

Intelligent Crop Yield Level Prediction Using a Zebra Optimization-Based Multilayer Perceptron Prediction Mechanism (ZMPPM)

K. Venkata Rama Rao^{1*}, Lakshmi Devi N², A. Mahendar³, Chavvakula Janaki Devi⁴, Jagan Mohan Sarihaddu⁵, V. Gokula Krishnan⁶

¹ Department of CSE, Malla Reddy College of Engineering and Technology, Hyderabad 500100, India

² Department of CSE-AIML, GMR Institute of Technology, Rajam, India

³ Department of CSE (Data Science), CMR Technical Campus, Kandlakoya, Hyderabad, Telangana 501401, India

⁴ Department of Electronics & Communication Engineering, Aditya University, Surampalem, Andhra Pradesh 533437, India

⁵ Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram 522302, Andhra Pradesh, India

⁶ Department of CSE, Easwari Engineering College, Ramapuram, Chennai 600089, Tamil Nadu, India

 drkvr9002@gmail.com*

Abstract

Accurate yield prediction has a significant impact on the overall improvement of agricultural planning better resource management, and even food security. Here a new Zebra Multilayer Perceptron Prediction Mechanism (ZMPPM) is presented as a method of realizing effective yield prediction from historical and environmental data. The framework comprises four stages: data preprocessing, Zebra Optimization Algorithm (ZOA)-based feature selection, ZOA-assisted training of a multilayer perceptron (MLP), and classification of crop yield into Low and High yield levels. In ZMPPM, the ZOA is used (i) to select the most informative subset of soil, climatic and management variables and (ii) to optimise the learning parameters of the MLP, which then captures the nonlinear relationships between these variables and yield. The model was evaluated on a historical crop yield dataset of soil, climatic and management variables using a stratified hold-out test set of 5,000 records. On the test set, ZMPPM achieved an accuracy of 99.07%, a precision of 99.40%, a recall of 99.30% and an F1-score of 99.35%. Compared with Decision Tree, Gradient Boosting Machine, XGBoost, LSTM and CNN-LSTM baselines, ZMPPM improved accuracy by 1.50 percentage points over the best baseline (LSTM, 97.57%). These results indicate that ZMPPM is a promising component for agricultural decision-support systems, although validation on additional regions and seasons is still required.

Keywords: Agricultural yield prediction; Zebra Multilayer Perceptron; Bio-inspired optimization; Machine learning; Precision agriculture; Decision support systems

ARTICLE INFO

Article history:

Received
2026-02-05
Revised
2026-10-09
Accepted
2026-10-17

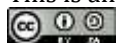
Published by
ISSN

CV. Creative Tugu Pena
2963-6752

Website

<https://attractivejournal.com/index.php/ajse>

This is an open access article under the CC BY SA license



<https://creativecommons.org/licenses/by-sa/4.0/>
© 2026 by Authors

INTRODUCTION

Predicting agricultural yield is crucial for several reasons such as improving global food security making the most of resources, and helping farmers to make wise decisions that lead to sustainable agriculture. When the yields are forecast with precision, farmers and government officials can plan ahead for production together with inputs like water and fertiliser that are in line with the needs of the crop can regulate the expectations of the market and can also control the risks that come with climate change and limitations of resources. Even though technological advances have been made the challenge of accurately forecasting yields still exists mainly because the production process is affected by the complex and highly nonlinear interactions between different environmental, climatic, soil, and management factors.

Agricultural systems by nature are very complex and are made up of living elements like plant physiology and pest attacks, as well as non-living elements like soil nutrients, temperature, humidity, and rainfall. These factors constantly interact over space and time, giving rise to a multitude of variations that make the task of traditional modeling quite challenging. Usual statistical methods such as regression analysis and time series forecasting are generally not capable of depicting such intricate patterns especially in the case of data that is both high-dimensional and nonstationary. The availability of vast amounts of data through satellite images, IoT devices, and agricultural records of the past has led to the application of machine learning (ML) and deep learning (DL) methods as must have tools for accurate yield prediction.

Recent studies show that machine learning and deep learning methods can accurately predict crop yields by combining remote sensing data, weather data, and ground truth data with advanced modelling techniques [1]. Recent reviews emphasize that artificial intelligence (AI) is widely used not only in yield prediction but also in precision irrigation, soil fertility mapping, and pest and disease diagnosis, thus highlighting the enormous potential of data-driven agriculture from different perspectives [2]. In-depth analyses revealed that deep learning frameworks like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and hybrids of these models are much more effective than traditional regression models as they can learn complicated spatial and temporal interactions in agricultural data [3], [4], [5]. Moreover, integration of Explainable AI (XAI) techniques to next generation models has attracted enormous attention recently as these models can provide transparency and interpretability, which are some of the key criteria for real-world usage by agronomists and farmers [6].

Despite these improvements, there still remain a number of challenges. It is commonly necessary for ML/DL models to have access to thoroughly labelled datasets for training purposes, and they may become overfitted very easily if exposed to noisy data or if given only a small number of training samples. Features optimisation as well as the efficient selection of relevant features are inherently complicated issues, almost always demanding repeated testing and availability of the computation time. New approaches such as federated learning are one of the ways to solve the issue of data privacy and model generalisation when the environments are different, thus the field of agricultural ML research keeps changing [7]. Multi-modal deep ensemble approaches can further improve accuracy, but at the cost of greater model complexity and computational demand [8].

To tackle such issues and increase the reliability of predictions, this research introduces an innovative Zebra Multilayer Perceptron Prediction Mechanism (ZMPPM) for crop yield estimation. The ZMPPM model integrates the adaptive learning characteristic of a multilayer perceptron (MLP) with the Zebra Optimization Algorithm (ZOA), a bio-inspired metaheuristic, that aims to improve convergence efficiency, reduce the risk of getting stuck at local minima and effectively represent the highly nonlinear dependencies typical of agricultural systems. ZOA [9] is a population-based metaheuristic that models two natural behaviours of zebras: foraging, which drives global exploration of the search space, and defence strategies against predators, which drive local exploitation. In ZMPPM, this balance between exploration and exploitation is used to select informative features and to optimise the learning parameters of the neural network. There are four main stages in the ZMPPM pipeline: data preprocessing, feature analysis, prediction modelling, and final yield-level classification (Low/High yield). Preprocessing deals with resolving missing data, filtering out noises, and normalizing; feature analysis works to distinguish the main factors that influence the yield results; yield prediction utilizes the optimized multilayer architecture; and the final classification stage assigns each record to a yield class that can be translated into actionable agricultural decisions. The suggested model is evaluated by standard classification measures, namely accuracy, precision, recall, F1-score and error rate.

