Iran Agricultural Research (2026) 45(AI Special Issue) 1-13

Shiraz University

Iran Agricultural Research

Journal homepage: https://iar.shirazu.ac.ir/?lang=en

Research Article

Machine learning approach for sugarcane yield prediction in southwest Iran: A radial basis function neural network

Alireza Ashtiani-Araghi[a*] , Abbas Rohani[b*] , Sina Sharifi[c]

[a] Department of Agrotechnology, Aburaihan College of Agricultural Technology, University of Tehran, I. R. Iran
[b] Department of Biosystems Engineering, Faculty of Agriculture, Ferdowsi University of Mashhad, I. R. Iran
[c] Department of Biosystems Engineering, Faculty of Agriculture, Shahid Chamran University of Ahvaz, I. R. Iran

ARTICLE INFO

Keywords:
Artificial intelligence
Harvest forecasting
Prediction model
Precision agriculture
RBF-NN

Received: 26 December 2025
Revised: 25 April 2026
Accepted: 29 April 2026

ABSTRACT- Crop yield prediction in terms of fresh weight is one of the many applications of machine learning in agriculture. Accurate yield prediction models are crucial as they guide growers in making appropriate decisions on what and when to cultivate under varying circumstances influenced by climatic and crop growth parameters, and market conditions. Production of sugarcane involves activities that heavily rely on accurate and timely cropping and harvest forecasting. This study aims to introduce a machine learning model, that estimates sugarcane yield based on various agronomic and field data collected in southwest Iran. Radial basis function neural network (RBF-NN), a shallow feedforward neural network distinguished for its simple structure, universal approximation, and fast learning speed, was employed to develop a yield prediction model. By utilizing datasets containing nine types of input variables and determining optimal values for network hyperparameters, different algorithms were examined for network training. The RBF-NN trained by Levenberg–Marquardt algorithm, containing 75 neurons in the hidden layer, bandwidth value of 0.9, while using 80% of the total data for training, achieved the highest accuracy with an efficiency of 92%, and least amount of estimation error of 4.77%. Further analysis revealed that crop harvest schedule, electrical conductivity of the soil, and crop variety have the most significant impacts on the estimation accuracy of the model. In terms of performance, RBF-NN may not always demonstrate a significant advantage over similar machine learning models. Yet, faster learning and quick convergence are its remarkable points, particularly when dealing with large datasets.

INTRODUCTION

Over the past two decades, there has been significant interest in applying machine learning (ML) across various scientific domains and industries because of its potential benefits and effectiveness. ML has proven successful in a wide range of research and practical applications, with a specific focus on prediction, clustering, and optimization issues (Ifaei et al., 2023). The use of artificial intelligence and ML has gained traction in food and agricultural production (Kutyauripo et al., 2023; Singh and Kaur, 2022; Waqas et al., 2025). Accordingly, since the beginning of the precision agriculture and smart farming era, ML intelligent solutions have been recognized as essential components of these evolving technologies (Akkem et al., 2023; Kuan et al., 2025; Shaikh et al., 2022).

Machine learning for crop yield prediction

Crop yield is a complex trait that is strongly influenced by the relationships among pre-planting, planting, and post-planting processes, as well as climatic and environmental factors, cultivars, and agricultural management practices employed during the plant production period (Guo et al., 2023; Saroj et al., 2021). Accurate and reliable yield prediction is a crucial component of agricultural decision support systems. Currently, ML prediction models are gaining significant attention for yield prediction and harvest forecasting for various agricultural crops. These models enable growers to make appropriate decisions based on climate factors, crop growth parameters, and market conditions.

Chlingaryan et al. (2018) concluded that rapid advancements in ML approaches, including artificial neural network (ANN) modeling, provide cost-effective and comprehensive solutions for better estimation of crop yield and environment states, ultimately leading to the improved decision-making. Elavarasan et al. (2018) provided an overview of various ML approaches, including supervised and unsupervised options used for harvest forecasting. They proposed ANN modeling as a popular solution for crop yield prediction under different farming conditions. Xu et al. (2019) discussed the pros and cons of traditional models for crop yield prediction, and mentioned that ML approaches, such as ANN,

*Corresponding Authors: Assistant Professor, Department of Agrotechnology, Aburaihan College of Agricultural Technology, University of Tehran, I. R. Iran; Professor, Department of Biosystems Engineering, Ferdowsi University of Mashhad, I.R. Iran
E-mail address: ar.ashtiani@ut.ac.ir; arohani@um.ac.ir
DOI:10.22099/iar.2026.54923.1733

Copyright© 2026 The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY-NC 4.0), which permits unrestricted use, distribution, and reproduction in any medium, as long as the original authors and source are cited. No permission is required from the authors or the publisher.

support vector machine (SVM), and random forest (RF), have demonstrated superior performance in crop yield prediction. van Klompenburg et al. (2020) conducted a comprehensive literature review on crop yield prediction using ML approaches. The review screened over 500 relevant studies, and found that temperature, precipitation, and soil type are the top three parameters addressed in the development of ML yield prediction models. Furthermore, ANN was the most frequently used ML tool across all surveyed studies, while convolutional neural network (CNN) was the most commonly employed deep learning (DL) algorithm. According to this study, ML approaches are applicable to almost all staple crops, such as wheat, rice, corn, and soybean, as well as high demand cash crops, such as cotton and sugarcane, if reliable time-series datasets with sufficient amount of required data are available. Bali and Singla (2022) found that ANN and adaptive neuro-fuzzy inference system (ANFIS) models provide the most accurate results among available solutions for crop yield prediction. The study also suggests that DL algorithms are promising choices for crop yield prediction, due to the ability of handling complex big data. Ghaffarian et al. (2022) introduced ANN modeling as the dominant ML approach for the management of crop production risk and harvest forecasting, because of flexibility and accuracy in making quick responses to crucial and decisive situations. This capability is more accentuated, when considering uncertainties in natural growth process of the crops. A systematic review analysis by Shawon et al. (2025) revealed that temperature, and type of soil and vegetation have been the most frequently employed features in the development of ML crop yield prediction models, whereas the most applied DL algorithms adopted in recent studies of intelligent yield prediction are CNN and long short-term memory (LSTM) recurrent network.

Machine learning for sugarcane yield pediction

Sugarcane (*Saccharum spp*.) is an industrial cash crop grown in tropical and subtropical regions because of its great potential to accumulate extremely high levels of sucrose in the stem. The main purpose of sugarcane cultivation is the commercial production of sugar. However, the by-products generated during sugarcane processing, such as bagasse and molasses, make it a significant source of biofuel production, particularly bioethanol, which is blended with gasoline and diesel fuels in various proportions for use in vehicle engines (Kunwer et al., 2022). According to the United Nations Food and Agriculture Organization (FAO), sugarcane is one of the top four crops, which together with all other crops, account for approximately half of the global primary crop production. The statistics provided by FAO (2024) report that sugarcane has been cultivated in over 80 countries with a total production of 1920 million tons. The same report indicates that approximately 8 million tons of sugarcane were harvested in Iran.

The estimation of sugarcane yield (SY) typically considers both the crop fresh weight and sucrose concentration in the cane stalk (Bocca et al., 2015). Sugarcane production includes activities that heavily rely on accurate and timely cropping and harvest forecasting. Overestimation or underestimation of SY may negatively affect farming and logistic issues, crop marketing and pricing, policy formulation, and the profitability of sugarcane production.

