Multivariate Regressive Gradient Spiral Optimized Deep Belief Learning For Crop Yield Prediction

Mrs.C.Karkuzhali[1], Dr.R.Padmapriya[2]
[1]Research Scholar RVS Arts and Science College, Sulur. Bharathiar University, janani14.csk@gmail.com
[2]Associate Professor, HOD of BCA Department, RVS Arts and Science College, Sulur, Bharathiar University, padmapriya@rvsgroup.com

Abstract

Agriculture plays an important role in financial system of many countries like India. Crop yield prediction in agriculture sector involves amount of crops estimation that are harvested from land. Different machine learning models face challenges like time-efficient crop yield prediction with improved accuracy level. In order to improve the crop yield prediction accuracy with minimal time consumption, a novel deep learning model called Multivariate Regressive Gradient Spiral Optimized Deep Belief Learning (MRGSODBL) method is developed. The proposed MRGSODBL method comprises of five methods namely data acquisition, preprocessing, feature selection, classification and hyperparmater tuning to enhance the accuracy with minimum prediction error. In the data acquisition phase, a number of data samples are collected from the dataset. After that, the proposed deep belief learning classifier is used for accurate yield prediction with minimal time consumption and error rate. The proposed deep belief learning classifier comprises of different layers, namely one input layer, one output layer and numerous hidden layers for crop yield prediction. Initially, the numbers of data samples are given to input layer of deep learning architecture. After that, the collected data are transmitted to the hidden layer 1. In that layer, data preprocessing is carried out to obtain the suitable format of the dataset by handling the missing data. Followed by, Censored Feature Regression Analysis is carried out in hidden layer 2 for selecting the suitable features to minimize the dimensionality of the dataset. Then the selected significant features are transmitted to the next hidden layer where the Multivariate Adaptive Regression Splines is employed in hidden layer 3 for crop yield prediction. Then, the adaptive gradient spiral optimization algorithm is employed in hidden layer 3 for tuning the hyperparameters of the deep belief learning classifier to minimize the error rate in the crop yield prediction. Finally, the accurate prediction results with minimal error are displayed at the output layer. Experimental evaluation considered the factors like crop yield prediction accuracy, precision, recall, specificity, root mean square error, F1 measure, and prediction time with respect to the number of data samples. The quantitatively analyzed results indicate that the proposed MRGSODBL method attains higher crop yield prediction accuracy with minimal time consumption when compared to conventional techniques.

Keywords: Agriculture, data preprocessing, hyperparameter tuning, censored feature regression, machine learning, multivariate adaptive regression spline decision tree

INTRODUCTION

Agriculture is the backbone of world economy across business for farmers. Farmers experienced the difficulties yielding crop in fields because of pesticides and crop varieties. An efficient deep learning (DL) and dimensionality reduction (DR) approach was introduced in [1] for crop yield prediction. DL approach performed preprocessing, dimensionality reduction and classification. But, the accuracy level was not improved by DL approach. An optimally driven Feedback Neural Network (FNN) approach was introduced in [2] to increase the yield prediction in intercropping systems. But, the prediction time was not reduced by FNN approach. AdaBoost algorithm with Gray Level Co-occurrence Matrix (AdaBoost GLCM) was introduced in [3] to improve the crop yield prediction accuracy. However, the designed AdaBoost GLCM failed to reduce the computational complexity. A new vegetation index termed triple red-edge index (TREI) was introduced in [4] to control the red-edge bands by Sentinel-2 satellite. TREI improved the crop yield prediction accuracy. But the time complexity remained unaddressed. A novel multi-modal learning approach was introduced in [5] to forecast the crop yield for different crops. A crop modeling approach was introduced in [6] with cutting-edge computational tools for crop production monitoring. The designed approach increased the rice yield prediction based on strengths and weaknesses. An advanced data-driven method was designed in [7] to study relationship between the yields

of different crops. MODIS-derived gross primary production and deep transfer learning was introduced in [8] to forecast the crop yield. The deep neural network and gated recurrent unit was employed to forecast the corn and soybean yields. Kernel Principal Component Analysis (KPCA) was introduced in [9] to extract the relevant feature. Least Absolute Shrinkage and Selection Operator (LESSO) Regression was used to perform efficient feature selection. A new deep neural networks framework was introduced in [10] to forecast the crop yields in Australia through considering climate change, fertilizer use, and crop area. The issues identified from the above literature are lesser prediction accuracy, higher prediction time consumption, higher computational complexity, higher computational cost, higher root mean square error, lesser precision, lesser recall, lesser f-measure and so on. In order to address these issues, a novel deep learning model called Multivariate Regressive Gradient Spiral Optimized Deep Belief Learning (MRGSODBL) method is introduced.

1.1 Research Contribution:

The main contribution of the article is given as,

➢ The proposed MRGSODBL method performed the crop yield prediction through data acquisition, data preprocessing, feature selection, classification and hyperparmater tuning to enhance the accuracy with minimum prediction error.

➢ To handle the missing data, the data preprocessing is carried out in suitable format of the dataset.

➢ To minimize the dimensionality, Censored Feature Regression Analysis is carried out in MRGSODBL method for selecting the suitable features.

➢ To increase the crop yield prediction accuracy, Multivariate Adaptive Regression Spline Decision Tree is employed in MRGSODBL method.

➢ To minimize the error rate in crop yield prediction, the adaptive gradient spiral optimization algorithm is employed for tuning hyperparameters of deep belief learning classifier.

➢ The experimental results indicate that proposed MRGSODBL method attains higher crop yield prediction accuracy with minimal time consumption when compared to conventional techniques.

1.2 Paper Organization:

The paper is organized into six various sections as follows: Section 2 reviews the related works. Section 3 explains the proposed MRGSODBL method with neat architecture diagram. The experimental setup and performance metrics description is given in section 4. The comparative analysis of different existing methods and proposed technique are discussed in section 5. Finally, Section 6 provides the conclusion of paper.

RELATED WORKS

An efficient deep learning (DL) and dimensionality reduction (DR) approach was designed in [11] for performing efficient crop yield prediction for Indian regional crops. The transfer-learning-based framework (TrG2P) was introduced in [12] with convolutional neural networks (CNN) to train prediction model with phenotypic and genotypic data. A random forest algorithm was designed in [13] with the standard statistical technique like linear regression on clean information. Multi-attribute weighted tree-based support vector machine (MAWT-SVM) method was introduced in [14] to forecast the crop yield. The genetic algorithm (GA) was employed to select the best features. An Augmented Reality (AR) based mobile application was introduced in [15] to obtain the information from the tomato farm in pre-harvest stage. Gorilla Troops Optimization with Deep Learning-based Crop Recommendation and Yield Prediction model (GTODL-CRYPM) was introduced in [16] for performing efficient crop recommendation and crop prediction. A meta-knowledge-guided framework was designed in [17] with different agricultural datasets. The designed framework discovered the meta-knowledge transfer in hyperparameter optimization scenarios. Gaussian kernel regression was designed in [18] for rice yield estimation from optical and SAR imagery with the ground truth data. A new hybrid feature selection approach was designed in [19] with optimized SVR model to improve the prediction accuracy efficiently. ICOA optimized the hyperparameters with higher prediction accuracy. An advanced crop yield

forecasting model was designed in [20] with deep reinforcement learning and remote sensing data. Hybrid Reinforcement-Supervised Learning (HRSL) model was employed for classification and decision-making. The cotton prediction framework termed Random Forest Extreme Gradient (RFXG) was introduced in [21] to predict the cotton production depending on observed data. Crop Yield Prediction Algorithm (CYPA) was designed in [22] through IoT technique in precision agriculture. Crop yield simulation simplified the cumulative impacts of field variables. A hybrid Deep Neural Network-Recurrent Neural Network (DNN-RNN) framework was designed in [23] with DNN and RNN concepts in temporal and spatial information. Independent Component Analysis (ICA) extracted the relevant features. An improved optimizer function (IOF) was introduced in [24] to carry out the accurate prediction with long short-term memory (LSTM). The manual data was gathered from the local agricultural departments for training. A new crop yield prediction system was designed in [25] with meteorological data, pesticide record and crop yield data. The classification-based crop yield prediction was introduced in [26] with Long Short-Term Memory (LSTM) and Attention Mechanism. Variance Inflation Factor Algorithm (VIF) selected the relevant features. An efficient method was designed in [27] for soil analysis and crop yield prediction in intelligent agriculture.

An expert system model termed precise ensemble expert system for crop yield prediction (PEESCYP) was introduced in [28] to forecast the best crop for agricultural land. Machine-learning technique was introduced in [29] to forecast the crop yield. Kalman filter algorithm was employed for data pre-processing. Linear Discriminant Analysis (LDA) was used for feature extraction. Bayesian spatially varying functional model (BSVFM) was designed in [30] to forecast the corn yield based on annual precipitation and multivariate functional predictors.

