Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 4s (2025)
https://www.jneonatalsurg.com

Radial Kernel Truncated Gradient Margin Boost Classification for Efficient Crop Yield Prediction

Mrs.C. Karkuzhali[1], Dr.R. Padmapriya[2]

[1]Research Scholar RVS Arts and Science College, Sulur. Bharathiar University

Email ID: janani14.csk@gmail.com

[2]Associate Professor, HOD of BCA Department, RVS Arts and Science College, Sulur. Bharathiar University

Email ID: padmapriya@rvsgroup.com

Cite this paper as: Mrs.C. Karkuzhali, Dr.R. Padmapriya, (2025) Radial Kernel Truncated Gradient Margin Boost Classification for Efficient Crop Yield Prediction. *Journal of Neonatal Surgery*, 14 (4s), 867- 882.

ABSTRACT

Agriculture involves cultivating land, growing crops, and raising animals for food, fiber, and other products essential for human life. In crop yield prediction, agriculture involves forecasting the amount of crop production from a given area of land. This process utilizes various methods, including historical data analysis, weather forecasting, soil conditions, and crop management practices. Accurate yield predictions help farmers make informed decisions about resource allocation, optimize crop management, and manage risks related to climate and market fluctuations. Several machine learning techniques have been developed, but timely yield prediction remains a challenging issue. A novel method called Radial Kernel Truncated Gradient Margin Boost Classification (RKTGMBC) has been developed for accurate crop yield prediction, achieving higher accuracy and lower time complexity. The main aim of the RKTGMBC method is to perform several processes such as data acquisition, preprocessing, and feature selection. Following this, crop yield prediction is performed using the selected features through an ensemble classification method. In the RKTGMBC method, the number of selected relevant features is used as input for the Truncated Gradient Margin Boost ensemble classification method. This method employs the radial basis kernel perceptron as a weak learner to analyze the data samples and provide final classification results. The Margin Boost ensemble classification method combines the results of the weak learners and applies the Truncated Gradient method to provide stable output classification results by minimizing or maximizing the margin to reduce error. In this way, accurate crop yield prediction is achieved with minimal computational time. Experimental evaluation considers factors such as crop yield prediction accuracy, precision, recall, F1 score, and prediction time with respect to the number of data samples. The quantitatively analyzed results indicate that the proposed RKTGMBC method achieves higher crop yield prediction accuracy with minimal computation time compared to conventional techniques.

Keywords: *Crop yield prediction, Margin Boost Classification, radial basis kernel perceptron, truncated gradient method*.

1. INTRODUCTION

Agriculture involves forecasting the amount of crop production that harvested from a given area of land. This process uses various methods such as historical data analysis, weather forecasting, soil conditions. Advanced machine learning algorithms are often employed to improve accuracy. Accurate yield predictions help farmers make informed decisions about resource allocation for effective agricultural planning, resource management, and ensuring food security. The AdaBoost algorithm with Gray Level Co-occurrence Matrix (AdaBoost GLCM) was developed in [1] aimed to improve the crop yield prediction accuracy by the means of feature selection. But, the designed AdaBoost GLCM did not reduce the computational complexity in crop yield prediction. A Random Forest Extreme Gradient (RFXG) method was developed in [2] for cotton yield prediction based on observed weather data to handle large-scale datasets. However, it failed to consider the influence of various meteorological indices, such as soil temperature and humidity, to further enhance prediction performance.

Machine learning and deep learning models were developed in [3] to enhance crop yield prediction and reduce the mean absolute error. However, it failed to handle larger datasets and incorporate more historically accurate environmental and weather data for each crop, making it difficult to identify the best-performing model. Machine Learning (ML) techniques and Random Forest Regression models were developed in [4] to estimate crop yield with high accuracy and minimal error. However, reducing the crop yield prediction time remained a significant challenge. A machine learning model using regression and a deep learning model were developed in [5] with the aim of forecasting agricultural yields with minimal error. But, it failed to improve the model's accuracy in crop production prediction when applied to a large dataset.

The attention-based convolutional neural network developed in [6] was designed to analyze and predict brinjal crop yield, improving detection performance. But, it failed to focus on optimizing the algorithm's efficiency through suitable assessment. A long short-term memory (LSTM) recurrent neural network and a 1-dimensional convolutional neural network (1DCNN) were developed in [7] with the aim of crop yield forecasting by automatically learning the features. However, it did not succeed in enhancing accuracy while minimizing time. The Extreme Gradient Boosting (XGBoost) model was designed in [8] for soybean yield prediction based on multidimensional feature engineering. But, it failed to provide operational and timely predictions for soybean yield.

The Crop Yield Prediction Algorithm (CYPA) was designed in [9] by using IoT for precision agriculture to improve the efficiency and accuracy of crop yield prediction. However, it faces significant challenges when applied to higher-dimensional datasets. A Long Short-Term Memory (LSTM) model was designed in [10] to predict crop yield accurately with minimal training error. But, a detailed evaluation of feature extraction was not conducted to enhance the performance of the prediction model. A hybrid approach using a deep learning model was designed in [11] to forecast corn yield at different growth phases and with various features. However, it failed to deliver accurate results due to time constraints. A hybrid machine learning model was developed in [12] using IoT for yield prediction, focusing on preprocessing and feature selection. But, it did not incorporate various parameters such as soil nutrients, soil quality, irrigated area, and agricultural points, which could have enhanced the system's accuracy. A Modified Multi-Layer Perceptron model was developed in [13] to create an effective model for accurately predicting maize crop yield. But, it did not incorporate meteorological factors that change over the seasons for achieving precise outcomes. A Convolutional Neural Network and Recurrent Neural Network (CNN-RNN) were developed in [14] with the aim of predicting cocoa crop yield. However, this approach did not improve accuracy compared to more advanced network architectures. A Deep learning techniques were developed in [15] with the aim of timely and accurate crop yield prediction. However, the integration of remote sensing data with crop growth models was not considered to enhance the predictive performance of the models.

1.1 Key contributions of the article

The key contributions of RKTGMBC method are listed below.

- To enhance crop yield prediction, a novel RKTGMBC method has been developed based on classification.
- To minimize the prediction time, the RKTGMBC method employs perform data preprocessing and relevant feature selection.
- To improve crop yield prediction accuracy and minimize the root mean square error, the RKTGMBC method utilizes the Margin Boost ensemble classification technique for precise classification, aided by a Radial Basis Kernel Perceptron. In addition, the truncated gradient method is applied to provide stable classification results and minimize errors.
- Finally, a comprehensive experimental assessment is carried out, incorporating a variety of performance metrics, to demonstrate the improvement of the RKTGMBC method over conventional methods.

1.2 Paper Organization

The paper is organized into six various sections as follows: Section 2 discusses related works. Section 3 describes the proposed RKTGMBC method with different processes. The detailed experimental setup and dataset description is presented in Section 4. Performance metrics description is presented in section 5. The comparison analyses of different methods are discussed in section 6 with various metrics. Finally, Section 7 provides the conclusion of paper.

2. RELATED WORKS

The integration of machine learning techniques was developed in [16], with climate and remote sensing data to provides more accurate yield predictions. But the time complexity of the yield predictions was not minimized. The multimodel ensemble (MME) using a particle filtering (PF) algorithm was designed in [17] for accurate, season-based crop yield prediction. An integration of the 2D-CNN and LSTM models was developed in [18] for crop yield prediction. However, it failed to consider climatic data for achieving better accuracy in forecasting crop yield. A new Partial Domain Adversarial Neural Network (PDANN) was developed in [19] to significantly improve crop yield in heterogeneous regions. But, it did not investigate integrating the PDANN model with farm management systems to enhance yield monitoring. A machine-learning-based maize yield prediction model was designed in [20] using domain adaptation for modern agriculture monitoring, helping to achieve food security and sustainability.

