predict yields for smaller regions [7]. The ability of Deep Learning (DL) based prediction models in agriculture was assessed for corn, grassland land-use intensity mapping, soybean mapping and land cover classification [8]. Remote Sensing (RS) technologies provide quick and cost-effective ways to collect data without direct contact with crops, and have been widely used to monitor crop conditions [9]. Analytical methods used for crop yield forecasting can be classified into three main categories: statistical models, machine learning methods and deep learning (DL) methods [10].

This research aims to address the existing limitations of crop yield prediction methods by implementing an KGAN-MDN, to overcome spatial-temporal, nonlinear and external environmental factors such as precipitation, temperature and soil conditions problems.

The main contribution of this paper as given as follows,

- The development of KGAN-MDN, first used a 3D CNN to extracts important feature. Next, graph attention network used to model, the proposed method as a node-level regression problem on a graph built from the distribution system
- After, Mixture density networks are used to predict multiple possible outcomes at once, along with their importance, which are represented as Gaussian distributions
- At last, a location-aware spatial attention graph network is used, which uses geospatial

* Corresponding author: priyanka.k@nmit.ac.in

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

knowledge to combine the features of nearby areas for crop yield prediction.

- The proposed KGAN-MDN used publicly available datasets such as Wheat yield prediction dataset and Crop yield prediction dataset, which helped to agriculture.

The rest of this paper is organized as follows: Section 2 presents the related work, section 3 presents the proposed methodology, the results and discussion present in section 4 and section 5 deals with a conclusion.

2 Related work

The related work section examines existing Machine Learning (ML) and Deep Learning (DL) approaches for crop yield prediction in agriculture.

Mengjia Qiao et al. [11] implemented a Knowledge-guided Spatial Temporal Attention Graph Network (KSTAGE), which uses prior knowledge. Start with 3D CNN to extract the initial spectral features in the crop. Then, a location-aware Spatial Attention Graph Network is used, which applies geospatial knowledge to combine the features of nearby areas for the final prediction of crop yield. However, predicting crop yield accurately and on time is still a difficult due to dynamic spatial-temporal, nonlinear and external environmental factors such as precipitation, temperature and soil conditions.

Muhammad Waqar et al. [12] presented Stacking Ensemble Model (SEM) and its Optimized variant (OSEM), which integrates real-world agricultural data. Leveraging synthetic data generated through Prophet model, based on features such as temperature, precipitation, pesticide usage, and historical yield, we demonstrated significant improvements in predictive accuracy and model generalization. However, integrating multiple base learners with SVD based dimensionality reduction and hyperparameter tuning increases computational cost due to the global scale of the dataset and the complexity of optimization, the framework demands significant computing resources, limiting scalability.

Fatma M. Talaat [13] explored Crop Yield Prediction Algorithm (CYPA) uses Internet Of Things (IoT) techniques in agriculture. This paper shows the effectiveness of CYPA by dividing the datasets into training and testing five models with the best hyperparameter settings for each machine learning method. It also introduces a new active learning-based algorithm that improves CYPA’s performance, which reduce the amount of labeled data required for training. However, the CYPA algorithm does not generalize well across different geographical regions, due to its reliance on data collected from a particular area, the model's performance was varied when applied to other areas with different environmental and agricultural condition.

Luka Jovanovic et al. [14] introduced Weight Agnostic Neural Networks (WANN) optimized by modified metaheuristic for crop yield prediction. WANNs make it possible to build lighter models by using shared weights and two-layer cooperative framework to optimize both the network design and the shared weights. Although WANNs are not presented as the final solution, they show strong potential for crop yield forecasting with lightweight architectures. By applying metaheuristic optimization, the training time can be greatly reduced and the overall performance of the model can be improved. However, the optimization process does not fully explore the capabilities of slower or faster converging algorithms due to the high cost of comparative analysis, the population sizes and number of iterations were limited.

O. Sri Nagesh et al. [15] explored the Grey Level Co-occurrence Matrix method is used for feature selection and for classification use of the K-Nearest Neighbor (KNN), AdaBoost, Decision Tree, Artificial Neural Network (ANN) algorithms. The data used in this paper was collected from different sources, including yield, pesticide use, rainfall, and average temperature. Predictions of agricultural output made using CNN are now considered some of the most accurate in the field. However, the model’s performance may be limited when applied to new or unseen regions due to the dataset being compiled from specific sources and regions, it does not fully capture the variability in agricultural conditions across different geographies.