As in other major agricultural crops, ML approaches have also effectively contributed to the development of SY prediction models. Everingham et al. (2007) and Everingham et al. (2009) conducted some of the pioneer studies exploring the potential of intelligent algorithms and ML solutions for SY predictive modeling. By making use of time-lagged recurrent networks, Obe and Shangodoyin (2010) developed prediction models to forecast sugarcane production in Africa. With input variables like rainfall, favorable market, government support, consumer and commercial demands, crop varieties, and farming practices, the highest accuracy achieved by the model was 85.7%. Kumar et al. (2015) evaluated the accuracy and effectiveness of multilayer perceptron neural networks (MLP-NN) in predicting SY with 96% accuracy in Indian farms. Using the big data obtained from industrial sugarcane mills, and applying specific processing techniques, such as tuning, feature selection, and feature extraction for improved results, Bocca and Rodrigues (2016) assessed the performance of various ML algorithms, including ANN, SVM, RF, regression tree (RT), and boosted RT in predicting SY. The derived models outperformed the multiple linear regression (MLR) model, and indicated over 90% accuracy in SY prediction. Everingham et al. (2016) accurately estimated annual variations in regional SY in northeast Australia using an RF classifier, based on simulated biomass data, seasonal climate prediction indices, and observational meteorological data. The RF algorithm achieved a classification rate, ranging from 86.4% to 95.5% accuracy in different dates of the cropping season. Employing datasets of soil, weather, crop, and agricultural practices, de Oliveira et al. (2017) developed models for sugar content estimation using the fusion of a feature selection algorithm and various ML regressors. It was found that RF outperformed SVM and RT models and displayed the highest accuracy. Taherei Ghazvinei et al. (2018) developed ML models, including ANN, genetic programming, and extreme learning machine (ELM) to predict growth status of sugarcane crop in Khuzestan province in Iran, with the information acquired from meteorological and field data. The ELM model developed with 8 inputs showed faster learning and higher accuracy compared to other alternatives. Using two specific types of decision trees (DTs), Zaki Dizaji et al. (2018) conducted a study to predict the SY in Iranian farms and evaluate the major factors affecting it. In an attempt by Medar et al. (2019), annual SY was estimated by SVM during a 3-step predictive modeling, according to long-term time series data of weather and soil attributes, and normalized difference vegetation index (NDVI). Using meteorological and crop management datasets, SVM, RF, and boosting ML approaches were adopted to predict SY in Brazilian fields by Hammer et al. (2020). This study also examined the hierarchical importance of the factors influencing the SY. Using a 15-year dataset of cultivating information of sugarcane, Zhou et al. (2023) developed a modified type of SVM combined with an optimization algorithm for predicting SY. The radial basis function (RBF) kernel was also evaluated as one of the applicable functions for model development. In another study by Zaki Dizaji et al. (2024), LSTM recurrent neural networking was examined for SY estimation, and the factors influencing the model accuracy were analyzed and evaluated. The feature extraction was used for processing the big data used in the study. They also compared the prediction performance of some other types of ML approaches like MLP-NN, SVM, and DTs with the LSTM model. Considering long-term weather data of various types, the ML approaches, mainly ANN, SVM and RF, were used to develop SY prediction models in different regions of India (Sridhara et al., 2024; Satpathi et al., 2025). In both studies, ANN exhibits the highest prediction accuracy and performance, and emerges as the more promising approach compared to other models.

Upon reconsideration of recently mentioned ML approaches for crop yield prediction, it becomes apparent that certain types of intelligent solutions are generally underutilized or disregarded for forecasting agricultural harvests. Radial basis function neural network (RBF-NN) exemplifies one of such cases, that has rarely been hired for predicting crop yields of any kind. RBF-NN is mainly recognized for its faster network learning, and may therefore be considered as a potential alternative to the slow convergence of MLP-NN, when extra-large datasets are used for modeling (Patan, 2019). As a recent adopted tool for crop yield predictive modeling, there are not ample examples of RBF-NN applications in this field across research datasets. However, the inherent potential of the RBF-NN, that is, universal approximation, faster learning, and high convergence speed (Sharif Ahmadian, 2016; Sharkawy, 2020), has been evaluated for crop yield estimation in several researches. The study of Mokarram and Bijanzadeh (2016) on barley, was one of the early surveys that utilized RBF-NN modeling for agricultural yield prediction using big data and numerous features. However, the RBF-NN model did not offer the best performance compared to other evaluated models of this study. Rocha and Dias (2019) proposed RBF-NN models for within-season prediction of durum wheat yield in Spain. They used a large number of climatic variables for model development, and claimed that RBF-NN models clearly outperform multivariate linear models used as benchmarks. RBF-NN was one of the several ML modeling tools examined by Joshua et al. (2021) for predicting paddy yield in India, but its performance was not considered the superior option. Kasthuri and Selvakumar (2021) studied the effectiveness of three ML approaches, including the RBF-NN to predict food grains production in India based on acquired time series data. According to Khalifani et al. (2022), RBF-NN was among the models examined for predicting the grain yield of sunflower under normal and salinity stresses. However, estimations provided by CNN model was more optimistic in this case. In the study by Parsaeian et al. (2022), multiple of ML approaches were investigated for yield prediction of oilseed crops. At the end, the RBF-NN model along with Gaussian process regression (GPR), exhibited the most accurate results. Following a different category of crop yield estimation methods, and using various vegetation indices extracted from remote sensing data, Souza et al. (2025) developed intelligent models for early prediction of peanut yield through RBF and MLP algorithms.

This study aims to investigate the effectiveness of the RBF-NN in developing SY prediction models with numerous and diverse inputs from different types and categories, as well as to discover the capabilities or restrictions associated with such models. The success or failure of RBF-NN models in handling large datasets of varying sizes and types and obtaining accurate and quick results greatly depends on the architecture details, employed algorithm, and hyperparameters of the network. Hence, the formulation for the proper selection of such elements to construct the optimal structure of the models which meet the demands in real situations, is the other objective of this study. To address the stated issues, the remainder of this article describes the different steps of data acquisition, planning, and performance analysis, which leads to the development of an optimal RBF-NN model for yield prediction of sugarcane in Khuzestan province of Iran, as the major area for sugarcane production and processing in the country. Using reliable field data, a dynamic adjustment approach, and RBF-NN modeling as a less-considered ML approach in this field, the current study attempts to provide secure decision criteria to determine the relative importance of production inputs and provide an accurate estimation of the final yield for sugarcane growers and decision/policy-makers in the sugar industry.

MATERIALS AND METHODS

Study location and data acquisition

Sugarcane cultivation in Iran is primarily limited to the southwestern region of the country in Khuzestan Province. The data required for this study were collected from over 1000 sugarcane fields owned by the Debel Khozaei Agro-industry Company (DKAC) in Karun County, located in the western part of the province at latitude 31° 8' N and longitude 48° 26' W (Fig. 1). Sugarcane cultivation in this region spans a vast area of approximately 12000 hectares annually. To collect the necessary data for modeling, 4-year stored documentary files were compiled and organized.

The selection of input variables is a pivotal part of model development, ensuring the robustness, generalizability, and proper functioning of yield prediction. Ideally, the input variables of the SY prediction model should reflect the climatic and agronomic features that drive the growth and yield of sugarcane stalks. Different categories of information, such as weather and climate data, agronomic and field data, socio-economic and marketing data, and remote sensing and satellite data, may be used as input variables for SY prediction models. However, the selection of the input variables for this study was based on data availability and manageability, scientifically established relevance, and experience of previous studies across production environments. Climate variables were not included in the model, mainly due to the lack of reliable data in the region and the inability to effectively manage them.

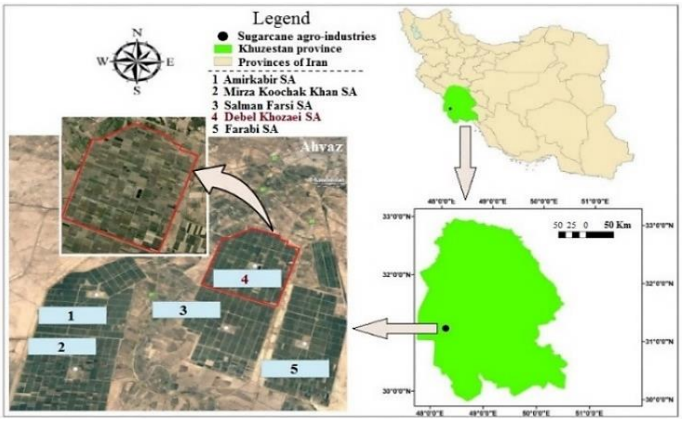

Fig. 1. Geographical coordinates of the surveyed area (SA: Sugarcane Agro-industry).