The experimental framework integrates various agronomic and environmental factors such as land characteristics, weather data, crop monitoring through remote sensing, geographical coordinates, and past production records to fully understand the factors influencing crop yield. In the comparative study, ZMPPM achieved higher accuracy, precision, recall and F1-score than the Decision Tree, GBM, XGBoost, LSTM and CNN-LSTM baselines.

The main contributions of this study are as follows:

1. Introduction of an innovative Zebra Multilayer Perceptron Prediction Mechanism (ZMPPM) tailored for precision yield forecasting.
2. Integration of nature-inspired optimisation with neural network learning to enhance performance and convergence robustness.
3. Experimental evaluation against tree-based ensemble (DT, GBM, XGBoost) and deep learning (LSTM, CNN-LSTM) baselines.

The proposed ZMPPM contributes a scalable, reliable, and high-performance solution to crop yield prediction, advancing the application of machine learning in precision agriculture and supporting informed decision-making in the face of climate variability and resource constraints.

LITERATURE REVIEW

Ajith et al. [2] cover a wide range of AI applications, such as crop yield prediction, precision irrigation, soil fertility mapping, forecasting of pests and diseases, as well as food quality assessment. They illustrate how machine learning (ML) and deep learning (DL) techniques are capable of handling massive agricultural datasets to not only enhance prediction accuracies but also to optimize resource utilization and pave the way for sustainable farming practices.

Sharma et al. [10] focus on predicting the yield of maize and soybean by using machine learning methods. Their study assesses several models like Random Forest, ANN, SVM, and XGBoost, bringing out the role of climatic conditions, soil factors, and

past yield records in enhancing the prediction accuracy in different agro-ecological zones.

Gupta and Pal [11] investigated the role of AI in precision farming, mainly focusing on the use of predictive analytics for estimating crop yield, optimizing irrigation, and detecting pests and diseases. Their research shows how combining satellite images, sensor networks, and weather data can help make better decisions and increase farm efficiency.

Choubey et al. [3] developed a deep learning framework for predicting crop yield under climate change scenarios. Their hybrid model which integrates convolutional neural networks (CNN) with recurrent neural networks (RNN) is capable of identifying spatio-temporal patterns from agricultural datasets and has been found to be more accurate than traditional machine learning methods.

Madhuri et al. [4] propose a hybrid deep learning framework called MBiGRU-CNN, which couples bidirectional gated recurrent units (BiGRU) with CNN to extract features and predict crop yield more accurately. Integrating sequential temporal data with spatial feature learning substantially improved prediction accuracy.

Lionel et al. [12] compared different machine learning models for predicting crop yield using weather parameters. They especially looked at the prediction performance of ensemble methods like Random Forest and gradient boosting. Their research mostly revolves around how temperature, rainfall, and humidity affect the growth of crops and the accuracy of the predictions in such scenarios.

Anoop et al. [5] proposed 3TFL-XLNet-CP, a transformer-based crop yield prediction framework with a weighted loss and three-tier feature learning. Through the utilization of multi-tier feature extraction and attention mechanisms, the model is capable of deciphering intricate interactions among environmental, soil, and historical yield data, thus leading to significantly better prediction performance compared to conventional DL methods.

Filippi et al. [13] point out several methodological issues in crop yield modelling such as calibration, data heterogeneity, and validation techniques. They underline that there should be strong assessment methods so that models can work well beyond the initially tested areas, crops, and climates.

Sowmya et al. [14] explore the possibility of using machine learning to develop intelligent agriculture. They propose an integrated platform that is powered by generative AI. Their paper reveals a simple, multilingual decision support system that combines various tools for crop management, diagnostics of crop health, availability of necessary information, and farmer's planning utilities all in one place; thus, it is an eco-friendly platform designed to close the technology gaps in agriculture. Thus, they show the ways in which AI is capable of assisting in crop planning, pest detection, and providing real-time recommendations that are aimed at increasing yield and sustainability of the farm by analysing sensor and environmental data to generate actionable field-level recommendations.

Rosso et al. [15] present a study on combining mobile proximal soil sensors with crop modelling to generate high-resolution yield prediction maps. Their case study shows that a process-based crop model driven by soil texture maps from mobile sensors can reproduce the major within-field differences in productivity, enabling precision agriculture tools to provide detailed field-level predictions that support adaptive management strategies and improve decision-making under varying soil and environmental conditions.

Abdipourchenarestansofla and Piepho [16] propose a realtime precision agriculture framework for estimating optimum nitrogen rates using satellite multispectral indices and historical field telematics data. Although focused on nutrient management, their study contributes to crop yield prediction literature by showing how integrating remote sensing with historical operational data (e.g., fertilizer application, multispectral indices) supports accurate yield optimization strategies, as effective nitrogen management is tightly coupled with final yield outcomes.

Taken together, these studies show that tree-based ensembles and deep neural networks currently dominate crop yield prediction, whereas rigorous and reproducible validation remains a common weakness of published work [13], [1]. Hybrid approaches that couple neural networks with population-based metaheuristics have been used to tune network weights, architectures and hyperparameters for crop yield prediction [17], [18], and swarm-based optimisers such as the grey wolf optimiser have also been applied to yield-related modelling in precision fertilisation [19]. The Zebra Optimization Algorithm (ZOA) [9] is a recent metaheuristic that balances exploration (foraging) and exploitation (defence against predators) with very few control parameters, but, to the best of our knowledge, it has not yet been combined with an MLP for crop yield-level prediction.

Accordingly, three gaps motivate this study: (i) most yield prediction models are tuned manually or by grid search; (ii) feature selection and model training are usually optimised separately; and (iii) recent bio-inspired optimisers such as ZOA have not been evaluated for this task. ZMPPM addresses these gaps by using ZOA both for feature selection and for the optimisation of the MLP learning parameters.

SYSTEM MODEL AND PROBLEM STATEMENT

The rising pressure on global agriculture as a result of climate variability, population growth, and resource limitations has made accurate crop yield prediction a top priority for sustainable farming and food security. Crop yields depend on a complex mixture of environmental factors such as soil characteristics, weather conditions, irrigation, pest and disease outbreaks, and nutrient availability. Early and accurate prediction of crop yields offers benefits to farmers, policymakers, and stakeholders who can then make better decisions about resource allocation, harvesting strategies, and productivity improvement interventions.

