Crop Yield Prediction by Integrating Et-DP Dimensionality Reduction and ABP-XGBOOST Technique

A. Radhika[1*] and Dr.M. Syed Masood[2]

[1*]Assistant Professor, Department of Computer Science and Engineering, B.S.A Crescent Institute of Science and Technology, Chennai. radhika.bucse@gmail.com

[2]Associate Professor, Department of Computer Applications, B.S.A Crescent Institute of Science and Technology, Chennai.

Received: August 11, 2022; Accepted: October 04, 2022; Published: November 30, 2022

Abstract

In the Indian economy, one of the most significant sectors is agriculture, which is the primary source of livelihood. In precision agriculture, the major intention is to ameliorate Crop Yield (CY) production along with quality by mitigating operational costs along with environmental pollution. To attain precision agriculture, Crop Yield Prediction (CYP) is highly significant. However, CYs become unpredictable owing to their reliance on huge dimension features like soil, climate, pesticides, diseases, et cetera. Therefore, for CYP, a Feature Selection (FS)-centric Machine Learning (ML) methodology is desired to provide precision agriculture along with mitigate the computational time. An Ensemble Threshold-centric Data Perturbation (DP) FS (ET-DPFS) dimensionality reduction centered Alpha-Beta Pruning centered Extreme Gradient Boosting (ABP-XGBOOST) CYP methodology is developed here. In this model, highly pertinent candidate features, which contain higher relevance regarding the class, lower redundancy amongst the features being selected, along with stay strong against noisy data, are selected. In the proposed work, firstly, by dealing with Nan values, missing values, outliers, categorical features, along with date-time variables, an exploratory evaluation is conducted over the data. After that, by utilizing the Analysis of Variance (ANOVA) test, the redundant features are taken away. Then, by employing the ET-DPFS, the relevant features are selected. ''3' base learner FS methodologies like Median Absolute Deviation-centric LASSO (MAD-LASSO), Coefficient Vector-centric Mayfly Optimization (CV-MO), and HE weight initialization-centered Relief (HEwint-Relief) are included in this ET-DPFS. Lastly, by deploying the ABP-XGBOOST, the features being selected are trained along with verified. Experiential evaluation displays that the proposed methodology is highly reliable than the prevailing methodologies by achieving precise prediction with lower Values for Mean Square Error (MSE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

Keywords: Agriculture, Crop Yield Prediction, Preprocessing, Analysis of Variance (ANOVA), Ensemble Threshold-based Data Perturbation Feature Selection (ET-DPFS), Coefficient Vector based Mayfly Optimization (CV-MO), Median Absolute Deviation based LASSO (MAD-LASSO), HE Weight Initialization based Relief (HEwint-Relief) and Alpha-Beta Pruning based Extreme Gradient Boosting (ABP-XGBOOST).

Journal of Internet Services and Information Security (JISIS), volume: 12, number: 4 (November), pp. 177-196
DOI: *10.58346/JISIS.2022.I4.013*

*Corresponding author: Assistant Professor, Department of Computer Science and Engineering, B.S.A Crescent Institute of Science and Technology, Chennai.

1 Introduction

For worldwide food production, CYP is more significant. To increase national food safety, the strategy makers depend on precise forecasting to provide prompt import and export choices [1, 2]. To breed superior varieties, seed producers must forecast the novel hybrids' performance in diverse conditions [3, 4]. Yield prediction also helps growers and farmers make better managerial and financial choices. But, due to several intricate factors, CYP is highly difficult. For exemplar, the genotype data is typically represented by high-dimensional marker data, which contains countless makers for every plant individual [5, 6]. Since the genetic markers may be influenced by numerous environmental situations and field management approaches, their effects should be evaluated [7].

Despite the fact that new research has revealed agricultural statistics data, some investigations have analyzed CYP regarding the historical data [8, 9]. Crop cultivation forecast is difficult as a result of the unrestricted usage of fertilizers containing potassium, micronutrients, and potassium. Soil texture, temperature, and rainfall are all agro-climatic features that affect the crop's production [10]. Agricultural input metrics differ from place to place and gathering these data over huge land areas is intricate. To do large-scale predictions the extensive datasets acquired can be employed Because of the nature of the situation, fresh ML techniques for farming arable land and exploiting the majority of restricted land resources are required [11]. For finding the best crop for a certain piece of land, agricultural experts have been investigating several anticipating approaches. Nowadays, the forecasting of suitable crops for farming is agriculture's important element, and ML algorithms have played a chief role in this prediction [12, 13]. The agricultural field stands to advantage deeply from correctly applied methodologies in this technology along with data science's epoch. ML methods like classification together with FS are crucial [14]. Choosing the largely significant characteristics from a dataset is called FS. It entails selecting a subset of relevant features as of a greater collection of original attributes by lowering their dimensionality regarding a preset benchmark, like categorization performance or class separability and it has an important function in ML applications. However, in producing correct outcomes, the FSs' scope is inefficient [15]. The CY is reliant on several diverse factors like seed type, fertilizer usage, weather, and climate because the precise forecasting of CY needs the compilation of numerous datasets and it causes a great dimensionality curse in identifying the extremely appropriate attributes [16, 17]. Furthermore, the existence of disturbances or outliers causes a significant error rate in CYP [18]. The forecast is erroneous because of the issue of baseline attributes instability and over-fitting. This paper created an ET-DP FS-based ABP-XGBOOST CYP model, for overcoming the aforesaid issues.

The remainder of this work is organised as follows: section 2 explains the associated works and their drawbacks on FS relating to the CYP. Section 3 demonstrates the suggested structure. Section 4 depicts the experimental design, the outcomes, and the key findings. Section 5 finally describes the conclusion.

2 Literature Survey

Dhivya Elavarasan *et al*. [19] designed a hybrid regression-centric algorithm, Reinforcement Random Forest (RF). For presented samples' effective utilization, the technique used reinforcement learning at each selection of a dividing attribute during the tree construction procedure. For choosing the significant variable for the node splitting mechanism in the system building along with to support effective training data usage, it evaluated the variable importance measure. The acquired outcomes described that the strategy carried out superior, with lower error measures together with enhanced accuracy of 92.2%. However, the method over-fitted the system.

Durai Raj Vincent *et al*. [20] described a hybrid feature extraction scheme that was an amalgamation of the Correlation-centered Filter (CFS) along with RF Recursive Feature Elimination (RFRFE) wrapper structure. Discovering the features' best subclass from a set of soil, groundwater, and climate properties to create CY anticipating ML system with larger accuracy and performance were the feature extraction methods' objectives. When analogized to the RF, gradient boosting, along with decision tree ML algorithms, the hybrid CFS together with RFRFE feature extraction's confirmation was superior. However, the selection technique causes informative data loss.

Ekaansh Khosla *et al*. [21] used Modular Artificial Neural Networks (MANN) to forecast the amount of monsoon rainfall. Then it utilized Support Vector Regression (SVR) to anticipate the main Kharif crops' quantity that may be produced regarding the rainfall information and place allotted to that specific crop. The MANN-SVR method was created for agriculture to boost agricultural productivity. The designed system was better in forecasting the occasions for Kharif crop production when analogized with the other ML methods. However, it was not accessible or stable enough to surpass different datasets.

Iniyan *et al*. [22] established Mutual information regarding a sophisticated ensemble regression method integrated in the prediction mechanism of CY on soybean and corn crops. It accomplished better forecasting accuracy regarding phenotypic variables. The sophisticated ensemble regression crop prediction system surpassed numerous ML and progressive learning techniques regarding the forecasted yield. Several regression accuracy metrics namely RMSE, MSE, along with MAE were entailed in performance computation. The forecasted results confirmed that crop and weather management metrics were more dominant than soil metrics in CYP. However, the methodology was mathematically intricate.

Chaoya dang *et al*. [23] used the method regarding Redundancy Analysis (RDA) to execute explanatory factors along with feature selection. The explanatory components' interpretation rates were estimated utilizing RDA's simple effects. For choosing the explanatory factor's features, the RDA's conditional effects were employed. Then, using SVR, RF Regression (RFR), along with Deep Neural Network (DNN) for the system, the autumn CY was separated into the training set along with testing sets with an 80/20 ratio. The outcomes showcased that the explanatory factor's interpretation rates varied as of 54.3% to 85.0% ($p = 0.002$), and relatively DNN executed by surpassing the RFR and SVR. But, the methodology was instructed with useful attributes, and a greater error rate was obtained.