A deep transfer learning model was designed in [21] for crop yield prediction with higher accuracy at different locations and scales. In [22], a deep transfer learning framework was developed for soybean yield prediction. However, it did not incorporate a regularized transfer learning model to enhance yield predictions across different regions. An integration of Generative Adversarial Networks (GANs) and Convolutional Neural Networks (CNNs) was developed in [23] for growth monitoring of winter wheat and yield estimation. However, the accuracy of yield prediction was not improved. The optimization of LSTM and Bi-LSTM models was developed in [24] for crop yield prediction with improved accuracy. However, the computational time for these deep learning models was higher.

A novel approach was developed in [25] for predicting crop yields by integrating a feature selection method with an optimized support vector regression model. This approach was designed to enhance both prediction accuracy and computational efficiency. A Bayesian spatially varying functional model (BSVFM) was developed in [26] to predict county-level corn yield. However, missing values at random increase the risk of inaccurate crop yield prediction. A novel prediction system based on machine learning was developed in [27] to forecast the yield of different crops at the country level based on weather data. A fuzzy hybrid ensemble classification model was developed in [28] using remote sensing data to enhance crop yield estimation with minimal processing time. A stochastic model was developed in [29] based on the Monte Carlo method for predicting rice and wheat crop yields. But, achieving better accuracy and minimizing error were major challenges. A novel prediction system based on machine learning was developed in [30] to forecast the yield of different crops at the country level based on weather data.

3. PROPOSAL METHODOLOGY

Agriculture plays a vital role in supporting the country's economy and satisfying a large portion of its food requirements. However, due to the significant climatic changes, has created challenges in maintaining a stable food supply chain. To address these challenges, various scientific methods have been integrated into agriculture to ensure a balance between food supply and demand. The unpredictable climate conditions make it increasingly difficult for farmers to adopt sustainable and adaptable practices. In response, modern technology and innovative farming techniques are becoming essential for predicting the crop yields. But accurately estimating crop production is therefore critical for identifying potential threats to food security. In this paper, a novel RKTGMBC method is employed for accurate crop yield prediction through the ensemble learning model. This RKTGMBC method combines the strengths of multiple classifiers, utilizing radial basis kernel-based techniques and margin boosting to improve predictive performance by analyzing the features. By integrating RKTGMBC models to handle complex patterns in agricultural data, farmers and decision-makers predicts yield outcomes more reliably, even during fluctuating environmental conditions.

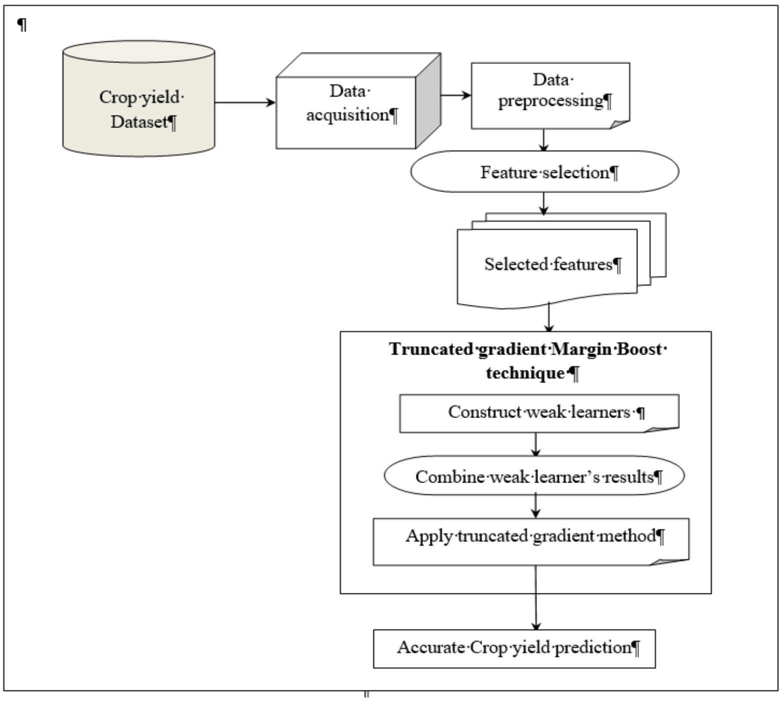

Figure 1 architecture of proposed RKTGMBC method

Figure 1 above demonstrates the architecture diagram of the proposed RKTGMBC method for accurate prediction of crop yield. The proposed method collects the number of features $f_1, f_2, \ldots f_n$ and the overall samples or instances or data 'D' are collected from the Crop Yield Prediction Dataset https://www.kaggle.com/datasets/patelris/crop-yield-prediction-dataset?resource=download. This dataset includes eight features and 28242 instances for accurate prediction of crop yield. After the data acquisition, the proposed method performs preprocessing, feature selection. With the number of selected features from the dataset is given to the truncated gradient Margin Boost technique for accurate classification of the data samples. Based on the data classification, the crop yield prediction is obtained in particular area.

The crop yield prediction process is explained briefly in the subsections below.

3.1 Preprocessing and feature selection

Preprocessing is a vital step that ensures the raw data samples is transformed into a clean, structured, and suitable format for accurate classification. Appropriate preprocessing improves model accuracy and performance by addressing issues such as missing data and normalization in the dataset. Missing data refers to the absence of values in a dataset.

Let us consider the number of data samples $D_1, D_2, D_3, \ldots D_m$ taken from the dataset and it given as input to the weak learners. The data samples and the selected significant features are arranged in the dataset as follows. The input dataset 'DS' is formulated in the form of matrix as given below.

$$M = \begin{bmatrix} f_1 & f_2 & \ldots & f_n \\ D_{11} & D_{12} & \ldots & D_{1n} \\ D_{21} & D_{22} & \ldots & D_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ D_{m1} & D_{m2} & \ldots & D_{mn} \end{bmatrix} \qquad (1)$$

Where, 'M' denotes an input matrix, where 'n' denotes a column represents the features $f_1, f_2, \ldots f_n$ and the overall data samples 'D' stored in the 'm' row respectively. The proposed technique utilizes the weighted average model to handle the missing data in the given dataset.

$$WA = \frac{\sum_{i=1}^{m} D_i * W_i}{\sum_{i=1}^{m} W_i} \qquad (2)$$

Where, WM refers to a weighted average results, D_i indicates a known data samples in the given dataset, W_i designates a weight assigned to data samples 'D_i'. Followed by, the normalization process is expressed as follows,

$$D_N = \frac{\sum_{i=1}^{m} D_i - \mu}{SD(D)} \qquad (3)$$

Where, D_N indicates a data normalization output, D_i indicates a data samples in dataset after handling the missing value, μ denotes a mean of feature value, $SD(D)$ indicates a standard deviation, m indicates a total number of data samples.

Followed by, significant feature selection process is carried out to minimize the classification time. Weighted Decay Regression method is employed for minimizes the squared residuals between the target variable and the input variables with a regularization term.

$$R = \arg\min \quad [|Y - \alpha M|^2 + \vartheta_2 \|\beta\|^{\,2}\,] \quad (4)$$

Where, R denotes a regression outcome, Y denotes a target variable, α indicates a weight parameter, M indicates an input data matrix, ϑ_2 ridge regularization parameter, $\|\beta\|^{\,2}$ indicates L2 norm (i.e. squared) of the coefficients vector 'β'. From the regression outcome, the features with minimal deviation from the target variable are considered more relevant to the target variable and discard the other features.

