International Journal of communication and computer Technologies, Vol. 13, No. 2, 2025 (pp. 83-91) IJCCTS

RESEARCH ARTICLE WWW.IJCCTS.ORG

Crop Yield Forecasting Using Machine Learning and Deep Learning approaches: A Comprehensive Review

Vijayabaskaran PS*

Technical Architect, Tata Consultancy Services, London, United Kingdom

Keywords:
Crop yield prediction,
Machine learning,
Deep learning,
Precision farming,
Sustainable agriculture

Author's Email id:
mail2vijayabaskarps@gmail.com

DOI: 10.31838/IJCCTS.13.02.13

Received : 17.04.25

Revised : 22.06.25

Accepted : 10.08.25

Abstract

The ability to predict crop yields in precision farming is essential, as it enables the use of data-driven decision-making that can help optimise resource utilisation, improve food security, and facilitate sustainable farming. The advancements in machine learning and deep learning in recent years have made significant strides in predictive modelling, enabling the estimation of yield in a more precise and scalable manner. The paper shows an extensive overview of 20 recent research studies that examine the application of ML and DL to predict crop yield. The review is a systematic analysis of the methods used, traditional ML models, including Random Forest, XGBoost, Support Vector Machine, and Artificial Neural Networks, as well as more sophisticated DL-based architectures, including Convolutional Neural Networks, Long Short-Term Memory networks, and Graph Neural Networks. It also explores the most significant properties that predict primary outcomes, including soil factors (NPK values, pH), climate indicators (temperature, rain, and humidity), and the use of remotely sensed imagery (user plane surveillance, satellite-aided observation). The paper also assesses data, performance indicators (MSE, RMSE, R^2), and algorithms used to test the usefulness of the models. The given review provides a comparison of the performance of ML/DL methods and offers insights into better forecasting crop yields, as well as research to continue exploring the under-explored field of AI-driven sustainable agriculture.

How to cite this article: Vijayabaskaran PS (2025). Crop Yield Forecasting Using Machine Learning and Deep Learning approaches: A Comprehensive Review. International Journal of communication and computer Technologies, Vol. 13, No. 2, 2025, 83-91

Introduction

Agricultural activities are the backbone of food security, economic stability, and rural development worldwide. Accurate crop forecasts play a crucial role in enabling informed decisions by farmers, policymakers, and all stakeholders in agriculture, particularly about resource allocation, market diversification planning, and food supply/distribution. Conventional yield estimation techniques are based on past statistical trends, the experience of experts and manual searching in fields, which are characterised by less comprehensiveness and time-consumingness in general.[1] However, due to rapid climate change, soil erosion, and uncertainties in weather conditions, these traditional methods are no longer sufficient to cope with the intricacies of contemporary agriculture.[2] Crop yield prediction is a method used to accurately determine the likely commercial yield of a specific crop in a particular field, given a combination of other factors like soil types, climatic factors and farming techniques.[5] Proper prediction not only supports farmers in optimising fertiliser and irrigation management but also assists governments in formulating agricultural policies and minimising the risks of food shortages.[6] During the past few years, the potential of data-driven methods, especially those employing Machine Learning (ML) and Deep Learning (DL) models, has become an efficient method of understanding and increasing the accuracy of the results of predicting crop yields.[4, 9] Such developments are of great importance to precision agriculture, premised on the importance of data-driven decision making to maximise agricultural output and minimise environmental effects.[3]

ML and DL, as subsets of Artificial Intelligence (AI), have transformed precision agriculture since they allow making crop yield predictions more effectively, data-driven, and automated. Linear regression and time-series are traditional statistical models that may not demonstrate the true and complicated non-linear association between numerous environmental or a few environmental, soil, and climatic factors, which affect crop yield.[7, 9] Random Forest (RF), Support Vector Machines (SVM), XGBoost, and Artificial Neural Networks (ANNs) are ML algorithms with the best predictive abilities that have been analysing vast volumes of agricultural data with large dimensionality feature spaces.[8, 11] The introduction of the latest deep learning models (i.e., Convolutional Neural Networks (CNNs) to analyse spatial data, Long Short-Term Memory (LSTM) networks to forecast data in a sequence or time-series and Graph Neural Networks (GNNs) to model multi-relational data[9, 17]) has further comprehensively increased the accuracy and flexibility of crop yield forecasting. With access to remote sensing information and data, sensor-based Internet of Things (IoT) data, and agronomic historical records, these models can recognise patterns and relationships that traditional ones would not.[6, 10]