A. Radhika et al. [24] designed dimensionality decrease regarding soft calculating techniques in data mining for the agriculture sector. For generating a decreased attribute's set, this method utilized Principle Component Analysis (PCA) dimension reduction in the vibrant surrounding. The designed technique's performance was extremely efficient for forecasting CY, as per the soil processing, dimensionality reduction, and forecasting through the DT algorithm. Regarding the number of inputs and variables, the predictions' accuracy was executed by dimensional reduction was organized. With less implementation time, the outcomes give in zero information loss. However, for large dimensions, the methodology was erroneous and also produced large computation time.

3 Proposed Crop Yield Prediction Framework

The CYP model, which supports the farmers to make better decisions about the perfect time to grow crops and what kinds of crops to be cultivated regarding environmental factors to get a better yield, is provided by the ML algorithm. Even then, owing to the higher dimensionality of features regarding harsh climatic changes, insufficient rainfall, inadequate crop nutrients, severity of crop disease, and poor soil nature, farmers all over the world weren't able to attain enough yields of their cultivated crops. To mitigate the farmer's workload along with financial investment on crops, an appropriate CYP is required regarding an accurate FS-centric ML methodology. Figure 1 exhibits the proposed ET-DP FS-centered ABP-XGBOOST CYP model.

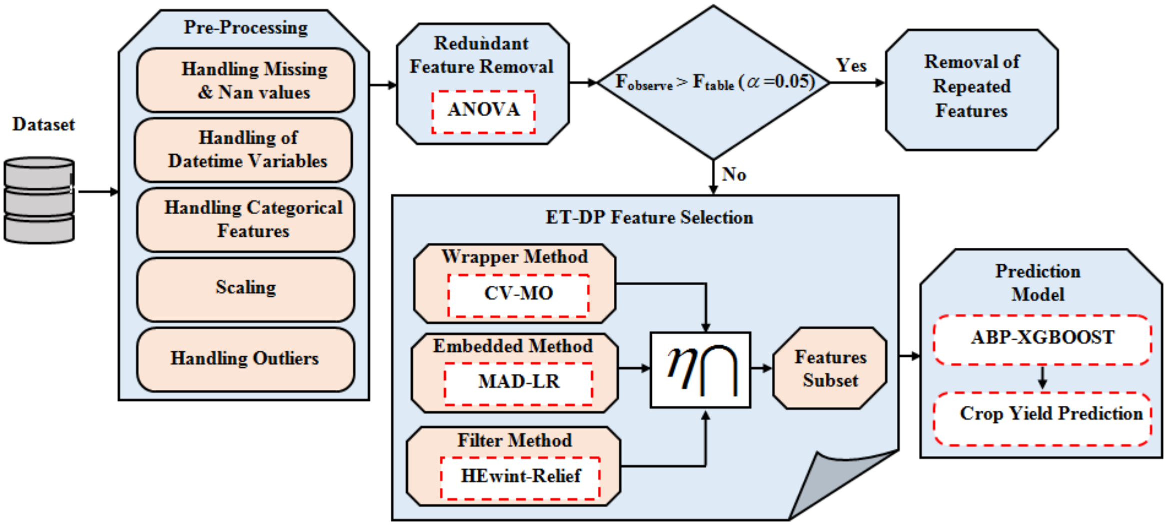


Figure 1: Proposed Framework for Crop Yield Prediction

3.1. Preprocessing

The process of transmuting unstructured data format into a structured data format is mentioned as pre-processing, which mitigates the probability of error. In pre-processing, by taking away the repeated data, noises occurred owing to outliers, missing values, et cetera, the data are cleaned. Handling of data time variables, handling of missing along with Nan values, handling outliers, handling categorical features, and scaling of data are done in pre-processing.

3.1.1. Handling of Missing and Nan Values

For a particular feature or set of features, "no data present" or "blank" represents the missing value; similarly, Nan values specify a "Not a Number" that is a component of a numeric data type, which is deduced as a value that is undefined or un-representable. There is a chance of removing useful data whilst removing the missing along with Nan values in the dataset; thus, the CYP's accurateness might be lowered. To get an effectual outcome, handling the missing along with Nan value is highly significant. For categorical along with numerical features, handling of missing values varies. The Nan values can be substituted with certain other variables or mode functions, Forward Fill, Backward Fill can be utilized for categorical features.

$$\Theta_{pre} = \chi_{fxn}(C^{Y}_{i\times j}) \qquad (1)$$

Where, the function that handles missing along with Nan values is specified as χ_{fxn} .Regarding the categorical along with numerical values, the methodologies are chosen; some of them are Mode, Media, Forward Fill, Backward Fill, Imputation, et cetera. The CY dataset with i^{th} row and j^{th} column is represented as $C^{Y}_{i\times j}$ .Lastly, the dataset devoid of Nan together with missing value is attained; in addition, framed into the data frame, that is to say,

$$C^{DS}_{n} = \left[C^{DS}_{1}, C^{DS}_{2}, C^{DS}_{3}, C^{DS}_{4}, ...C^{DS}_{n}\right] \tag{2}$$

3.1.2. Handling of Date-time Variables

For CY, some useful data are comprised by date along with time, which provides valuable information that is utilized to enumerate the variability in yield explicated by genetics along with space-time factors, examine CY trends in space along with time, and analyze how Spatio-temporal data could be integrated. It is also employed in modelling together with prediction yield. Nevertheless, to make date-time variables into useful information, some transformations are required. Some useful data about the CY, which supports augmenting the model's accuracy, may get lost by the removal of data-time variables.

$$\Theta_{DT} = \aleph_{Datetime}(C^{DS}_{n}) \tag{3}$$

Where, the function that handles date-time variables is signified as $\aleph_{Datetime}$.

3.1.3. Handling of Categorical Feature

In a dataset, several significant data are hidden along with masked by the categorical feature. The ML methodologies don’t have the potency to read categories; thus, handling categorical features is highly significant. First, the categories are transmuted into numerical variables; then, they are utilized by the ML methodologies.

$$\Theta_{Cat} = \Im_{categorical}(C^{DS}_{n}) \tag{4}$$

Where, the function that handles the categorical features is specified as $\Im_{categorical}$. The function may be one hot encoding or KDD CUP Orange or Label encoder or mapping function.

3.1.4. Scaling

The feature scaling is performed to acquire the dataset feature of the same unit, that is, to standardise the range of independent variables or data features. To scale down the features in the same range, the Minmax Scaler is utilized. The feature vector betwixt a range of 0 and 1 is computed by the Minmax Scaler, which is expressed as,

$$\Theta_{Sca} = \frac{C_{i\times j} - \min\left(C_{i\times j}\right)}{\max\left(C_{i\times j}\right) - \min\left(C_{i\times j}\right)} \tag{5}$$

Therefore, to attain healthier data, the data is pre-processed; thus, mitigating the error rate.

3.1.5. Handling Outliers

In some cases, extreme values are possessed by the dataset. These values, which are beyond the expected range or unlike the other data, are mentioned as outliers. Generally, by understanding and also by taking away these outlier values, the ML modelling along with model skills can be enhanced. Here, to find outliers, a Quantile score is produced. It handles sampling fluctuations along with differentiation in open distribution, handles skewed distribution, and achieves the finest elimination of outliers than the prevailing methodologies. The Quantile score is proffered as the absolute value of the individual feature value (C_i) minus the median value ($\tilde{C}$) by its inter quantile range (IQR). It is formulated as,

$$\Theta_{out} = \frac{\left| C_i - \tilde{C} \right|}{I_U - I_L} \tag{6}$$

Where, the upper and lower quantiles are denoted as I_U and I_L. It is computed as,

$$IQR = Q_3 - Q_1 \tag{7}$$

$$I_U = Q_3 + 1.5(IQR) \tag{8}$$

$$I_L = Q_1 - 1.5(IQR) \tag{9}$$

Where, the inter quantile range is specified as IQR, the 3rd quantile range of 2nd half is signified as Q_3 and the 2nd quantile range of the 1st half is notated as Q_1.The median value is represented by the 2nd and 1st half. It is gauged by partitioning the data into '2' half; 1st half is lesser than the median value; 2nd half is larger than the median value for which the median value is analyzed.