2.1 Problem statement and solution

To capture dynamic spatial-temporal, nonlinear and external environmental factors such as precipitation, temperature and soil conditions, due to its graph performance. To overcome this problem proposed Knowledge guided Graph Attention network with Mixture Density Network (KGAN-MDN) is used to predict crop yield with more flexibility.

3 Proposed Methodology

This research proposed KGAN-MDN for predicting crop yield in agriculture. First, crop yield prediction dataset and Wheat yield prediction dataset are used and then applying pre-processing. Finally, the proposed method used to predict the crop yield in agriculture. Figure 1 shows the overall architecture of the proposed method for crop yield prediction.

3.1 Dataset Acquisition

All experiments are conducted using two datasets, namely, the crop yield prediction dataset and wheat yield prediction dataset, which is collected from agriculture. A detailed description of the dataset is provided below:

3.1.1 Wheat yield prediction dataset

This research used a dataset contains 1,266 healthy and (Stip Rust, Septoria) disease-infected images. From the dataset, training used 80% of the images, 10% used for validation, and 10% used for testing. Data augmentation was applied during training to create additional images for the model. The results show that the model works effectively in detecting wheat leaf diseases, specifically Strip Rust and Septoria [16].

3.1.2 Crop yield prediction dataset

This dataset was collected from two different sources. Pesticides tonnes and crop yield datasets are gathered from Food and Agriculture Organization (FAO), while climatic data containing average temperature and rainfall datasets are extracted using the World Data Bank. Pesticides data contain information about different pesticides being used globally from 1990 to 2016. Temperature data consists of different average temperature values from year 1743 to 2013. Rainfall has data of average rainfall per year during 1985 to 2017. At last, the yield data consists of different crops produced globally with their yield value from 1961 to 2016. From this, 70% for training, 10% for testing and 20% for validation [17].

3.2 Data pre-processing

After data collection, the data cannot be used directly in an analysis because it leads to erroneous results. Therefore, the pre-processing steps are discussed below:

3.2.1 Data cleaning

After collecting the data, data cleaning is performed using missing information and removing' unusual data, which affect the model's result during the prediction.

3.2.2 Normalization

After data cleaning, normalization is applied to make the data dimensionless and have similar distributions.

3.3 Prediction using Knowledge guided Graph Attention Network with Mixture Density Network (KGAN-MDN)

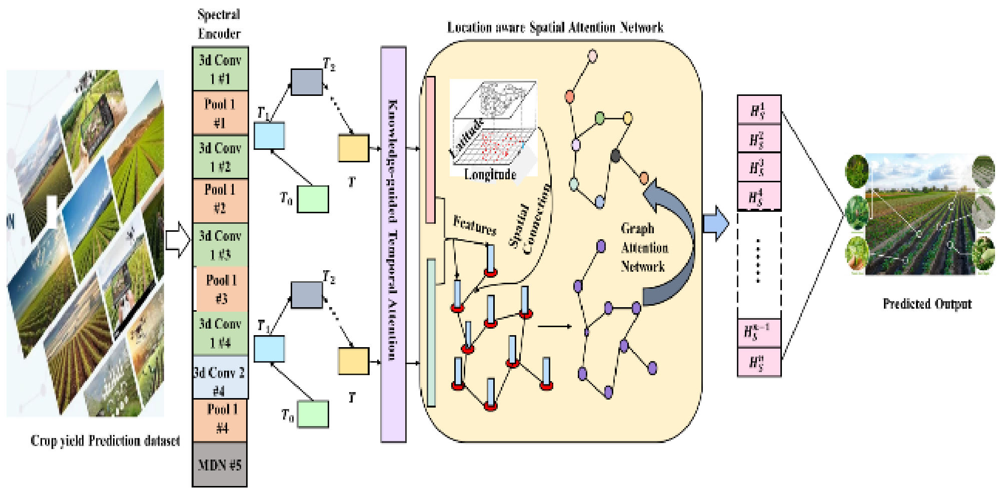

Fig. 1. Architecture diagram of proposed knowledge guided graph attention network with mixture density network (KGAN-MDN).

This section discussed overall architecture of the proposed KGAN-MDN and presented in Figure 1. The proposed KGAN-MDN contains of three key parts such as first, spectral encoder network, knowledge guided mixture density network and location-aware spatial attention graph network.