In this study, four major groups of sugarcane agronomic and field data with significant roles in plant growth and development were used as input variables for SY prediction modeling. These groups include crop, fertilizer, soil, and water management attributes, respectively. The selected groups were further divided into nine subgroups for final modeling (Table 1). The list of input variables includes both numerical, and categorical variables, that are soil texture, crop variety, harvest schedule, and crop age. The input variable "crop age" indicates whether the harvested cane belongs to the first cultivated crop, or to the first, second, or third ratoon (regrowth) of the sugarcane plant. The harvest schedule refers to each of the eight months of the year in which sugarcane harvesting takes place. SY, in terms of fresh weight harvested per unit area, is the single output of the model, the information of which was gathered from the cane fields surveyed across the DKAC.

Radial basis function neural network (RBF-NN)

As a feedforward network with shallow architecture, the RBF-NN is a generalized version of kernel methods, which employs a Gaussian kernel for operation (Kung, 2014). The RBF-NN uses a combination of radially symmetric functions to estimate the existing input-output relationships (Fig. 2). The single hidden layer of the network receives input data directly from the input layer without conducting any computational weight operation, and transforms them into a higher-dimensional space. This transformation often enables linear separability, making RBF-NN a powerful tool for data analysis (Aggarwal, 2018). The hidden layer of the RBF-NN comprises a number of neurons that are typically activated by non-linear Gaussian functions, whereas the neurons of the output layer are linearly activated (Faris et al., 2017). Each hidden neuron in the RBF-NN has a mean prototype vector ($\bar{\mu}$) and a bandwidth value (σ). The initial bandwidth is set using heuristic rules, with $\bar{\mu}$ representing the center of the neuron and σ its width of influence. The output of the *i*th neuron of the RBF-NN hidden layer, which is characterized by the mean prototype vector ($\bar{\mu}_i$) and bandwidth (σ_i), is calculated using the following Gaussian activation function (Aggarwal, 2018):

$$\emptyset_i(\bar{X}) = \exp\left(-\frac{\|\bar{X}-\bar{\mu}_i\|^2}{2\sigma_i^2}\right) \qquad \text{Eq. (1)}$$

where $\bar{X}$ represents the input training vector. This function computes the distance between the input training vector and mean prototype vector of the neuron. The output of the function is used to determine the contribution of the neuron to the final prediction. In the RBF-NN, hidden neurons typically have the same bandwidth value, and bias neurons can be implemented in the output layer. The output of a hidden neuron is a non-linear function of the Euclidean norm between the input training vector and the mean prototype vector, which is then passed to the linear activation function in the output layer.

Table 1. Input variables of sugarcane yield prediction model

Variable		**Details/unit**
Crop	Age	1st crop-1st, 2nd, 3rd ratoon
	Variety	CP57-614, CP69-1062, SP70-1143
	Harvest schedule	Month of harvest
Fertilizer	Nitrogen (N)	$kg.ha^{-1}$
	Phosphate (P)	
Soil	Texture	Clay loam, loam, sandy clay loam, sandy loam, silt clay, and silt clay loam
	Electrical conductivity (EC)	$dS.m^{-1}$
Water	Irrigation volume	$m^3.ha^{-1}$
	Irrigation frequency	Number of times plants are irrigated

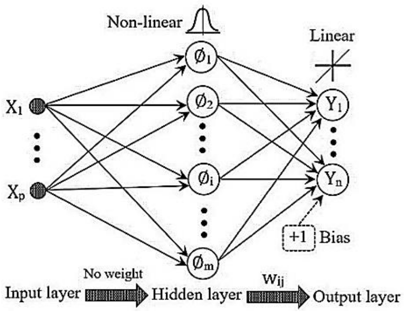

Fig. 2. Diagram of RBF-NN architecture.

The value predicted by the *j*th neuron of the output layer (y_j) is determined by the computational weight assigned to the connection between the *i*th neuron of the hidden layer and the *j*th neuron of the output layer (w_{ij}), as described in Eq. (2), where "m" represents the total number of neurons in the hidden layer.

$$y_j = \sum_{i=1}^{m} w_{ij} \,.\, \emptyset_i(\bar{X}) \qquad \text{Eq. (2)}$$

Training algorithm and data processing

To accomplish the modeling, thirteen training algorithms were examined to develop the SY prediction model based on the agronomic and field data listed in Table 1. These algorithms were applied to train the RBF-NN using the specific training functions defined in MATLAB. A summary of each algorithm, including additional details about each training option, is presented in Table 2. To ensure more uniformity of the feature values present in the datasets, initial screening and removal of outlier data were performed. Additionally, considering the minimum and maximum values of the numerical data collected in the input subgroups, data processing was performed using min-max normalization. One-hot encoding was used to convert the categorical input variables for neural network modeling.

Performance evaluation

Performance evaluation metrics, including the root mean square error (RMSE), mean absolute percentage error (MAPE), and efficiency (EF), that are represented by Eq. (3), Eq. (4), and Eq. (5), were used to assess the effectiveness of the RBF-NN prediction models. The RMSE provides a measure of the average difference between the actual (y_{sa}) and predicted (y_{sp}) values of SY under given farming conditions. The MAPE and EF are used to evaluate the percentage difference, and to measure the ratio between the total variation in the predicted values and the total variation in the actual values, respectively. $\bar{y}_{sp}$ is the mean predicted value. The best performance of the prediction model is achieved when the RMSE and MAPE metrics reach a minimum value, while the EF value approaches 100%.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{sa} - y_{sp}\right)^2} \qquad \text{Eq. (3)}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_{sa} - y_{sp}}{y_{sa}}\right| \times 100 \qquad \text{Eq. (4)}$$

$$\text{EF} = (1 - \frac{\sum_{i=1}^{n}\left(y_{sa} - y_{sp}\right)^2}{\sum_{i=1}^{n}\left(y_{sa} - \bar{y}_{sp}\right)^2}) \times 10 \qquad \text{Eq. (5)}$$

A linear regression curve was also used to assess the fitting status between the actual values observed and the predicted values estimated by the model.

RESULTS AND DISCUSSION

Optimal hyperparameters

We embark on the design of the RBF-NN by first identifying the optimal values for a couple of hyperparameters "number of hidden neurons" and "bandwidth value". Number of neurons in the hidden layer of the shallow neural networks, such as MLP-NN and RBF-NN plays a crucial role in determining the performance of the yield prediction model, and should be decided based on recurring trials and appraisals. This requires careful consideration of various factors, such as the complexity of the problem, quality of the training data, and available computational resources. Insufficient number of neurons in the hidden layer of the RBF-NN may lead to underfitting, which occurs when the model fails to capture the nuances of the data. On the contrary, excessive number of hidden neurons may lead to overfitting, which makes the model too complex and unable to generalize to new data. To ensure utmost accuracy, generalizability, and computational efficiency of the RBF-NN model for SY prediction, selecting the appropriate number of hidden neurons was conducted, and the effect of their number on network performance was evaluated. Fig. 3 displays the variations of RMSE and EF, as the number of hidden neurons increases during network training and testing. Results indicate that as the number of hidden neurons increases from 3 to 75, the RMSE and EF decreases and increases, respectively. However, the changes observed in RMSE and EF become negligible, when number of neurons exceeds 63. It was also found that increasing the number of hidden neurons beyond 75 did not significantly improve the performance of the network, and the EF may not exceed 92%. The complexity of the SY function for the independent variables studied is probably the reason why the efficiency of the model did not increase beyond 92%. Consequently, 75 was selected as the optimal number of hidden neurons for developing the RBF-NN model for SY prediction. This number is significantly higher than those observed in similar studies conducted for SY ANN modeling, for example, Kumar et al. (2015) and Satpathi et al. (2025), as well as for other types of yield prediction (Sahoo et al., 2023), in all of which much fewer hidden neurons are used in MLP-NN and RBF-NN architectures, possibly at the expense of slower convergence.