Lastly, healthier data are obtained in pre-processing by removing the probability of errors. Additionally, to proceed with further processing, the features are formulated within the data frame. It is expressed as,

$$\Theta_{pre}^{Y} = \left[C_1, C_2, C_3, C_4, C_5, ...C_n \right]_{i \times j} \qquad \text{n} \tag{10}$$

3.2. Redundant Data Removal

Features that possess the same distribution along with zero variance are detected by utilizing this redundant data removal process. Data complexity along with overrate noise that brings about imprecise CYP may occur owing to the existence of '2' features with the same distribution. Initially, the ANOVA test is conducted. To verify whether the means of 2 or more features are crucially varied from every single other, the ANOVA test is utilized. By verifying amongst 2 or more features, the level of variance within the dataset is evaluated with the aid of this test. Null hypothesis and alternative hypothesis are the set of hypotheses utilizing which the ANOVA test is initialized.

$$H_0 : The\ two\ groups\ \mathrm{var}iance\ are\ significantly\ same$$
$$H_1 : The\ two\ groups\ \mathrm{var}iance\ are\ not\ significantly\ same \tag{11}$$

The hypothesis is estimated by analyzing the Correction term (T), which is gauged by dividing the sum of squares of features by total features in the dataset (N). It is expressed as,

$$\mathrm{T} = \frac{\left(\sum c\right)^2}{N} \tag{12}$$

After that, by subtracting the difference of the sum of the squares of features $\left(\sum c\right)^2$ from the correction term (T), the Sum of squares total is analyzed. It is formulated as,

$$\chi^{sst} = \sum c^2 - \mathrm{T} \tag{13}$$

Next, the Sum of squares betwixt features (χ^{ssb}) is estimated by the difference of variation of feature means as of the total grand mean (n) to the correction term. It is given as,

$$\chi^{ssb} = \frac{\left(\sum c\right)^2}{n} - \mathrm{T} \tag{14}$$

Afterwards, by calculating the difference amongst the Sum of squares and the sum of squares betwixt features, the Sum of individual features within the dataset is determined. It is expressed as,

$$\chi^{sfd} = \chi^{ssw} - \chi^{ssb} \tag{15}$$

By dividing the sum of squares between features with degrees of freedom ($r-1$), the mean of the Sum of squares betwixt features is analyzed. It is formulated as,

$$\chi^{mssb} = \frac{\chi^{ssb}}{r-1} \tag{16}$$

Where, categories of features are represented as k .

By dividing the Sum of features within the dataset with degrees of freedom ($N-r$), the mean of the Sum of squares within features is estimated. It is expressed as,

$$\chi^{mssw} = \frac{\chi^{ssw}}{N-r} \tag{17}$$

Lastly, F-Ratio is attained, which is expressed as,

$$f = \frac{\chi^{mssb}}{\chi^{mssw}} \tag{18}$$

The ratio of χ^{mssb} and χ^{mssw} is recognized to pursue the F distribution. Consequently, to attain statistical conclusions, the F value obtained as of the observed data with critical values at an error level of 0.05 in the F table is analogized. At last, the hypothesis significance is confirmed with a 95% confidence interval regarding the F-value; then, the similar features with the same data spread are taken away.

$$C^{T}_{M \times N} = \Omega''\left[c^{1}_{m \times n}, c^{2}_{m \times n}, c^{3}_{m \times n} .... c^{t}_{m \times n}\right] \tag{19}$$

Where, the sampled data's similarity measures are specified as Ω'' ; then the similarly distributed features are taken away regarding the measures.

3.3. Feature Selection

To obtain outcomes with higher accurateness in prediction along with classification, appropriate feature engineering together with the FS process is regarded as the backbone for any ML technique. For CY, the dataset's dimensionality is mitigated by the FS methodology by choosing highly pertinent data. FS is highly significant in the prediction of CY also because owing to the reliance of CY on numerous factors like weather, climate, fertilizer used, soil, seed variety, et cetera that brings about higher complications in detecting relevant features, the prediction of CYs indulge with various datasets collection. The problem of baseline feature volatility that is to say discrepancies regarding alterations in the input data is a complication that exists still even though several methodologies have been produced. In fact, there is a chance to occur inaccurate prediction, if the selection process's outcome is highly sensitive to variations in the set of training instances. Subsequently, the inaccurate prediction might also occur owing to the over-fitting issue. An ET-DPFS is developed to overcome the aforementioned complications. The dataset is sub-sampled into numerous partitions by the developed methodology; thus, providing the common features utilizing the intersection operation. A threshold methodology was developed in which the feature intersection is decided regarding the multiple FS if there are n FS methodologies; after that, the feature, which is common in the entire n methodologies is prioritized first by the outcome produced by the intersection.

$$F(C_i) = \bigcap_{z=1}^{Z} F_{z,i} \tag{20}$$

Rule 2: Sometimes, the outcomes generated by the FS methodology might be relevant data or irrelevant data. In that case, to differentiate them, the threshold (η) value is utilized. If the outcomes are greater than the threshold value, then the feature obtained is pondered as a relevant feature, or else it is considered as an irrelevant feature.

$$F(C_i) = \eta \bigcap_{z=1}^{Z} F_{z,i} \tag{21}$$

Therefore, the ET-DPFS functions regarding the aforementioned '2' rules(i) Lasso regression from embedded methodology, (ii) Mayfly optimization from wrapper methodology, and (iii) Relief technique from filter methodology are the '3' FS mechanisms contained by the ET-DPFS.Identifying the most relevant features regarding the dependent variables is the intention of these methodologies. It is expressed as,

$$Obj(c_i) = [\varphi_1 c_1 + \varphi_2 c_2 + \varphi_3 c_3 + \varphi_4 c_4 + \varphi_5 c_5 + ... + \varphi_n c_n] \tag{22}$$

Where, the fitness value in wrapper methodology, slope in embedded methodology, and weights in filer methodology are signified as $\varphi_1, \varphi_2, \varphi_3 ...., \varphi_n$.

3.3.1 Feature Selection Technique using the Wrapper Method

The mayflies' social behaviour particularly from their mating process is the factor on which the FS model 1 for choosing the highly relevant feature relied. It was assumed that mayflies are already adults after hatching from the egg; also, the fittest mayflies survive regardless of how long they live. The solution for the problem is specified by every single mayfly in the search space. However, the previous Mayfly optimization gets stuck into the local optimal solution; thus, whilst updating the male along with the female mayfly's position during positive attraction constant value, the search speed might get reduced gradually. The CV-MO is developed to avoid the trap down. To retain the balance betwixt local optimum and global optimum, this methodology employs a coefficient vector, which proffers the minimum along with the maximum value.

Firstly, male along with female mayfly's populaces are created randomly. Then, every single fly is specified as a candidate solution for every single problem in the search space that is to say *a d*-dimensional vector $c = c_1, c_2, c_3...., c_n$; subsequently, the performance is analyzed on objective function ($Obj(c_i)$).

For every single mayfly, the velocity $\gamma = \gamma_1, \gamma_2, \gamma_3...., \gamma_n$ specifies the position change along with flying direction for the individual in conjunction with social flying experience. Particularly, each mayfly modifies its trajectory toward its own best position ($\aleph_i^p$) thus far, as well as the best position reached by any mayfly in the swarm thus far ($\aleph_i^g$).