METHODOLOGY

Crop yield prediction is the process of forecasting the amount of crop that will be produced in a given area. It is an agricultural practice that can help farmers and farming businesses predict crop yield in a particular season when to plant a crop, and when to harvest for better crop yield. To predict crop yield accurately, dissimilar machine learning models and deep learning models are often used. These models analyze the historical data, weather patterns, and various other agricultural parameters to predict the crop yield. Several factors influence crop yield such as climate conditions, soil health, irrigation, seed quality, pest management, and farming practices. Different researchers carried out their research on crop yield prediction. Many classification and clustering techniques are introduced for effective crop yield prediction. But, the accuracy level was not improved and time complexity was not reduced by existing methods. In order to address these issues, a new method called Multivariate Regressive Gradient Spiral Optimized Deep Belief Learning (MRGSODBL) method is developed. The architecture diagram of MRGSODBL method is illustrated in the figure 1.
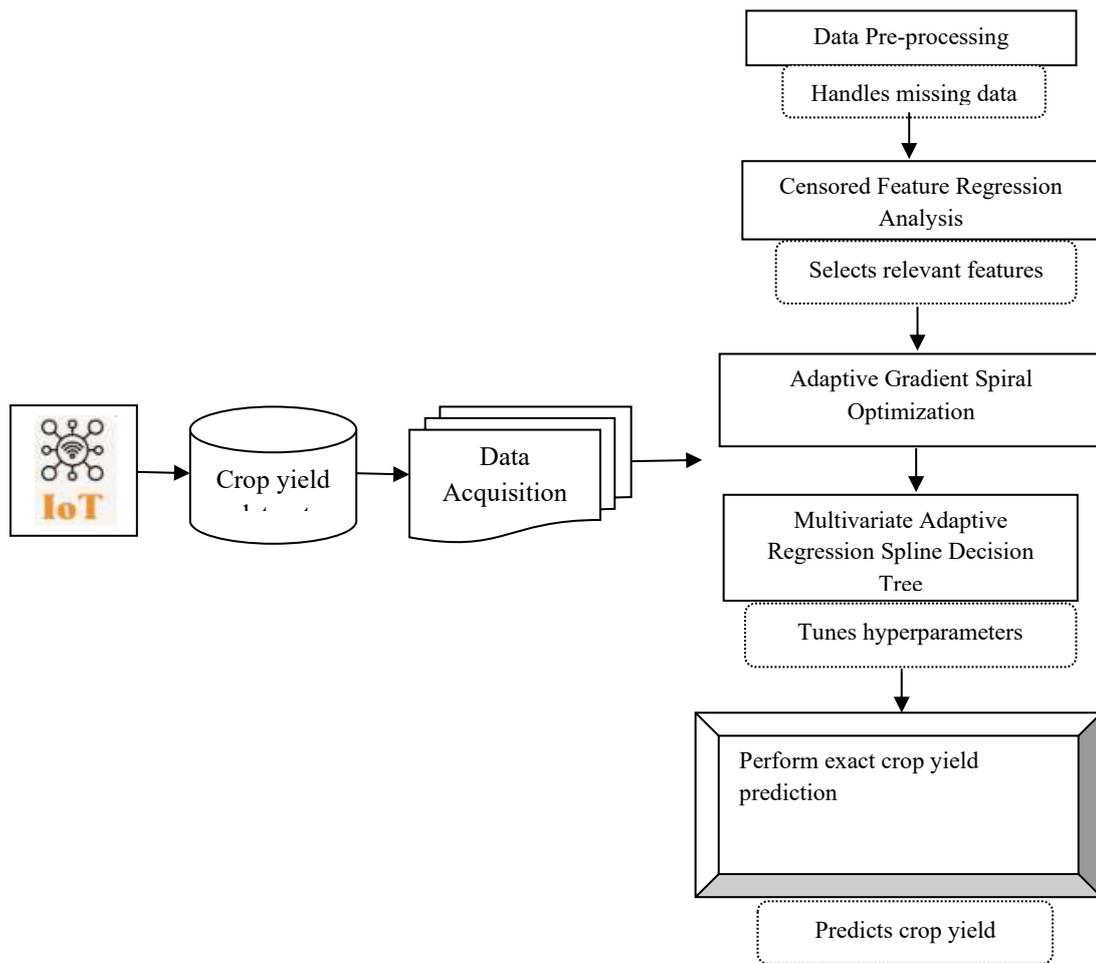
Deep Belief Network Classifier

Figure 1 Architecture diagram of proposed MRGSODBL method

Figure 1 demonstrates the architecture diagram of the proposed MRGSODBL method for accurate crop yield prediction in the agricultural sector. MRGSODBL method includes four different processes namely data acquisition, data pre-processing, feature selection and optimization for accurate crop yield prediction. The different processes of the MRGSODBL method are explained in the following subsections.

3.1 Data Collection

The first process in the proposed MRGSODBL method is data collection. For conducting the experiment, crop yield prediction dataset is used from Kaggle. The URL of the dataset is given as https://www.kaggle.com/datasets/patelris/crop-yield-prediction-dataset?resource=download. The dataset includes eight features and 28242 data instances. For the experimental consideration, the numbers of data samples ranging from 2500 to 25000 is considered as an input. The eight features.

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

Multivariate Regressive Gradient Spiral Optimized Deep Belief Learning

Deep Belief Network (DBN) is a deep learning model used to enhance the crop yield prediction accuracy with minimum processing time. DBN model increases the feature selection and classification performance for managing the huge data. DBN architecture comprised the many layers of Restricted Boltzmann Machines (RBMs) in hierarchical manner. The proposed DBN reduce the errors during the training process for crop yield prediction. The learning process is partitioned into training phase and fine-tuning phase. In training phase, each layer of Deep Belief Network process comprised the input data samples and transmitted to the next layer. In fine-tuning phase, the network hyperparameters get adjusted through spiral search optimization method. The deep belief network consists of training set $\{AD, Y\}$. 'AP' represent the training agricultural data samples. The deep belief network includes number of features and data samples '$DP = \{ DP_1, DP_2, DP_3, \dots DP_n \}$' collected from the dataset and a label 'Y' representing their category. The input data samples is related to agricultural data points ' $AD_1, AD_2, \dots, AD_n$' and added with the bias value 'B'. The neuron activation probability is illustrated in the visible layer ' P'. It is expressed as follows,

$$PV = A\left(\sum_{i=1}^{n} AD_i * w_{vs} \right) + B_v \quad (1)$$

From (1), 'PV' symbolizes the neuron activation probability in the visible layer. 'A' symbolize the sigmoid activation function, 'AD_i' represents the input data samples. 'w_{vs}' denote the weights in visible layer. '$*$' denotes the convolution operator, B_v indicates a bias of visible layer. The input data is transferred into the hidden layer where data preprocessing is carried out.

3.2 Data Pre-processing

Data preprocessing is an important process where the machine learning model learns from high-quality and structured dataset for accurate yield prediction. In preprocessing step, the proposed MRGSODBL method addresses the missing data issues in given dataset through neighbor imputation process. The nearest neighbor imputation process employs the nearest neighbors to fill in missing entries. The missing data imputation process is expressed as,

$$DI = \frac{\sum_{v=1}^{K} S_{nv} \, \delta_v}{\sum_{v=1}^{K} \delta_v} \qquad (2)$$

From (2), 'DI' represent the data imputation process. 'S_{nv}' represents the observed neighboring known data sample values. 'δ_v' symbolizes the weights assigned to the neighboring known data sample values. After finding the missing data, the determined values are refined by normalization process. The refinement process stabilizes the continuous and categorical variables to capture the fundamental structure and relationships within data. This in turn, the mean of each known value is computed as,

$$Mean = \frac{1}{K} \sum_{v=1}^{K} S_v \qquad (3)$$

From (3), 'K' represent the number of neighboring data samples. Subsequently, the normalization method is used to rescale the data into standard normal distribution. It is formulated as,

$$ND = \frac{(S_{miss} - mean)}{SD} \qquad (4)$$

From (4), 'ND' denotes the normalization of corresponding missing values, 'SD' denotes a standard deviation. The missing values are replaced with the mean value to reduce the deviation. Finally, the imputed values refine dataset for enhancing the accuracy of crop yield prediction. In hidden layer, the probability of neuron is measured as,

$$PH = A\left(\sum_{i=1}^{n} AD_i * w_{vs\_h} \right) + B_h \qquad (5)$$

From (5), 'PH' symbolizes the neuron activation probability of hidden layer. 'A' symbolize the sigmoid activation function. 'w_{vs_h}' denotes the weights between the visible layer and hidden layer, ' $*$' represents the convolution operator. 'B_h' denotes a bias of hidden layer.