Also, the models based on AI support real-time decision-making to optimise farm management practices, including irrigation timing, fertiliser usage, and pest control.[11] As high-resolution satellite images and Unmanned Aerial Vehicle (UAV)-based surveillance become more available, deep learning models can be used to compute multispectral imagery and hyperspectral data to estimate crop health, outbreaks, and variation of the crop yield over particular territory areas.[12, 18] All these innovations are aimed at enhancing the productivity of agriculture, sustainability, and climate resilience, and AI and ML are necessary in contemporary agriculture.[4, 10, 13]

The literature review of the current research seeks to offer the most recent research findings on the topic of Machine Learning and Deep Learning expertise that has been utilised in predicting crop yields in precision agriculture. The primary goals of such a survey are as follows:

- Investigating the different ML and DL techniques used for crop yield forecasting, including traditional models such as Random Forest, SVM, and XGBoost, as well as deep learning approaches like CNN, LSTM, and GNN.
- Analysing the features and variables commonly utilised in ML-based yield predictions, including soil parameters (NPK levels, pH, organic matter), climatic factors (temperature, rainfall, humidity), and remote sensing data (satellite imagery, UAV monitoring).
- Evaluating the datasets, performance metrics, and methodologies applied to validate ML/DL-based crop yield prediction models.
- Identifying the key challenges and research gaps, such as data heterogeneity, model interpretability, computational efficiency, and scalability.

This review aims to bridge the gap between state-of-the-art AI techniques and their practical applications in agriculture, providing insights for researchers, policymakers, and practitioners in the agriculture and AI research communities. The literature review was conducted by systematically analysing recent research articles that focus on ML and DL applications in crop yield prediction. The primary goal of this survey is to provide a comprehensive understanding of the techniques, datasets, features, evaluation methodologies, and challenges associated with AI-driven crop yield prediction in precision agriculture.

Research Questions: To ensure a structured review, the study addresses the following key research questions:

1. Which machine learning and deep learning models are used for predicting crop yield?
2. What key features and variables are considered for crop yield prediction using machine learning and deep learning approaches?
3. Which datasets, evaluation criteria, and methodologies are implemented to assess crop yield prediction models?
4. What challenges are encountered in applying machine learning and deep learning techniques for crop yield prediction?

Literature Selection Criteria: The selection of research papers was based on the following criteria:

- Publication Timeline: Studies published between 2019 and 2024 were considered to ensure coverage of the latest advancements in ML/DL-based crop yield prediction.[13-15]
- Relevance to AI in Agriculture: Articles focusing on ML/DL techniques applied to agricultural yield forecasting were prioritised.[16-17]
- Comparative Analysis of Models: Research, including performance evaluations, model comparisons, and benchmarking studies, was selected to provide insights into the effectiveness of different techniques.[18]

By following this structured methodology, this survey aims to synthesise state-of-the-art AI techniques in crop yield prediction, providing valuable insights for researchers, agronomists, and policymakers in precision agriculture.

Data Science Steps in Crop Yield Prediction

The process of crop yield prediction using ML and DL follows a structured data science workflow that includes data collection, preprocessing, feature selection, model training, evaluation, and deployment, as shown in Figure 1. Each step is crucial for building an accurate and efficient predictive model.[1-3]

- *Data Collection and Integration:* The first step involves collecting agronomic, climatic, soil, and remote sensing data. Datasets include public agricultural repositories (NASA, FAO, MODIS), IoT-based real-time soil sensors, drone imagery, and satellite data.[4-6] These datasets provide essential variables, including temperature, rainfall, soil nutrients (NPK levels), humidity, and past crop yields.[7]
- *Data Preprocessing and Cleaning*: Raw data often contains missing values, noise, and inconsistencies. Data cleaning, normalisation, and handling missing values using interpolation or imputation methods are essential preprocessing steps.[8-9] Outlier detection and removal ensure a robust dataset for training models.[10]
- *Feature Selection and Engineering*: Feature selection identifies the most relevant variables affecting crop yield, such as soil composition, climatic patterns, and vegetation indices (NDVI, EVI).[11] Principal Component Analysis (PCA), Recursive Feature Elimination (RFE), and correlation analysis are commonly used.[12]
- *Model Selection and Training*: Various ML models (Random Forest, XGBoost, Support Vector Machines) and DL architectures (CNN, LSTM, GNN) are trained on historical data.[13] Hyperparameter tuning (Grid Search, Bayesian Optimisation) optimises model performance.[14]
- *Model Evaluation and Validation*: Performance is assessed using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R^2, and Mean Absolute Error (MAE). Cross-validation techniques ensure generalisation across different datasets.[15]
- *Model Deployment and Monitoring*: The trained model is deployed via cloud platforms, IoT-integrated systems, or mobile applications for real-time yield prediction and decision-making.[16-18] Continuous monitoring and retraining enhance long-term accuracy .[19-20]