At this instant, a male mayfly's movement is analyzed regarding its own experience along with its neighbours' experience. Let the mayfly's current position i in the search space at time step t is specified as c_i^t, by adding a velocity γ_i^{t+1} to the current position, the position is changed. It is expressed as,

$$c_i^{t+1} = c_i^t + \gamma_i^{t+1} \tag{23}$$

With $c_i^0 \sim U(c_{\min}, c_{\max})$

The male mayflies' velocity gets down during nuple dance. Consequently, a male mayfly's velocity i is computed as,

$$\gamma_{ij}^{t+1} = \gamma_{ij}^t + \alpha_1 e^{-\beta R_P^2}\left(\aleph_{ij}^p - c_{ij}^t\right) + \alpha_2 e^{-\beta R_P^2}\left(\aleph_{ij}^g - c_{ij}^t\right) \tag{24}$$

$$\alpha_1 = 1/4 \log\left(\ell + \frac{1}{\ell_{\max}}\right) q \tag{25}$$

$$\alpha_2 = 2.h \tag{26}$$

Where, the mayfly's velocity in time step dimension is specified as γ_{ij}^t, the position of mayfly i in dimension j at time step t is signified as c_{ij}^t, positive attraction coefficient vectors utilized to scale the contribution of the cognitive and social component are represented as α_1 and α_2, respectively, the maximum iteration is indicated as $\ell_{\max}$, the coefficient vector ranging between -1 and 1 is denoted as q, the random number that lies betwixt [0,1] is illustrated as h .Moreover, the best position mayfly i that had been ever visited is notated as $\aleph_i^p$. Regarding minimization problems, the personal best position $\aleph_{ij}^p$ at the next time step $t+1$ is computed as,

$$\aleph_i^p = \begin{cases} c_i^{t+1}, & if\ f\left(c_i^{t+1}\right) < f\left(\aleph_i^p\right) \\ is\ kept\ the\ same, & otherwise \end{cases} \tag{27}$$

Where, the objective function is illustrated as $f: Rn \rightarrow R$. It analyzes a solution's quality. The global best position $gbest$ at time step t is proffered as,

$$\begin{aligned}\aleph_i^g &\in \left\{\aleph_1^p, \aleph_2^p, \aleph_3^p, ...\aleph_n^p \mid f\left(best^c\right)\right\} \\ &= \min\left\{f\left(\aleph_1^p\right), f\left(\aleph_2^p\right), f\left(\aleph_3^p\right), ..., f\left(\aleph_n^p\right)\right\}\end{aligned} \tag{28}$$

Where, the total number of male mayflies in the swarm is specified as n. Lastly, β is a fixed visibility coefficient utilized in eq. (7). It is utilized to restrict a mayfly's visibility to others. The Cartesian distance betwixt c_i and $\aleph_i^p$ is signified as R_p along with the Cartesian distance betwixt c_i and $\aleph_i^g$ is R_G. These distances are computed as,

$$\left\|c_i - \Im_i\right\| = \sqrt{\sum\left(c_{ij} - \Im_{ij}\right)^2} \tag{29}$$

Where, the j^{th} element of mayfly i is notated as c_{ij} and $\Im_i$ corresponds to $\aleph_i^p$ or $\aleph_i^g$.

The best mayflies in the swarm must continue to perform their unique up-and-down nuptial dance for the algorithm to work flawlessly. As a result, the best mayflies must constantly adjust their velocities, which are measured as,

$$\gamma_{ij}^{t+1} = \gamma_{ij}^t + D.R \tag{30}$$

In which case, the nuptial dance coefficient and a random value in the range [-1, 1] is described as D and R. To the algorithm, a stochastic element is introduced by this up and down movement.

At this moment, the female mayflies' movement is calculated. However, the female flies don't gather like male flies rather they fly towards males to breed. The position is altered by adding a velocity γ_i^{t+1} to the i current position by assuming x_i^t as the current position of the female mayfly i in the $y$search space at a time step t. It is computed as,

$$x_i^{t+1} = x_i^t + \gamma_i^{t+1} \tag{31}$$

With $x_i^0 \sim U\left(x_{\min}, x_{\max}\right)$

Now, the best female flies' attraction gets attracted towards the best male. Hence, their velocities are computed regarding minimization problems as,

$$\gamma_i^{t+1} = \begin{cases} \gamma_{ij}^t + \alpha_2 e^{-\beta R_{mf}^2}\left(c_{ij}^t - x_{ij}^t\right) & if\ \ f(x_i) > f(c_i) \\ \gamma_i^{t+1} + fl * R & if\ \ f(x_i) \le f(c_i) \end{cases} \tag{32}$$

Where, the velocity of female mayfly i in dimension $j = 1,2,3,4...,n$ at time step t is proffered as γ_{ij}^t, the position of female mayfly i in dimension j at time step t is defined as γ_{ij}^t, the positive attraction coefficient vector, which is analyzed utilizing Eqn. (25 & 26) along with fixed visibility coefficient, is illustrated as α_2 and β, and the Cartesian distance betwixt male and female mayflies is symbolized as R_{mf}. Lastly, a random walk coefficient is signified as fl, when a female is not attracted by a male, it flies at random and a random value in the range [- 1, 1] is indicated as R.

After that, the mating of the fly takeovers is a crossover operator. From the male populace, one parent is chosen and the other is selected as of the female populace. The process by which females are attracted to males and the process by which parents are selected are similar. The selection is done either randomly or regarding their fitness function. Whilst considering the fitness function, The best female crosses with the best male, the second-best female crosses with the second-best male, and so the following are the ‘2’ offspring produced by the crossover.

$$\Phi_1 = \wp * f(c_i) + (1-\wp) * f(x_i) \tag{33}$$

$$\Phi_2 = \wp * f(x_i) + (1-\wp) * f(c_i) \tag{34}$$

Where, the male parent is specified as $f(c_i)$, the female parent is signified as $f(x_i)$ and a random value within a particular range is denoted as $\wp$. Offspring’s initial velocities are set to be zero. Lastly, the worst solution is substituted with the best ones and the pbest and gbest values are updated. The most relevant features are provided by the wrapper methodology to predict the CY, which is formulated as,

$$FS_1^w = \left[c_1^w, c_2^w, c_3^w, c_4^w, c_5^w c_6^w, c_7^w ... c_n^w\right] \tag{35}$$

3.3.2 Feature Selection Technique using Embedded Method

The embedded methodology performs like FS along with classification. The MAD-LR is developed here. The complexity like the number of absolute size of the sum of all coefficients in the model whilst minimizing the residual sum of squares is reduced by this approach. By utilizing MAD as a tuning parameter, the problem of noise level along with time computation is also conquered. The variance is minimized by the MAD-LR. It also conducts variable selection. The estimated regression coefficient is shrunk to zero by the LR approach. The MAD-LR estimate is proffered as,

$$Z'_{lasso} = \arg\min_{z} \left\{ \frac{1}{2} \sum_{i=1}^{N} \frac{\left(\mathrm{Y}_i - \mathrm{Z}_0 - \sum_{j=1}^{p} c_{ij} \mathrm{Z}_j \right)^2}{Median\left|\mathrm{Y}_i - \tilde{\mathrm{Y}}_i\right|} + \lambda \sum_{j=1}^{p} \left|\mathrm{Z}_j\right| \right\} \tag{36}$$

It is also estimated as,

$$Z'_{lasso} = \arg\min_{z} \left\{ \sum_{i=1}^{N} \frac{\left(\mathrm{Y}_i - \mathrm{Z}_0 - \sum_{j=1}^{p} c_{ij} \mathrm{Z}_j \right)^2}{Median\left|\mathrm{Y}_i - \tilde{\mathrm{Y}}_i\right|} \right\},$$

$$= subject\,to\,\lambda \sum_{j=1}^{p} \left|\mathrm{Z}_j\right| \le t, \tag{37}$$

Where, the actual value is represented as Y_i, the slope is denoted as Z_0, the learning rate is indicated as λ and the input features are defined as c_{ij}, all the error values’ median value is proffered as $\tilde{\mathrm{Y}}_i$. The predicted value is represented by the term $\left(\hat{\mathrm{Y}} = -\left(\mathrm{Z}_0 + \sum_{j=1}^{p} c_{ij} \mathrm{Z}_j \right) \right)$.

MAD-LR converts every single coefficient by a constant component λ, truncating at zero. As a result, it is a predictive variable selection tool for regression. The residual sum of squares exposed to the absolute value of the coefficients is reduced less than a constant. By choosing highly relevant features, the model’s accurateness is ameliorated with the aid of LASSO. Regarding correlation, the redundant data are eliminated by the LASSO. It selects only one amongst them and shrinks the others to zero. By shrinking some coefficients to zero, the estimates’ variability is reduced; thus, producing effortlessly interpretable models. Lastly, to predict the CY, the most relevant features are provided by the embedded methodology. It is formulated as,

$$FS_2^E = \left[c_1^E, c_2^E, c_3^E, c_4^E, c_5^E, c_6^E, c_7^E \ldots c_n^E\right] \tag{38}$$

3.3.3 Feature Selection Technique using the Filter Method

Rather than the model, the variables are selected by the filter-centric FS methodology. They operate fundamentally in association with the variable that is to be predicted. The HEwint-Relief feature weighing algorithm is developed, which offers several weights regarding the relevancy along with categories. By utilizing this technique, the problem of higher computational complexity whilst analyzing the connection betwixt characteristics by mutual information is conquered. To calculate the initial weight along with to avoid the slowdown of the computational time, the HE normal distribution weight initialization scheme is utilized by this approach. In this developed approach, the relevance of features along with categories is centered on the features’ ability to differentiate betwixt close-sample. At first, by utilizing the HE normal distribution weighs initialization mechanism, the weights are initialized, which is notated as,

$$\begin{aligned} &w[c]^t \approx N[0,\sigma] \\ &\sigma = \sqrt{2/c_{in}} \end{aligned} \tag{39}$$

Where, current weights are illustrated as $w[c]^t$, the normal distribution is described as $N[0,\sigma]$with standard deviation and mean as 0.