3.3 Censored Feature Regression Analysis

After performing data preprocessing, the feature selection process is carried out in MRGSODBL method. Feature selection process is used to select the most relevant features for accurate crop yield prediction. MRGSODBL method uses the Hoover Censored Regression Analysis (HCRA) to choose the relevant features from pre-processed dataset by selecting the final feature subsets depending on the specified censoring point or threshold.
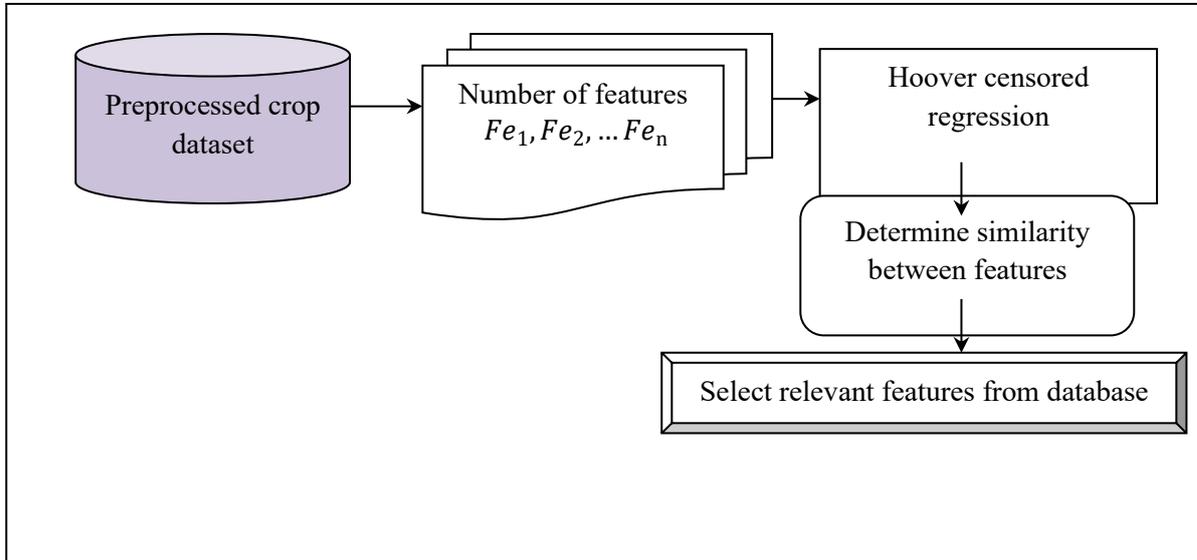
for accurate crop yield prediction. HCRA is a machine learning technique employed for measuring the relationship between features to attain the dependent variable (i.e., features censored above or below certain threshold). In HCRA, accurate feature selection results are attained for certain values and other values are removed because of limitations described by censoring point. The censoring point acts as the threshold with boundary between different values. HCRA performed left-censoring and right-censoring. Left-censoring occurs when the similarity value is lesser than specific threshold. The right-censoring occurs when observed value is higher than certain threshold. First, the number of features '$Fe_1, Fe_2, \dots Fe_n$' are collected from the given dataset. Hoover index similarity is formulated by,

$$\pi = \frac{1}{2} \frac{\sum_i |Fe_i - Fe_j|}{\sum_i Fe_i} \qquad (6)$$

From (6), 'π' symbolizes the similarity coefficient. 'Fe_i' represent the extracted feature. 'Fe_j' indicates the pre-stored feature. HCRA provides the value ranging from '0' to '+1'. '+1' denotes the higher similarity and '0' symbolizes the low similarity. Censored regression is a class of models where hoover similarity coefficient is censored above or below a certain threshold.

$$CR = \begin{cases} \pi > T \text{ ; relevant features} \\ \pi < T \text{ ; irrelavant features} \end{cases} \qquad (7)$$

From (7), 'CR' indicates a censored regression output. 'T' denotes a threshold. 'RC' denotes a Ruzicka similarity coefficient.
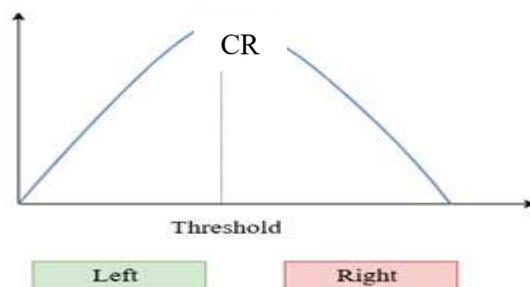
Figure 3 Censored Regression Analysis

Figure 3 illustrates the diagrammatic representation of censored regression analysis. When the coefficient value is greater than threshold (i.e., right censoring), the relevant feature is selected. When the similarity value is less than the threshold (i.e. left censoring), the irrelevant feature are removed. This is in turn the hoover censored regression analysis identifies the relevant features. The relevant feature is employed for accurate crop yield prediction. The algorithm of hoover censored regression based feature selection is given below,

// Algorithm 1: Hoover Censored Regression based Feature Selection
Input: Dataset 'Ds', features $Fe_1, Fe_2, \dots Fe_n$
Output: Select relevant features
Begin 1: Collect the number of features '$Fe_1, Fe_2, \dots Fe_n$' 2: For each features 3: Measure the hoover similarity coefficient 4. End for 5. if ($\pi > T$) then 6. features is said to be a relevant 7. else 8. features is said to be a irrelevant 9. end if 10. Return (relevant features) End

Algorithm 1 illustrates the process of relevant feature selection to improve the accurate crop yield prediction with minimum time consumption. The preprocessed dataset is considered as an input to the feature selection process. For every feature in input dataset, hoover similarity is determined. The hoover censored regression function examines the similarity value and differentiate the relevant from irrelevant features through threshold (i.e., the censoring point). Lastly, the relevant features are chosen to increase the crop yield prediction performance.

3.4 Multivariate Adaptive Regression Splines

Multivariate Adaptive Regression Splines (MARS) is a non-parametric regression analysis method used in proposed MRGSODBL method. MARS are employed for efficient classification of crop yield data in high-dimensional database. MARS is used to identify the non-linear relationships between dependent variable and independent variables without any assumptions. MARS framework in proposed MRGSODBL method avoids the overfitting problems and categorizes the crop yield data. MARS based crop yield data classification is formulated as,

$$Actual_{Out} = f(x) = \alpha_0 + \sum_{m=1}^{M} \alpha_m \prod_{k=1}^{P} S_p(CR_{km} - \tau_{km}) + \varepsilon \qquad (8)$$

From (8), 'α_0' symbolizes the main of basis function. 'S_p' denotes value '1' or '-1' when the data is to right of knot point or left of knot point. 'CR_{km}' represent the predictor variable. 'τ_{km}' represent the knot point of predictor variable. 'α_0' symbolizes the main of basis function. 'α_m' symbolizes the coefficient of basis function. 'M' represent the maximum of basis function.

3.5 Fine Tuning

Fine-tuning process is an important process used in the deep learning idea where the classifier model is optimized for particular task. The fine tuning process is performed to refine the weights for improving the prediction performance. In fine tuning process, the error rate is computed depending on the squared difference between actual and predicted output. In fine tuning process, error rate is determined depending on squared difference between actual and predicted classification output. It is determined as,

$$E = [Actual_{Out} - Predicted_{Out}]^2 \qquad (9)$$

From (9), 'E' represents the error rate, '$Actual_{out}$' symbolizes the actual classification output. '$Predicted_{out}$' represent the predicted classification output. For minimizing error rate, the proposed MRGSODBL method used adaptive gradient method to adjust the weight. It is formulated as,

$$weight^{new} = weight^{old} - \eta \left[\frac{\partial E}{\partial weight^{old}} \right] \qquad (10)$$

From (10), '$weight^{new}$' represent the new weight. '$weight^{old}$' symbolizes the current weight. 'η' describes the learning rate. '$\frac{\partial E}{\partial weight^{old}}$' represent the first-order derivative algorithm for identifying the local minimum of error rate through updating current weight '$weight^{old}$'.