Table 2. Specifications of the algorithms examined for training of the RBF-NN model

Algorithm	Function	Description
Bayesian regularization	trainbr	Prevents overfitting via a penalty term that encourages weights to be close to zero
Levenberg-Marquardt	trainlm	Minimizes error function using the Levenberg-Marquardt optimization method, which is fast and efficient for small to medium-sized datasets
BFGS Quasi-Newton	trainbfg	Approximates the Hessian matrix of the error function using a quasi-Newton optimization method. Suitable for large datasets and networks with many parameters
One-step secant	trainoss	Similar to BFGS, but uses a simplified Hessian matrix approximation. Suitable for large datasets and networks with many parameters
Resilient backpropagation	trainrp	Adapts learning rate with weight back-tracking, making it suitable for noisy datasets and networks with many parameters
Gradient descent	traingd	Updates weights in the direction of the negative gradient of the error function. Simple and intuitive, but slow to converge and sensitive to learning rate
Gradient descent with adaptive learning rate	traingda	Similar to gradient descent, but the learning rate is adaptively adjusted based on the sign of the gradient. Suitable for datasets with varying gradients
Gradient descent with momentum	traingdm	Uses a momentum term to speed up convergence and avoid getting stuck in local minima. It is a good choice for noisy datasets and networks with many parameters
Gradient descent with adaptive learning rate and momentum	traingdx	Combines adaptive learning rate and momentum techniques to speed up convergence and avoid local minima
Scaled conjugate gradient	trainscg	Uses a scaled conjugate gradient optimization method to avoid overshooting the minimum. Suitable for large datasets and networks with many parameters
Conjugate gradient backpropagation with Polak-Ribiére updates	traincgb	Variation of the standard conjugate gradient backpropagation algorithm that uses the Polak-Ribiére method to compute the search direction during weight updates Suitable for small to medium-sized datasets and networks with many parameters
Conjugate gradient (Fletcher-Reeves)	traincgf	Uses a conjugate gradient optimization method with Powell-Beale restarts. Suitable for large datasets and networks with many parameters
Conjugate gradient (Polak-Ribiére)	traincgp	Uses a conjugate gradient optimization method with Polak-Ribiére updates. Suitable for small to medium-sized datasets and networks with many parameters

The performance and complexity of the RBF-NN model also depend on the bandwidth (σ) hyperparameter. The bandwidth value determines the extent to which the input data points influence the network output. Appropriate bandwidth values are essential to ensure that the network can capture complex and non-linear patterns in the data. To achieve the best possible performance of the RBF-NN while avoiding the pitfalls of overfitting or underfitting, careful selection of the optimal bandwidth value was also performed. Accordingly, the effect of the initial bandwidth on the network performance was analyzed. To this end, the domain of the bandwidth value was increased from 0.05 to 1, and the effect of this surging trend on the performance of the network was evaluated according to the RMSE and EF.

Fig. 4 displays the variations in the RMSE and EF as the bandwidth value changes during network training and testing. The results show that increasing the bandwidth value toward 1 improves the EF but inversely decreases the RMSE. The optimal bandwidth was achieved at a value of 0.9 for the initiation of the network, where the EF of the model was approximately 91%. The trends of the curves shown in Fig. 3 and Fig. 4 imply that the RMSE and EF values relevant to the training and testing of the RBF-NN are relatively close to each other. This indicates the ability of the network to provide an accurate prediction of SY in terms of the study variables.

Best training algorithm

Selecting an appropriate training algorithm is critical when designing an ANN. The proper choice of algorithm depends on several factors, including the size and complexity of the training dataset and the desired level of accuracy. The results of the RMSE and EF analyses presented in Fig. 5 reveal that the Levenberg–Marquardt (trainlm), Bayesian regulation (trainbbr), and scaled conjugate gradient (trainscg) are ranked as the top three algorithms with superior performance for training the RBF-NN prediction model. This finding is consistent with those of other studies that used different training algorithms for neural network modeling of yield prediction in various crops, such as Garg et al. (2016), who used six different algorithms to train backpropagation neural networks for estimating wheat production and identified trainlm and trainbbr functions as the most efficient algorithms. Similarly, Haghverdi et al. (2018) successfully applied the trainlm function with hyperbolic tangent and linear transfer functions as the best algorithm to predict cotton lint yield through the phenology of crop indices. Belouz et al. (2022) found that the trainlm training function was the most effective option for the yield prediction of greenhouse tomatoes when using MLP-NN for modeling. However, in the study by Kumar et al. (2015), in which ten backpropagation training algorithms were examined for SY ANN modeling, the performance achieved by the trainlm training function was not among the ones with prominent performance, while the model developed by the trainbfg training function was the selected model with the lowest RMSE.

Size of training dataset

When designing ANNs, it is necessary to select an appropriate size for the training dataset to achieve accuracy and enhance the generalization ability of the model. The training dataset should capture the complexity and variability of the data while avoiding being too small or too large.

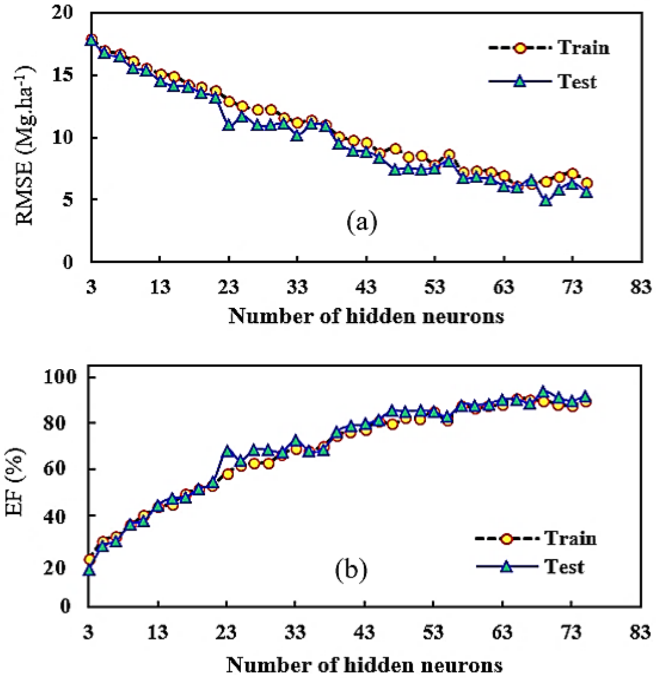

Fig. 3. Effect of the number of hidden neurons on (a) root mean square error (RMSE) and (b) efficiency (EF) of sugarcane yield prediction model.

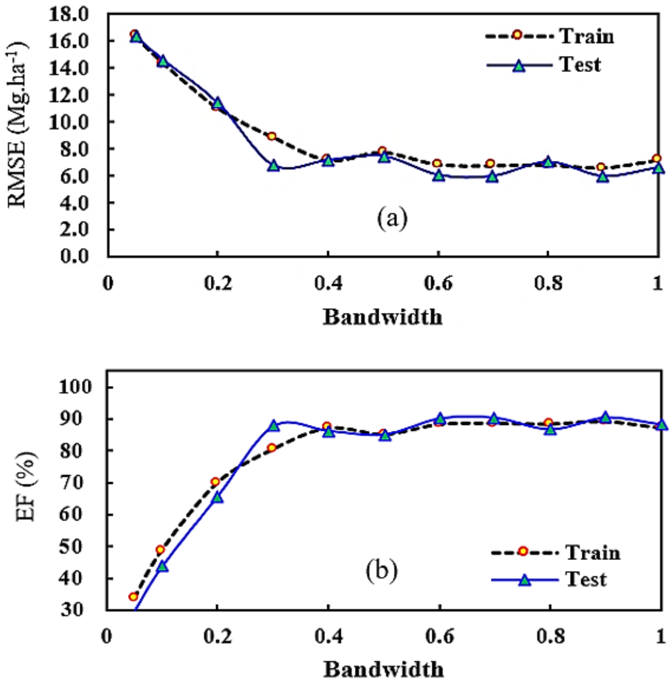

Fig. 4. Effect of the bandwidth value on (a) root mean square error (RMSE) and (b) efficiency (EF) of sugarcane yield prediction model.

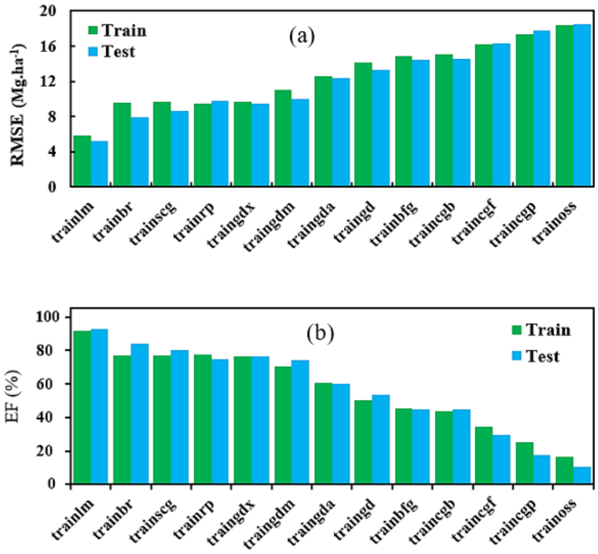

Fig. 5. Performance of different training algorithms for sugarcane yield prediction model according to (a) root mean square error (RMSE) and (b) efficiency (EF).