After that, from the training sets D, a sample data point κ is selected randomly by the approach; subsequently, the nearest neighbour is searched, that is to say, the near hit search for the nearest neighbour ς from the similar samples with κ and let $\hbar$ be the non nearest neighbour. For every single feature, the weights’ updation is specified as,

$$w[c]^{t+1} = w[c]^t - diff\left(c,\kappa,\varsigma\right)/m + diff\left(c,\kappa,\hbar\right)/m \tag{40}$$

Where, function $diff\left(feature, ins\tan ce1, ins\tan ce2\right)$ computes the difference amongst the feature of the ‘2’ varied samples that are discrete along with continuous features is notated as,

For discrete features:

$$diff\left(c,I_1,I_2\right) = \begin{cases} 0 & Value(c,I_1) = Value(c,I_2) \\ 1 & others \end{cases} \tag{41}$$

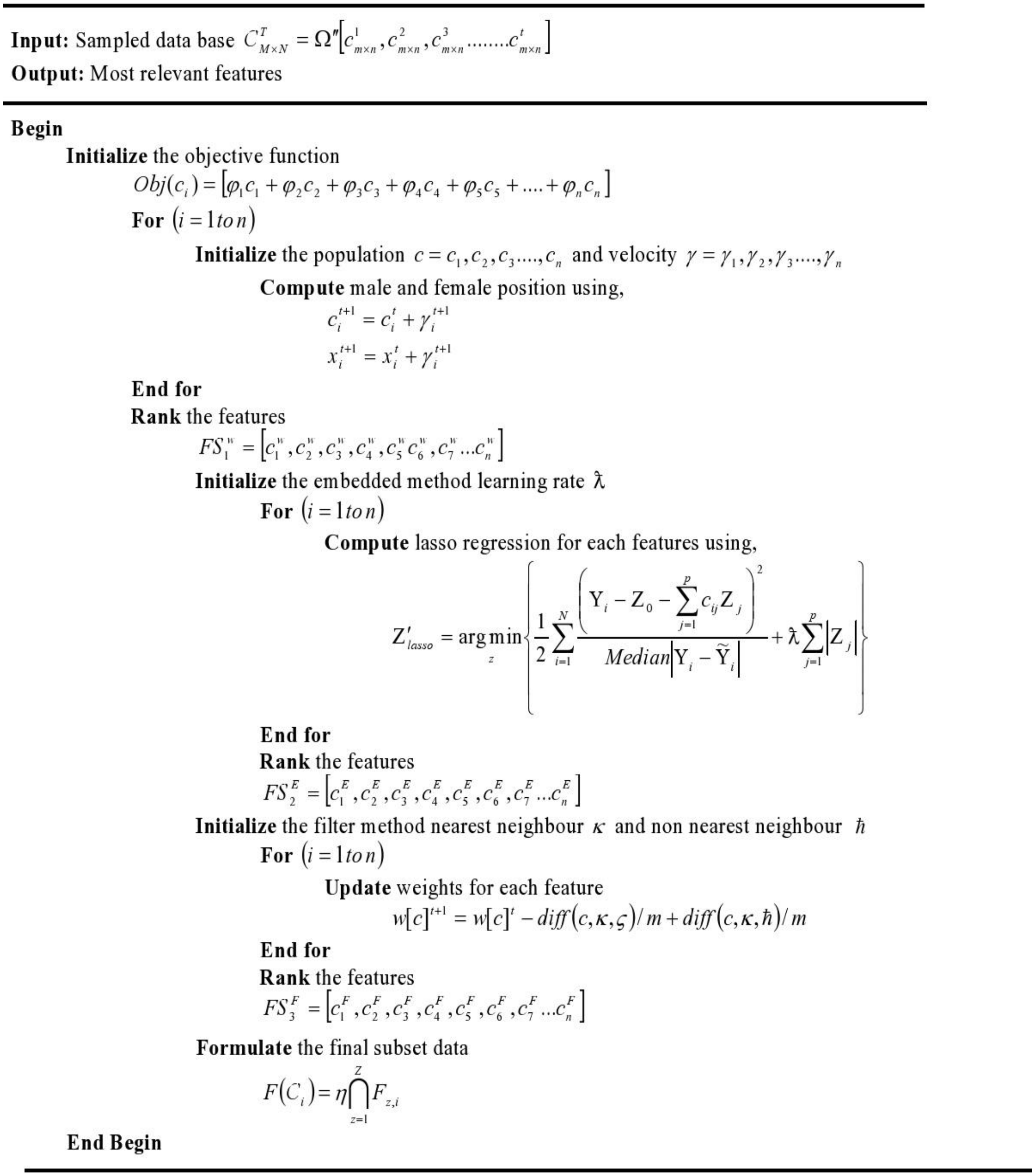
For continuous feature:

$$diff\left(c,I_1,I_2\right) = \frac{Value(c,I_1) - Value(c,I_2)}{\max(c) - \min(c)} \tag{42}$$

The feature’s weights are sorted in descending orders after being updated; then, the feature with the highest weight entails a highly relevant feature for better classification capability. Consequently, the dimensionality curse problem is addressed by the proposed ET-DPFS technique. Therefore, the highly relevant features are ordered and are expressed as,

$$FS_3^F = \left[c_1^F, c_2^F, c_3^F, c_4^F, c_5^F, c_6^F, c_7^F \ldots c_n^F\right] \tag{43}$$

According to the rules, the relevant feature is chosen; then, the feature being selected is inputted into the prediction model. Figure 2 exhibits the proposed ET-DPFS's pseudo-code.



```
Input: Sampled data base $C^T_{M\times N} = \Omega''\left[c^1_{m\times n}, c^2_{m\times n}, c^3_{m\times n} \ldots\ldots c^t_{m\times n}\right]$
Output: Most relevant features

Begin
    Initialize the objective function
        $Obj(c_i) = \left[\varphi_1 c_1 + \varphi_2 c_2 + \varphi_3 c_3 + \varphi_4 c_4 + \varphi_5 c_5 + \ldots + \varphi_n c_n\right]$
        For $(i = 1 \, to \, n)$
            Initialize the population $c = c_1, c_2, c_3, \ldots, c_n$ and velocity $\gamma = \gamma_1, \gamma_2, \gamma_3, \ldots, \gamma_n$
                Compute male and female position using,
                    $c_i^{t+1} = c_i^t + \gamma_i^{t+1}$
                    $x_i^{t+1} = x_i^t + \gamma_i^{t+1}$
        End for
        Rank the features
            $FS_1^w = \left[c_1^w, c_2^w, c_3^w, c_4^w, c_5^w c_6^w, c_7^w \ldots c_n^w\right]$
            Initialize the embedded method learning rate $\bar{\lambda}$
                For $(i = 1 \, to \, n)$
                    Compute lasso regression for each features using,
                        $Z'_{lasso} = \arg\min_z \left\{ \frac{1}{2} \sum_{i=1}^{N} \frac{\left( Y_i - Z_0 - \sum_{j=1}^{p} c_{ij} Z_j \right)^2}{Median\left|Y_i - \widetilde{Y}_i\right|} + \bar{\lambda} \sum_{j=1}^{p} \left|Z_j\right| \right\}$
                End for
                Rank the features
                $FS_2^E = \left[c_1^E, c_2^E, c_3^E, c_4^E, c_5^E, c_6^E, c_7^E \ldots c_n^E\right]$
            Initialize the filter method nearest neighbour $\kappa$ and non nearest neighbour $\hbar$
                For $(i = 1 \, to \, n)$
                    Update weights for each feature
                        $w[c]^{t+1} = w[c]^t - diff(c, \kappa, \varsigma)/m + diff(c, \kappa, \hbar)/m$
                End for
                Rank the features
                $FS_3^F = \left[c_1^F, c_2^F, c_3^F, c_4^F, c_5^F, c_6^F, c_7^F \ldots c_n^F\right]$
            Formulate the final subset data
                $F(C_i) = \eta \bigcap_{z=1}^{Z} F_{z,i}$
    End Begin
```

Figure 2: Pseudo Code for Proposed ET-DPFS

3.4. Prediction Model

The prediction model gives a prediction after being trained on the selected features, that is to say, regarding the climate, soil nature, et cetera, the crop apt for cultivation is predicted. For crop prediction, the ABP-XGBOOST methodology is being developed. The issues like computational complexity owing to the nodes' depth and the over fitting along under fitting problems are conquered by this methodology. The alpha-beta pruning that is utilized to XGCOOST is employed by this model to prune the whole tree leaves or entire sub-tree, which doesn't influence the final decision; in addition, it is primarily accountable for making the algorithm slow.