Spiral Search algorithm is employed in the proposed MRGSODBL method to minimize the error and improve the crop yield prediction accuracy through identifying the optimal weight value. Spiral Search algorithm is a meta-heuristic algorithm inspired by natural spiral patterns. The spiral search algorithm is to create the logarithmic spirals for exploration and exploitation in the search process. Spiral Search algorithm carried out the diversification capabilities to perform the global exploration. The growth behavior carried out the search around current solution for performing the local exploitation. The spiral search algorithm employed the several spiral models in deterministic systems. Search points followed the path of logarithmic spiral toward the central point. Spiral Search algorithm is described as the current best solution. The global best solution is recognized through updating the common center.
In Spiral Search algorithm of MRGSODBL method, the multiple search points are taken as the number of weights. At first, the populations of multi points (i.e., weights) are initialized in the search space. For every point, the fitness is computed depending on the error rate. It is formulated as,
$$fitness = \arg\min(E) \qquad (11)$$

From (11), '$\arg\min$' denotes the argument of minimum function. The point with higher fitness among population is chosen as middle point. The position of the next search point gets updated as,
$$Search_{i+1} = Search_i + S_k M * |Search_i - Mp_i| \qquad (12)$$

From (12) '$Search_{i+1}$' denotes an update next search point position. '$Search_i$' indicates the current position of point. 'S_k' denotes the step rate value ranging from 0 to 1. 'M' represent the rotation matrix (i.e., identity matrix). '$|Search_i - Mp_i|$' symbolizes the difference between current position of the search point '$Search_i$' and 'Mp_i' represents the position of middle point. Then, the center point gets updated as,
$$Mp_{i+1} = \begin{cases} Mp_{i+1} ; & if\ ff(Mp_i) > ff(Mp_i) \\ Mp_i ; & otherwise \end{cases} \qquad (13)$$

From (13), 'Mp_{i+1}' represent the updated position of center point. 'Mp_i' symbolizes the current position of center point. When fitness of updated position of center point 'Mp_{i+1}' is greater than current position of center point 'Mp_i', the position gets updated as 'Mp_{i+1}'. Otherwise, the previous position 'Mp_i' is chosen as the final one. By this manner, an optimal weight is selected to attain higher classification accuracy and reduce the error rate in crop yield prediction. Lastly, the precise crop yield prediction is obtained at the output layer. Algorithm 2 illustrates the hyper parameter tuning using spiral search optimization.

Algorithm 2: Spiral Search Optimized Hyperparameter Tuning
Input: Weights
Output: Prediction accuracy and Prediction time
Begin Step 1:   Initialize the populations of search points (i.e. weights) Step 2:   for each weight Step 3:   Calculate the fitness Step 4:   While (t < max_ iteration) Step 5:   Select the center point with high fitness

Step 6:	Update the next search point based on the selected center point
Step 7:	$if\ (fitness(Mp_{i+1}) > fitness(Mp_i))$ then
Step 8:	Update the center point as a $fitness(Cp_{i+1})$
Step 9:	else
Step 10:	Update the center point as a 'Mp_i'
Step 11:	End if
Step 12:	t= t+1
Step 13:	End while
Step 14:	End for
Step 15:	Obtain the optimal weight
Step 16:	Obtain the final classification outcomes at the output layer
End	

Algorithm 2 illustrates the spiral search optimization algorithmic process involved in predicting the crop yield. The fine-tuning process is carried out through spiral search optimization. At first, the population of search points (i.e., weights) is initialized in the search space. The fitness of each point is determined based on error value. After that, the position of following search point is updated and the position of center point is tuned. This process gets iterated until the spiral search optimization reaches maximum iteration't'. Through iteration, the spiral search algorithm finds the optimal weight value to minimize the error. Finally, the crop yield prediction result is identified to minimize error rate at the output layer.

EXPERIMENTAL ANALYSIS

The proposed MRGSODBL method is implemented using python high level programming language with Crop Yield Prediction Dataset taken from https://www.kaggle.com/datasets/patelris/crop-yield-prediction-dataset?resource=download. The metrics involved for evaluating the performance of proposed MRGSODBL method are crop yield prediction accuracy, precision, recall, f-Measure, specificity, root mean square error and prediction time. Crop yield prediction accuracy is defined as the ratio of accurately identifying the crop yield from total number of data points. Consequently, the crop yield prediction accuracy is formulated as,

$$CYPA = \left(\frac{Tp+Tn}{Tp+Tn+Fp+Fn}\right) * 100 \qquad (14)$$

From (14), '$CYPA$' represent the crop yield prediction accuracy. 'Tp' symbolizes the true positive. 'Tn' denotes the true negative. 'Fp' symbolizes the false positive. 'Fn' denotes the false negative. It is computed in terms of percentage (%).
Precision is computed as the ratio of true positives to sum of true positives and false positives. Precision symbolizes the proportion of correctly classified the crop yield from total number of data samples. The precision is computed as,

$$PR = \left(\frac{Tp}{Tp+Fp}\right) \qquad (15)$$

From (15), 'PR' symbolizes the precision. 'Tp' symbolizes the true positive. 'Fp' denotes the false positive. Recall is defined as ability to correctly classify all data samples in a dataset. Recall is the ratio of true positive prediction to the sum of true positives and false negatives. It is formulated as,

$$RC = \left(\frac{Tp}{Tp+Fn}\right) \qquad (16)$$

From (16), 'RC' represent the recall. 'Tp' symbolizes the true positive. 'Fn' represent the false negative.
F-Measure refers to the harmonic mean both precision and recall to calculate the performance of crop yield prediction. The F1-measure is calculated as follows,

$$F1 - measure = 2 * \left(\frac{PR*RC}{PR+RC}\right) \qquad (17)$$

From (17), 'PR' symbolizes the precision. 'RC' represent the recall rate.

Root Mean Square Error (RMSE) is used to estimate the prediction accuracy. RMSE symbolizes the square root of average squared difference between predicted values and actual values to the total number of data samples. It is computed as,

$$RMSE = \left[\sqrt{\frac{(Y_{act}-Y_{pre})^2}{m}} \right] \qquad (18)$$

From (18), '$RMSE$' symbolizes the root mean square error. 'Y_{act}' represent the data samples for crop yield prediction. 'Y_{pre}' denotes the data samples for predicted crop yield results.

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

$$CYPT = \sum_{i=1}^{m} D_i * TM(CYP) \qquad (19)$$

From (19), '$CYPT$' represent the crop yield prediction time depending on patient data 'D_i'. The actual time consumed in crop yield prediction is represented by '$TM(CYP)$'. It is computed in terms of milliseconds (ms).

Specificity refers to the ability of a classification model to correctly classify the crop yield. It is mathematically computed as,

$$Spe = \frac{TN}{TN+FP} \qquad (20)$$

From (20), 'Spe' symbolizes the specificity. 'TN' represents the true negative and 'FP' indicates the false positive.

PERFORMANCE METRIC ANALYSIS
In this section, performance of the proposed MRGSODBL method, existing deep learning (DL) and dimensionality reduction (DR) approach [1], existing optimally driven Feedback Neural Network (FNN) approach [2] and AdaBoost GLCM [3] are evaluated with various metrics like crop yield prediction accuracy, precision, recall, root mean square error, crop yield prediction time with different number of samples. Table 2 shows the crop yield prediction accuracy comparison for four different methods.

Table 2 Comparison of Crop Yield Prediction Accuracy

| Number of samples	Crop yield prediction accuracy (%)			
	DL-DR Approach [1]	Optimally driven FNN approach [2]	AdaBoost GLCM [3]	Proposed MRGSODBL method
2500	91.2	83.86	88	99.20
5000	93.14	82.58	87.56	95.24
7500	94.74	84.28	88.52	98.18
10000	92.15	87.14	87.96	97.24
12500	93.58	86.64	88.21	98.34
15000	91.17	84.17	89.45	99.25
17500	92.34	83.67	88.22	98.64
20000	93.48	89.14	87.45	97.41
22500	94.12	87.45	88.45	95.68
25000	93.58	82.58	87.63	96.79
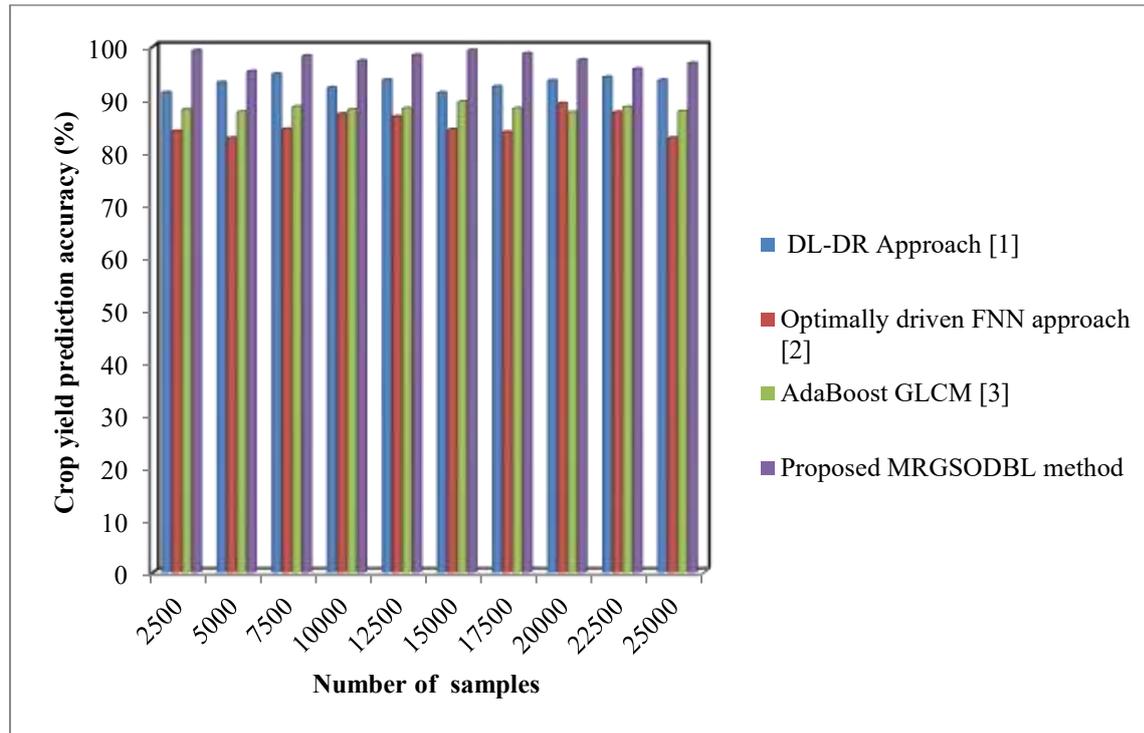
Figure 4 Measurement of Crop Yield Prediction Accuracy