3.2 Truncated gradient Margin Boosting Classification

Truncated gradient Margin Boosting is an ensemble method used in machine learning and to enhance the performance and robustness of predictive models by combining multiple weak learners to create a strong learner. The main aim of boosting technique is to reduce errors in the weak learner. A weak learner is a base cassation model provides less accurate classification results. **A strong learner** is a model that has high predictive accuracy and performs well and achieves robust and reliable performance in classification. Therefore, the proposed RKTGMBC method utilizes the truncated gradient margin boosting technique for accurate classification of the data samples. Compared to other boosting technique, the proposed margin Boosting is a specialized variant of the boosting algorithm designed to enhance the performance of predictive models by focusing on improving margins and addressing issues related to error.

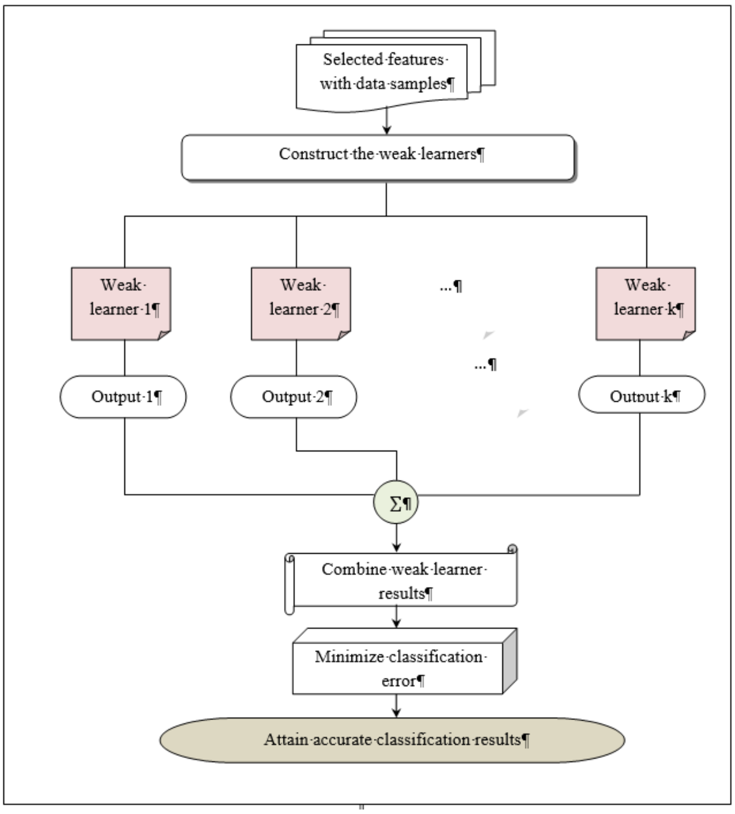

Figure 2 Structure of margin boost classification method

Figure 2 illustrates the process of the margin boost classification method. This proposed boosting ensemble technique first constructs 'k' weak learners, which are base classifiers using a radial basis function (RBF) kernel perceptron. The algorithm uses the selected significant features from the training dataset $\{D_i, Y\}$ as input for the weak learners. In this training set, D_i represents the input training samples, and Y represents the output labels for the ensemble classification methods.

Then the Radial basis function (RBF) kernel perceptron is used as weak learner for classifying the data samples by transforming the input space into a higher-dimensional feature space where the data samples are linearly separable. The perceptron is a simple linear base classifier used for classification tasks by measuring the similarity between the data samples through the Radial basis kernel function. The perceptron is used for combining a set of weights with the input vector.

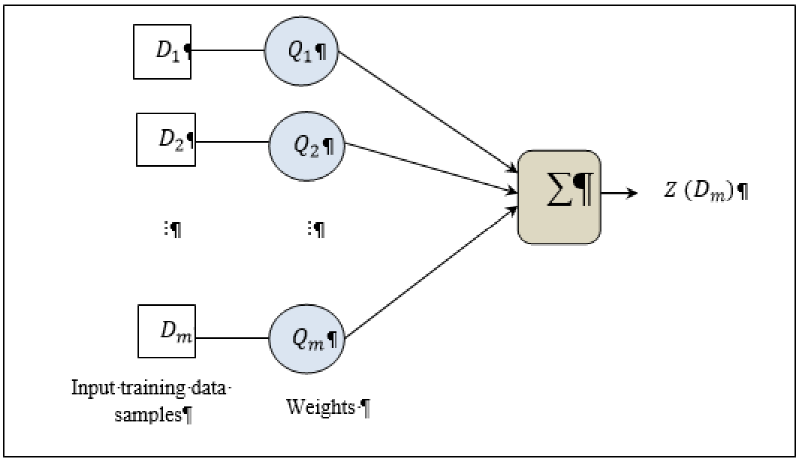

Figure 3 process of Radial basis function (RBF) kernel perceptron

Figure 3 depicts the process of Radial basis function (RBF) kernel perceptron and it iteratively improves a classification performance by running it on training data samples 'D_i'', then updating the model whenever it determines an incorrect classification. At first, it initializes weight '$Q_i = 0$' and loss threshold '$L = \vartheta$'. Then, perceptron determines a weighted sum of similarities between input data samples 'D_i' and testing data samples 'D_T' is given below

$$Z\ (D_m) = sgn \sum_{i=1}^{m} Q_i\ \ \varphi(D_i, D_T) \qquad (5)$$

Where, '$Z\ (D_m)$' represents the perceptron classification result, '$\varphi(D_i, D_T)$' denotes the radial basis kernel function that determines the similarity between input training data samples 'D_i' and testing data samples 'D_T'. Here, 'Q_i' indicates the weights for the training input data samples and 'sgn' is a sign function which evaluates whether the classified output is positives or negative.

The Radial basis kernel is measured between the two data samples in input space.

$$\varphi(D_i,\ D_T)\ = \exp\left[-0.5 * \sum_{i=1}^{m}\left(\frac{|D_i - D_T|^2}{\sigma^2}\right)\right] \qquad (6)$$

Where, $\varphi(D_i,\ D_T)$ denotes an outcome of Radial basis kernel between the training data samples 'D_i' such as temperature, rainfall, year of crop yield, particular crop and area etc and testing data samples (D_T), 'σ 'refers to a standard deviation between the two data samples. The kernel function provides the similarity results from '0' to '1. From the (1), the classification outcome is obtained using current weights. Compute the loss function for each classification output.

$$L = \eta\ (Z_T\ (D_m) - \ Z(D_m)) \qquad (7)$$

Where, L indicates a loss rate, η denotes a learning rate, $Z_T\ (D_m)$ refers to the target output of the data samples, $Z(D_m)$ denotes a predicted classification output. If the loss is reached to threshold ϑ', meaning the prediction is correct. Otherwise, the prediction is not correct. It states that if the loss function is not equal to threshold, this condition trigger weight updates in the learning algorithm.

$$Q_{new} = Q_i + \eta\ \left(Z_T\ (D_m) - \ Z(D_m)\right) * D_i \qquad (8)$$

Where, Q_{new} indicates a updated weight, Q_i refers to current weight, η indicates a learning rate, $Z_T\ (D_m)$ refers to the target output of the data samples, $Z(D_m)$ denotes a predicted class output, D_i denotes a input data samples. This process is repeated for each misclassified data samples until the model converges to correctly classify all the data samples.