To evaluate the prediction performance of the RBF-NN, training datasets containing 50%, 60%, 70%, and 80% of the total data were used. Table 3 presents the evaluation results of the SY prediction model for different data partitions based on the performance metrics RMSE, MAPE, and EF. In all datasets, the best performance was achieved when 80% of the total data were used for training. This is fully consistent with the findings of Kumar et al. (2015), Sridhara et al. (2024), and Satpathi et al. (2025), in which using over 75% of the total data for training displayed the best performance in SY ANN modeling. The results show that although the prediction errors of the model increase as the size of the training dataset decreases, the amount of this increase is not significant. In the “All” dataset, the prediction error of the model in terms of MAPE varied from 4.77% to 6.79%, which falls within an acceptable range. Table 3 also includes the statistical comparison between the actual and predicted SY datasets in terms of the mean, variance, and statistical distribution at significance level of 5%. The *P*-value obtained from the statistical test of the equality of the mean, variance, and distribution between the actual and predicted SY datasets confirmed no significant difference between them. Although the *P*-value decreased as the size of the training dataset decreased, it remained greater than or equal to 0.05. Based on these findings, it can be concluded that the RBF-NN can successfully establish the relationship between SY and all nine independent input variables during the training of the entire dataset.

Strong agreement between the actual and predicted data serves as an indicator of the model’s ability to effectively identify the underlying patterns within the input data, which will ultimately lead to accurate predictions. Conversely, a poor agreement between the actual and predicted data indicates that the model has a tendency to overfit or underfit the training data, resulting in inaccurate predictions. Fig. 6 illustrates the evaluation of the agreement between the actual and predicted SY data for different sizes of the training dataset during training and testing. Reducing the percentage size of the training dataset from 80% to 50% of the total data led to a decrease in the existing agreement between the actual and predicted datasets. The best agreement was achieved with a training dataset size of 80%. This is evident from the highest values of the coefficient of determination (R^2) in the 80% training dataset, which were 0.92 and 0.93 during training and testing, respectively. In addition, the slope and intercept of the regression line between the actual and predicted SY values were approximately 0.92 and 5.74, and 0.93 and 5.45 during training and testing, respectively. However, obtaining more accurate fitting functions for crop yields requires access to more comprehensive and diverse datasets.

A comparison with similar studies conducted on SY estimation using ML approaches illustrates that the RBF-NN is an acceptable choice. The results of the study conducted by Zaki Dizaji et al. (2018) in Khuzestan, which aimed to predict SY using classification and regression trees (CART) and chi-squared automatic interaction detection (CHAID), indicated that the estimation accuracies achieved by these supervised ML algorithms were to some extent lower than that of the RBF-NN. Sharifi et al. (2020) showed that the overall performance of the RBF-NN model developed for SY prediction was considerably higher than that of the MLP-NN model developed using the same datasets. In another study by Zaki Dizaji et al. (2024) in Khuzestan sugarcane fields with supplementary feature engineering methods, the performance of LSTM, MLP-NN, SVM, and DTs were evaluated to predict SY. The LSTM network was the only model that showed slightly greater superiority over the RBF-NN model in terms of estimation performance. The outperformance of RBF-NN for yield prediction of crops other than sugarcane has been confirmed in other studies (Rocha and Dias, 2019), while opposing results may also be found (Sahoo et al., 2023).

Table 3. Evaluation of prediction accuracy across different sizes of training datasets

Size of training dataset (% of the total data)	Train			Test			All		
	RMSE ($Mg.ha^{-1}$)	MAPE (%)	EF (%)	RMSE ($Mg.ha^{-1}$)	MAPE (%)	EF (%)	RMSE ($Mg.ha^{-1}$)	MAPE (%)	EF (%)
80	5.82	4.90	92	5.24	4.26	93	5.71	4.77	92
70	5.97	5.41	91	6.11	5.71	90	6.03	5.53	91
60	6.99	6.36	89	6.23	5.73	89	6.62	6.05	89
50	7.14	6.02	88	6.77	5.86	90	6.76	6.79	89
	P_m	P_v	P_d	P_m	P_v	P_d	P_m	P_v	P_d
80	0.97	0.22	89	0.94	0.57	96	1.00	0.18	83
70	0.95	0.08	54	0.91	0.39	99	0.99	0.05	46
60	0.90	0.20	5	0.87	0.38	13	0.99	0.12	5
50	0.90	0.16	56	0.85	0.21	61	0.98	0.06	22

*P_m, P_v, and P_d represent the *P*-values obtained from statistical tests at significance level of 5% to compare the mean, variance, and distribution between the predicted and actual sugarcane yield datasets, respectively.

** RMSE: Root mean square error; MAPE: Mean absolute percentage error; EF: Efficiency.

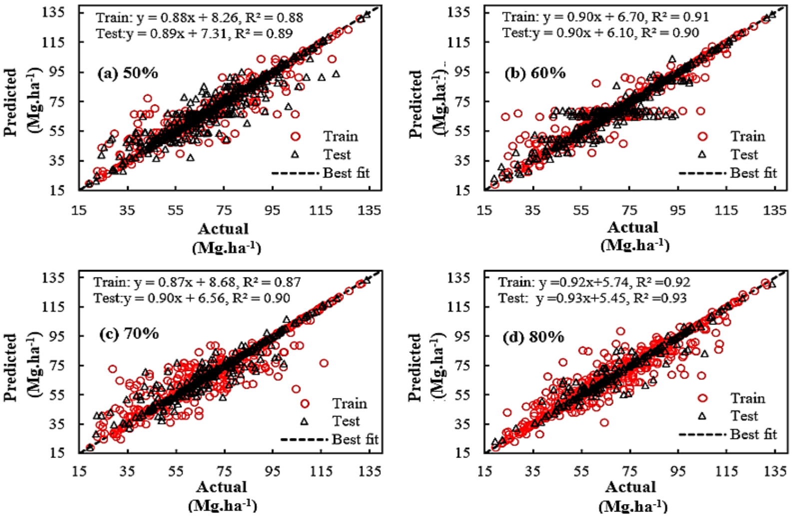

Fig. 6. Agreement between actual and predicted sugarcane yield in different sizes of the training dataset.

Insights from sensitivity analysis

Sensitivity analysis is a tool that identifies critical variables driving the performance of a certain model. It provides a comprehensive understanding of how each individual input variable impacts the model performance, and analyzes the effect of the change or removal of the inputs on the outcome of interest. By performing such analysis on the final RBF-NN model, it is possible to identify crucial key variables and optimize their values to improve the accuracy later. This step is vital for developing effective yield prediction models and gaining insights into complex cropping systems. Table 4 presents the results of the final analysis for the input variables of the selected RBF-NN prediction model with all nine independent input variables listed in Table 1. The first row represents the information of the original model when all variables are included. Other rows, each showing the effect of the change of a certain input variable with other variables remained fixed, can be used to compare the outputs based on the estimation of performance evaluation metrics. The performance evaluation analysis is based on the metrics employed in this study over training, testing, and "All" datasets. Information shown in Table 4 is arranged in ascending order of MAPE for "All" dataset. Examining the results based on MAPE and EF indicates that the minimum increase in prediction error, that is the lowest drop in model accuracy, occurs when the input variables phosphate (P) and nitrogen (N) fertilizers alter, while the opposite is true about the month of harvest, EC of the soil, and crop variety at the bottom of the list. This implies that the latter variables have greater contributions to the output of the model, and higher impacts on the accuracy of the yield estimation. This is partly compatible with the findings of de Oliveira et al. (2017) for sugarcane, in which fertility and most soil attributes were of lower importance in model contribution compared to crop and weather variables. Additionally, crop variety has been found to be among the high-significant input variables in SY prediction models of Zaki Dizaji et al. (2018) and Sharifi et al. (2020). However, since a limited number of input variables are used for the RBF-NN modeling in this study, and considering that most soil attributes, such as texture and EC are theoretically interrelated to other soil and water features, for example, soil fertility and field capacity, more comprehensive and detailed studies are essentially needed to verify the obtained results. It should be noted that these findings are relative comparisons, and whole input variables of the final model have important roles in predicting the SY. It is obvious that the prediction error of the model increases, if any of the available input variables are considerably changed or ignored. It is worth noting that using all of the variables of the model may not also guarantee the accurate estimation of the yield.