Step1: Firstly, the base model is built. Types of crops, which describe a multiclass classification model, are the base model's outputs. Regarding the outputs, the base model's probability is gauged as,

$$\Pr(Bm) = \frac{\sum_{i=1}^{n} c_i}{n} \tag{44}$$

The following equation computes the residual regarding the base model's probability.

$$\operatorname{Re} s_i = \sum_{i=1}^{n} \mathrm{Y}_i - \Pr(Bm) \tag{45}$$

Regarding every single feature, the trees are built after detecting the residual. The features are branched and the corresponding residuals are separated regarding the categories.

Step2: At this moment, the similarity weight of the split (let it be (s_1) and (s_2)) and the root node are, (R_n)

$$Sw = \frac{\sum (\operatorname{Re} s)^2}{[\Pr(1-\Pr)+\Omega]} \tag{46}$$

Where, $\Omega = 0$ or it can be selected by utilizing the hyper parameter tuning.

Step3: Here, the tree's gain is computed as,

$$gain = Sw(s_1) + Sw(s_2) - Sw(R_n) \tag{47}$$

To discover the best rood node for splitting, the gain is established. The remaining features will be split (binary split) regarding the best split.

Step4: Here, the Alpha-beta pruning $(\alpha - \beta)$ decides whether the trees' splitting should be continued or not. Essentially α is provided by the highest value so far detected at any point along the path of maximizer that is initialized with $-\infty$ and β describes the lowest value so far established at any point along the path of maximizer that is initiated with $+\infty$. Initially, the maximum best node is estimated.

Firstly, the minimum gain node is analyzed,

$$eva = \min(node, depth, alpha, beta) \tag{48}$$

The maximum gain node is assessedby utilizing,

$$\begin{aligned} Maxeva &= \max(Maxeva, eva) \\ \alpha &= \max(\alpha, Maxeva) \end{aligned} \tag{49}$$

After that, the alpha value is analogized that is $if\left(\beta <= \alpha\right) \;\; return \;\; Maxeva$

Now, the minimum node is analyzed.

$$eva = \min(\, node, depth, alpha, beta) \tag{50}$$

The maximum gain node is appraised by utilizing,

$$\begin{aligned} Mineva &= \min(\, Mineva, eva) \\ \beta &= \min(\, \beta, eva) \end{aligned} \tag{51}$$

Next, the beta value is analogized that is $if\left(\beta <= \alpha\right) \;\; return \;\; Mineva$

Step5: Now, for gauging the new data, the model estimates by employing,

$$\begin{aligned} \Pr{}^{+} &= \frac{1}{1+e^{-x}} \\ x &= [Bs_{out} + \sum_{i=1}^{n} \varepsilon(T_i)] \\ Bs_{out} &= \log\left(\frac{\Pr}{1-\Pr}\right) \end{aligned} \tag{52}$$

Regarding the new probability, the residual is computed again and the steps from step 2 to step 5 are repeated until it attains a lesser value.

Therefore, the CY's accurate prediction is provided by the developed methodology regarding the features being selected.

4 Results and Discussion

Using diverse parameters, the proposed framework for forecasting CY is authenticated and analogized with the prevailing approaches to determine its efficiency. The work is executed using PYTHON, which is derived from free open sources.

4.1. Performance Analysis Proposed Feature Selection Technique based on Computation Time

To acquire the optimal features for validating a low-error prediction model, the proposed ET-DPFS is evaluated using the computation time metrics. Mutual information FS (MIFS), Hybrid filter, Wrapper FS, Ridge FS, and Extra tree FS are the few prevailing approaches that are correlated with the acquired outcome. Figure 3 shows a pictorial depiction of the computing time for the ET-DPFS model.

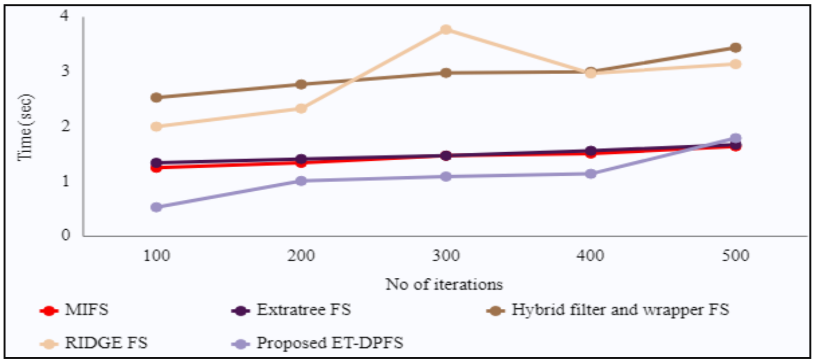


Figure 3: Graphical Demonstration of Computation Time for Proposed ET-DPFS Technique

The proposed ET-DPFS has a computation time that varies from 0.54s-1.8s for iterations of 100 to 500 is presented in figure 3. The proposed FS methodology for dimensionality decrease takes lesser time when analogized to the prevailing techniques that manage to acquire a time value varying from 1.26s to 3.45s. The FS technique's convergence rate is low and large error rate, owing to the long computing time. In relation to that, the prevailing techniques have a more error rate and are inferior to the proposed FS.

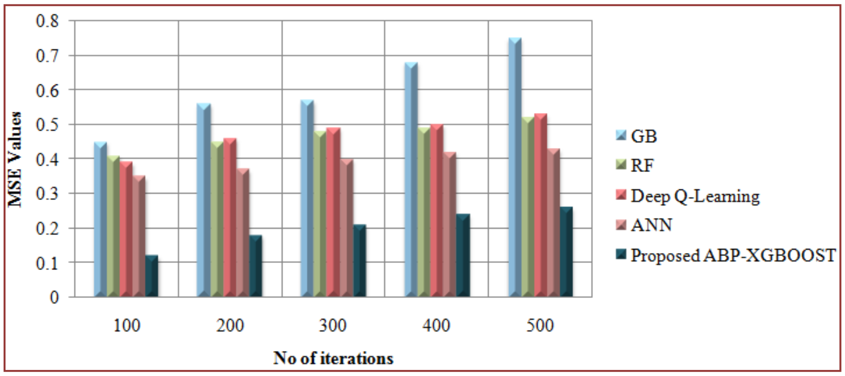
4.2. Evaluation of Prediction Model based on Various Metrics

Several metrics like Mean Absolute Percentage Error (MAPE), R-Squared, MSE, accuracy, RMSE, and MAE together with the prevailing approaches like RF, Gradient Boosting (GB), ANN, and Deep Q-Learning, are utilized to compute the ABP-XGBOOST method for forecasting the CY. The metrics' estimation is calculated below.