Figure 4 illustrates the performance analysis of crop yield prediction accuracy using four different methods namely proposed MRGSODBL method, existing deep learning (DL) and dimensionality reduction (DR) approach [1], existing optimally driven Feedback Neural Network (FNN) approach [2] and AdaBoost GLCM [3]. For every method, the simulation of 10 runs is carried out with performed with 25000 data samples from input dataset. When increasing the number of data samples, the accuracy of crop yield prediction is found to be improved. But, the crop yield prediction accuracy using MRGSODBL method is found to be increased when compared to [1], [2] and [3] respectively. In the first iteration with 2500 data samples, the crop yield prediction accuracy for the MRGSODBL method was found to be 99.20%. Similarly, the crop yield prediction accuracy for methods [1] [2] and [3] was found 91.20%, 83.36% and 88% respectively. The average results obtained from proposed MRGSODBL method are compared to the existing method results. The comparison shows that the crop yield prediction accuracy performance using the proposed MRGSODBL method is significantly increased by 5%, 15% and 11% when compared to the existing methods [1], [2] and [3]. This is due to the application of feature selection and classification. The proposed deep belief network classifier model selects the relevant features for increasing the crop yield prediction accuracy. Table 3 shows the precision comparison for four different methods.

Table 3 Comparison of Precision

Number of samples	Precision			
	DL-DR Approach [1]	Optimally driven FNN approach [2]	AdaBoost GLCM [3]	Proposed MRGSODBL method
2500	0.914	0.877	0.885	0.994
5000	0.925	0.825	0.886	0.984
7500	0.945	0.865	0.895	0.997

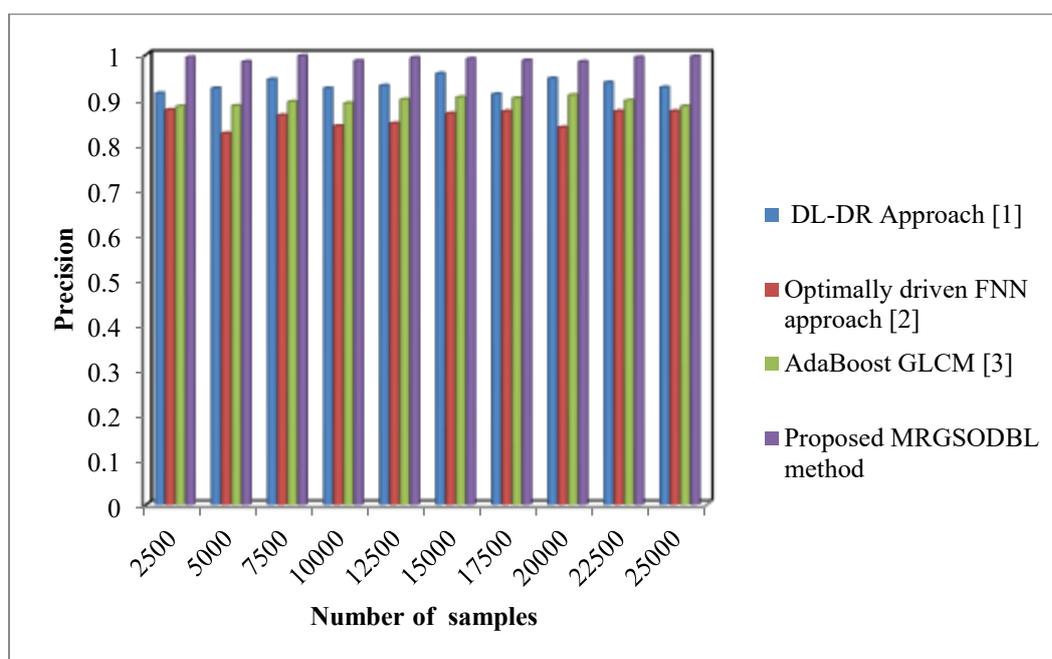|				
10000	0.925	0.841	0.892	0.986
12500	0.931	0.847	0.9	0.993
15000	0.958	0.869	0.905	0.991
17500	0.912	0.874	0.903	0.987
20000	0.947	0.838	0.91	0.984
22500	0.938	0.874	0.898	0.994
25000	0.927	0.874	0.885	0.996

Figure 5 Measurement Analysis of Precision

Figure 5 illustrates the graphical illustration of precision for four different methods namely proposed MRGSODBL method, existing DL-DR approach [1], existing optimally driven FNN approach [2] and AdaBoost GLCM [3]. In the figure, the horizontal axis represents the number of data samples and the vertical axis denotes the precision performance. The proposed MRGSODBL method outperforms better than existing methods [1], [2] and [3]. In experiment conducted with 2500 data samples, the precision was found to be 0.994 for the proposed MRGSODBL method, precision was found to be 0.914, 0.877 and 0.885 for the three existing methods [1], [2] and [3] respectively. The improvement is achieved through the application of censored feature regression analysis to minimize the dimensionality of the dataset and multivariate adaptive regression splines for crop yield prediction. The adaptive gradient spiral optimization algorithm tunes the hyperparameters of deep belief learning classifier to increase precision. Normally, the ten performance result analysis demonstrates that the precision obtained using the proposed MRGSODBL method is significantly improved by 6% compared to [1], 15% compared to [2] and 11% compared to [3]. Table 4 shows the recall comparison for four different methods.

Table 4 Comparison of Recall

Number of samples	Recall			
	DL-DR Approach [1]	Optimally driven FNN approach [2]	AdaBoost GLCM [3]	Proposed MRGSODBL method
2500	0.961	0.953	0.939	0.99

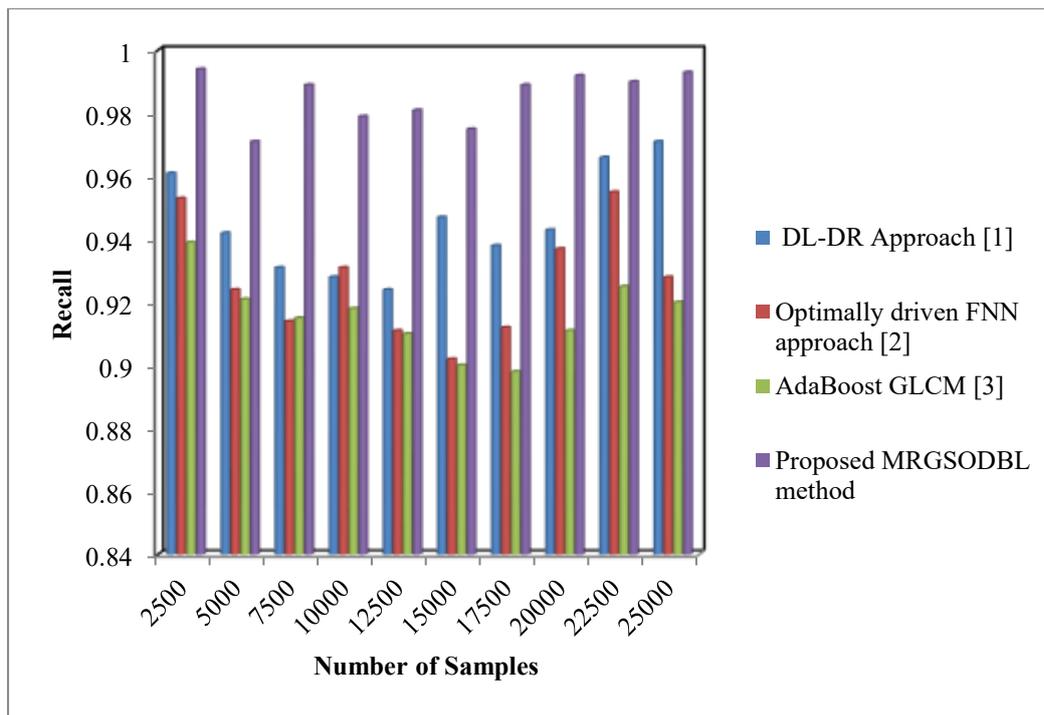| 5000 | 0.942 | 0.924 | 0.921 | 0.971 |
7500	0.931	0.914	0.915	0.989
10000	0.928	0.931	0.918	0.979
12500	0.924	0.911	0.91	0.981
15000	0.947	0.902	0.9	0.975
17500	0.938	0.912	0.898	0.989
20000	0.943	0.937	0.911	0.992
22500	0.966	0.955	0.925	0.990
25000	0.971	0.928	0.92	0.993

Figure 6 Measurement Analysis of Recall

Figures 6 illustrate the recall analysis versus number of data samples ranging from 2500 to 25000 from the dataset. The performance of recall was determined using four methods namely the proposed MRGSODBL method, existing DL-DR approach [1], existing optimally driven FNN approach [2] and AdaBoost GLCM [3]. The horizontal axis denotes the number of data samples while the vertical axis denotes the recall values. The experimental result reveals that the MRGSODBL method attained higher recall performance when compared to the other deep learning methods. Different recall values were attained for every method depending on number of input data samples. The observed results for the MRGSODBL method were compared against three existing techniques. The comparison results indicate that recall for crop yield prediction by 4%, 6% and 8% than the existing [1] [2] and [3] respectively. Table 5 shows the root mean square error comparison for four different methods.