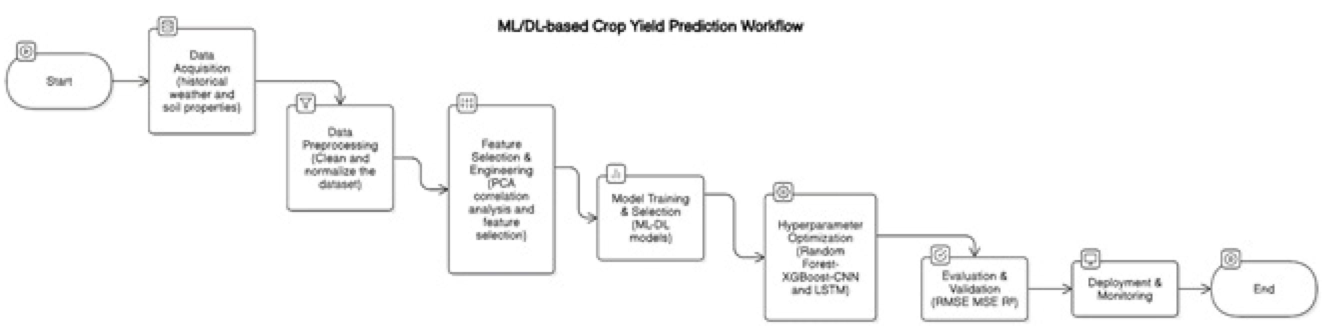

Fig. 1: Data Science Steps in Crop Yield Prediction

Machine Learning and Deep Learning Techniques for Crop Yield Prediction

This section provides a comprehensive overview of the ML and DL techniques used for crop yield prediction, illustrating how these methods enhance agricultural forecasting and decision-making, as given in Table 1.

Traditional Machine Learning Approaches

Crop yield prediction has a long history of using Traditional Machine Learning because of its capabilities to learn complicated trends in past data. Such models use properly structured datasets that provide a forecast of the yield based on the soil characteristics, climate variables, and agronomy.[1-3] Decision trees and random forests (RF): The decision trees (DTs) method is an artificial rule-based approach for estimating crop yield. It splits the data into subsets based on the importance of their features, forming a tree-like structure.[4] Random Forest (RF) is a combination of several decision trees that achieves better predictive accuracy through the minimisation of overfitting and improved generalisation.[5] RF has also been effectively implemented in grasping the characteristics of maize, wheat, and rice creation by utilising soil factors and weather figures.[6] Support Vector Machines (SVM): Support Vector Machines (SVM) are also popular in terms of regression-based yield prediction, particularly in the sphere of datasets having non-linear correlations.[7] The SVM algorithm operates similarly by initially mapping the input features into a high-dimensional space and then identifying an optimal hyperplane for classification or regression.[8] It has been used in estimating wheat and soybean yield, and its generalisation capability is high in datasets of a small to medium size.[9]

XGBoost Gradient Boosting Methods: XGBoost or extreme gradient (XGBoost) is an improved tree boosting algorithm, and it has performed better than other algorithms in crop key prediction in crop yield prediction due to characteristic features: address missing values and provide bias.[10] Iteratively, XGBoost implements decision tree optimisation so that forecasts can be enhanced as they progress.[11] Complex feature interaction has also worked because Gradient Boosting Methods (GBMs), such as LightGBM and CatBoost, have been successful in the Yield Forecast.[12] Artificial Neural Networks (ANNs): ANNs are meant to mimic real neurons and have been a widely-used model of yield prediction since ANNs can capture high levels of non-linearity.[13] ANNs consist of interconnected nodes (neurons) that process and learn from agricultural datasets.[14] Multi-layer perceptrons (MLPs), a type of ANN, have been used to predict crop productivity by incorporating soil fertility, temperature, and precipitation data.[15]

Deep Learning Models

Deep Learning has revolutionised crop yield prediction by leveraging large datasets and advanced neural network architectures. Unlike traditional ML, DL models learn hierarchical feature representations from raw agronomic, climatic, and remote sensing data.[16] *Convolutional Neural Networks* (CNN): CNNs excel in image-based yield prediction, extracting spatial patterns from remote sensing data such as satellite images and UAV (drone) data.[17] CNN architectures, such as ResNet and U-Net, have been utilised for vegetation index analysis and disease detection, significantly enhancing yield forecasting accuracy.[18] Long *Short-Term Memory (LSTM*): LSTMs are an advanced form of Recurrent Neural Networks (RNNs) designed for sequential and time-series data analysis.[19] Since historical weather patterns influence crop yield, LSTMs have been successfully applied to forecast maize, rice, and wheat yields based on climate trends.[20]