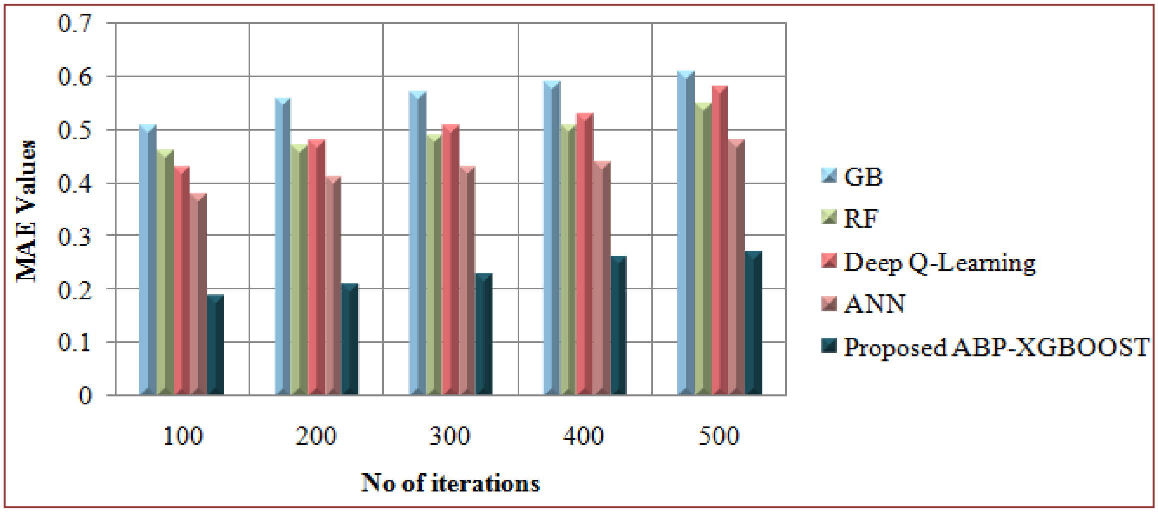
Table1: Evaluation of Proposed ABP-XGBOOST based on Accuracy

Techniques/iterations	**100**	**200**	**300**	**400**	**500**
GB	65	68	62	63	75
RF	68	71	70	69	79
Deep Q-Learning	69	72	73	74	75
ANN	70	72	73	75	78
Proposed ABP-XGBOOST	82	85	88	89	91

The accuracy levels attained by the proposed and prevailing techniques at every iteration are described in table 1. With the exception of the GB and RF algorithms, which have few fluctuations during the 300 to 400th iterations, the accuracy score slowly increases with every iteration that can be noticed from the table. However, when analogized with the prevailing techniques, which reach a total accuracy of 65%-79%, the ABP-XGBOOST strategy obtains a greater accuracy varying from 82% to 91% for each iteration. Thus, the ABP-XGBOOST has a good track record of accurately forecasting CY with a less error rate. Figure 4 describes the MSE and MAE's pictorial estimation for the ABP-XGBOOST methodology.



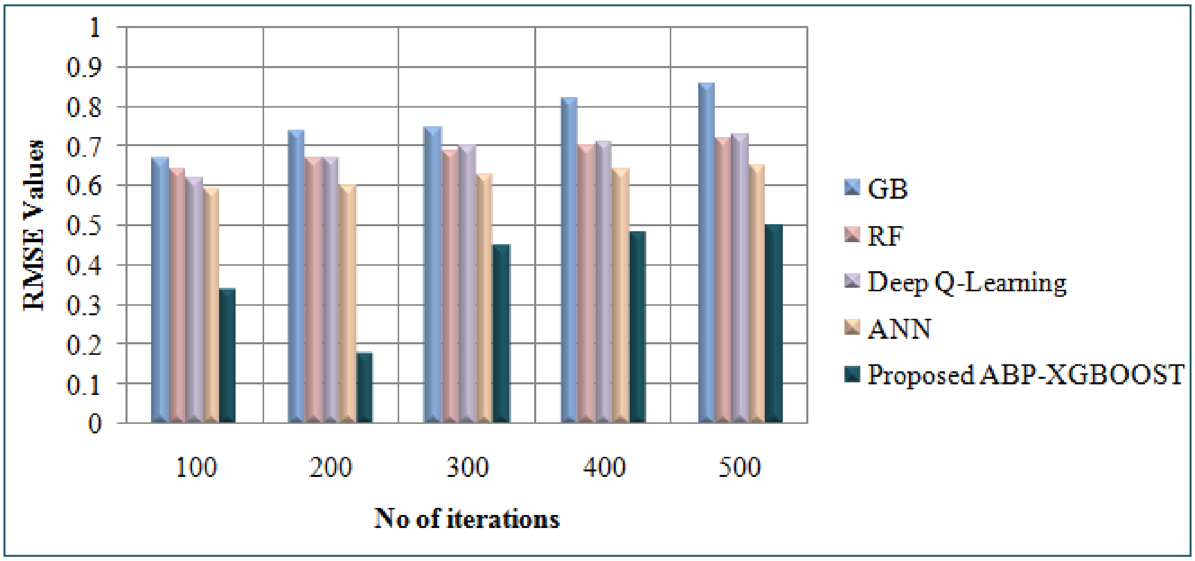
(a)



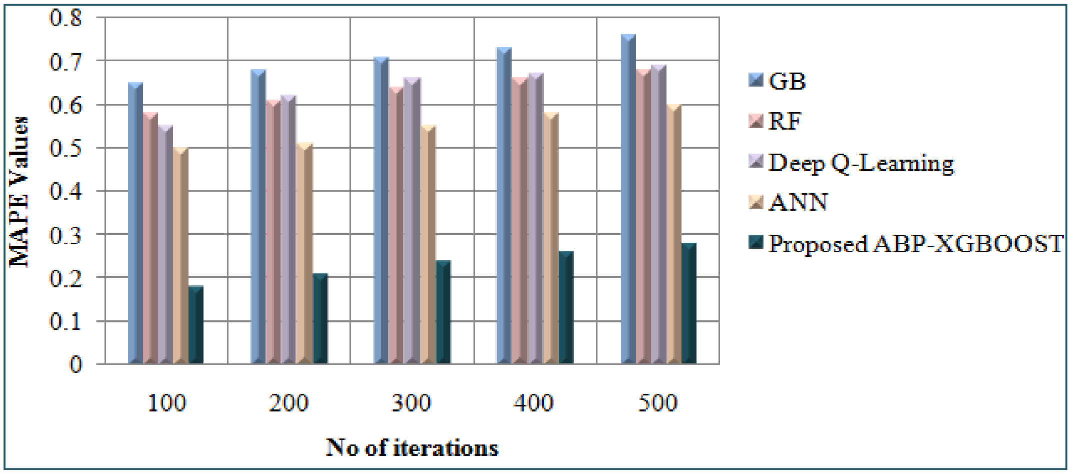
(b)

Figure 4: Graphical Analysis of Proposed ABP-XGBOOST based on (a) MSE(b) MAE

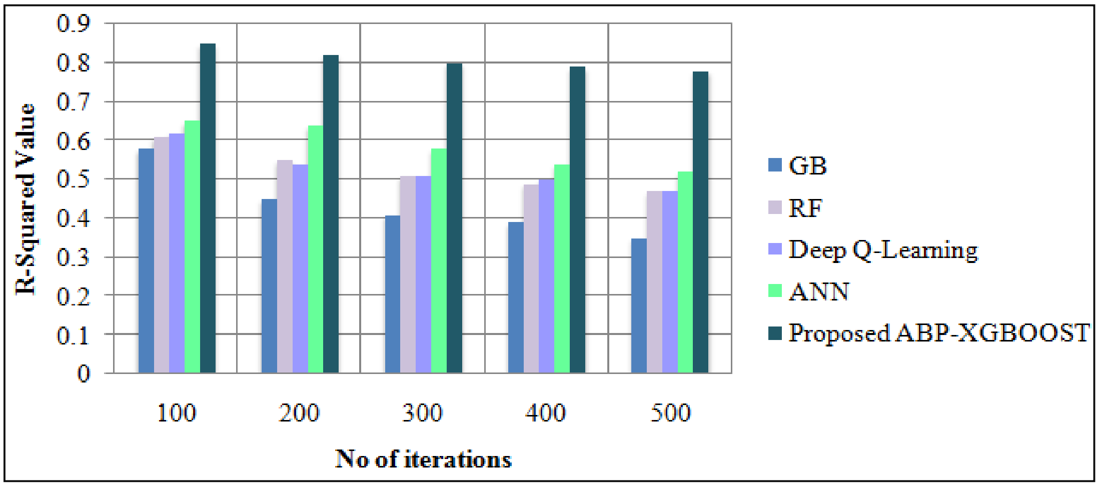
The proposed techniques' MSE along with MAE value verification are showcased in figure 4. MSE is determined by dividing the total number of observations by the squared sum of the absolute difference between the original and projected values. If the MSE has greater value, then it is a poor prediction model. So, as a result, the prevailing approaches get an MSE value of 0.35 and 0.75, which indicates the poorest model, whereas the ABP-XGBOOST model achieves an MSE value that varies from 0.12 to 0.26, as demonstrated in figure 4(a). Similarly, the sum of the absolute difference betwixt the original and forecasted values divided by the total of the entire observations is termed MAE. To accomplish superior prediction, MAE's value must be maintained less, just like MSE. Therefore, the ABP-XGBOOST technique accomplishes an MAE value ranging betwixt 0.19 - 0.27 and it is comparatively small as compared to the prevailing techniques, which obtains an overall MAE of 0.38 to 0.61 as depicted in figure 4(b). In figure 5, the ABP-XGBOOST methodology is evaluated using R-Square, RMSE, and MAPE.