Table 5 Comparison of Root Mean Square Error

Number of samples	Root Mean Square Error			
	DL-DR Approach [1]	Optimally driven FNN approach [2]	AdaBoost GLCM [3]	Proposed MRGSODBL method
2500	0.153	0.181	0.240	0.096
5000	0.148	0.168	0.175	0.068

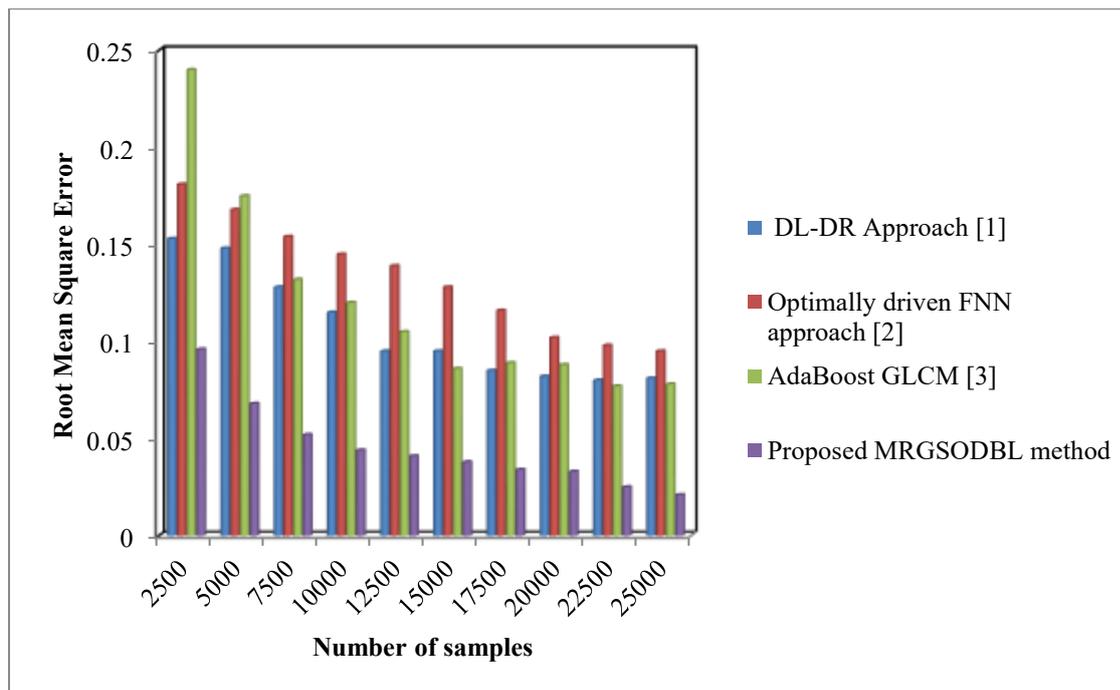| 7500	0.128	0.154	0.132	0.052
10000	0.115	0.145	0.120	0.044
12500	0.095	0.139	0.105	0.041
15000	0.095	0.128	0.086	0.038
17500	0.085	0.116	0.089	0.034
20000	0.082	0.102	0.088	0.033
22500	0.080	0.098	0.077	0.025
25000	0.081	0.095	0.078	0.021

Figure 7 Measurement Analysis of Root Mean Square Error

In Figure 7, the performance analysis of root mean square error is examined with respect to the number of sample ranging from 2500 to 25000. Four methods were applied namely the proposed MRGSODBL method, existing DL-DR approach [1], existing optimally driven FNN approach [2] and AdaBoost GLCM [3] in crop yield detection. The horizontal axis indicates the number of data samples, while the vertical axis represents the performance of root mean square error. The results illustrates that the MRGSODBL method achieves minimal root mean square error when compared to the other three existing deep learning methods. Let us consider the number of data samples to be 2500 in the first iteration. By applying the MRGSODBL method, the root mean square error was found to be 0.096, 0.153 for [1], 0.181 for [2] and 0.24 for the [3]. The overall performance outcomes were obtained for each method with varying number of data samples. The results reveal that root mean square error performance of MRGSODBL method is minimized by 59% when compared to [1], 67% when compared to the [2] and 63% when compared to the [3]. This performance of root mean square error is reduced by applying the gradient spiral optimized deep belief network. The spiral search optimization model determines the optimal weight to reduce the root mean square error. Table 6 shows the crop yield prediction time comparison for four different methods.

Table 6 Comparison of Crop Yield Prediction Time

	Crop Yield Prediction Time (ms)

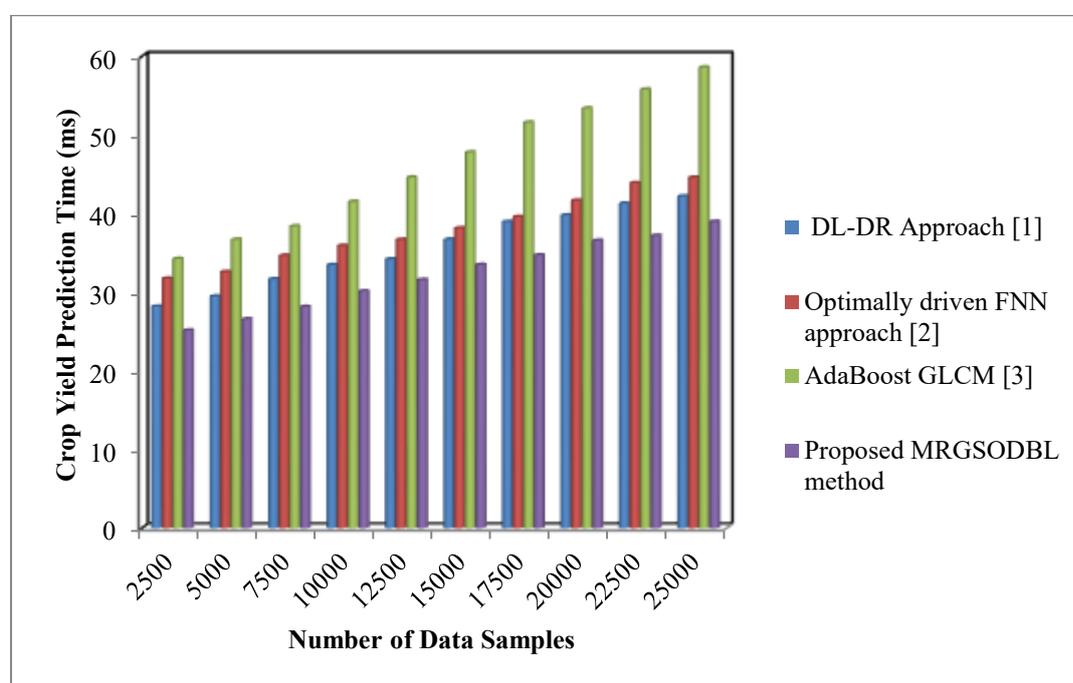| Number of samples	DL-DR Approach [1]	Optimally driven FNN approach [2]	AdaBoost GLCM [3]	Proposed MRGSODBL method
2500	28.17	31.74	34.25	25.15
5000	29.48	32.58	36.7	26.58
7500	31.67	34.67	38.4	28.14
10000	33.46	35.91	41.5	30.12
12500	34.19	36.68	44.6	31.58
15000	36.71	38.14	47.8	33.50
17500	38.95	39.58	51.6	34.71
20000	39.78	41.67	53.4	36.58
22500	41.29	43.89	55.8	37.19
25000	42.19	44.58	58.6	38.97

Figure 8 Measurement Analysis of Crop Yield Prediction Time

Figure 8 illustrates the performance results of crop yield prediction time across three methods namely proposed MRGSODBL method, existing DL-DR approach [1], existing optimally driven FNN approach [2] and AdaBoost GLCM [3] in crop yield detection. Crop yield prediction time for all four methods increases with increasing number of samples. Notably, the MRGSODBL method reveals a significant reduction in crop yield prediction when compared to the conventional methods [1] and [2]. For instance, in the first iteration with 2500 data samples, the MRGSODBL method achieved a crop yield prediction time of $25.15ms$, whereas methods [1], [2] and [3] consumed 28.17 ms, $31.74ms$ and $34.25ms$ respectively. A comparison of the overall results indicates that the MRGSODBL method reduces crop yield prediction time by 9%, 15% and 30% compared to [1], [2] and [3] respectively. This is because of censored feature regression analysis to minimize the dimensionality of the dataset with minimum time consumption. The adaptive gradient spiral optimization algorithm tunes the hyperparameters of deep belief learning classifier with minimum time complexity. Table 7 shows the F1-measure comparison for four different methods.