In order to improve the accuracy of classification ad minimize the false negative rate, combining all the weak learners' results. The output of strong classifier is expressed as follows,

$$Y = \sum_{r=1}^{k} Z_k\ (D_m) \qquad (9)$$

Where, 'Y' denotes the output of strong classier,$Z_k\ (D_m)$ denotes an output of 'k^{th}' weak learner results. Followed by, each weak learner are weighted based on their error rate as given below,

$$\varphi_k\ = \frac{1}{2}\ln\left(\frac{1-L_k}{L_k}\right) \qquad (10)$$

Where, L_k denotes a classification loss or error rate, ln denotes a natural logarithm, φ_k indicates a weight assigned to the k^{th} weak learner. Iteratively, the boost classifier defines the margin for each weak learner results as follows,

$$m_k = \frac{Y\ \sum_{h=1}^{k} Z_k\ (D_m)\ W_k}{\sum |W_k|} \qquad (11)$$

Where, m_k denotes a margin of the weak learner 'k', Y denotes a target output classification results, $Z_k\ (D_m)$ predicted classification results of weak learner, W_k indicates a weight assigned to the k^{th} weak learner. After that, the margin truncation process is performed. By this definition, the margin is positive if the data samples are correctly classified and negative if the example is incorrectly classified. Truncated margin helps in controlling the influence of extreme values of the margin, making the classification process more stable, preventing overfitting.

To avoid extreme values in the margin, this may lead to misclassification, truncate the margin using a threshold or truncation parameter. The truncated margin introduces a limit to ensure that margins not exceed a certain level, either positively or negatively. The truncated gradient method is expressed as follows,

$$T(m_k, \beta) = \begin{cases} arg\ max(0, m_k - \beta)\ , & if\ \ m_k > 0 \\ arg\ min(0, m_k + \beta)\ , & if\ \ m_k \leq 0 \end{cases} \qquad (12)$$

Where, $T(m_k, \beta)$ denotes a truncated gradient method, β indicates a threshold or truncation parameter used to controls the degree to which the values of margin 'm_k' are truncated or limited. When the margin is positive it means the data samples is correctly classified, $arg\ max(0, m_k - \beta)$ denotes maximize the between 0 and $m_k - \beta$. It truncates the margin by ensuring that if margin 'm_k' is less than or equal to 'β', it is set to zero. If Margin 'm_k' is greater than β, the margin is reduced by β but remains positive. When the margin is negative means the data samples misclassified or on the boundary,arg min denotes an minimize between 0 and$m_k + \beta$. By limiting the minimum value to either positive or negative, this margin truncation is used to the classification model for making the learning process more stable, preventing and overfitting. This stabilization helps the model converge to a more optimal solution, thereby reducing both training and validation errors. Based on classification accurate crop yield prediction is performed. The algorithm of RKTGMBC method is described as follows,

Algorithm 1 : Radial Kernel Truncated Gradient Margin Boost Classification

Input: Dataset 'DS', selected features $f_1, f_2, \ldots f_n$, data samples $D_1, D_2, D_3, \ldots D_m$,

Output: Improve crop yield prediction accuracy

Begin

1. **for** each data samples D_i
2. Construct 'k' number of weak learners
3. **End for**
4. **for** each training data samples D_i
5. **for each** testing data samples D_T
6. Measure weighted sum of similarities using (6)
7. Obtain the classification output
8. **End for**
9. **End for**
10. **For each** classification output
11. Measure the loss using (7)
12. **If** $(L = \vartheta)$ **then**
13. Obtain accurate classification results
14. **else**
15. Update the weight using (8)
16. Go to step 6
17. **End if**
18. **End for**
19. **C**ombine all weak classifier results into strong $Y = \sum_{r=1}^{k} Z_k\ (D_m)$
20. Assign weights to weak classifier 'φ_k'
21. **For** each $Z_k\ (D_m)$
22. Define margin using 'm_k'
23. **If** $(m_k\ is\ positive\ )$ **then**
24. Data samples are correctly classified
25. **else**
26. Data samples incorrectly classified
27. **End if**
28. **End for**
29. Apply truncated gradient method '$T(m_k, \beta)$'
30. Controls the values of margin
31. Obtain strong classification results
32. **End for**

End

Algorithm 1 describes the process of ensemble classification to enhance crop yield prediction accuracy while minimizing time complexity. Initially, the ensemble classifier constructs 'k' weak learners for each input data sample. Each weak learner measures the weighted similarity between the training and testing data samples using the radial basis kernel function. Based on the similarity measure, classification results are obtained. For each classified result, the loss rate is computed. If the estimated loss value equals the threshold value, the classifier provides accurate classification results. Otherwise, the weights are updated, and the similarity measure is repeated to re-evaluate the loss value against the threshold. This process generates classification results. After classifying the data samples, the results from the weak learners are combined to form a strong classifier. Similar weights are assigned to a set of weak classifiers based on their loss values. For each set of classifiers, margins are assigned to provide accurate classification results through the truncated gradient method. The ensemble classifier identifies the best weak learner with the minimum loss. Based on these classification results, the crop yield prediction is performed accurately.

4. EXPERIMENTAL SETUP

Experimental evaluation of the proposed RKTGMBC method and AdaBoost GLCM [1], RFXG [2] are implemented using Python high level programming language. To conduct the experiment, crop yield prediction dataset is collected from the https://www.kaggle.com/datasets/patelris/crop-yield-prediction-dataset?resource=download. This dataset includes eight features and 28242. For the experimental consideration, the numbers of data samples are taken in the ranges from 2500, 5000, 7500 ….25000. Table 1 given below provides the features description.

Table 1 features description

S.No	features	Description
1.	S.No	Serial number
2.	Area	Country
3.	Item	Crops
4.	Year	Year of crop yield
5.	hg/ha_yield	Crop yield
6.	Pesticides_tonnes	Pesticides used per tonne
7.	Average_rain_fall_mm_per_year	Average rain fall
8.	avg_temp	Average Temperature

5. EVALUATION METRICS

In this section, various metrics, including crop yield prediction accuracy, precision, recall, root mean square error, crop yield prediction time are described with the mathematical formulation.

Crop yield prediction accuracy: It is measured as the ratio of accurately predicting the crop yield from the total number of data samples. Therefore, accuracy is formulated as follows:

$$CYPA = \left(\frac{Tp+Tn}{Tp+Tn+Fp+Fn}\right) * 100 \quad (13)$$

Where, $CYPA$ denotes a an crop yield prediction accuracy, Tp indicates the true positive, Tn denotes the true negative, Fp represents the false positive, Fn represents the false negative. It is measured in percentage (%).

Precision: It is measured as the ratio of true positives to the sum of true positives and false positives, indicating the proportion of correctly classified the crop yield from the total number of data samples. The precision is computed as,

$$\boldsymbol{PR} = \left(\frac{\boldsymbol{Tp}}{\boldsymbol{Tp+Fp}}\right) \quad (14)$$

Where, PR denotes a precision, Tp denotes the true positive, Fp represents the false positive.

Recall: it refers to the ability of a model to correctly classify all samples in a dataset. It is the ratio of true positive predictions to the sum of true positives and false negatives. It is mathematically formulating as follows,

$$RC = \left(\frac{\boldsymbol{Tp}}{\boldsymbol{Tp+Fn}}\right) \quad (15)$$

Where, RC denotes a recall, Tp indicates the true positive, Fn represents the false negative.

Root Mean Square Error (RMSE): it is a metric used to evaluate the accuracy of a predictive model. It represents the square root of the average squared differences between the predicted values and the actual values to the total number of data samples. It is mathematically computed as follows,

$$\boldsymbol{RMSE} = \left[\sqrt{\frac{(Y_{act} - Y_{pre})^2}{m}} \right] \qquad (16)$$

Where, $\boldsymbol{RMSE}$ indicates an root mean square error, $\boldsymbol{Y_{act}}$denotes the data samples for which actual crop yield prediction, $\boldsymbol{Y_{pre}}$ denotes a data samples for which predicted crop yield results.

Crop yield prediction time: It is measured as an amount of time taken by algorithm for predicting crop yield. The time is mathematically formulated as follows,

$$CYPT = \sum_{i=1}^{m} D_i * TM(CYP) \qquad (17)$$

Where, $CYPT$ denotes a crop yield prediction time based on the patient data 'D_i' and the actual time consumed in crop yield prediction denoted by '$TM(CYP)$. It is measured in terms of milliseconds (ms).