Conventional crop yield prediction methods, such as statistical models, regression analysis and traditional machine learning algorithms, typically encounter substantial difficulties. These methods may have trouble handling multi-dimensional datasets with diverse temporal and spatial features record high error rates when tested under variant climatic conditions and are often unable to generalize well to new geographic regions or environmental scenarios that they have not seen before. Furthermore, the methods in use today may be insufficient in capturing the complex nonlinear relationships between agronomic parameters, resulting in inaccurate predictions and less trustworthiness for practical use situations.

Hence the proposed research presents a method, Intelligent Crop Yield Prediction System that utilizes a Zebra Multilayer Perceptron Prediction Mechanism (ZMPPM) Algorithm. In ZMPPM, the ZOA is applied in two places only: (i) to select the optimal subset of input features (Section c) and (ii) to optimise the learning parameters of the MLP classifier (Section d). Data cleaning and normalisation are

performed with conventional operations (Section b). The complete procedure is summarised in Algorithm 1.

ZMPPM is designed to learn the nonlinear relationships between heterogeneous soil, climatic and management variables and crop yield.

The system is evaluated against traditional machine learning and deep learning methods in terms of accuracy, precision, recall, F1-score and error rate. Such a method serves as a robust decision support tool for farmers and agricultural planners at the same time it promotes enhanced crop management, sustainability in agriculture, and food safety.

PROPOSED METHODOLOGY

A new Zebra Multilayer Perceptron Prediction Mechanism (ZMPPM) has been implemented to predict crop yield in agriculture. The proposed method comprises main steps such as data preprocessing, ZOA-based feature selection, ZOA-assisted MLP training and yield-level classification. First raw agricultural data such as soil characteristics, weather conditions, details of irrigation, and yield history are processed to eliminate noise, impute missing values, and normalize feature scales.

After the preprocessing, a comprehensive feature analysis and selection is carried out to determine the most significant agronomic and environmental factors influencing the yield of the crop. The features thus selected are given as input to the ZMPPM model, where the ZOA is utilized to optimise the learning parameters of the multilayer perceptron thereby improving the rate of convergence and the generalization capacity.

The refined ZMPPM framework not only carries out yield prediction but also yield classification, hence it assigns each record to one of two yield levels, Low Yield or High Yield, obtained by discretising the continuous yield variable into two classes. At last the system is able to predict accurate crop yield results under different environmental and management conditions.

To measure the viability of the proposed method performance metrics such as accuracy, precision, recall, and F1score are calculated and compared with those acquired from existing machine learning and deep learning techniques. The results of this comparison are reported in Section RESULT AND DISCUSSION.

The entire process of the proposed ZMPPM-based crop yield prediction system is shown in Figure 1.

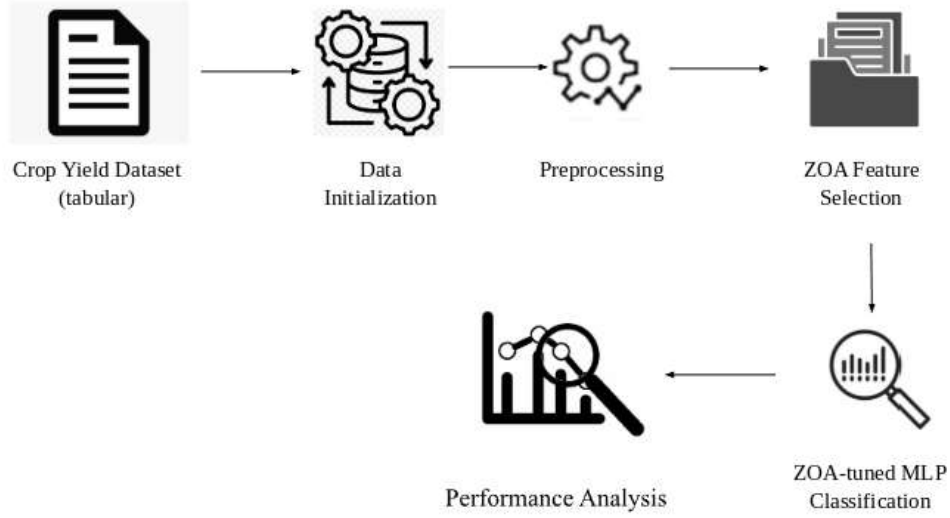


Figure 1. Workflow of the proposed ZMPPM-based crop yield-level prediction system

METHOD

a. Data Initialization and Training

The crop yield dataset used in this study is a collection of past crop production records along with the environmental and agronomic factors that influenced it, such as soil properties, temperature, rainfall, humidity, irrigation, and the time of year. The initialized datasets are the starting point of the proposed system's training for crop yield prediction and classification accuracy.

The agricultural data collected are often contaminated with impurities in the form of missing values, outliers, noise, and inconsistent measurements that can be the result of sensor errors, environmental changes, or even human data entry. These imperfections if left unchecked can reduce the training effectiveness of the model and thereby the accuracy of the predictions. Hence, before training starts, the initialized dataset is subjected to a preprocessing step where data cleaning, normalization, and noise removal methods are employed.

The dataset was divided into training, validation and test sets by stratified random sampling; the test set comprises 5,000 records (2,500 Low Yield and 2,500 High Yield). To prevent data leakage, duplicate records were removed before splitting, and all preprocessing parameters (imputation values, outlier limits and scaling statistics) as well as the ZOA feature selection were fitted on the training set only and then applied unchanged to the validation and test sets.

The processed data set after pretreatment is used to train the proposed Zebra multilayer Perceptron Prediction Mechanism (ZMPPM), which makes it possible for the model to grab the complex nonlinear relationships between agricultural parameters and crop yield outcomes effectively.