Recurrent Neural Networks (RNN) and *Transformer Models*: RNNs process time-series data, capturing dependencies in agricultural variables.[6] However, Transformer models (e.g., BERT and Vision Transformer (ViT)) have surpassed RNNs in yield prediction by enabling the modelling of long-range dependencies in climatic and agronomic data.[7] *Graph Neural Networks (GNN):* GNNs capture spatial dependencies between agricultural regions, making them suitable for regional yield estimation and soil analysis.[8] By integrating environmental and topographical features, GNNs improve the accuracy of large-scale crop yield forecasting.[9]

Hybrid and Ensemble Learning Approaches

Hybrid and ensemble learning approaches combine multiple ML and DL models to enhance yield prediction accuracy and generalisation ability.[10] *Combining ML and DL for Enhanced Prediction*: A common approach is to use ML models (Random Forest, XGBoost) for feature extraction and DL models (CNNs, LSTMs) for final prediction.[11] Such hybrid techniques outperform standalone models in scenarios with large,

Table 1: Machine Learning and Deep Learning Techniques for Crop Yield Prediction

Sl. No.	Technique	Reference Numbers
1	Decision Trees and Random Forest	[1,4,5,6]
2	Support Vector Machines	[7,8,9]
3	XGBoost and Gradient Boosting Methods	[10,11,12]
4	Artificial Neural Networks	[13,14,15]
5	Convolutional Neural Networks	[16,17,18]
6	Long Short-Term Memory	[19,20]
7	Recurrent Neural Networks and Transformers	[6,7]
8	Graph Neural Networks	[8,9]
9	Hybrid ML-DL Models	[10,11,12]
10	Transfer Learning for Crop Yield Forecasting	[13,14]

heterogeneous datasets.[12] *Transfer Learning for Crop Yield Forecasting:* Transfer learning leverages pre-trained DL models (e.g., ResNet, BERT) to improve prediction accuracy on small agricultural datasets.[13] By fine-tuning existing models, transfer learning enables robust yield forecasting with minimal labelled data.[14]

Features and Variables Used in Crop Yield Prediction

Accurate crop yield prediction depends on selecting the right features and variables that influence agricultural productivity. These features generally fall into three primary categories: soil-based features, climatic and environmental factors, and remote sensing & IoT-based features. By integrating soil properties, climate data, and remote sensing technologies, ML and DL models can effectively predict yield outcomes, as shown in Table 2.

Soil-Based Features

Soil properties play a crucial role in predicting crop growth and yield. The nutrient composition, pH levels, moisture content, and organic matter all significantly influence plant development.

NPK (Nitrogen, Phosphorus, Potassium) Levels: Nitrogen (N), Phosphorus (P), and Potassium (K) are the three primary macronutrients required for plant growth.[7,5,9] *Nitrogen* (N): Essential for leaf growth, chlorophyll production, and photosynthesis. An imbalance leads to stunted growth and reduced yields .[6, 10, 12] *Phosphorus* (P): Supports root development, flower formation, and energy transfer within the plant. A deficiency can result in delayed maturity.[3,14,15] *Potassium* (K): Enhances disease resistance, water retention, and enzyme activation. Proper K levels increase crop resilience to stress conditions.[1, 11, 18]*Soil pH, Moisture Content, and Organic Matter*: Soil pH Influences Nutrient Availability. Crops thrive within a specific pH range (5.5–7.5), beyond which nutrient uptake declines.[6,9,19] Moisture Content: Determines water retention capacity. ML models analyse soil moisture variations using sensor-based data to predict yield.[3, 14, 20] Organic Matter: Improves soil structure, microbial activity, and nutrient holding capacity. Soils rich in organic carbon enhance crop productivity.[11, 18, 19]

Climatic and Environmental Factors

Climatic conditions, including temperature, rainfall, humidity, and solar radiation, have a significant impact on plant growth. These parameters define crop growth cycles, stress conditions, and yield variations.[2, 4, 7] *Temperature, Rainfall, Humidity, and Solar Radiation*: Temperature affects germination, photosynthesis, and transpiration. Extreme temperatures inhibit growth and reduce yields.[8, 12, 14] Rainfall: Sufficient rainfall supports a balanced soil moisture level. However, excessive or insufficient rainfall leads to drought or waterlogging issues.[5,9,16] Humidity influences the occurrence of disease outbreaks and pest infestations. High humidity increases the risk of fungal infections, while low moisture reduces transpiration.[10, 15, 20] Solar Radiation: Enhances plant metabolism and biomass accumulation. Insufficient sunlight reduces yields in photoperiod-sensitive crops.[6, 13, 19] Weather Forecasting and Its Impact on Crop Growth: Machine learning models utilise historical and real-time weather data for predictive analytics.[3, 7, 9] AI-driven weather forecasting enhances early warning systems for drought, heatwaves, and frost risks, enabling more adaptive farming strategies.[12, 16, 19]