6. PERFORMANCE METRIC ANALYSIS

In this section, performance of the RKTGMBC method and AdaBoost GLCM [1] RFXG [2] are evaluated with various metrics, including crop yield prediction accuracy, precision, recall, root mean square error, crop yield prediction time with different number of samples.

Table 2 comparison of crop yield prediction accuracy

Number of samples	Crop yield prediction accuracy (%)		
	RKTGMBC	AdaBoost GLCM	RFXG
2500	95.2	88	90.4
5000	94.89	87.56	89.05
7500	95	88.52	90
10000	95.06	87.96	89.56
12500	94.89	88.21	90.74
15000	94.25	89.45	90.98
17500	95.12	88.22	90.33
20000	95	87.45	89.52
22500	95.89	88.45	90
25000	94.87	87.63	89.44

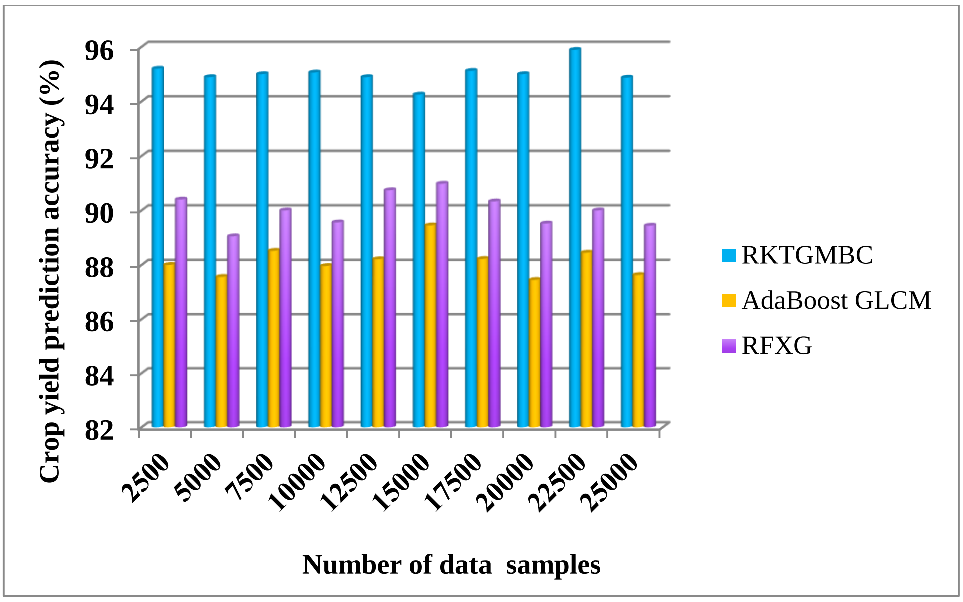

Figure 4 graphical illustration of crop yield prediction accuracy

Figure 4 depicts the graphical illustration of crop yield prediction accuracy versus the number of data samples collected from the dataset for accurate crop yield prediction. The number of data samples considered on the horizontal axis, ranges from 2500 to 25000, while the accuracy of three methods namely RKTGMBC method and AdaBoost GLCM [1] RFXG [2] are shown on the vertical axis. The obtained overall results indicate that the crop yield prediction accuracy of the RKTGMBC method is higher compared to the existing methods [1] and [2]. For instance, with 2500 data samples, the RKTGMBC method achieved an accuracy of 95.2%, whereas the existing methods [1] and [2] achieved accuracies of 88% and 90.4%, respectively. Similar a variety of results were observed across different data samples. The average value of ten various indicates that the RKTGMBC method improves the performance of accuracy by 8% and 9% when compared to the existing methods [1] and [2], respectively. This is owing to the RKTGMBC method utilizes the truncated gradient margin boost ensemble classification method. This method utilizes the radial basis kernel perceptron as a weak learner to perform data analyzes based on the similarity measure between the training and testing data samples. After that ensemble classification method combines the results of the weak learners and applies the Truncated Gradient method to adjust the margin level for each classifier results to provide stable output classification results. This in turn enhances the accuracy in crop yield prediction.

Table 3 comparison of precision

Number of samples	Precision		
	RKTGMBC	AdaBoost GLCM	RFXG
2500	0.954	0.885	0.908
5000	0.948	0.886	0.902
7500	0.958	0.895	0.907
10000	0.942	0.892	0.908
12500	0.95	0.9	0.909
15000	0.957	0.905	0.915
17500	0.946	0.903	0.918
20000	0.955	0.91	0.927
22500	0.948	0.898	0.914
25000	0.956	0.885	0.907

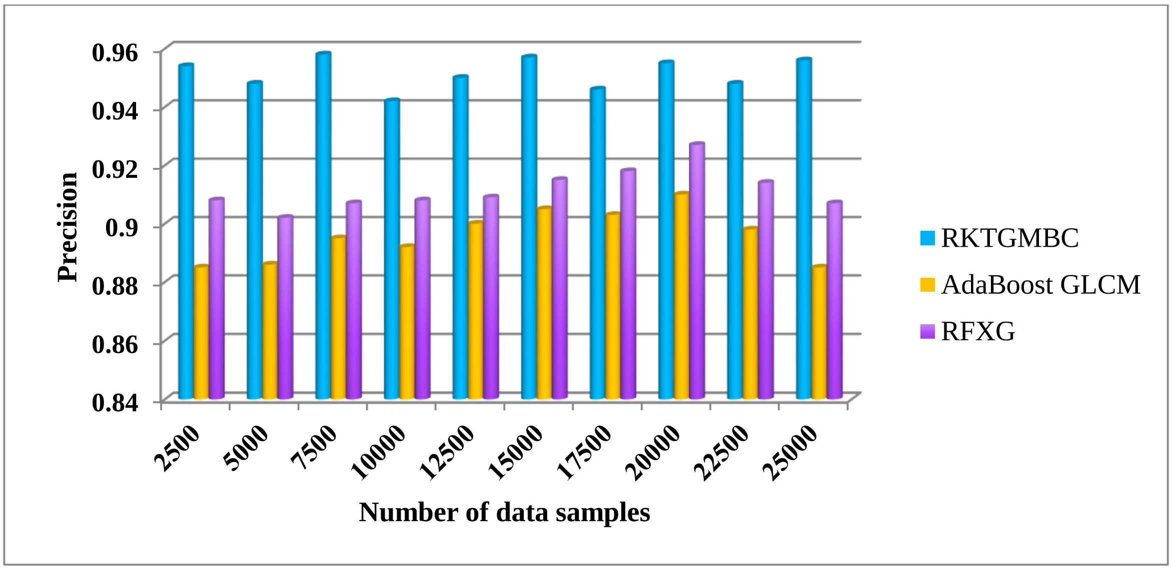

Figure 5 graphical illustration of precision

Figure 5 shows a graphical illustration of the precision of using three methods namely proposed RKTGMBC method, AdaBoost GLCM [1] and RFXG [2]. In the figure, the horizontal axis represents the number of data samples, while the vertical axis indicates the performance of precision. The results observed indicate that the proposed RKTGMBC method outperforms the existing methods [1] and [2]. In experiment conducted with 2500 data samples, the precision was found to be 0.954 for the proposed RKTGMBC method, precision was found to be 0.885 and 0.908 for the two existing methods [1] and [2], respectively. Overall, the analysis of ten performance results demonstrates that the precision achieved using the proposed RKTGMBC method is significantly enhanced by 6% compared to [1] and 4% compared to [2]. The improvement is achieved through the application of a margin-boost ensemble classifier model for predicting crop yield. This model analyzes features using a radial basis kernel function, resulting in classification with higher true positive rates and minimized false positive rates, ultimately enhancing precision.