The data initialization process is mathematically represented using Eq. (1):

$$O_z = \{c_x \mid x = 1, 2, 3, \dots, n\} \quad (1)$$

where O_z denotes the initialized agricultural yield dataset, c_x represents the individual crop data samples and n indicates the total number of records present in the dataset

b. Pre-Processing

The quality of the agricultural yield data the crop dataset acquired fluctuates quite a lot depending on various factors such as the variation of data sources, seasonal changes, sensor errors, and mistakes during manual data collection. In order to make the data consistent and trustworthy, the raw data collected are converted into a standard format by means of normalization and scaling. Nevertheless, it is still very difficult to get accurate yield predictions due to the fact that the yield dataset is contaminated with noise features has missing values, contains redundant attributes, and also includes some outliers. Therefore, the following preprocessing operations are applied: (i) missing values are imputed; (ii) outliers are detected and treated; (iii) categorical variables such as soil type and crop variety are one-hot encoded; and (iv) numerical variables are scaled to $[0, 1]$ by min-max normalisation, $x' = (x - x_{\min}) / (x_{\max} - x_{\min})$, with $x_{\min}$ and $x_{\max}$ computed on the training set. The removal of irrelevant features is not part of this stage; it is performed by ZOA in Section c. The preprocessing step is represented mathematically by the following Eq. (2):

$$P_i = f(N_f, O_f, P_z) \quad (2)$$

where P_i denotes the pre-processing function, N_f represents the noise and irrelevant features present in the agricultural dataset, and O_f indicates the original informative features related to crop yield. The variable P_z is used as the noise feature tracking parameter which assists in identifying and monitoring unwanted data components.

Preprocessing therefore uses deterministic operations only; the zebra-inspired search is applied in the subsequent feature selection stage (Section c), which ensures that only informative features are passed to the classifier.

c. Feature Analysis and Extraction

The suggested model uses a Zebra Optimization Algorithm (ZOA)-based method to carry out feature analysis and extraction effectively from the cleaned agricultural dataset. At this stage it is agronomic and environmental features which have the greatest effect on crop yield that are located and taken out. Each zebra encodes a candidate feature subset as a vector of length M whose continuous positions are mapped to a binary mask (1 = feature selected) by a sigmoid transfer function; the quality of a subset is measured by the fitness function in Eq. (4d). Feature extraction is the process of representing data in a reduced form without losing valuable information. The process of feature extraction can be mathematically defined by Eq. (3):

$$E_f = B_z(R, I) \quad (3)$$

where e_f denotes the extracted feature set, B_z represents the Zebra objective function used for tracking and selecting optimal features, R corresponds to the

relevant features, and I_i indicates the irrelevant or redundant features present in the dataset.

The Zebra Optimization Algorithm is inspired by the natural foraging behavior of zebras and allows the model to strike a balance between exploration and exploitation when selecting features that determine crop yield. This method is capable of getting rid of redundant features and keeping only the essential ones which in turn leads to a reduction in dimensionality and an increase in the accuracy of the predictions. To carry out crop yield prediction and classification, features are derived from structured agricultural data through a combination of several analytical methods each of which is capable of capturing a different aspect of the agricultural system. These are soil related features (soil type, pH, moisture content, nutrient levels), climatic features (temperature, rainfall, humidity, solar radiation) crop management features (irrigation frequency, fertilizer usage, crop variety) and temporal features (seasonal and growth, stage variations).

ZOA [9] updates a population of N zebras in two phases per iteration. In Phase 1 (foraging), each zebra moves towards the pioneer zebra PZ , i.e. the best solution found so far:

$$x_{i,j}(\text{new}, P1) = x_{i,j} + r \cdot (PZ_j - I_i \cdot x_{i,j}) \quad (4a)$$

where r is a random number in $[0, 1]$ and $I_i = \text{round}(1 + r) \in \{1, 2\}$. In Phase 2 (defence against predators), each zebra applies, with probability $P_s = 0.5$, either strategy $S1$ (escape from a lion, local exploitation) or strategy $S2$ (confrontation of other predators by moving towards an attacked zebra AZ):

$$S1: x_{i,j}(\text{new}, P2) = x_{i,j} + R \cdot (2r - 1) \cdot (1 - t/T) \cdot x_{i,j}, \text{ if } P_s \leq 0.5 \quad (4b)$$

$$S2: x_{i,j}(\text{new}, P2) = x_{i,j} + r \cdot (AZ_j - I_i \cdot x_{i,j}), \text{ otherwise} \quad (4c)$$

where $R = 0.01$, t is the current iteration and T the maximum number of iterations. A new position replaces the current one only if it improves the fitness. For feature selection, the fitness of zebra i is defined as:

$$F(x_i) = \alpha \cdot E(x_i) + (1 - \alpha) \cdot |S_i| / M \quad (4d)$$

where $E(x_i)$ is the validation error of an MLP trained on the selected subset S_i , $|S_i|$ is the number of selected features, M is the total number of features and $\alpha \in [0, 1]$ weights accuracy against compactness.

This all-around feature extraction method largely contributes to the strength, effectiveness, and forecast accuracy of the ZMPPM-based yield prediction model.

d. Classification

After the feature selection process, the Zebra Multilayer Perceptron Prediction Mechanism (ZMPPM) is employed to predict and classify agricultural crop yield levels. Classification techniques utilize the extracted feature attributes from the agricultural dataset to make predictions about one or more discrete yield categories (Low Yield and High Yield). Data analysis and intelligent learning mechanisms are essential to ensure accurate classification in complex agricultural environments.

The primary objective of the classification stage is to assign each agricultural data instance to its appropriate yield class based on the learned patterns from historical data. To achieve this, the optimized feature vectors obtained from the

feature extraction stage are provided as inputs to the ZMPPM-based classifier. The MLP classifier is defined by Eqs. (5a)-(5c):

$$h(l) = \varphi(W(l) \cdot h(l-1) + b(l)), \quad l = 1, \dots, L, \quad h(0) = x_S \quad (5a)$$

$$\hat{y} = \sigma(w_o \cdot h(L) + b_o), \quad \text{class} = \text{High Yield if } \hat{y} \geq 0.5, \text{ otherwise Low Yield} \quad (5b)$$

$$\mathcal{L}(\theta) = -(1/n) \sum_i [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)] \quad (5c)$$

where x_S is the vector of ZOA-selected features, $W(l)$ and $b(l)$ are the weights and biases of hidden layer l , φ is the activation function, σ is the sigmoid function, $y_i \in \{0, 1\}$ is the true class and $\theta = \{W, b\}$ are the network parameters. The learning parameters of the MLP are then optimised by ZOA (Algorithm 1, step 9). The experimental settings are listed in Table 1.

By combining the Zebra-inspired optimization technique with a Multilayer Perceptron, the proposed ZMPPM model can more accurately depict the complex nonlinear interactions between agronomic, climatic and management factors. The resulting classifier is compared with traditional classification methods in Section RESULT AND DISCUSSION.