The proposed KGAN-MDN architecture diagram steps are given below:

- First, the input is collected from crop yield prediction dataset. Next, Spectral encoder, which uses 3D convolution and pooling layers to extract spatial, spectral and temporal features from the input. It generates at different resolutions such as T, T_1, etc. (capture important information from the leaf). it provides a full probability distribution of potential yields, not just a single point estimate. After, MDN used to provide a full probability distribution of potential yields, not just a single point estimate.
- Next, Knowledge-guided Temporal Attention used to identify the most informative images within the time series and enabling the model to focus on the most relevant seasonal data.
- Location-aware Spatial Attention Network, used to takes the features from the temporal attention module and integrates with geographical coordinates (latitude and longitude). Next, Graph Attention Network (GAT) used to learn and utilized the relationship between different parts of the spatial locations.
- The final output, the model aggregates the processed information to produce a single yield estimate for the entire field and more powerfully, a map showing predicted yield variations across the field.

3.3.1 Spectral encoder network

Crop yield prediction is mostly difficult one due to the multiple variability and to effectively capture the spectral information embeddings hidden in RSIs. 3D CNNs are used to extracts 2D filters into 3D enabling to better handle hyperspectral and multispectral inputs shown in Equation (1):

$$f_{ij}^{xyz} = \emptyset\left(\sum_m \sum_{a=0}^{A_i-1} \sum_{b=0}^{B_i-1} \sum_{c=0}^{C_i-1} W_{ijm}^{abc} V_{(i-1)m}^{(x+a)(y+b)(z+c)} + b_{ij}\right) \tag{1}$$

Where, $\emptyset$ represent activation function, b_{ij} represent bias parameter and W_{ijm}^{abc} denotes weight parameter.

3.3.2 Knowledge-guided with Mixture Density Networks (MDNs)

MDNs is built in combination with two main components: a Neural Network and a Gaussian Mixture Model. The output of an MDN consists of the parameters of several Gaussian probability density functions, specifically their means, standard deviations, and mixing coefficients. This design helps deal with the multi-value mapping problem, which occurs in inverse modeling. This problem arises because different design parameter values can lead to similar optical responses. Mathematical representation is given below Equations (2) and (3):

$$p(y|x) = \sum_{k=1}^{K} \Pi_k(X)\phi_k(y|x) \tag{2}$$

$$= \sum_{k=1}^{K} \Pi_k(X)\phi_k(y, \mu_k(x), \sigma_k(x)) \tag{3}$$

Where, Π_k (x) represents mixing coefficient weights and ϕ_k denotes pdfs. A Gaussian probability density function (pdf) is used because combining enough of them (as mixed Gaussians) is closely approximate even very complex data distributions shown in Equations (4) and (5):

$$\phi_k(y, \mu_k(x), \sigma_k(x)) \tag{4}$$

$$\prod_{c=1}^{N}\left(\frac{1}{\sqrt{2\pi\phi_{k,c}(x)}} exp\left(-\left(\frac{y_c-\mu_{k,c}(x)}{\sqrt{2\sigma_{k.c}(x)}}\right)^2\right)\right) \tag{5}$$

3.3.3 Location-aware spatial attention graph network with Graph Attention Network

In crop yield prediction, the prediction units are not independent but influenced by spatial dependencies. These dependencies near units share similar weather conditions, and their predictions are closely related to the location. Therefore, instead of analyzing each unit on its own, it is better to consider these spatial connections. To handle this to use a location-aware spatial attention graph network based on GAT. This network uses location information and combines data from nearby units to produce accurate crop yield predictions. Spatial attention is introduced to capture these dynamic spatial dependencies. Specifically, the pairwise distance L_{ij}between node i and node j is computed.

$$L_{ij} = exp\left[\frac{||dist(v_i,v_j)_2^2}{2l_s^2}\right] \tag{6}$$

Where in Equation (6), $dist(v_i, v_j) = loc_i - loc_j)|$ represent the distance metric between node I and node j.

$$S_{ij} = \frac{exp(L_{ij})}{\sum_{k=1}^{N} exp(L_{ik})} \tag{7}$$

Where in Equation (7), Element S_{ij} represents the pairwise correlation between nodes i and j. The formula is as follows in Equation (8):

$$S_{ij} = \begin{cases} 1 - S_{ij}, \; S_{ij} < \lambda \\ \quad 0, \; Otherwise \end{cases} \tag{8}$$

Where λ denotes the hyper-parameter, and k represents the connections in the spatial graph G_s.

4 Result and Discussion

This section describes the experiments conducted using various machine learning and deep learning approaches applied to crop yield forecasting. The results are discussed in detail below:

4.1 Experiments

All experiments are conducted on 64-bit Intel Core i7-6900K CPU3. 20GHz with 64GB of RAM and an NVIDIA GeForce 1080 GPU using CUDA version 9.0.176.