Table 4 comparison of recall

Number of data samples	**Recall**		
	RKTGMBC	**AdaBoost GLCM**	**RFXG**
2500	0.976	0.939	0.952
5000	0.96	0.921	0.942
7500	0.968	0.915	0.935
10000	0.972	0.918	0.928
12500	0.965	0.91	0.932
15000	0.958	0.9	0.92
17500	0.968	0.898	0.924
20000	0.962	0.911	0.932
22500	0.975	0.925	0.945
25000	0.97	0.92	0.94

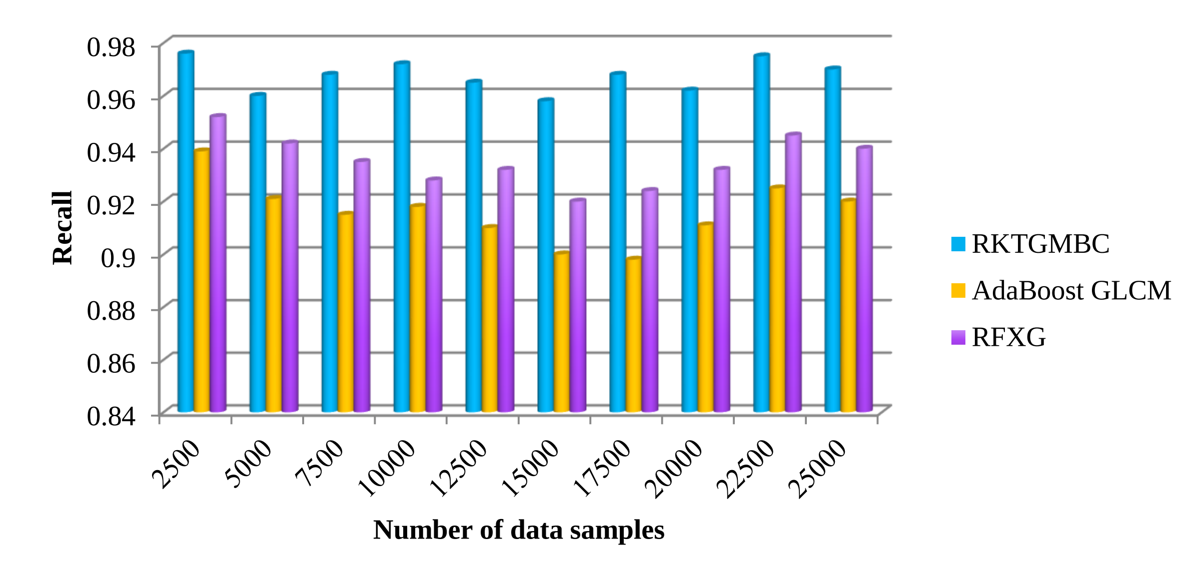

Figure 6 graphical illustration of recall

Figure 6 illustrates the performance outcomes of recall against the number of data samples, ranging from 2500 to 25000 taken from dataset. To calculate recall, three methods were employed namely proposed RKTGMBC method, AdaBoost GLCM [1] and RFXG [2]. The horizontal axis indicates the number of data samples, while the vertical axis indicates recall. The experimental results demonstrate that the RKTGMBC method achieved improved recall compared to the other two existing techniques. For each method, a variety of results were observed with different counts of input samples. The observed results of the RKTGMBC method model were compared with the existing techniques. The overall comparison shows that the performance of recall using RKTGMBC method in accurately predicting the crop yield is enhanced by 6% compared to [1] and 3% compared to [2] respectively.

Table 5 comparison of root mean square error

Number of data samples	**Root mean square error**		
	RKTGMBC	**AdaBoost GLCM**	**RFXG**
2500	0.096	0.24	0.192
5000	0.072	0.175	0.154
7500	0.057	0.132	0.115
10000	0.049	0.120	0.104
12500	0.045	0.105	0.082
15000	0.046	0.086	0.073
17500	0.036	0.089	0.073
20000	0.035	0.088	0.074
22500	0.027	0.077	0.066
25000	0.032	0.078	0.066

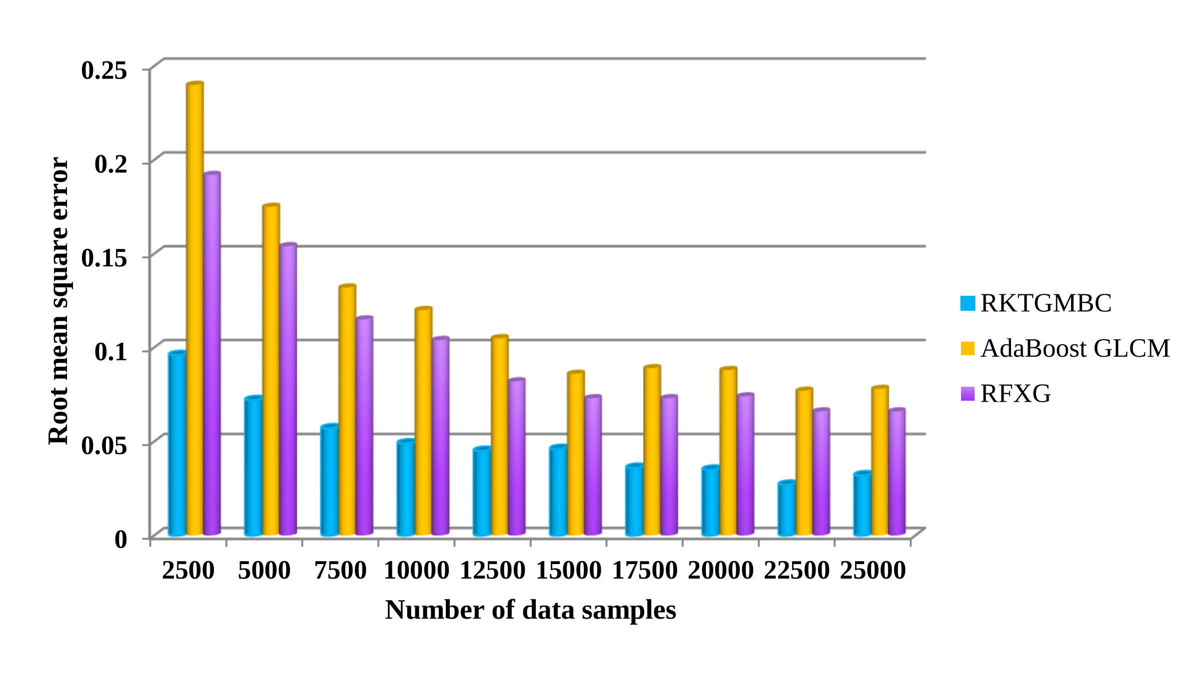

Figure 7 graphical illustration of root mean square error

In Figure 7, the graphical illustrate of root mean square error (RMSE) are illustrated versus the number of data samples, ranging from 2500 to 25000. Three methods namely the RKTGMBC method, AdaBoost GLCM [1] and RFXG [2] are employed to evaluate RMSE. The horizontal axis represents the number of data samples, while the vertical axis represents the RMSE. The results demonstrate that the RKTGMBC method achieves minimal RMSE compared to the other two conventional ensemble methods. Let us consider the number of data samples to be 2500 in the first run. By applying the RKTGMBC method, the RMSE was found to be 0.096, 0.24 for [1] and 0.192 for the [2]. Similar performance outcomes were obtained for each method with varying number of data samples. The overall comparison reveals that the RMSE performance in accurately predicting the crop yield is minimized by 58% compared to [1] and by 50% compared to the [2] when applying the RKTGMBC method. This improved performance is achieved through the application of the truncated gradient method to the results of weak learners. This method adjusts the margin of classification results to minimize the error rate during the classification process. This, in turn, minimizes the RMSE.