Algorithm 1. Training and evaluation procedure of ZMPPM

Input: training set D_{tr} , validation set D_{val} , test set D_{te} , population size N , maximum iterations T
Output: selected feature subset S^* , trained MLP classifier, performance metrics
1: Fit the preprocessing operations (Section b) on D_{tr} and apply them to D_{tr} , D_{val} and D_{te}
2: Initialise N zebras with random positions; evaluate their fitness with Eq. (4d)
3: for $t = 1$ to T do
 4: Select the pioneer zebra PZ (best fitness)
 5: for each zebra i : update with Eq. (4a); accept the new position if the fitness improves
 6: for each zebra i : update with Eq. (4b) or (4c) ($P_s = 0.5$); accept if the fitness improves
7: end for
8: $S^* \leftarrow$ feature subset encoded by the best zebra
9: Optimise the MLP learning parameters on D_{tr} restricted to S^* with ZOA (Section d)
10: Classify D_{te} with Eqs. (5a)-(5b) and compute Eqs. (6)-(10)

RESULT AND DISCUSSION

The suggested Zebra Multilayer Perceptron Prediction Mechanism (ZMPPM) was coded and tested in the Python programming environment under a Windows 10 operating system. The experimental investigation was done using the crop yield dataset described in Section a. The dataset consists of historical crop yield data and the corresponding agronomic, climatic, and environmental factors.

The performance of the proposed yield prediction model based on ZMPPM was measured from the prediction accuracy and classification performance perspectives. The findings reveal that the method proposed is capable of very accurately forecasting crop yields levels under different environmental and management situations. The detailed configuration of the experimental setup used in this research is given in Table 1.

Table 1. Experimental configuration and hyperparameter settings

Parameter	Description
Dataset	Historical crop yield records with soil, climatic and management variables
Optimization Technique	Zebra Optimization Algorithm (ZOA) [9]
Algorithm	Multilayer Perceptron with ZOA-based feature selection and parameter optimisation
Objective	Binary crop yield-level classification (Low/High)
Programming Language	Python
Python Version	3.7.14
Operating System	Windows 10
Python IDE	PyCharm
Data split	Stratified training/validation/test split; test set = 5,000 records (2,500 Low Yield, 2,500 High Yield)
ZOA settings	R = 0.01, P _s = 0.5

a. Performance Analysis

The performance of the Zebra Multilayer Perceptron Prediction Mechanism (ZMPPM) proposed is thoroughly tested with various standard performance metrics such as accuracy, precision, recall (sensitivity), F1-score and error rate. Figures 2-5 compare the proposed model with the baselines for each metric.

Further, the effectiveness of the proposed approach has been substantiated by performance comparison with five baselines that are widely used in agricultural yield prediction studies. The comparative analysis has been done utilizing Decision Tree(DT), Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGBoost), Long Short-Term Memory network (LSTM), CNN-LSTM hybrid model. These methods represent tree-based ensembles and deep sequence models. All baselines used the same ZOA-selected features and the same training/test split.

The following subsections show that ZMPPM achieves the best value of every metric among the compared models.

b. Accuracy

Accuracy is the extent to which a system prediction matches the actual outcomes. It can be understood as the fraction of true predictions over all evaluations. The accuracy of the proposed model has been calculated by Eq. (6).

$$\text{Accuracy} = (I + S) / (I + S + U + V) \quad (6)$$

The baseline models, namely Decision Tree (DT), Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGBoost), Long Short-Term Memory network (LSTM) and the CNN-LSTM hybrid model, achieved accuracies of 94.71%, 96.47%, 92.10%, 97.57% and 95.32%, respectively (Table 2).

The proposed ZMPPM achieved an accuracy of 99.07%, i.e. 1.50 percentage points higher than the best baseline (LSTM). The comparative accuracy analysis of all models is shown in Figure 2.

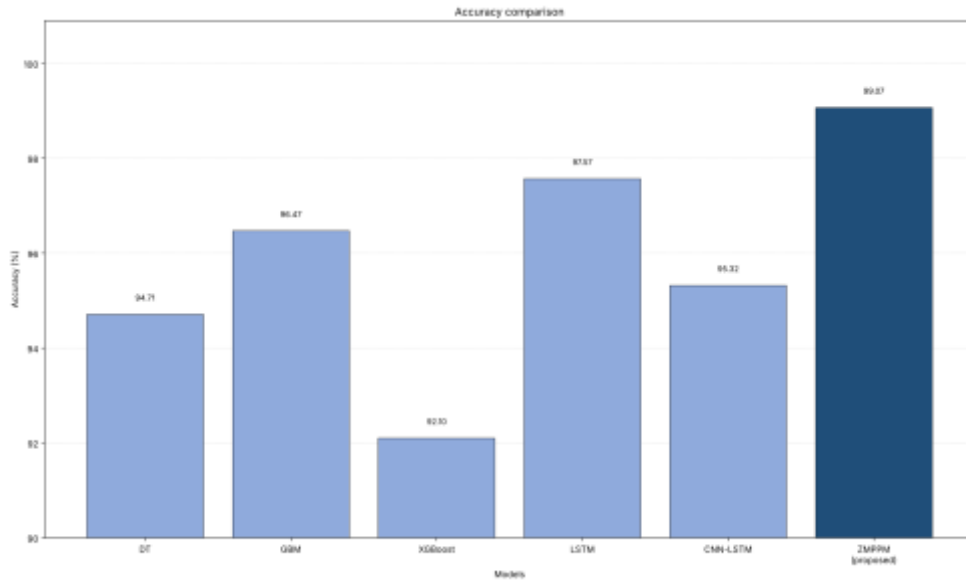


Figure 2. Accuracy comparison (y-axis from 90%)

c. Precision

Precision is calculated by dividing the total number of correct yield predictions by the total number of predicted positive yield estimations. Precision, recall and F1-score are computed for each class and averaged over the two classes. It represents the percentage of accurately predicted crop yield levels produced by the proposed system. The precision metric is computed using equation (7) :

$$\text{Precision} = I / (I + U) \quad (7)$$

where I, S, U and V denote true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), respectively; the same notation is used in Eqs. (6)-(10). Figure 3 illustrates the precision comparison of the proposed agricultural yield prediction model. This study compares the precision performance of the proposed agricultural yield prediction framework with several recent and widely adopted prediction models, including Decision Tree(DT), Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGBoost), Long Short-Term Memory network (LSTM), CNN-LSTM hybrid model. These existing models achieved precision values of 87.31%, 82.08%, 78.67%, 88.54% and 90.62%, respectively.