4.2 Evaluation metrics

To assess the performance of crop yield prediction, the following recognition parameters were considered is as follows:

4.2.1 Mean Squared Error (MSE)

MSE is used to measure mean square of the difference between the predicted yield and the actual yield. It is commonly used in regression tasks to determine the accuracy of the model. Equation (9) is used to calculate MSE.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{9}$$

4.2.2 Root Mean Squared (RMSE)

RMSE is the under root of MSE. It gives magnitude of the error in the same scale as actual data, with lower value means high performance. It is calculated using Equation (10).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{10}$$

4.2.3 Mean Absolute Error (MAE)

MAE examines mean of the absolute difference among the predicted and actual yield. It provides simple mean of the absolute errors, making it easier to understand the evaluation of the model. To calculate MAE, Equation (11) is used.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i| \quad (11)$$

4.2.4 Mean Absolute Percentage Error (MAPE)

MAPE calculates percentage of mean absolute error among the predicted and actual output. We can get a percentage for comparing the performance across different datasets using MAPE. Equation (12) is used to calculate MAPE.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \quad (12)$$

4.2.5 Coefficient of Determination (R^2)

R^2 calculates proportion of the variance of the depen dent feature from the independent feature. It ranges from 0 to 1, with values closed to 1 indicates that the model shows a greater part of variance. Equation (13) is used to find.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad (13)$$

4.3 Performance Analysis

This section used to evaluate performance of the proposed KGAT-MDN are compared with existing methods namely Deep Convolution Neural Networks (DCNN), CNN, Deep Belief Network (DBN) and Bi-GRU. These models are compared based on MSE, RMSE, MAE and R^2. In the first experiment, the prediction results on the Crop yield prediction dataset are compared with existing techniques and the results are summarized in Table 1. Sequentially, the result of the wheat yield prediction dataset is compared with the existing techniques and it summarized in Table 2. The proposed method achieved the highest recognition rate in crop yield prediction dataset.

Table 1. Performance analysis for proposed method using crop yield prediction dataset.

Models	RMSE	R^2	MAPE	MAE	MSE
DCNN	1.54	0.22	1.45	1.56	1.67
CNN	0.99	0.34	1.10	1.03	0.672
DBN	1.80	0.51	2.65	0.39	1.895
Bi-GRU	0.83	0.45	0.56	2.67	1.75
Proposed KGAT-MDN	0.50	0.95	0.11	0.20	0.235

Table 2. Performance analysis for proposed method using wheat yield prediction dataset.

Models	RMSE	R^2	MAPE	MAE	MSE
DCNN	2.76	0.18	1.76	2.78	0.71
CNN	1.68	0.27	1.93	1.83	1.89
DBN	0.79	0.82	1.98	1.28	0.52
Bi-GRU	1.06	0.76	0.93	1.59	1.57
Proposed KGAT-MDN	0.45	0.90	0.37	1.15	0.374

4.4 Computational complexity

This section evaluates the complexity of the KGAN-MDN, which higher compared to existing methods such as DCNN, CNN, DBN and BiGRU are approximately 18740 seconds. The KGAN-MDN demonstrated 10743 seconds, memory usage also low and hyperparameter are upto 6. Table 3 present the comparison of hyperparameter, memory usage and training time across existing methods.

Table 3. Comparison of training time, memory usage and hyperparameter for crop yield prediction.

Models	Hyperparameter	Memory (in MB)	Training time (in seconds)
DCNN	9.2	0.74	18740
CNN	6.2	0.85	16675
DBN	7.3	0.95	14380
Bi-GRU	10.5	0.57	13600
Proposed KGAT-MDN	5.9	0.30	10743

4.5 Statistical Analysis

This section validates statistical, using of both Wilcoxon signed-rank and Friedman tests based on MAE values across 10-folds. The Friedman test yielded a highly significant result (x^2 (5) = 48.63, p < 0.00001), confirming meaningful performance differences among the models. Additionally, the Wilcoxon signed-rank test demonstrated a statistically significant advantage of KGAN-MDN over existing model (p = 0.00195).

4.6 Ablation study

Ablation experiments are used to examine performance of proposed method and assessing the individual features that affect a model's performance across crop yield prediction dataset and wheat yield prediction dataset. Based on the proposed method, generated three cases for the conduction of ablation study. The first case considers baseline GAT, KSTAGE with GAT, KSTAGE with MDN, and the KGAT-MDN model. The outcome of these ablation study is presented in Table 4 and Table 5.