Table 6 comparison of crop yield prediction time

Number of data samples	**Crop yield prediction time (ms)**		
	RKTGMBC	**AdaBoost GLCM**	**RFXG**
2500	27.75	34.25	33.25
5000	29.6	36.7	34.3
7500	31.5	38.4	36.7
10000	33.7	41.5	39.8
12500	35.9	44.6	42.3
15000	38.5	47.8	45.8
17500	40.3	51.6	49.7
20000	42.9	53.4	51.9
22500	48.5	55.8	54.2
25000	50	58.6	57

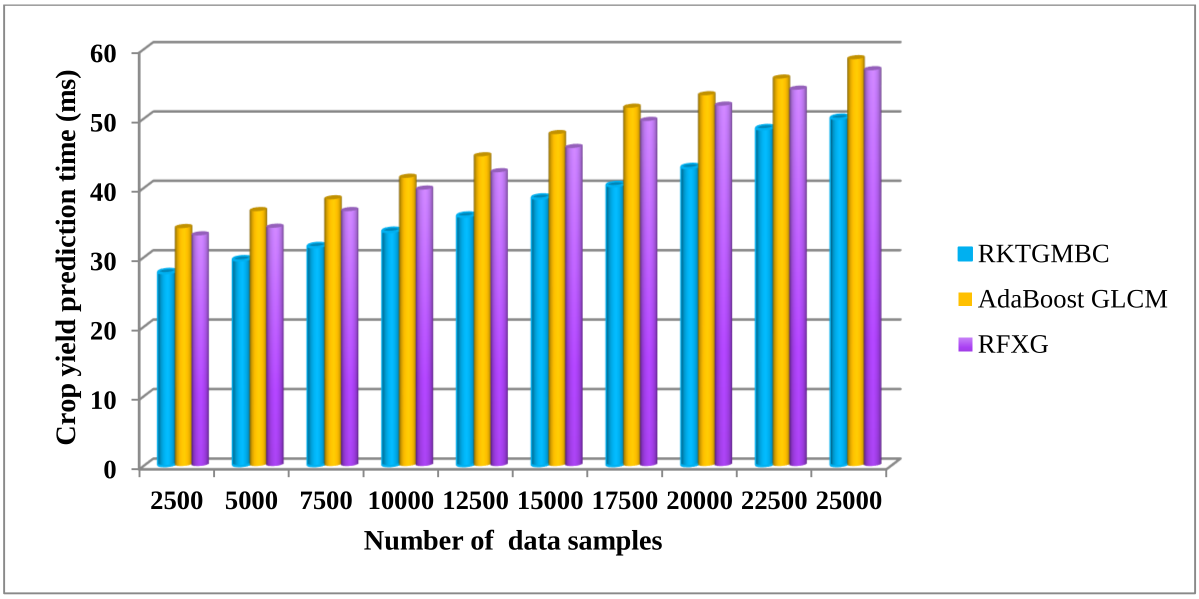

Figure 8 graphical illustration of crop yield prediction time

Figure 8 illustrates the performance analysis of crop yield prediction time using three methods namely RKTGMBC method, AdaBoost GLCM [1] and RFXG [2]. For each method, a simulation of 10 runs was performed with 25000 distinct images collected from the dataset. From the above figure, increasing the number of data samples, the time incurred for crop yield prediction was also found to be increased. However, the prediction time using the RKTGMBC method was found to be minimized when compared to [1] and [2] respectively. In the first iteration with 2500 data samples, the prediction time for the RKTGMBC method was found to be $27.75ms$. Similarly, the time consumption for methods [1] and [2] was found to be $34.25\ ms$ and $33.25\ ms$, respectively. The overall results obtained from the RKTGMBC method are compared to the results of the existing methods. The comparison illustrates that the performance of crop yield prediction time using the RKTGMBC method is significantly minimized by 18% and 15% compared to the existing methods [1] and [2]. The reason behind this is to perform feature selection. The proposed ensemble classifier model utilizes the selected significant features to enhance the accuracy of crop yield prediction. This process minimizes the time consumption involved in crop yield prediction.

7. CONCLUSION

Accurately predicting crop yields, particularly in large areas, is essential for addressing global food security challenges. This paper introduces an ensemble learning method called RKTGMBC, focusing on precise crop yield estimation with minimal time consumption in the agriculture domain. The RKTGMBC method first performs data preprocessing by handling missing data, normalization, and selecting significant features, which reduces the time complexity of crop yield prediction. Subsequently, the Truncated Gradient Margin Boost ensemble classification method is employed to classify data samples based on the radial basis kernel perceptron, providing final crop yield prediction results. A comprehensive experimental assessment is conducted using various performance metrics, such as crop yield prediction accuracy, precision, recall, root mean square error, and prediction time concerning the number of data samples. The overall performance results demonstrate that the proposed RKTGMBC method achieves improved accuracy with minimal error and time consumption compared to conventional ensemble methods.

REFERENCES

[1] O. Sri Nagesh, Raja Rao Budaraju, Shriram S. Kulkarni, M. Vinay, Samuel-Soma M. Ajibade, Meenu Chopra, Malik Jawarneh, Karthikeyan Kaliyaperumal, "Boosting enabled efficient machine learning technique for accurate prediction of crop yield towards precision agriculture", Discover Sustainability, Springer, Volume 5, 2024, Pages 1-9. https://doi.org/10.1007/s43621-024-00254-x

[2] Syed Tahseen Haider, Wenping Ge, Jianqiang Li, Saif Ur Rehman, Azhar Imran, Mohamed Abdel Fattah Sharaf, And Syed Muhammad Haider, "An Ensemble Machine Learning Framework for Cotton Crop Yield Prediction Using Weather Parameters: A Case Study of Pakistan", IEEE Access , Volume 12, 2024, Pages 124045 – 124061. DOI: 10.1109/ACCESS.2024.3454511

[3] Kavita Jhajhariaa , Pratistha Mathur, Sanchit Jaina , Sukriti Nijhawan, "Crop Yield Prediction using Machine

Learning and Deep Learning Techniques", Procedia Computer Science, Elsevier, Volume 218, 2023, Pages 406-417. https://doi.org/10.1016/j.procs.2023.01.023

[4] Bharati Panigrahi, Krishna Chaitanya Rao Kathala, M. Sujatha, "A Machine Learning-Based Comparative Approach to Predict the Crop Yield Using Supervised Learning With Regression Models", Procedia Computer Science, Elsevier, Volume 218, 2023, Pages 2684-2693. https://doi.org/10.1016/j.procs.2023.01.241

[5] Priyanka Sharma, Pankaj Dadheech,Nagender Aneja, Sandhya Aneja, "Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning", IEEE Access, Volume 11, 2023, Pages 111255 – 111264. DOI: 10.1109/ACCESS.2023.3321861

[6] M Venkateswara Rao, Y. Sreeraman, Srihari Varma Mantena, Venkateswarlu Gundu, D. Roja, Ramesh Vatambeti, "Brinjal crop yield prediction using shuffled shepherd optimization algorithm based ACNN-OBDLSTM model in smart agriculture", Journal of Integrated Science and Technology, Volume 12, Issue 1, 2024, Pages 1-7. https://pubs.thesciencein.org/journal/index.php/jist/article/view/a710

[7] Dilli Paudel, Allard de Wit, Hendrik Boogaard, Diego Marcos, Sjoukje Osinga, Ioannis N. Athanasiadis, "Interpretability of deep learning models for crop yield forecasting", Computers and Electronics in Agriculture, Elsevier, Volume 206, 2023, Pages 1-14. https://doi.org/10.1016/j.compag.2023.107663