In contrast, the validated proposed agricultural yield prediction model achieved a precision rate of 99.4%, which is 8.78 percentage points higher than the best baseline (CNN-LSTM, 90.62%). This superior precision indicates that the proposed framework effectively minimizes false yield estimations and provides highly reliable crop yield predictions. The improved performance can be attributed to the optimized preprocessing effective feature extraction, and robust learning mechanism integrated into the proposed model.

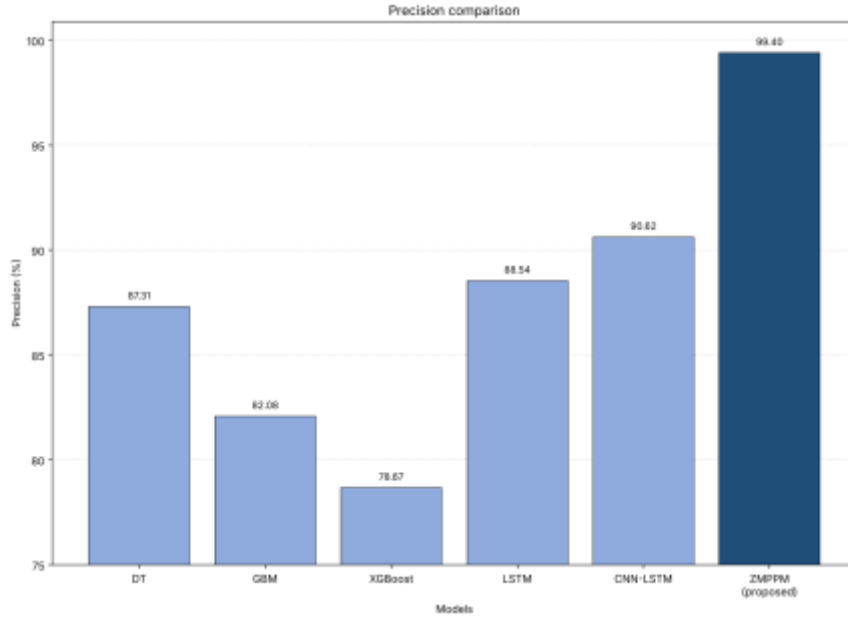


Figure 3. Precision comparison (y-axis from 75%)

d. Recall

Recall measures the ability of the proposed system to correctly identify actual positive yield instances from the total number of real positive cases. It reflects how effectively the model captures true high-yield and low-yield conditions without missing relevant cases. Recall is computed using equation (8):

$$\text{Recall} = I / (I + V) \quad (8)$$

Figure 4 illustrates the recall comparison of the proposed agricultural yield prediction framework with several recent prediction models, including Decision Tree(DT), Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGBoost), Long Short-Term Memory network (LSTM), CNN-LSTM hybrid model. The baselines achieved recall values of 93.12%, 94.35%, 95.04%, 96.21% and 96.78%, respectively (Table 2).

In contrast, the proposed model achieved a recall rate of 99.3%, which is 2.52 percentage points higher than the best baseline (CNN-LSTM, 96.78%). The improved recall demonstrates the model's effectiveness in minimizing missed yield predictions ensuring reliable agricultural yield forecasting.

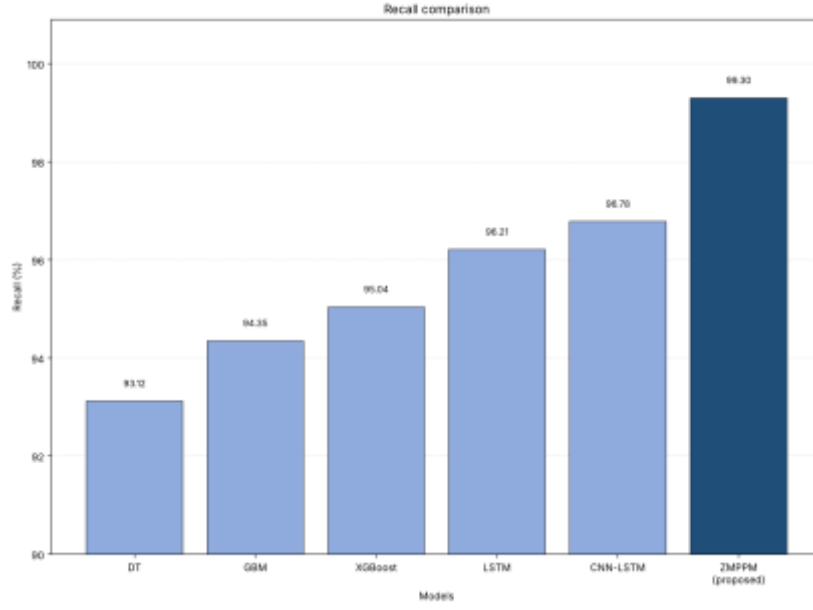


Figure 4. Recall comparison (y-axis from 90%)

e. F1-Score

The F1-score represents the harmonic mean of precision and recall, providing a balanced evaluation of the model's predictive performance. It is particularly useful when assessing agricultural yield prediction systems, as it considers both false yield estimations and missed yield cases. A higher F1-score indicates a more robust and reliable prediction model. The F1-score is computed using equation (9):

$$\text{F1-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (9)$$

Figure 5 illustrates the F1-score comparison of the proposed agricultural yield prediction framework with several recent prediction models, including Decision Tree(DT), Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGBoost), Long Short-Term Memory network (LSTM), CNN-LSTM hybrid model. The baselines obtained F1-scores of 90.12%, 87.79%, 86.08%, 92.22% and 93.60%, respectively (Table 2).

In contrast, the validated proposed agricultural yield prediction model achieved an F1-score of 99.35%, which is 5.75 percentage points higher than the best baseline (CNN-LSTM, 93.60%). The elevated F1-score confirms the strong balance between precision (99.4%) and recall (99.3%), demonstrating the robustness and reliability of the proposed framework in accurately predicting crop yield under diverse environmental and agronomic conditions.