Table 4. Sensitivity analysis of RBF-NN model for sugarcane yield estimation: Insights from the results

Variable	Train			Test			All		
	RMSE ($Mg.ha^{-1}$)	MAPE (%)	EF (%)	RMSE ($Mg.ha^{-1}$)	MAPE (%)	EF (%)	RMSE ($Mg.ha^{-1}$)	MAPE (%)	EF (%)
All variables (Model)	5.82	4.90	92	5.24	4.26	93	5.71	4.77	92
All-P fertilizer	6.45	6.38	90	5.83	5.69	91	6.15	6.04	91
All-N fertilizer	7.27	6.44	88	6.52	5.64	88	6.91	6.03	88
All-Crop age	7.26	6.74	88	6.26	6.01	89	6.78	6.38	89
All-Irrigation volume	6.63	6.42	90	7.10	6.58	86	6.87	6.50	88
All-Soil texture	7.02	6.75	89	7.93	7.35	83	7.49	7.05	86
All-Irrigation frequency	7.76	7.84	86	7.06	6.94	86	7.42	7.39	86
All-Crop variety	8.16	7.86	85	7.28	7.17	86	7.73	7.52	85
All-Soil EC	7.87	8.39	86	7.27	7.28	86	7.58	7.84	86
All-Month of harvest	8.77	9.06	82	7.86	7.65	83	8.33	8.36	83

* RMSE: root mean square error; MAPE: Mean absolute percentage error; EF: Efficiency.

CONCLUSION

Review of the literature reveals that the most common intelligent tools successfully used for crop yield prediction modeling are ANNs. However, it seems that the potential of less conventional and more basic ANNs such as RBF-NN has not been adequately addressed for the modeling of crop yield prediction thus far. Therefore, the current study utilized various agronomic and field data through a dynamic adjustment approach to develop an RBF-NN model for sugarcane harvest forecasting in southwest Iran. Attempts to design a reliable RBF-NN model began with exploring the best training algorithm, and identifying the optimal values for hyperparameters of the network. This was followed by a thorough evaluation of model predictive abilities, and presentation of the results of sensitivity analysis. Similar to many previous studies, the Levenberg-Marquardt training algorithm produced the most promising outcome in RBF-NN modeling, in terms of RMSE and MAPE minimization, and EF maximization. Proper selection of network hyperparameters, including the number of hidden neurons and initial bandwidth value, is an important issue in designing the RBF-NN, as these are the key values that regulate the complexity and performance of the model. Another factor to consider in the development of the RBF-NN model is the optimal percentage of the total data used in the training dataset, which was determined to be 80% in this study. Accordingly, 9-75-1 RBF-NN model with the initial bandwidth value of 0.9 was selected for predicting SY. The EF obtained by the final model was 92%, with lowest estimation errors being 5.71 $Mg.ha^{-1}$ for RMSE, and 4.77% for MAPE. The prediction accuracy based on R^2 of the linear fitting between actual and predicted values was 0.93. Further analysis was conducted to examine how the output of the model responds to variations in different input variables. Results indicated that the month in which the sugarcane harvesting occurs, the EC of the soil in sugarcane fields, and crop variety have greater impacts on the accuracy and ability of model estimation. The opposite was true for chemical fertilizers, probably due to the complexity of the SY function, and ambiguities about how the soil structure, nutrients, water, and sugarcane varieties interact to affect the final yield performance. Compared to other ML models developed for SY prediction, the performance of the current model was acceptable, but not the highest in some metrics. However, it is accurate enough to meet the demands of reliable decision-making.

Generally, one possible reason for slightly lower performance of ML yield prediction models could be the lack of long-term data for model training due to the fairly short duration of data acquisition, as well as the failure to ideally apply supplementary data processing techniques like data tuning, feature selection, and feature extraction. It is noteworthy to mention that in the context of crop yield estimation, where high-dimensional heterogeneous massive data often span multiple seasons, locations, and management systems, the utility of implementing such techniques lies in their combined impact on the robustness, accuracy, and interpretability of the ML prediction model as discussed in Bocca and Rodrigues (2016), Charoen-Ung and Mittrapiyanuruk (2018), and Zaki Dizaji et al. (2024). Crop yield prediction using various hybrid models (Oikonomidis et al., 2022; Hernández Hernández et al., 2025) can also be a promising option that may warrant further exploration in predicting SY in future national studies. Limited number of input variables and exclusion of influential climate data for modeling are additional constraints that may potentially pose challenges in the performance of the current RBF-NN model. Therefore, additional studies using larger amount of data acquired from long-term surveys, a wider variety of data types, and deeper consideration of the soil-water-crop nexus and their functional interrelationships are recommended for more reliable and generalizable results. In general, the RBF-NN model generated favorable and promising results in predicting SY. However, while RBF-NN may not always demonstrate a significant advantage over similar ML models in crop yield prediction, the faster learning and quick convergence it displays are its remarkable points in data mining purposes. This is especially useful when dealing with extra-large datasets that contain numerous input variables and data types gathered during long-term surveys in complicated situations.

Numerical and categorical data obtained from field survey measurements and weather station records are not the only types of data available for estimating SY. ML models can predict SY by analyzing other types of datasets, such as remote sensing data, aerial digital images, and real-time in-situ data provided by satellites, drones, and other monitoring technologies (Som-Ard et al., 2021; Cardoso et al., 2022; Amarasingam et al., 2022; de França e Silva et al., 2024; Vinayaka and Prasad, 2024). The essence of such digital data and the corresponding procedures used to analyze them for ML prediction modeling differ from those supporting conventional data types. However, as a significant segment of crop yield estimation strategies, it is highly advisable to focus more on studies utilizing the digital data types mentioned for the development of ML yield prediction models in Iranian sugarcane farming. Moreover, because sugarcane is also cultivated in northern regions of Iran, mainly Mazandaran province, under totally different climate and conditions, and new plans are on the horizon for spreading sugarcane cultivation in some other southern regions of the country, for example, Fars, Bushehr, Hormozgan, and Sistan and Baluchestan, promoting similar ML modeling studies for SY monitoring and forecasting is suggested for these regions.

Ultimately, the information obtained from this study may be useful in managing sugarcane fields, as well as in logistic and financial planning for different departments of the sugar industry and its allied market. However, caution should be exercised when generalizing these results to other regions and different cropping and climatic conditions, as additional studies are needed to verify their applicability. This study provides valuable insights into the performance of the RBF-NN and its potential implications for future research in this field. This may offer primary decision criteria for sugarcane growers, sugar millers, industries involved with sugarcane by-products, and other potential market participants in the future.

FUNDING

This research did not receive any specific grants from any funding agencies in the public, commercial, or not-for-profit sectors.

CRediT AUTHORSHIP CONTRIBUTION STATEMENT

Conceptualization: Alireza Ashtiani-Araghi and Sina Sharifi; Methodology: Abbas Rohani; Software: Abbas Rohani; Validation: Abbas Rohani; Formal analysis: Abbas Rohani; Investigation: Alireza Ashtiani-Araghi and Sina Sharifi; Resources: Alireza Ashtiani-Araghi; Data curation: Sina Sharifi; Writing—original draft preparation: Alireza Ashtiani-Araghi; Writing—review and editing: Alireza Ashtiani-Araghi and Abbas Rohani; Visualization: Alireza Ashtiani-Araghi and Abbas Rohani; Supervision: Abbas Rohani; Project administration: Abbas Rohani; Funding acquisition: None.

DECLARATION OF COMPETING INTEREST

The authors declare no conflicts of interest.

ETHICAL STATEMENT

Not applicable.