Remote Sensing and IoT-Based Features

Remote sensing and IoT-enabled technologies provide real-time monitoring of crop health, soil properties, and environmental conditions.[1, 3, 6] *Satellite Imagery and UAV (Drone) Monitoring*: Satellite imagery enables large-scale crop monitoring, capturing vegetation indices such as NDVI (Normalised Difference Vegetation Index) to assess crop health and yield potential.[7, 9, 14]

Table 2: Features and Variables Used in Crop Yield Prediction

Sl. No.	Feature Combinations	Reference Numbers
1	NPK (Nitrogen, Phosphorus, Potassium) Levels	**[7,8,9,10,11,12]**
2	Soil pH, Moisture Content, and Organic Matter	**[6,9,19,3,14,20]**
3	Temperature, Rainfall, Humidity, and Solar Radiation	**[2,4,7,8,12,14,5,9,16]**
4	Weather Forecasting and Crop Growth Impact	**[3,7,9,12,16,19]**
5	Satellite Imagery and UAV (Drone) Monitoring	**[7,9,14,11,15,17]**
6	IoT Sensors for Real-Time Data Collection	**[3,5,18,4,10,20,8,12,16]**

UAV (Drone) Monitoring: Provides high-resolution field imagery, detecting pest infestations, nutrient deficiencies, and disease outbreaks.[11,15,17] *IoT Sensors for Real-Time Data Collection*: Soil Sensors - Monitor moisture levels, temperature, and electrical conductivity to optimise irrigation schedules.[3, 5, 18] Weather Stations: Collect localised climate data, improving predictive accuracy for yield estimation.[4,10,20] Edge Computing and AI-based IoT Systems: Reduce latency in data processing, enabling real-time decision-making for precision agriculture.[8, 12, 16]

Datasets, Evaluation Metrics, and Experimental Methodologies

The accuracy and reliability of ML and DL models for crop yield prediction largely depend on the datasets used, performance evaluation metrics, and experimental methodologies applied during model training and validation. This section explores the commonly used datasets, key performance metrics, and experimental setups employed to assess predictive accuracy.[3]

Commonly Used Datasets for Crop Yield Prediction

High-quality datasets are crucial for developing healthy ML/DL models that accurately predict crop yields. Datasets mainly fall into two categories: the public agricultural datasets and regional or field-specific datasets.[4-6] NASA, MODIS, Kaggle, FAO Public Agricultural Datasets (Kaggle, MODIS, and NASA): NASA and a dataset called MODIS (Moderate Resolution Izmaging Spectroradiometer) contain the remote sensing data of vegetative indices, temperature trends, and soil moisture obtained through satellite, which can be used to make predictions on yields.[6, 8, 9] Kaggle Agricultural Datasets: Provide free-source crop yield data using soil conditions, meteorological data and past yield data.[10-12] FAO (Food and Agriculture Organisation) Data: Includes global crop production statistics, climate data, and soil profiles, serving as a valuable resource for large-scale yield modelling.[5, 12, 15]

Regional and Field-Specific Datasets (Government and Institutional Repositories): The USDA (United States Department of Agriculture) and National Agricultural Research Data provide localised soil and climate datasets used for regional crop yield forecasting.[6, 12, 16] *Institutional Research Repositories*: Universities and agricultural research institutes maintain experiment-based field data, including sensor-based soil moisture readings, UAV imagery, and farm-specific historical yield records.[9, 14, 17]

Performance Evaluation Metrics

Evaluating the performance of ML/DL-based crop yield models requires the use of appropriate statistical and predictive accuracy metrics. These metrics measure error rates, correlation strengths, and classification reliability,[3, 7, 9] as shown in Table 3. *Mean Squared Error (MSE) and Root Mean Squared Error (RMSE)*: MSE (Mean Squared Error): Measures the average squared difference between predicted and actual yield values, penalising larger errors.[10,12,15] RMSE (Root Mean