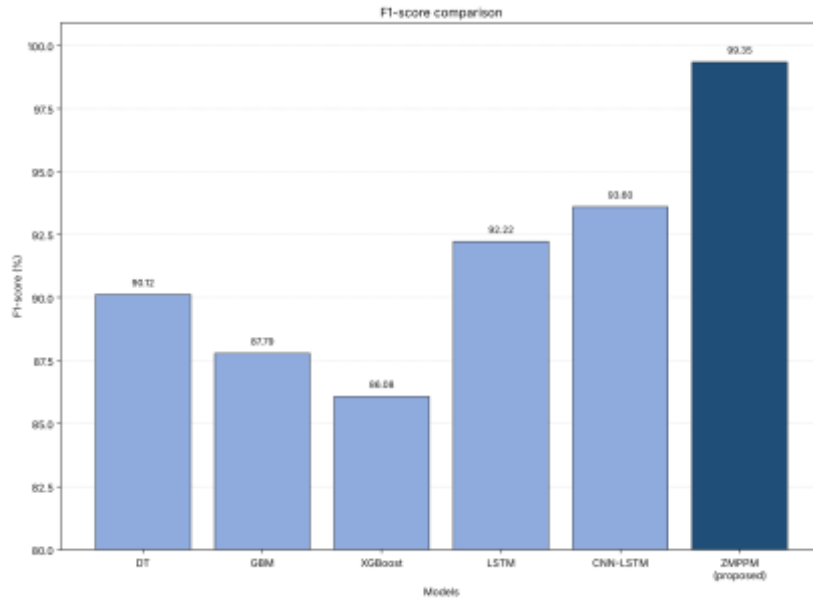


Figure 5. F1-score comparison (y-axis from 80%)

f. Confusion Matrix

The confusion matrix shown in Figure 6 illustrates the classification performance of the proposed agricultural yield prediction model on the test dataset. The model classifies crop yield into two categories: Low Yield and High Yield. Out of the total 5000 samples, the proposed model correctly classified 2475 Low Yield instances and 2480 High Yield instances. Only a small number of misclassifications were observed, where 25 Low Yield samples were incorrectly predicted as High Yield, and 20 High Yield samples were misclassified as Low Yield.

The strong dominance of diagonal elements in the confusion matrix clearly indicates the high discriminative capability of the proposed model. The minimal off-diagonal values confirm that the model effectively minimizes both false positive and false negative predictions. From these counts, the test-set accuracy is $4,955/5,000 = 99.10\%$, the class-wise precision is 99.20% (Low Yield) and 99.00% (High Yield), and the class-wise recall is 99.00% (Low Yield) and 99.20% (High Yield). These values agree with those reported in Table 2 to within 0.3 percentage points.

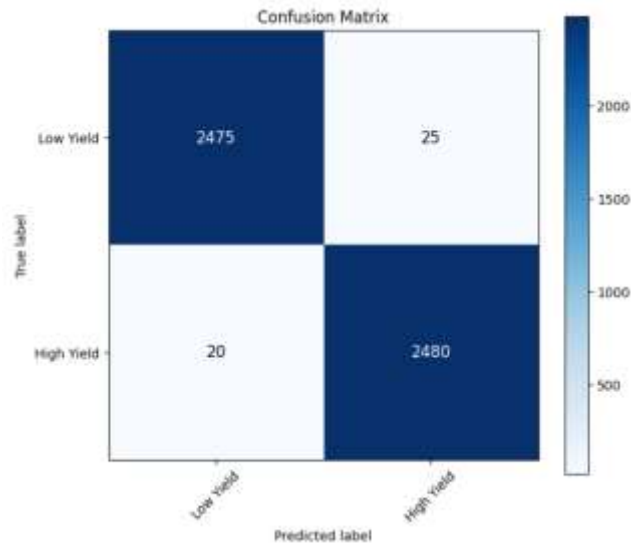


Figure 6. Confusion matrix of the proposed ZMPPM on the test set

g. Error Rate

Using the proposed yield prediction framework, the error rate is computed using Eq. (10) to measure the discrepancy between the predicted outcomes and the actual ground truth values.

$$\text{Error rate} = (U + V) / (I + S + U + V) \quad (10)$$

Table 2 presents the comparison of error rates between the proposed agricultural yield prediction model and the machine learning and deep learning baselines, including Decision Tree(DT), Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGBoost), Long Short-Term Memory network (LSTM), CNN-LSTM hybrid model.

The conventional models achieved error rates of 5.29%, 3.53%, 7.90%, 2.43% and 4.68%, respectively. In contrast, the proposed model demonstrated a significantly lower error rate of 0.93%, confirming its superior prediction capability and robustness. The reduced error rate highlights the effectiveness of the proposed framework in minimizing incorrect yield predictions. All metrics of all models are summarised in Table 2.

Table 2. Performance comparison of the proposed ZMPPM and the baseline models (%)

Methods	Accuracy	Precision	Recall	F1-score	Error rate
Decision Tree	94.71	87.31	93.12	90.12	5.29
Gradient Boosting Machine (GBM)	96.47	82.08	94.35	87.79	3.53
Extreme Gradient Boosting (XGBoost)	92.10	78.67	95.04	86.08	7.90
Long Short-Term Memory network (LSTM)	97.57	88.54	96.21	92.22	2.43
CNN-LSTM hybrid model	95.32	90.62	96.78	93.60	4.68
ZMPPM (proposed)	99.07	99.40	99.30	99.35	0.93

Note: F1-scores are the harmonic mean of the reported precision and recall.

h. Discussion

In summary, the proposed model achieved the best values of all reported metrics: an accuracy of 99.07%, a precision of 99.40%, a recall of 99.30%, an F1-score of 99.35% and an error rate of 0.93% (Table 2).

Compared with the strongest baseline (LSTM, accuracy 97.57%), ZMPPM increased accuracy by 1.50 percentage points and reduced the error rate from 2.43% to 0.93%. We attribute this improvement mainly to (i) the removal of redundant and noisy variables by ZOA before training and (ii) the global search of ZOA, which reduces the risk of poor local minima compared with gradient-only training. These results should nevertheless be interpreted with caution: they were obtained on a single dataset with a discretised two-class target, the very high accuracy may partly reflect the separability of the Low/High classes defined by the chosen threshold, and ZOA increases the training time. Validation on other regions, crops and seasons, and with a continuous yield target, is required before operational use.

CONCLUSION

This study proposed ZMPPM, a crop yield-level prediction method in which the Zebra Optimization Algorithm selects informative features and optimises the learning parameters of a multilayer perceptron. The method was implemented in Python and evaluated on a historical crop yield dataset using accuracy, precision, recall (sensitivity), F1-score and error rate. ZMPPM achieved an accuracy of 99.07%, a precision of 99.40%, a recall of 99.30%, an F1-score of 99.35% and an error rate of 0.93%, outperforming the DT, GBM, XGBoost, LSTM and CNN-LSTM baselines; its accuracy was 1.50 percentage points higher than that of the best baseline (LSTM). The study has several limitations: a single dataset was used, the yield target was discretised into two classes, ZOA increases training time compared with gradient-based training, and spatial or temporal hold-out validation has not yet been performed. Future work will evaluate ZMPPM on multi-region and multi-season datasets, extend it to continuous yield regression (RMSE, MAE, R^2) and add explainability analysis (e.g. SHAP) to support agronomic interpretation.