DATA AVAILABILITY

The data supporting the findings of this study are available from the corresponding authors upon reasonable request.

ACKNOWLEDGMENTS

The authors express their sincere gratitude to the administrative managers and technical staff of the Debel Khozaei Agro-industry Company (DKAC) for their generous cooperation and support in providing the diverse data and supplementary information used in this study.

REFERENCES

Aggarwal, C. C. (2018). *Neural networks and deep learning: A textbook* (1st ed.). Springer.

Akkem, Y., Biswas, S. K., & Varanasi, A. (2023). Smart farming using artificial intelligence: A review. *Engineering Applications of Artificial Intelligence*, *120*, 105899. https://doi.org/10.1016/j.engappai.2023.105899

Amarasingam, N., Salgadoe, A. S. A., Powell, K., Gonzalez, L. F., & Natarajan, S. (2022). A review of UAV platforms, sensors, and applications for monitoring of sugarcane crops. *Remote Sensing Applications: Society and Environment*, *26,* 100712. https://doi.org/10.1016/j.rsase.2022.100712

Bali, N., & Singla, A. (2022). Emerging trends in machine learning to predict crop yield and study its influential factors: A survey. *Archives of Computational Methods in Engineering*, *29,* 95-112. https://doi.org/10.1007/s11831-021-09569-8

Belouz, K., Nourani, A., Zereg, S., & Bencheikh, A. (2022). Prediction of greenhouse tomato yield using artificial neural networks combined with sensitivity analysis. *Scientia Horticulturae*, *293,* 110666. https://doi.org/10.1016/j.scienta.2021.110666

Bocca, F. F., & Rodrigues, L. H. A. (2016). The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. *Computers and Electronics in Agriculture*, *128,* 67-76. https://doi.org/10.1016/j.compag.2016.08.015

Bocca, F. F., Rodrigues, L. H. A., & Arraes, N. A. M. (2015). When do I want to know and why? Different demands on sugarcane yield predictions. *Agricultural Systems*, *135,* 48-56. https://doi.org/10.1016/j.agsy.2014.11.008

Cardoso, L. A. S., Farias, P. R. S., & Soares, J. A. C. (2022). Use of unmanned aerial vehicle in sugarcane cultivation in brazil: A review. *Sugar Tech, 24*(6), 1636-1648. https://doi.org/10.1007/s12355-022-01149-9

Charoen-Ung, P., & Mittrapiyanuruk, P. (2018). Sugarcane yield grade prediction using random forest with forward feature selection and hyper-parameter tuning. In *International Conference on Computing and Information Technology* (pp. 33-42). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-93692-5_4

Chlingaryan, 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. https://doi.org/10.1016/j.compag.2018.05.012

de França e Silva, N. R., Chaves, M. E. D., Luciano, A. C. D. S., Sanches, I. D. A., de Almeida, C. M., & Adami, M. (2024). Sugarcane yield estimation using satellite remote sensing data in empirical or mechanistic modeling: A systematic review. *Remote sensing*, *16*(5), 863. https://doi.org/10.3390/rs16050863

de Oliveira, M. P. G., Bocca, F. F., & Rodrigues, L. H. A. (2017). From spreadsheets to sugar content modeling: A data mining approach. *Computers and Electronics in Agriculture, 132*, 14-20. https://doi.org/10.1016/j.compag.2016.11.012

Everingham, Y., Inman-Bamber, N., Thorburn, P., & McNeill, T. (2007). A bayesian modelling approach for long lead sugarcane yield forecasts for the Australian sugar industry. *Australian Journal of Agricultural Research*, *58*(2), 87-94. https://doi.org/10.1071/AR05443

Everingham, Y. L., Smyth, C. W., & Inman-Bamber, N. G. (2009). Ensemble data mining approaches to forecast regional sugarcane crop production. *Agricultural and Forest Meteorology*, *149*(3-4), 689-696. https://doi.org/10.1016/j.agrformet.2008.10.018

Everingham, Y., Sexton, J., Skocaj, D., & Inman-Bamber, G. (2016). Accurate prediction of sugarcane yield using a random forest algorithm. *Agronomy for Sustainable Development*, *36*(2), 27. https://doi.org/10.1007/s13593-016-0364-z

Elavarasan, D., Vincent, D. R., Sharma, V., Zomaya, A. Y., & Srinivasan, K. (2018). Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Computers and Electronics in Agriculture*, *155,* 257-282. https://doi.org/10.1016/j.compag.2018.10.024

FAO. (2024). *World Food and Agriculture. Statistical Yearbook 2024.* Rome, Italy.

Faris, H., Aljarah, I., & Mirjalili, S. (2017). Evolving radial basis function networks using moth–flame optimizer. In Samui, P., Sekhar, S., and Balas, V.E. (Eds.), *Handbook of neural computation* (pp. 537-550). Academic Press. https://doi.org/10.1016/B978-0-12-811318-9.00028-4

Garg, B., Kirar, N., Menon, S., & Sah, T. (2016). A performance comparison of different back propagation neural networks methods for forecasting wheat production. *CSI Transactions on ICT*, *4*(2), 305-311. https://doi.org/10.1007/s40012-016-0096-x

Ghaffarian, S., van der Voort, M., Valente, J., Tekinerdogan, B., & de Mey, Y. (2022). Machine learning-based farm risk management: A systematic mapping review. *Computers and Electronics in Agriculture*, *192*, 106631. https://doi.org/10.1016/j.compag.2021.106631

Guo, S., Zhang, Z., Zhang, F., & Yang, X. (2023). Optimizing cultivars and agricultural management practices can enhance soybean yield in Northeast China. *Science of the Total Environment*, *857*(2), 159456. https://doi.org/10.1016/j.scitotenv.2022.159456

Haghverdi, A., Washington-Allen, R. A., & Leib, B. G. (2018). Prediction of cotton lint yield from phenology of crop indices using artificial neural networks. *Computers and Electronics in Agriculture*, *152*, 186-197. https://doi.org/10.1016/j.compag.2018.07.021

Hammer, R. G., Sentelhas, P. C., & Mariano, J. C. (2020). Sugarcane yield prediction through data mining and crop simulation models. *Sugar Tech*, *22*(2), 216-225. https://doi.org/10.1007/s12355-019-00776-z

Hernández Hernández, G. C., Gómez Gómez, J., & Jiménez-Cabas, J. (2025). Predictive models based on artificial intelligence to estimate crop yield: A literature review. *Agriculture*, *15*(23), 2438. https://doi.org/10.3390/agriculture15232438

Ifaei, P., Nazari-Heris, M., Tayerani Charmchi, A.S., Asadi, S. & Yoo, C. K. (2023). Sustainable energies and machine learning: An organized review of recent applications and challenges. *Energy*, *266*, 126432. https://doi.org/10.1016/j.energy.2022.126432

Joshua, V., Priyadharson, S. M., & Kannadasan, R. (2021). Exploration of machine learning approaches for paddy yield prediction in eastern part of Tamilnadu. *Agronomy*, *11*(10), 2068. https://doi.org/10.3390/agronomy11102068

Kasthuri, V., & Selvakumar, S. (2021). Forecasting foodgrains production using arima model and neural network. *American Journal of Neural Networks and Applications*, *7*(2), 30-37. https://doi.org/10.11648/j.ajnna.20210702.12

Khalifani, S., Darvishzadeh, R., Azad, N., & Rahmani, R. S. (2022). Prediction of sunflower grain yield under normal and salinity stress by RBF, MLP and, CNN models. *Industrial Crops and Products*, *189*, 115762. https://doi.org/10.1016/j.indcrop.2022.115762

Kuan, Y. N., Goh, K. M., & Lim, L. L. (2025). Systematic review on machine learning and computer vision in precision agriculture: Applications, trends, and emerging techniques. *Engineering Applications of Artificial Intelligence*, *148*, 110401. https://doi.org/10.1016/j.engappai.2025.110401

Kumar, S., Kumar, V., & Sharma, R. K. (2015). Sugarcane yield forecasting using artificial neural network models. *International Journal of Artificial Intelligence and Applications*, *6*(5), 51-68. https://doi.org/10.5121/ijaia.2015.6504

Kung, S. Y. (2014). *Kernel methods and machine learning* (1st ed.). UK: Cambridge University Press.