Table 7 Comparison of F1-measure

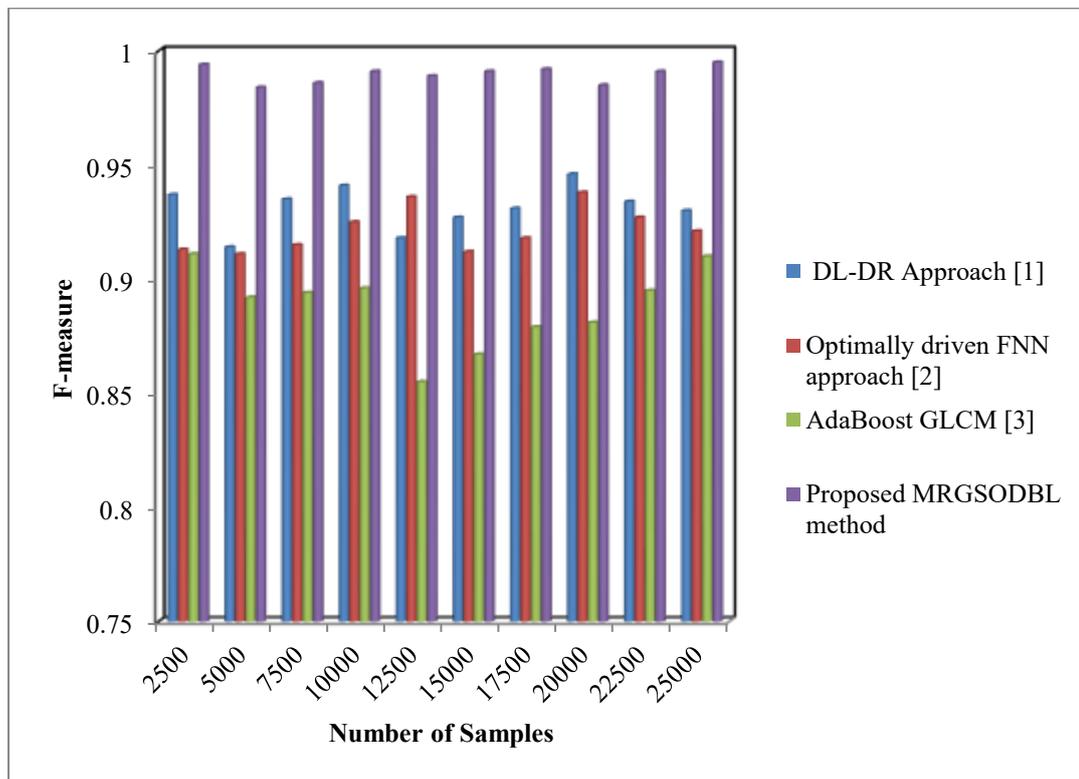| Number of samples	F1-measure			
	DL-DR Approach [1]	Optimally driven FNN approach [2]	AdaBoost GLCM [3]	Proposed MRGSODBL method
2500	0.937	0.913	0.911	0.994
5000	0.914	0.911	0.892	0.984
7500	0.935	0.915	0.894	0.986
10000	0.941	0.925	0.896	0.991
12500	0.918	0.936	0.855	0.989
15000	0.927	0.912	0.867	0.991
17500	0.931	0.918	0.879	0.992
20000	0.946	0.938	0.881	0.985
22500	0.934	0.927	0.895	0.991
25000	0.930	0.921	0.910	0.995

Figure 9 Measurement Analysis of F1-Measure

Figures 9 show the graphical analysis of the F1 measure versus numbers of data samples ranging from 2500 to 25000. The observed results indicate that the proposed MRGSODBL method achieved improved F1 score performance when compared to existing models. Each method was computed with ten different iterations. The average results of F1 measure of the performance of increased by 6%, 7% and 12% when compared to [1], [2] and [3] respectively. This enhancement is because of using proposed MRGSODBL to improve the performance of precision and recall during crop yield prediction. Table 8 shows the specificity comparison for four different methods.

Table 8 Comparison of Specificity

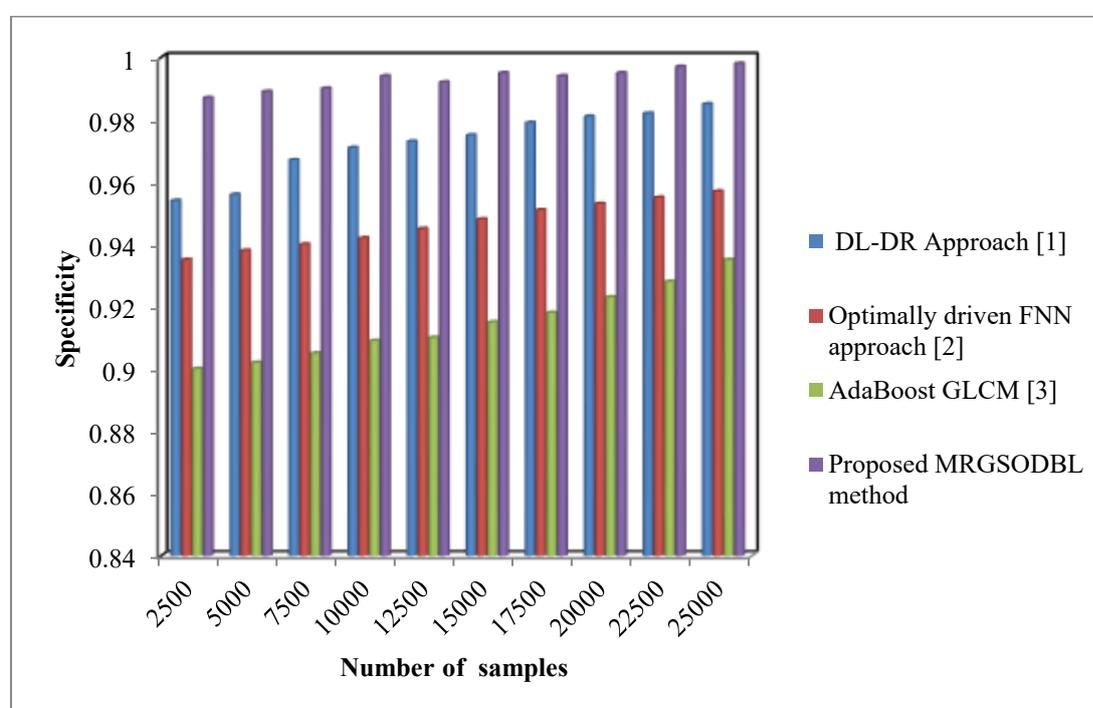| Number of samples	Specificity			
	DL-DR Approach [1]	Optimally driven FNN approach [2]	AdaBoost GLCM [3]	Proposed MRGSODBL method
2500	0.954	0.935	0.90	0.987
5000	0.956	0.938	0.902	0.989
7500	0.967	0.940	0.905	0.990
10000	0.971	0.942	0.909	0.994
12500	0.973	0.945	0.910	0.992
15000	0.975	0.948	0.915	0.995
17500	0.979	0.951	0.918	0.994
20000	0.981	0.953	0.923	0.995
22500	0.982	0.955	0.928	0.997
25000	0.985	0.957	0.935	0.998

Figure 10 Measurement Analysis of Specificity

Figure 10 shows the performance results of specificity versus number of data samples ranging from 2500 to 25000. In order to calculate the specificity value, four methods were employed namely the proposed MRGSODBL method, existing DL-DR approach [1], existing optimally driven FNN approach [2] and AdaBoost GLCM [3]. The horizontal axis indicates the number of data samples, while the vertical axis denotes the specificity. The result shows that the MRGSODBL method achieved better results in specificity when compared to the other three existing methods. Let us consider the number of images to be 2500 in the first run. The specificity was found to be 0.954, 0.935 and 0.900 for [1], [2] and [3] while specificity of proposed MRGSODBL method was 0.987. Different performance results were attained for each method with varying number of data samples. The overall comparison shows that the specificity of proposed MRGSODBL method for crop yield prediction is enhanced by 2% compared to [1], 5% compared to the [2] and 9% when compared to [3]. This is because of using fine tuning process using gradient spiral optimization through minimizing incorrect predictions of the crop yield.