REFERENCES

- [1] S. M. Shawon, F. B. Ema, A. K. Mahi, F. L. Niha, and H. T. Zubair, "Crop yield prediction using machine learning: An extensive and systematic literature review," *Smart Agricultural Technology*, vol. 10, Art. no. 100718, 2025, doi: 10.1016/j.atech.2024.100718.
- [2] S. Ajith, S. Vijayakumar, and N. Elakkiya, "Yield prediction, pest and disease diagnosis, soil fertility mapping, precision irrigation scheduling, and food quality assessment using machine learning and deep learning algorithms," *Discover Food*, vol. 5, no. 1, Art. no. 67, 2025, doi: 10.1007/s44187-025-00338-1.
- [3] M. Choubey, Y. K. Gupta, A. Dubey, and R. Prasad, "Deep learning for precision agriculture: Predicting crop yields in a changing climate," in *AgriTech Revolution*. Singapore: Springer Nature Singapore, 2025, pp. 253–275, doi: 10.1007/978-981-95-1268-3_14.

- [4] J. Madhuri, M. Indiramma, and N. Nagarathna, "M-Bi-GRU-CNN: a hybrid deep learning model with optimized feature selection for enhanced crop yield prediction," *Multimedia Tools and Applications*, vol. 84, no. 32, pp. 39787–39811, 2025, doi: 10.1007/s11042-025-20747-9.
- [5] G. L. Anoop, C. Nandini, and E. Naresh, "3TFL-XLnet-CP: A novel transformer-based crop yield prediction framework with weighted loss based 3-tier feature learning model," *SN Computer Science*, vol. 6, no. 3, Art. no. 275, 2025, doi: 10.1007/s42979-025-03778-9.
- [6] R. N. V. J. Mohan, P. S. Rayanoothala, and R. P. Sree, "Next-gen agriculture: integrating AI and XAI for precision crop yield predictions," *Frontiers in Plant Science*, vol. 15, Art. no. 1451607, 2025, doi: 10.3389/fpls.2024.1451607.
- [7] A. Abouaomar, M. El Hanjri, A. Kobbane, A. Laouiti, and K. Nafil, "Hierarchical federated learning for crop yield prediction in smart agricultural production systems," in *Proc. 2025 12th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, 2025, pp. 1–6, doi: 10.1109/WINCOM65874.2025.11313439.
- [8] A. D. Yewle, L. Mirzayeva, and O. Karakuş, "Multi-modal data fusion and deep ensemble learning for accurate crop yield prediction," *arXiv preprint arXiv:2502.06062*, 2025, doi: 10.48550/arXiv.2502.06062.
- [9] E. Trojovská, M. Dehghani, and P. Trojovský, "Zebra Optimization Algorithm: A new bio-inspired optimization algorithm for solving optimization algorithm," *IEEE Access*, vol. 10, pp. 49445–49473, 2022, doi: 10.1109/ACCESS.2022.3172789.
- [10] R. K. Sharma et al., "Maize and soybean yield prediction using machine learning methods: a systematic literature review," *Discover Agriculture*, vol. 3, no. 1, Art. no. 64, 2025, doi: 10.1007/s44279-025-00215-6.
- [11] G. Gupta and S. K. Pal, "Applications of AI in precision agriculture," *Discover Agriculture*, vol. 3, no. 1, Art. no. 61, 2025, doi: 10.1007/s44279-025-00220-9.
- [12] B. M. Lionel, R. Musabe, O. Gatera, and C. Twizere, "A comparative study of machine learning models in predicting crop yield," *Discover Agriculture*, vol. 3, no. 1, Art. no. 151, 2025, doi: 10.1007/s44279-025-00335-z.
- [13] P. Filippi, S. Y. Han, and T. F. A. Bishop, "On crop yield modelling, predicting, and forecasting and addressing the common issues in published studies," *Precision Agriculture*, vol. 26, no. 1, Art. no. 8, 2025, doi: 10.1007/s11119-024-10212-2.
- [14] B. J. Sowmya et al., "Leveraging machine learning for intelligent agriculture," *Discover Internet of Things*, vol. 5, no. 1, Art. no. 33, 2025, doi: 10.1007/s43926-025-00132-6.
- [15] P. Rosso et al., "Combining mobile proximal soil sensors and a crop model to produce high spatial resolution yield prediction maps," *Precision Agriculture*, vol. 26, no. 5, Art. no. 79, 2025, doi: 10.1007/s11119-025-10274-w.
- [16] M. Abdipourchenarestansofla and H.-P. Piepho, "In season estimation of economic optimum nitrogen rate with remote sensing multispectral indices and historical telematics field-operation data," *Precision Agriculture*, vol. 26, no. 2, Art. no. 34, 2025, doi: 10.1007/s11119-025-10224-6.

- [17] L. Jovanovic et al., "Evaluating the performance of metaheuristic-tuned weight agnostic neural networks for crop yield prediction," *Neural Computing and Applications*, vol. 36, no. 24, pp. 14727–14756, 2024, doi: 10.1007/s00521-024-09850-4.
- [18] T. K. N. Fariz and S. S. Basha, "Enhanced crop yield prediction using a hybrid artificial neural network with coati optimization algorithm," *Results in Engineering*, vol. 28, Art. no. 107529, 2025, doi: 10.1016/j.rineng.2025.107529.
- [19] C. Chen, X. Wang, H. Chen, C. Wu, M. Mafarja, and H. Turabieh, "Towards precision fertilization: Multi-strategy grey wolf optimizer based model evaluation and yield estimation," *Electronics*, vol. 10, no. 18, Art. no. 2183, 2021, doi: 10.3390/electronics10182183.

Copyright Holder :

© K. Venkata Rama Rao, Lakshmi Devi N, A. Mahendar, Chavvakula Janaki Devi, Jagan Mohan Sarihaddu, V. Gokula Krishnan (2026)

First Publication Right :

© Asian Journal Science and Engineering

This article is under:

CC BY SA