Table 4. Ablation study for proposed method in crop yield prediction dataset.

Models	RMSE	R^2	MAPE	MAE	MSE
Baseline GAT	1.54	0.34	1.76	16.87	2.68
KSTAGE - GAT	1.45	0.25	2.92	3.67	1.78
KSTAGE - MDN	0.98	0.56	1.34	2.94	0.97
Proposed KGAT-MDN	0.50	0.95	0.11	0.20	0.235

Table 5. Ablation study for proposed method in wheat yield prediction dataset.

Models	RMSE	R^2	MAPE	MAE	MSE
Baseline GAT	2.56	0.25	1.63	1.87	1.35
KSTAGE - GAT	1.89	0.73	1.87	1.37	1.56
KSTAGE - MDN	0.82	0.36	0.93	2.56	2.72
Proposed KGAT-MDN	0.45	0.90	0.37	1.15	0.374

4.7 Comparative analysis

In this section represent comparison of proposed KGAT-MDN with the state-of-art models such as DCNN, CNN, DBN and Bi-GRU. In comparison, the KGAN-MDN model when implemented with GAT outperformed the existing models. The KGAN-MDN achieves a high RMSE of 0.50 for Crop yield prediction dataset compared to SEM-OSEM. Tables 6, 7 presents a comparison of MSE, RMSE, MAE, R^2, between the existing models and KGAN-MDN in crop yield prediction dataset and wheat yield prediction dataset

Table 6. Comparative analysis of proposed method using crop yield prediction dataset.

Models	RMSE	R^2	MAPE	MAE	MSE
SEM-OSEM [12]	NON	0.990	6.78	0.185	NON
L2-WANN-MSRSA [14]	0.117725	0.4825	NON	0.076147	0.013859
Proposed KGAT-MDN	0.50	0.95	0.11	0.20	0.235

Table 7. Comparative analysis of proposed method using wheat yield prediction dataset.

Models	RMSE	R^2	MAPE	MAE	MSE
KSTAGE [11]	0.57	0.88	11.20	NON	NON
Proposed KGAT-MDN	0.45	0.90	0.37	1.15	0.374

5 Discussion

This paper introduces a proposed KGAN-MDN to predict crop yield in agriculture, using 3D CNN, graph attention network, mixture density networks and location-aware spatial attention graph networks used geospatial knowledge. The performance including an RMSE of 0.50, MAPE of 0.11, and R^2 of 0.95, when the crop yield prediction dataset was used. Similarly, the wheat yield prediction datasets obtained 0.45, 0.37, and 0.90, R^2.These improvements are attributed, the proposed KGAT-MDN first uses a 3D Convolutional Neural Network (CNN) to extract important features. Next, the graph attention network was used to model the proposed method as a node-level regression problem on a graph built from the distribution system. Subsequently, mixture density networks are used to predict several possible outcomes simultaneously, including their respective importance as Gaussian distributions. Finally, a location-aware spatial attention graph network is used, which uses geospatial knowledge to combine the features of nearby areas for final crop yield prediction. This research the proposed KGAN-MDN achieved a better result compared to existing models.

6 Conclusion

In this research proposed a KGAN-MDN, used to predict the crop yield in agriculture, which demonstrated significant improvements in predictive result and model generalization. The proposed KGAT-MDN first used a 3D Convolutional Neural Network (CNN) to extracts important feature. Next, graph attention network used to model, the proposed method as a node-level regression problem on a graph built from the distribution system. After, Mixture density networks are used to predict multiple possible outcomes at once, along with their importance, which are represented as Gaussian distributions. At last, location-aware Spatial Attention Graph Network is used, which uses geospatial knowledge to combine the features of nearby areas used for the final result. Statistical analysis performance is evaluating of the proposed KGAN-MDN using 10-fold cross-validation by statistical significance tests including the Wilcoxon signed-rank and Friedman tests. For future, aim to expand this framework by integrating multimodal datasets, including soil quality, satellite imagery, humidity, CO2 levels and economic indicators. These advancements will further strengthen the applicability of model in agricultural planning and climate resilient farming strategies to support global food security.