(a)



(b)



(c)

Figure 5: Graphical Analysis of Proposed ABP-XGBOOST based on (a) RMSE(b) MAPE (c) R-Square

The proposed method's RMSE validation is shown in figure 5(a). The MSE's square root is utilized to calculate the RSME value. Lesser the RMSE value, superior the prediction of CY. The ABP-XGBOOST accomplishes an RMSE value of 0.34 to 0.50 while, the prevailing approaches obtain an RMSE value of 0.59 to 0.86, which is comparatively poorer than the proposed technique.

The MAPE metrics, which measure the forecasting techniques' accuracy, are described in figure 5(b). For producing high-demand crops, the lesser MAPE value is used. As per this, for 500 iterations, the ABP-XGBOOST accomplishes a MAPE value of 0.28, while the prevailing approaches attain a MAPE of 0.60 - 0.76, indicating a greater error rate along with less prediction accuracy.

The R-Squared value for the ABP-XGBOOST technique is discussed in figure 5(c). The coefficients' closest determination to the original values is defined as R-Square. R-Square values range from 0 to 1, with 0 indicating the model that does not fit the dataset and 1 indicating a model that fits the specified data flawlessly. With respect to this, the ABP-XGBOOST have an R-Squared value of 0.78-0.85, but the prevailing techniques have an R-Squared value of 0.35-0.65 and it is comparatively less than the ABP-XGBOOST technique. Therefore, by completely fitting the data points, the ABP-XGBOOST model provides an accurate CYP.

5 Conclusion

Climate, soil, pest along with fertilizers, groundwater, irrigation applied, crop rotation, et cetera are several inter-reliant factors on which the CYP relies; thus, CYP is highly significant. The CY's prediction accuracy is mitigated by evaluating the complex structured data. Additionally, the system's overall efficacy is minimized owing to the issue of higher sensitivity, understating, overstating, data loss, et cetera. An ET-DPFS-centered ABP-XGBOOST CYP model is developed to overcome the aforementioned complications. A strong feature ranking along with feature subset selection is provided by the proposed model by utilizing DP methodologies. Larger together with smaller datasets are handled by the ET-DPFS by retaining overstating along with understating. Regarding the features, a highly relevant FS along with classification provides a novel model selection technique by integrating the classification performance along with strength in the analysis. Experiential outcomes displayed that for CYP, the proposed methodology attains an average accuracy of 87%; then, predicts the CY with an average MSE, MAE, and RMSE of 0.202, 0.232, and 0.39, respectively within a lower computation time of 1.122 secs. Therefore, the proposed model's overall performance remains higher than the prevailing methodologies.

References

[1] Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, *177*, 1-18.

[2] Shidnal, S., Latte, M.V., & Kapoor, A. (2021). Crop yield prediction: two-tiered machine learning model approach. *International Journal of Information Technology*, *13*(5), 1983-1991.

[3] Crane-Droesch, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*, *13*(11), 1-13.

[4] Nigam, A., Garg, S., Agrawal, A., & Agrawal, P. (2019). Crop yield prediction using machine learning algorithms. *In IEEE Fifth International Conference on Image Information Processing (ICIIP)*, 125-130.

[5] Abbas, F., Afzaal, H., Farooque, A.A., & Tang, S. (2020). Crop yield prediction through proximal sensing and machine learning algorithms. *Agronomy*, *10*(7), 1046.

[6] Medar, R., Rajpurohit, V.S., & Shweta, S. (2019). Crop yield prediction using machine learning techniques. *In IEEE 5th International Conference for Convergence in Technology*, 1-5.

[7] Gandhi, N., Armstrong, L.J., Petkar, O., & Tripathy, A.K. (2016). Rice crop yield prediction in India using support vector machines. *In IEEE 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 1-5.

[8] Kumar, R., Singh, M.P., Kumar, P., & Singh, J.P. (2015). Crop Selection Method to maximize crop yield rate using machine learning technique. *In IEEE international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM)*, 138-145.

[9] Kumar, Y.J.N., Spandana, V., Vaishnavi, V.S., Neha, K., & Devi, V.G.R.R. (2020). Supervised machine learning approach for crop yield prediction in agriculture sector. *In IEEE 5th International Conference on Communication and Electronics Systems (ICCES)*, 736-741.

[10] Lasso, E., Corrales, D.C., Avelino, J., de Melo Virginio Filho, E., & Corrales, J.C. (2020). Discovering weather periods and crop properties favorable for coffee rust incidence from feature selection approaches. *Computers and Electronics in Agriculture*, *176*, 1-11.

[11] Bocca, F.F., & Rodrigues, L.H.A. (2016). The effect of tuning, feature engineering, and feature selection in data mining applied to rainfed sugarcane yield modelling. *Computers and electronics in agriculture*, *128*, 67-76.

[12] PS, M.G. (2019). Performance evaluation of best feature subsets for crop yield prediction using machine learning algorithms. *Applied Artificial Intelligence*, *33*(7), 621-642.

[13] PS, M.G., & Bhargavi, R. (2019). Selection of important features for optimizing crop yield prediction. *International Journal of Agricultural and Environmental Information Systems (IJAEIS)*, *10*(3), 54-71.

[14] Tiwari, P., & Shukla, P. (2020). Artificial neural network-based crop yield prediction using NDVI, SPI, VCI feature vectors. *In Information and Communication Technology for Sustainable Development*, Springer, Singapore, 585-594.

[15] Tiwari, P., & Shukla, P. (2019). A hybrid approach of TLBO and EBPNN for crop yield prediction using spatial feature vectors. *Journal of Artificial Intelligence*, *1*(2), 45-58.

[16] Nevavuori, P., Narra, N., & Lipping, T. (2019). Crop yield prediction with deep convolutional neural networks. *Computers and electronics in agriculture*, *163*, 1-9.

[17] Gopal, P.M., & Bhargavi, R. (2019). A novel approach for efficient crop yield prediction. *Computers and Electronics in Agriculture*, *165*, 1-9.

[18] Manjula, A., & Narsimha, G. (2015). XCYPF: A flexible and extensible framework for agricultural Crop Yield Prediction. *In IEEE 9th international conference on intelligent systems and control (ISCO)*, 1-5.

[19] Elavarasan, D., & Vincent, P.M. (2021). A reinforced random forest model for enhanced crop yield prediction by integrating agrarian parameters. *Journal of Ambient Intelligence and Humanized Computing*, *12*(11), 10009-10022.

[20] Elavarasan, D., Vincent PM, D.R., Srinivasan, K., & Chang, C.Y. (2020). A hybrid CFS filter and RF-RFE wrapper-based feature extraction for enhanced agricultural crop yield prediction modelling. *Agriculture*, *10*(9), 1-27.

[21] Khosla, E., Dharavath, R., & Priya, R. (2020). Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression. *Environment, Development and Sustainability*, *22*(6), 5687-5708.

Iniyan, S., & Jebakumar, R. (2022). Mutual information feature selection (MIFS) based crop yield prediction on corn and soybean crops using multilayer stacked ensemble regression (MSER). *Wireless Personal Communications*, *126*(3), 1935-1964.

[22] Dang, C., Liu, Y., Yue, H., Qian, J., & Zhu, R. (2021). Autumn crop yield prediction using data-driven approaches:support vector machines, random forest, and deep neural network methods. *Canadian Journal of Remote Sensing*, *47*(2), 162-181.

[23] Radhika, A., & Masood, M.S. (2021). Effective dimensionality reduction by using soft computing method in data mining techniques. *Soft Computing*, *25*(6), 4643-4651.

[24] Manfredi, S., Ceccato, M., Sciarretta, G., & Ranise, S. (2022). Empirical Validation on the Usability of Security Reports for Patching TLS Misconfigurations: User-and Case-Studies on Actionable Mitigations. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, *13*(1), 56-86.