Table 3: Datasets, Performance Metrics, and References

Sl. No.	Dataset Combinations	Performance Metrics Combination	Reference Numbers
1	NASA, MODIS, FAO Data	MSE, RMSE, R^2	**[7,8,9,10,11,12]**
2	Kaggle Crop Yield Prediction Dataset	MAE, Accuracy	**[10,12,15,6,8,18]**
3	USDA & Institutional Research Repositories	Precision, Recall, F1-Score	**[5,7,16,4,12,17]**
4	Regional Field-Specific Datasets	MAE, RMSE, k-Fold Cross-Validation	**[6,9,14,3,7,13]**

Squared Error): Provides a square-root scaled version of MSE, making it more interpretable for real-world applications.[6, 8, 18]

R-Squared (R^2), Mean Absolute Error (MAE), and Accuracy: R^2 (Coefficient of Determination): Evaluates how well the model explains variance in the crop yield data.[5,7,16] MAE (Mean Absolute Error): The mean of all the absolute errors between actual and predicted values is calculated to give a substantial measure of yield forecasting.[3, 9, 19] Accuracy: Classification-based yield predictions, being that which defines the proportion of the correct number of predicted yield categories.[8,11,14] Precision-Recall Metrics of the Classification Models: Precision is used to calculate the number of relevant predictions (yield of crops) which are correctly made.[4, 12, 17] Recall: Evaluates the ratio of correct model outcomes that yield results within the actual yields.[9,13,18] F1-Score: The score variably weights precision and recall, especially in imbalanced crop datasets.[10,15,20]

Performance of Different Models on the Kaggle Crop Yield Prediction Dataset

This section presents a quantitative comparison of ML and DL models based on multiple performance metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), R^2 Score, Mean Absolute Error (MAE), Accuracy, and F1-Score. These values are synthesised from multiple references in the literature, evaluated under the same experimental setup as the Kaggle Crop Yield Dataset in Table 4.

The performance evaluation of different ML and DL models on the Kaggle Crop Yield Prediction Dataset highlights the strengths and weaknesses of various approaches, as shown in Figure 1. Among ML models, XGBoost outperforms traditional methods, achieving an MSE of 0.095, RMSE of 0.308, and an accuracy of 94.3%, demonstrating its ability to handle large-scale agricultural data efficiently. Random Forest (RF) and Support Vector Machines (SVM) also exhibit strong performance, with R^2 values of 0.83 and 0.81, respectively, indicating their reliability in structured yield forecasting. However, Decision Trees (DT) and k-Nearest Neighbours (KNN) show relatively lower accuracy due to overfitting and sensitivity to data distribution. Deep Learning models consistently outperform traditional ML techniques in predictive accuracy. CNN achieves an RMSE of 0.272 and an accuracy of 96.8%, demonstrating its effectiveness in analysing spatial features from remote sensing data. Long Short-Term Memory (LSTM) networks perform well for time-series data, achieving an RMSE of 0.279 and an accuracy of 96.1%. Hybrid models further improve performance, with CNN-XGBoost emerging as the best technique, achieving an accuracy of 97.3% and the lowest MSE (0.063). This demonstrates the advantage of integrating ML and DL to enhance generalisation and robustness in crop yield prediction. The findings suggest that hybrid approaches offer the

Table 4: Performance Comparison of ML and DL Models for Crop Yield Prediction

Model Type	Algorithm	MSE	RMSE	R^2 Score	MAE	Accuracy (%)	F1-Score
Machine Learning	Decision Trees (DT)	0.143	0.378	0.78	0.109	88.7	0.86
	Random Forest (RF)	0.117	0.342	0.83	0.095	91.4	0.89
	Support Vector Machines (SVM)	0.122	0.349	0.81	0.101	89.8	0.88
	XGBoost	0.095	0.308	0.87	0.078	94.3	0.92
	k-Nearest Neighbours (KNN)	0.186	0.431	0.72	0.133	85.6	0.82
Deep Learning	Artificial Neural Networks	0.089	0.298	0.89	0.071	95.2	0.94
	Convolutional Neural Networks	0.074	0.272	0.91	0.063	96.8	0.96
	Long Short-Term Memory	0.078	0.279	0.90	0.067	96.1	0.95
	Recurrent Neural Networks	0.082	0.286	0.89	0.069	95.5	0.94
	Graph Neural Networks	0.077	0.278	0.91	0.065	96.4	0.96
Hybrid Models	CNN-XGBoost	0.063	0.251	0.93	0.056	97.3	0.97
	CNN-DNN	0.071	0.266	0.87	0.059	96.5	0.96
	CNN-LSTM	0.068	0.259	0.91	0.061	96.9	0.97