CONCLUSION

Crop yield forecasting is important one for agricultural planning, resource allocation and decision-making. The proposed MRGSODBL method uses deep belief network for predicting the crop yield during cultivation. The data preprocessing in MRGSODBL method handles the missing data. Censored feature regression analysis selects the suitable features to minimize the dimensionality. Multivariate adaptive regression splines performs the crop yield prediction. This process helps to increase the accuracy of the crop yield prediction. The fine tuning process of MRGSODBL method reduces the root mean square error of crop yield prediction. A comprehensive performance analyses is carried out using various performance metrics, such as crop yield prediction accuracy, precision, recall, F1-score, specificity, root means square error and crop yield prediction time across different number of data samples. The overall quantitatively analyzed results show that the proposed MRGSODBL method enhances accuracy while reducing time as well as error when compared to conventional deep learning methods.

REFERENCES

[1] Leelavathi Kandasamy Subramaniam, Rajasenathipathi Marimuthu, "Crop yield prediction using effective deep learning and dimensionality reduction approaches for Indian regional crops", e-Prime - Advances in Electrical Engineering, Electronics and Energy, Elsevier, Volume 8, 2024, Pages 1-9

[2] Amna Ikram and Waqar Aslam, "Enhancing Intercropping Yield Predictability Using Optimally Driven Feedback Neural Network and Loss Functions", IEEE Access, Volume 12, 2024, Pages 162769 – 162787

[3] 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.

[4] Dhahi Al-Shammari, Brett M. Whelan, Chen Wang, Robert G.V. Bramley and Thomas F.A. Bishop "Assessment of red-edge based vegetation indices for crop yield prediction at the field scale across large regions in Australia", European Journal of Agronomy, Elsevier, Volume 164, March 2025, Pages 1-15

[5] Francisco Mena, Deepak Pathak, Hiba Najjar, Cristhian Sanchez, Patrick Helber, Benjamin Bischke, Peter Habelitz, Miro Miranda, Jayanth Siddamsetty, Marlon Nuske, Marcela Charfuelan, Diego Arenas, Michaela Vollmer and Andreas Dengel, "Adaptive fusion of multi-modal remote sensing data for optimal sub-field crop yield prediction", Remote Sensing of Environment, Elsevier, Volume 318, 1 March 2025, Pages 1-15

[6] Seungtaek Jeong, Jonghan Ko, Jong-oh Ban, Taehwan Shin and Jong-min Yeom, "Deep learning-enhanced remote sensing-integrated crop modeling for rice yield prediction", Ecological Informatics, Elsevier, Volume 84, December 2024, Pages 1-15

[7] Mohammad Amin Razavi, A. Pouyan Nejadhashemi, Babak Majidi, Hoda S. Razavi, Josué Kpodo, Rasu Eeswaran, Ignacio Ciampitti and P.V. Vara Prasad, "Enhancing crop yield prediction in Senegal using advanced machine learning techniques and synthetic data", Artificial Intelligence in Agriculture, Elsevier, Volume 14, December 2024, Pages 99-114

[8] Shahid Nawaz Khan, Dapeng Li and 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, July 2024, Pages 1-15

[9] T. Sudhamathi and K. Perumal, "Ensemble regression based Extra Tree Regressor for hybrid crop yield prediction system", Measurement: Sensors, Elsevier, Volume 35, October 2024, Pages 1-15

[10] Haydar Demirhan, "A deep learning framework for prediction of crop yield in Australia under the impact of climate change", Information Processing in Agriculture, Elsevier Volume 12, Issue 1, March 2025, Pages 125-138

[11] Leelavathi Kandasamy Subramaniam and Rajasenathipathi Marimuthu, "Crop yield prediction using effective deep learning and dimensionality reduction approaches for Indian regional crops", e-Prime - Advances in Electrical Engineering, Electronics and Energy, Elsevier, Volume 8, June 2024, Pages 1-15

[12] Jinlong Li, Dongfeng Zhang, Feng Yang, Qiusi Zhang, Shouhui Pan, Xiangyu Zhao, Qi Zhang, Yanyun Han, Jinliang Yang, Kaiyi Wang and Chunjiang Zhao, "TrG2P: A transfer-learning-based tool integrating multi-trait data for accurate prediction of crop yield", Plant Communications, Elsevier, Volume 5, Issue 7, 8 July 2024, Pages 1-18

[13] Letícia Bernabe Santos, Donna Gentry, Alex Tryforos, Lisa Fultz, Jeffrey Beasley and Thanos Gentimis, "Soybean yield prediction using machine learning algorithms under a cover crop management system", Smart Agricultural Technology, Elsevier, Volume 8, August 2024, Pages 1-15

[14] M. Rajakumaran, G. Arulselvan, S. Subashree and R. Sindhuja, "Crop yield prediction using multi-attribute weighted tree-based support vector machine", Measurement: Sensors, Elsevier, Volume 31, February 2024, Pages 1-15

[15] Balaji Prabhu B V, Shashank R, Shreyas B and Omkar Subbaram Jois Narsipura, "ARIA: Augmented Reality and Artificial Intelligence enabled mobile application for Yield and grade prediction of tomato crops", Procedia Computer Science, Elsevier, Volume 235, 2024, Pages 2693-2702

[16] A. Punitha and V. Geetha, "Gorilla troops optimization with deep learning based crop recommendation and yield prediction", International Journal of Cognitive Computing in Engineering, Elsevier, Volume 5, 2024, Pages 494-504

[17] Muhammad Hanif Tunio, Jian Ping Li, Xiaoyang Zeng, Faijan Akhtar, Syed Attique Shah, Awais Ahmed, Yu Yang and Md Belal Bin Heyat "Meta-knowledge guided Bayesian optimization framework for robust crop yield estimation", Journal of King Saud University - Computer and Information Sciences, Elsevier, Volume 36, Issue 1, January 2024, Pages 1-15

[18] Yeshanbele Alebele, Wenhui Wang, Weiguo Yu, Xue Zhang, Xia Yao and Yongchao Tian "Estimation of Crop Yield from Combined Optical and SAR Imagery using Gaussian Kernel Regression", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Volume 14, October 2021, Pages 10520 - 10534

[19] Mahmoud Abdel-salam, Neeraj Kumar and Shubham Mahajan, "A proposed framework for crop yield prediction using hybrid feature selection approach and optimized machine learning", Neural Computing and Applications, Springer, Volume 36, August 2024, Pages 20723–20750

[20] Yogita Rahulsing Chavan, Brinthakumari Swamikan, Megha Varun Gupta, Sunil Bobade and Anu Malhan, "Enhanced Crop Yield Forecasting using Deep Reinforcement Learning and Multi-source Remote Sensing Data", Remote Sensing in Earth Systems Sciences, Springer, Volume 7, 2024, Pages 426–442

[21] 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

[22] 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

[23] T. Sudhamathi and K. Perumal, "A novel hybrid DNN-RNN framework for precise crop yield prediction", International Journal of System Assurance Engineering and Management, Springer, November 2024, 2024, Pages 1-15

[24] Usharani Bhimavarapu, Gopi Battineni and Nalini Chintalapudi, "Improved Optimization Algorithm in LSTM to Predict Crop Yield", Computers, Volume 12, Issue 1, 2023, Pages 1-19

[25] Md Jiabul Hoque, Md. Saiful Islam, Jia Uddin, Md. Abdus Samad, Beatriz Sainz De Abajo, Débora Libertad Ramírez Vargas, "Incorporating Meteorological Data and Pesticide Information to Forecast Crop Yields Using Machine Learning", IEEE Access, Volume 12, Pages 47768 – 47786

[26] Nandini Geddlehally Renukaradya, Kishore Gopala Rao and Anand Babu Jayachandra, "Classification and Estimation of Crop Yield Prediction in Karnataka using LSTM with Attention Mechanism", International Journal of Intelligent Systems and Applications in Engineering, Volume 12, Issue 3, 2024, Pages 1-15

[27] Vivek Parganiha and Monika Verma "An Effective Soil Analysis and Crop Yield Prediction Based on Optimized Light GBM in Smart Agriculture", Journal of Agronomy and Crop Science, Wiley Online Publications, Volume 210, Issue 4, August 2024, Pages 1-15

[28] Deeksha Tripathi and Saroj K. Biswas, "Design of a precise ensemble expert system for crop yield prediction using machine learning analytics", Journal of Forecasting, Wiley Online Publications, Volume 43, Issue 8, December 2024, Pages 3161-3176

[29] Swati Vashisht, Praveen Kumar and Munesh Chandra Trivedi, "Crop Yield Prediction Using Improved Extreme Learning Machine", Communications in Soil Science and Plant Analysis, Taylors and Francis, Volume 54, Issue 1, 2023, Pages 1-21

[30] Yeonjoo Park, Bo Li and 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 70-83