[8] Yuanchao Li, Hongwei Zeng, Miao Zhang, Bingfang Wu, Yan Zhao, Xia Yao, Tao Cheng, Xingli Qin, Fangming Wu, "A county-level soybean yield prediction framework coupled with XGBoost and multidimensional feature engineering", International Journal of Applied Earth Observation and Geoinformation, Elsevier, Volume 118, 2023, Pages 1-21. https://doi.org/10.1016/j.jag.2023.103269

[9] Fatma M. Talaat, "Crop yield prediction algorithm (CYPA) in precision agriculture based on IoT techniques and climate changes", Neural Computing and Applications, Springer, Volume 35, 2023, Pages 17281–17292. https://doi.org/10.1007/s00521-023-08619-5

[10] Usharani Bhimavarap , Gopi Battineni and Nalini Chintalapudi, "Improved Optimization Algorithm in LSTM to Predict Crop Yield", Computers, Volume 12, Issue 1, 2023, Pages 1-19. https://doi.org/10.3390/computers12010010

[11] Yiting Ren, Qiangzi Li, Xin Du, Yuan Zhang, Hongyan Wang, Guanwei Shi and Mengfan Wei , "Analysis of Corn Yield Prediction Potential at Various Growth Phases Using a Process-Based Model and Deep Learning", Plants , Volume 12, Issue 3, 2023, Pages 1-19. https://doi.org/10.3390/plants12030446

[12] Akanksha Gupta & Priyank Nahar, "Classification and yield prediction in smart agriculture system using IoT", Journal of Ambient Intelligence and Humanized Computing, Volume 14, 2023, Pages 10235–10244. https://doi.org/10.1007/s12652-021-03685-w

[13] Shakeel Ahmed, "A Software Framework for Predicting the Maize Yield Using Modified Multi-Layer Perceptron", Sustainability, Volume 15, Issue 4, 2023, Pages 1-19. https://doi.org/10.3390/su15043017

[14] Sunday Samuel Olofintuyi, Emmanuel Ajayi Olajubu, Deji Olanike, "An ensemble deep learning approach for predicting cocoa yield", Heliyon, Volume 9, Issue 4, 2023, Pages 1-20. https://doi.org/10.1016/j.heliyon.2023.e15245

[15] Shitong Zhou, Lei Xu and Nengcheng Chen, "Rice Yield Prediction in Hubei Province Based on Deep Learning and the Effect of Spatial Heterogeneity", Remote Sensing, Volume 15, Issue 5, 2023, Pages 1-17. https://doi.org/10.3390/rs15051361

[16] Muhammad Ashfaq, Imran Khan, Abdulrahman Alzahrani, Muhammad Usman Tariq, Humera Khan, and Anwar Ghani, "Accurate Wheat Yield Prediction Using Machine Learning and Climate-NDVI Data Fusion", IEEE Access, Volume 12, 2024, Pages 40947 – 40961. DOI: 10.1109/ACCESS.2024.3376735

[17] Hossein Zare, Tobias KD Weber, Joachim Ingwersen, Wolfgang Nowak, Sebastian Gayler, Thilo Streck, "Within-season crop yield prediction by a multi-model ensemble with integrated data assimilation", Field Crops Research, Elsevier, Volume 308, 2024, Pages 1-16. https://doi.org/10.1016/j.fcr.2024.109293

[18] Seyed Mahdi Mirhoseini Nejad , Dariush Abbasi-Moghadam , Alireza Sharifi , Nizom Farmonov , Khilola Amankulova , and Mucsi Lászl´z, "Multispectral Crop Yield Prediction Using 3D-Convolutional Neural Networks and Attention Convolutional LSTM Approaches", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing , Volume 16, 2022, Pages 254 – 266. DOI: 10.1109/JSTARS.2022.3223423

[19] Yuchi Ma, Zhengwei Yang, Qunying Huang and Zhou Zhang, "Improving the Transferability of Deep Learning Models for Crop Yield Prediction: A Partial Domain Adaptation Approach", Remote Sensing, Volume 15, Issue 18, 2023, Pages 1-18. https://doi.org/10.3390/rs15184562 ,

[20] Rhorom Priyatikanto, Yang Lu, Jadu Dash, Justin Sheffield, "Improving generalisability and transferability of machine-learning-based maize yield prediction model through domain adaptation", Agricultural and Forest

Meteorology, Elsevier, Volume 341, 2023, Pages 1-16. https://doi.org/10.1016/j.agrformet.2023.109652

[21] Shahid Nawaz Khan, Dapeng Li, Maitiniyazi Maimaitijiang, "Using gross primary production data and deep transfer learning for crop yield prediction in the US Corn Belt", International Journal of Applied Earth Observation and Geoinformation, Elsevier, Volume 131, 2024, Pages 1-12. https://doi.org/10.1016/j.jag.2024.103965

[22] Florian Huber, Alvin Inderka and Volker Steinhage, "Leveraging Remote Sensing Data for Yield Prediction with Deep Transfer Learning", Sensors, Volume 24, Issue 3, 2024, Pages 1-18. https://doi.org/10.3390/s24030770

[23] Jingqi Zhang, Huiren Tian, Pengxin Wang, Kevin Tansey, Shuyu Zhang, Hongmei Li, "Improving wheat yield estimates using data augmentation models and remotely sensed biophysical indices within deep neural networks in the Guanzhong Plain, PR China", Computers and Electronics in Agriculture, Elsevier, Volume 192, 2022, Pages 1-13. https://doi.org/10.1016/j.compag.2021.106616

[24] V. Kiran Kumar, K. V. Ramesh & V. Rakesh, "Optimizing LSTM and Bi-LSTM models for crop yield prediction and comparison of their performance with traditional machine learning techniques", Applied Intelligence, Springer, Volume 53, 2023, Pages 28291–28309. https://doi.org/10.1007/s10489-023-05005-5

[25] Mahmoud Abdel-salam, Neeraj Kumar, Shubham Mahajan, "A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning", Neural Computing and Applications, Springer, 2024, Pages 1-28. https://doi.org/10.1007/s00521-024-10226-x

[26] Yeonjoo Park, Bo Li & Yehua Li, "Crop Yield Prediction Using Bayesian Spatially Varying Coefficient Models with Functional Predictors", Journal of the American Statistical Association, Volume 118, Issue 541, 2023, Pages 1-14. https://doi.org/10.1080/01621459.2022.2123333

[27] Lontsi Saadio Cedric, Wilfried Yves Hamilton Adoni, Rubby Aworka, Jérémie Thouakesseh Zoueu, Franck Kalala Mutombo, Moez Krichen, Charles Lebon Mberi Kimpoloa, "Crops yield prediction based on machine learning models: Case of West African countries", Smart Agricultural Technology, Elsevier, Volume 2, 2022, Pages 1-14. https://doi.org/10.1016/j.atech.2022.100049

[28] Qazi Mudassar Ilyas, Muneer Ahmad and Abid Mehmood, "Automated Estimation of Crop Yield Using Artificial Intelligence and Remote Sensing Technologies", Bioengineering, Volume 10, Issue 2, 2023, Pages 1-24. https://doi.org/10.3390/bioengineering10020125

[29] Akram Javadi, Mohammad Ghahremanzadeh, Maria Sassi, Ozra Javanbakht, Boballah Hayati, "Impact of Climate Variables Change on the Yield of Wheat and Rice Crops in Iran (Application of Stochastic Model Based on Monte Carlo Simulation)", Computational Economics, Springer, Volume 63, 2024, Pages 983–1000. https://doi.org/10.1007/s10614-023-10389-0

[30] Alexandros Oikonomidis, Cagatay Catal & Ayalew Kassahun, "Hybrid Deep Learning-based Models for Crop Yield Prediction", Applied Artificial Intelligence, Volume 36, Issue 1, 2022, Pages 1-18. https://doi.org/10.1080/08839514.2022.2031823