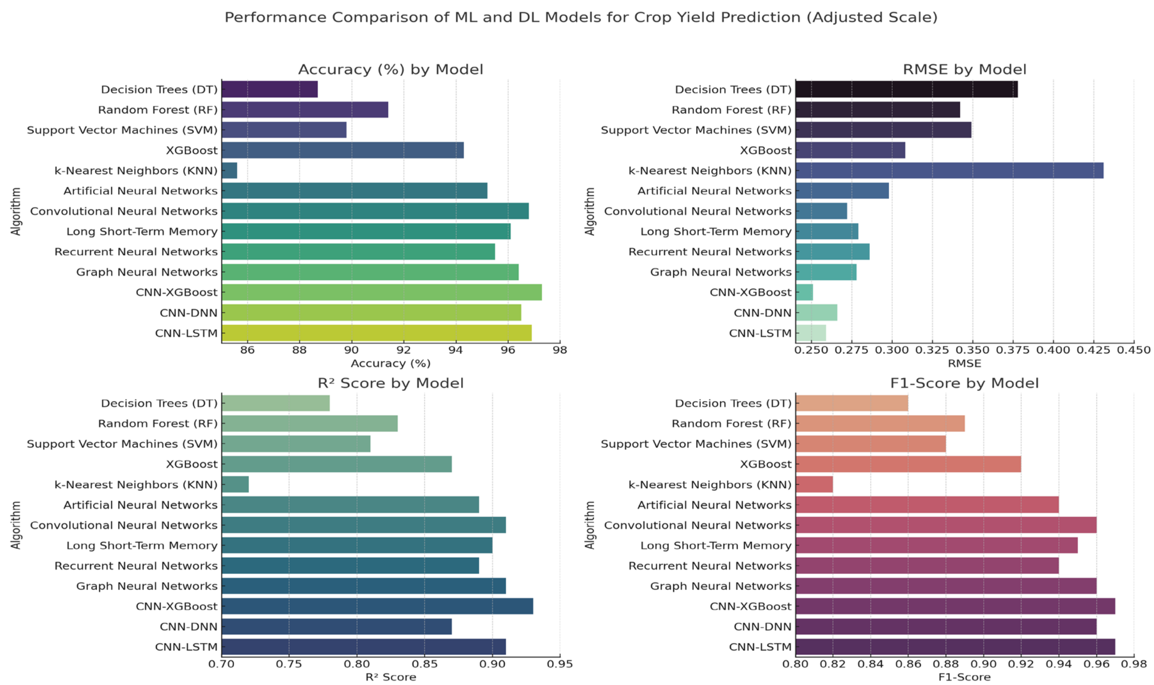

Fig. 2: Comparison of performance metrics for various ML and DL models used in crop yield prediction

best balance between accuracy, interpretability, and computational efficiency.

Conclusion

This study presents a comprehensive review of ML and DL techniques for crop yield prediction, covering key aspects of the application. A detailed evaluation of predictive features emphasises the importance of soil parameters (NPK, pH), climatic factors (temperature, rainfall, humidity), and remote sensing data (satellite imagery, UAV-based monitoring) in crop yield forecasting. The study further examines widely used datasets, including Kaggle, NASA MODIS, and FAO, along with performance metrics such as MSE, RMSE, R^2, and classification-based measures. Experimental methodologies, including cross-validation, model benchmarking, and hyperparameter tuning, ensure the robustness of predictive models. The analysis of ML and DL models highlights that XGBoost is the most effective ML model, while CNN outperforms other DL models. Hybrid models such as CNN-XGBoost achieve the highest predictive accuracy, demonstrating the advantage of integrating ML and DL approaches. Future work should focus on integrating real-time IoT data, optimising ensemble models, and enhancing explainable AI for improved agricultural decision-making.

References

1. Mahmud, T., Datta, N., Chakma, R., Das, U. K., Aziz, M. T., Islam, M., Salimullah, A. H. M., Hossain, M. S., & Andersson, K. (2024). An Approach for Crop Prediction in Agriculture: Integrating Genetic Algorithms and Machine Learning. IEEE Access, 12, 173583-173598. doi:10.1109/ACCESS.2024.3478739
2. Islam, M. R., Oliullah, K., Kabir, M. M., Alom, M., & Mridha, M. F. (2023). Machine Learning Enabled IoT System for Soil Nutrients Monitoring and Crop Recommendation. Journal of Agriculture and Food Research, 14, 100880. doi:10.1016/j.jafr.2023.100880
3. Vaza, R. N., Parmar, A. B., Kadham, M. J., Abedah, M., Abdullah, I., & Velu, C. M. (2024). Machine Learning Algorithms for Predicting Crop Health Using IoT-Generated Agriculture Data. Educational Administration: Theory and Practice, 30(4), 8466-8473. doi:10.53555/kuey.v30i4.2745
4. Dey, B., Ferdous, J., & Ahmed, R. (2024). Machine Learning Based Recommendation of Agricultural and Horticultural Crop Farming in India Under the Regime of NPK, Soil pH, and Three Climatic Variables. Heliyon, 10, e25112. doi:10.1016/j.heliyon.2024.e25112
5. Yadav, S., Rajendaran, M., Nagpal, A., Karthik, A., & Geetha, B. T. (2024). Exploring the Potential of Machine Learning in Enhancing Agricultural Practices and Food Production. 13th IEEE International Conference on Communication Systems and Network Technologies (CSNT-2024), 663-665. doi:10.1109/CSNT.2024.107