References

1. L.K. Subramaniam, R. Marimuthu, Crop yield prediction using effective deep learning and dimensionality reduction approaches for Indian regional crops. e-Prime Adv. Electr. Eng. Electron. Energy **8**, 100611 (2024). https://doi.org/10.1016/j.prime.2024.100611
2. S. Bregaglio, F. Ginaldi, E. Raparelli, G. Fila, S. Bajocco, Improving crop yield prediction accuracy by embedding phenological heterogeneity into model parameter sets. Agric. Syst. **209**, 103666 (2023). https://doi.org/10.1016/j.agsy.2023.103666
3. S.D. Gawade, A. Bhansali, S. Chopade, U. Kulkarni, Optimizing crop yield prediction with R2U-Net-AgriFocus: A deep learning architecture leveraging satellite imagery and agro-environmental data. Expert Syst. Appl. **296**, 128942 (2026). https://doi.org/10.1016/j.eswa.2025.128942
4. D. Paudel, A. De Wit, H. Boogaard, D. Marcos, S. Osinga, I.N. Athanasiadis, Interpretability of deep learning models for crop yield forecasting. Comput. Electron. Agric. **206**, 107663 (2023). https://doi.org/10.1016/j.compag.2023.107663
5. S. Mahalakshmi, A.J. Anand, P. Partheeban, Soil and crop interaction analysis for yield prediction with satellite imagery and deep learning techniques for the coastal regions. J. Environ. Manag. **380**, 125095 (2025). https://doi.org/10.1016/j.jenvman.2025.125095
6. A. Kaur, P. Goyal, R. Rajhans, L. Agarwal, N. Goyal, Fusion of multivariate time series meteorological and static soil data for multistage crop yield prediction using multi-head self attention network. Expert Syst. Appl. **226**, 120098 (2023). https://doi.org/10.1016/j.eswa.2023.120098
7. U. Nirosha, G. Vennila, Enhancing crop yield prediction for agriculture productivity using federated learning integrating with graph and recurrent neural networks model. Expert Syst. Appl. **289**, 128312 (2025). https://doi.org/10.1016/j.eswa.2025.128312
8. S. Stiller, K. Grahmann, G. Ghazaryan, M. Ryo, Improving spatial transferability of deep learning models for small-field crop yield prediction. ISPRS Open J. Photogramm. Remote Sens. **12**, 100064 (2024). https://doi.org/10.1016/j.ophoto.2024.100064
9. D. Al-Shammari, B.M. Whelan, C. Wang, R.G. Bramley, T.F. Bishop, Assessment of red-edge based vegetation indices for crop yield prediction at the field scale across large regions in Australia. Eur. J. Agron. **164**, 127479 (2025). https://doi.org/10.1016/j.eja.2024.127479
10. H. Demirhan, A deep learning framework for prediction of crop yield in Australia under the impact of climate change. Inf. Process. Agric. **12**, 125-138 (2025). https://doi.org/10.1016/j.inpa.2024.04.004
11. M. Qiao, X. He, X. Cheng, P. Li, Q. Zhao, C. Zhao, Z. Tian, KSTAGE: A knowledge-guided spatial-temporal attention graph learning network for crop yield prediction. Inf. Sci. **619**, 19-37 (2023). https://doi.org/10.1016/j.ins.2022.10.112
12. M. Waqar, Y.W. Kim, Y.C. Byun, A stacking ensemble framework leveraging synthetic data for accurate and stable crop yield forecasting. IEEE Access. **13**, 136909-136926 (2025). https://doi.org/10.1109/ACCESS.2025.3591802
13. F.M. Talaat, Crop yield prediction algorithm (CYPA) in precision agriculture based on IoT techniques and climate changes. Neural Comput. Appl. **35**, 17281-17292 (2023). https://doi.org/10.1007/s00521-023-08619-5
14. L. Jovanovic, M. Zivkovic, N. Bacanin, M. Dobrojevic, V. Simic, K.K. Sadasivuni, E.B. Tirkolaee, Evaluating the performance of metaheuristic-tuned weight agnostic neural networks for crop yield prediction. Neural Comput. Appl. **36**, 14727-14756 (2024). https://doi.org/10.1007/s00521-024-09850-4
15. O.S. Nagesh, R.R. Budaraju, S.S. Kulkarni, M. Vinay, S.S.M. Ajibade, M. Chopra, M. Jawarneh, K. Kaliyaperumal, Boosting enabled efficient machine learning technique for accurate prediction of crop yield towards precision agriculture. Discov. Sustain. **5**, 78 (2024). https://doi.org/10.1007/s43621-024-00254-x
16. The Wheat yield prediction dataset link: https://www.kaggle.com/datasets/shaikasif89/wheat-yeild/data
17. The crop yield prediction dataset link: https://www.kaggle.com/datasets/patelris/crop-yield-prediction-dataset