Kunwer, R., Pasupuleti, S. R., Bhurat, S. S., Gugulothu, S. K., & Rathore, N. (2022). Blending of ethanol with gasoline and diesel fuel–A review. *Materials Today: Proceedings*, *69*, 560-563. https://doi.org/10.1016/j.matpr.2022.09.319

Kutyauripo, I., Rushambwa, M., & Chiwazi, L. (2023). Artificial intelligence applications in the agrifood sectors. *Journal of Agriculture and Food Research*, *11*, 100502. https://doi.org/10.1016/j.jafr.2023.100502

Medar, R. A., Rajpurohit, V. S., & Ambekar, A. M. (2019). Sugarcane crop yield forecasting model using supervised machine learning. *International Journal of Intelligent Systems and Applications*, *10*(8), 11. https://doi.org/10.5815/ijisa.2019.08.02

Mokarram, M., & Bijanzadeh, E. (2016). Prediction of biological and grain yield of barley using multiple regression and artificial neural network models. *Australian Journal of Crop Science*, *10*(6), 895-903. https://doi.org/10.21475/ajcs.2016.10.06.p7634

Obe, O., & Shangodoyin, D. (2010). Artificial neural network based model for forecasting sugar cane production. *Journal of Computer Science*, *6*(4), 439.

Oikonomidis, A., Catal, C., & Kassahun, A. (2022). Hybrid deep learning-based models for crop yield prediction.

Applied Artificial Intelligence, *36*(1), 2031822. https://doi.org/10.1080/08839514.2022.2031823

Parsaeian, M., Rahimi, M., Rohani, A., & Lawson, S. S. (2022). Towards the modeling and prediction of the yield of oilseed crops: A multi-machine learning approach. *Agriculture, 12*(10), 1739. https://doi.org/10.3390/agriculture12101739

Patan, K. (2019). *Robust and fault-Tolerant control: Neural-network-based solutions* (1st ed.). Springer.

Rocha, H., & Dias, J. M. (2019). Early prediction of durum wheat yield in Spain using radial basis functions interpolation models based on agroclimatic data. *Computers and Electronics in Agriculture*, 157, 427-435. https://doi.org/10.1016/j.compag.2019.01.018

Sahoo, M., Dey, S., Sahoo, S., Das, A., Ray, A., Nayak, S., & Subudhi, E. (2023). MLP (multi-layer perceptron) and RBF (radial basis function) neural network approach for estimating and optimizing 6-gingerol content in Zingiber officinale Rosc. in different agro-climatic conditions. *Industrial Crops and Products, 198*, 116658. https://doi.org/10.1016/j.indcrop.2023.116658

Saroj, R., Soumya, S. L., Singh, S., Sankar, S. M., Chaudhary, R., Yashpal, Saini, N., Vasudev, S., & Yadava, D. K. (2021). Unraveling the relationship between seed yield and yield-related traits in a diversity panel of Brassica juncea using multi-traits mixed model. *Frontiers in Plant Science, 12,* 651936. https://doi.org/10.3389/fpls.2021.651936

Satpathi, A., Chand, N., Setiya, P., Ranjan, R., Nain, A. S., Vishwakarma, D. K., Saleem, K., Obaidullah, A. J., Yadav, K. K., & Kisi, O. (2025). Evaluating statistical and machine learning techniques for sugarcane yield forecasting in the Tarai region of North India. *Computers and Electronics in Agriculture, 229*, 109667. https://doi.org/10.1016/j.compag.2024.109667

Shaikh, T. A., Rasool, T., & Lone, F. R. (2022). Towards leveraging the role of machine learning and artificial intelligence in precision agriculture and smart farming. *Computers and Electronics in Agriculture, 198*, 107119. https://doi.org/10.1016/j.compag.2022.107119

Sharifi, S., Monjezi, N., & Hafezi, N. (2020). Performance of multilayer perceptron neural network models and radial-based functions in estimation of sugar-cane crop yield. *Journal of Agricultural Science and Sustainable Production, 30*(4), 213-228. (In Persian) https://doi.org10.22034/saps.2020.12313

Sharif Ahmadian, A. (2016). *Numerical models for submerged breakwaters:Coastal hydrodynamics &morphodynamics*. UK: Butterworth-Heinemann.

Sharkawy, A. N. (2020). Principle of neural network and its main types. *Journal of Advances in Applied & Computational Mathematics, 7*(1), 8-19. https://doi.org/10.15377/2409-5761.2020.07.2

Shawon, S. M., Ema, F. B., Mahi, A. K., Niha, F. L., & Zubair, H. T. (2025). Crop yield prediction using machine learning: An extensive and systematic literature review. *Smart Agricultural Technology, 10,* 100718. https://doi.org/10.1016/j.atech.2024.100718

Singh, P., & Kaur, A. (2022). A systematic review of artificial intelligence in agriculture. In Poonia, R.C., Singh, V. and Nayak, S.R. (Eds.). *Deep learning for sustainable agriculture*, (pp. 57-80). Academic press. https://doi.org/10.1016/B978-0-323-85214-2.00011-2

Som-Ard, J., Atzberger, C., Izquierdo-Verdiguier, E., Vuolo, F., & Immitzer, M. (2021). Remote sensing applications in sugarcane cultivation: A review. *Remote Sensing*, *13*(20), 4040. https://doi.org/10.3390/rs13204040

Souza, J. B. C., de Almeida, S. L. H., de Oliveira, M. F., dos Santos Carreira, V., de Brito Filho, A. L., dos Santos, A. F., & da Silva, R. P. (2025). Generalization of peanut yield prediction models using artificial neural networks and vegetation indices. *Smart Agricultural Technology, 11*, 100873. https://doi.org/10.1016/j.atech.2025.100873

Sridhara, S., Soumya, B. R., & Kashyap, G. R. (2024). Multistage sugarcane yield prediction using machine learning algorithms. *Journal of Agrometeorology, 26*(1), 37-44. https://doi.org/10.54386/jam.v26i1.2411

Taherei Ghazvinei, P., Hassanpour Darvishi, H., Mosavi, A., Yusof, K. b. W., Alizamir, M., Shamshirband, S., & Chau, K.-w. (2018). Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network. *Engineering Applications of Computational Fluid Mechanics, 12*(1), 738-749. https://doi.org/10.1080/19942060.2018.1526119

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. https://doi.org/10.1016/j.compag.2020.105709

Vinayaka & Prasad, P. R. C. (2024). AI-enhanced remote sensing applications in Indian sugarcane research: A comprehensive review. *Sugar Tech*, *26*(3), 609-628. https://doi.org/10.1007/s12355-024-01409-w

Waqas, M., Naseem, A., Humphries, U. W., Hlaing, P. T., Dechpichai, P., & Wangwongchai, A. (2025). Applications of machine learning and deep learning in agriculture: A comprehensive review. *Green Technologies and Sustainability, 3*(3), 100199. https://doi.org/10.1016/j.grets.2025.100199

Xu, X., Gao, P., Zhu, X., Guo, W., Ding, J., Li, C., Zhu, M., & Wu, X. (2019). Design of an integrated climatic assessment indicator (ICAI) for wheat production: A case study in Jiangsu Province, China. *Ecological Indicators, 101,* 943-953. https://doi.org/10.1016/j.ecolind.2019.01.059

Zaki Dizaji, H., Monjezi, N., & Sheikhdavoodi, J. (2018). Investigating effective factors on sugarcane production performance to increase the production of sugarcane using data mining. *Iranian Journal of Biosystem Engineering, 49*(3), 501-511. (In Persian) https://doi.org/10.22059/ijbse.2018.248601.665021

Zaki Dizaji, H., Shirini, K., Taheri Hajivand, A., & Monjezi, N. (2024). Modelling variables affecting the yield of sugarcane fields using deep recurrent neural network. *Iranian Journal of Biosystem Engineering*, *55*(2), 93-108. (In Persian) https://doi.org/10.22059/ijbse.2025.378958.665557

Zhou, Y., Pan, M., Guan, W., Fu, C., & Su, T. (2023). Predicting sugarcane yield via the use of an improved least squares support vector machine and water cycle optimization model. *Agriculture, 13*(11), 2115. https://doi.org/10.3390/agriculture13112115