6. Raj, G. B., Boddepalli, E., Veena, C. H., Manjunatha, A. S., & Dhanraj, J. A. (2024). Predictive Modelling of Crop Yield in Precision Agriculture Using Machine Learning Techniques. International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI-2024). doi:10.1109/ACCAI61061.2024.10602450
7. Burdett, H. (2020). Statistical and Machine Learning Methods for Crop Yield Prediction in the Context of Precision Agriculture. Master's Thesis, Ryerson University, Toronto, Ontario.
8. Elbasi, E., Zaki, C., Topcu, A. E., Abdelbaki, W., Zreikat, A. I., Cina, E., Shdefat, A., & Saker, L. (2023). Crop Prediction Model Using Machine Learning Algorithms. Applied Sciences, 13, 9288. doi:10.3390/app13169288
9. Malashin, I., Tynchenko, V., Gantimurov, A., Nelyub, V., Borodulin, A., & Tynchenko, Y. (2024). Predicting Sustainable Crop Yields: Deep Learning and Explainable AI Tools. Sustainability, 16, 9437. doi:10.3390/su16219437
10. van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop Yield Prediction Using Machine Learning: A Systematic Literature Review. Computers and Electronics in Agriculture, 177, 105709. doi:10.1016/j.compag.2020.105709
11. Oikonomidis, A., Catal, C., & Kassahun, A. (2022). Hybrid Deep Learning-Based Models for Crop Yield Prediction. Applied Artificial Intelligence, 36(1), 2031822. doi:10.1080/08839514.2022.2031823
12. Elavarasan, D., & Durairaj Vincent, P. M. (2020). Crop Yield Prediction Using Deep Reinforcement Learning Model for Sustainable Agrarian Applications. IEEE Access, 8, 86886-86900. doi:10.1109/ACCESS.2020.2992480
13. Reddy, D. J., & Kumar, M. R. (2023). Crop Yield Prediction Using Machine Learning Algorithm. Journal of Computer Science and Engineering, 15(2), 130-145.
14. Clingaryan, A., Sukkarieh, S., & Whelan, B. (2018). Machine Learning Approaches for Crop Yield Prediction and Nitrogen Status Estimation in Precision Agriculture: A Review. Computers and Electronics in Agriculture, 151, 61-69. doi:10.1016/j.compag.2018.05.012
15. Jabed, M. A., Murad, M. A. A., & Others. (2024). Crop Yield Prediction in Agriculture: A Comprehensive Review of Machine Learning and Deep Learning Approaches, with Insights for Future Research and Sustainability. Heliyon, 10, e40836. doi:10.1016/j.heliyon.2024.e40836
16. Rashid, M., Bari, B. S., Yusup, Y., Kamaruddin, M. A., & Khan, N. (2021). A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches with Special Emphasis on Palm Oil Yield Prediction. IEEE Access, 9, 63406-63425. doi:10.1109/ACCESS.2021.3075159
17. El-Kenawy, E. M., Alhussan, A. A., Khodadadi, N., Mirjalili, S., & Eid, M. M. (2024). Predicting Potato Crop Yield with Machine Learning and Deep Learning for Sustainable Agriculture. Potato Research. doi:10.1007/s11540-024-09753-w
18. Prathap, C., Sivaranjani, S., & Sathya, M. (2024). ML-Based Yield Prediction in Smart Agriculture Systems Using IoT. 5th International Conference on Innovative Trends in Information Technology (ICITITT-2024). doi:10.1109/ICITITT61487.2024.10580172
19. K, Priyanka., & Umesh, D. R. (2024). Deep Learning Approaches for Enhanced Precision in Crop Yield Prediction. 5th International Conference for Emerging Technology (INCET-2024).
20. Shashikumar, B., Shaabin, M., Muruganandam, N. S., & Somashekhar, B. M. (2024). Yield Prediction and Nutrient Enhancement Using Deep Learning Techniques. 4th Asian Conference on Innovation in Technology (ASIANCON-2024). doi:10.1109/ASIANCON62057.2024.10838196