Knowledge-Based Systems 232 (2021) 107460

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

A modified genetic algorithm and weighted principal component analysis based feature selection and extraction strategy in agriculture

K. Aditya Shastry *, Sanjay H.A.

Nitte Meenakshi Institute of Technology, Bengaluru 64, India

ARTICLE INFO

Article history:
Received 1 June 2021
Received in revised form 28 July 2021
Accepted 31 August 2021
Available online 3 September 2021

Keywords:
Feature selection
Feature extraction
Hybrid
Genetic Algorithm
Weighted-Principal Component Analysis

ABSTRACT

Data pre-processing is a technique that transforms the raw data into a useful format for applying machine learning (ML) techniques. Feature selection (FS) and feature extraction (FeExt) form significant components of data pre-processing. FS is the identification of relevant features that enhances the accuracy of a model. Since, agricultural data contain diverse features related to climate, soil, fertilizer, FS attains significant importance as irrelevant features may adversely impact the prediction of the model built. Likewise, FeExt involves the derivation of new attributes from the prevailing attributes. All the information that the original attributes possess is present in these new features minus the duplicity. Keeping these points in mind, this work proposes a hybrid feature selection and feature extraction strategy for selecting features from the agricultural data set. A modified-Genetic Algorithm (m-GA) was developed by designing a fitness function based on "Mutual Information" (MutInf), and "Root Mean Square Error" (RtMSE) to choose the best features that affected the target attribute (crop yield in this case). These selected features were then subjected to feature extraction using "weighted principal component analysis" (wgt-PCA). The extracted features were then fed into different ML models viz. "Regression" (Reg), "Artificial Neural Networks" (ArtNN), "Adaptive Neuro Fuzzy Inference System" (ANFIS), "Ensemble of Trees" (EnT), and "Support Vector Regression" (SuVR). Trials on 3 benchmark and 8 real-world farming datasets revealed that the developed hybrid feature selection and extraction technique performed with significant improvements with respect to Rsq^2, RtMSE, and "mean absolute error" (MAE) in comparison to FS and FeExt methods such as Correlation Analysis (CA), Singular Valued Decomposition (SiVD), Genetic Algorithm (GA), and wgt-PCA on "benchmark" and "real-world" farming datasets.

1. Introduction

Data Pre-processing is an essential phase that needs to be performed before applying any ML techniques. It is a phase that transforms the original raw data into a format which can be useful to the machine learning algorithms. Data pre-processing consists of various steps like data cleaning, feature selection, feature extraction to name a few [1]. The agricultural datasets usually comprise of many features that may not be useful to the prediction task. In such a scenario, feature selection and feature extraction form important tasks that need to be performed to improve the algorithmic accuracy [2].

FS is a Data pre-processing technique where manual or automatic choice of appropriate attributes is performed. It is the determination of relevant attributes that enhance the accuracy of a model [3]. These features contribute to the variable to be predicted. Basically, it removes features which are irrelevant or may not impact the feature to be predicted. It forms one of the fundamental concepts in ML which impacts the model's performance. Feature selection is very important since irrelevant or unnecessary attributes will negatively influence the performance of the ML model. If data has many irrelevant features, the accuracy of ML model decreases. Advantages of using feature selection are highlighted below [4]:

- Overfitting is reduced: When irrelevant features are removed, the noise will be removed, and the model will be able to perform better on the test data also.
- Reduces training time: When the number of dimensions in the data is reduced, the time taken by the ML model will reduce and it becomes computationally faster.
- Improves accuracy of model: The accuracy of the model increases when the irrelevant features are eliminated from the dataset.

Several feature selection techniques exist. Some of the popular techniques are:

* Corresponding author.
E-mail addresses: adityashastry.k@nmit.ac.in (K. Aditya Shastry), sanjay.ha@nmit.ac.in (Sanjay H.A.).

https://doi.org/10.1016/j.knosys.2021.107460
0950-7051/

- Filter Methods: Here, the feature selection happens before the ML algorithm is applied. Correlation is one of the popular examples of this technique [5].
- Wrapper Methods: Here subset of features is used, and model is trained on them. "forward selection", "backward feature elimination" and "recursive feature elimination" are three popular examples of this technique.
  - In "forward selection", initially the model has no features. In each loop, new features are added that best enhances the model's performance.
  - In "backward elimination", initially the model starts with all attributes, and it eliminates the feature which is less important at every loop that enhances the model's performance. This continues until no performance improvement is observed [6].
  - "Recursive Feature elimination" finds the optimal subset of attributes by using "greedy based approach". Models are constructed repetitively by choosing best or the worst attributes at each loop. The next model is created using the attributes which are left over. This process continues until all the attributes are tried and tested. The attributes are then ranked based on the order of their exclusion [7].
- "Embedded Methods": These techniques integrate both the "wrapper" and "filter" approaches. They are employed by techniques that possess their own built-in FS methods [8].

Agriculture data contain diverse features related to climate, soil, fertilizer. In such a context, FS attains significant importance as irrelevant features may adversely impact the prediction of the model built [3]. In our work, we have used the filter methods for FS since the other two techniques are computationally expensive [7]. Specifically, a modified genetic algorithm (m-GA) was developed for FS by designing a new fitness function based on MutInf, and RtMSE,

FeExt is a technique of decreasing the number of attributes from the original data set where new features are constructed from the prevailing attributes. The data in original attributes are present in these new attributes minus the duplicity [9]. It is useful since it reduces the number of features in the raw data without loss of information. It reduces the amount of duplicate or redundant information that may exist in the raw data. FeExt can speed up the ML algorithm and its steps of generalization. "Principal Component Analysis" (PrCA) represents one of the well-known methods in FeExt. It transforms the original data from a lower to a higher dimension so that the task of ML can be performed effectively. In this work, wgt-PCA was employed which is an improved version of the original PrCA. wgt-PCA uses the weighted covariance matrix which gives emphasis on training records that are very near to the test record and diminishes the impact of other training records [10].

By combining feature selection and feature extraction techniques the benefits of both the techniques may be reaped. While FS can remove irrelevant features from the original data and FeExt can remove features without loss of information [11].

In view of this, a combination of FS and FeExt approach for agricultural data prediction is presented in this effort. The designed technique consists of 3 stages. Pre-processing of information forms the 1st stage in which prediction of missing values and the normalization of information is performed. Missing values are predicted for real world data using "mean imputation" method. Subsequently, the normalization of the information between 0 and 1 is performed to decrease the dominance of features which are high valued over the small-valued features during the prediction task. In the feature selection phase, we designed a modified genetic algorithm (m-GA) with an improved fitness function consisting of weights, MutInf and RtMSE. During the FeExt phase, wgt-PCA is employed on the attributes obtained from the GA. That is, we are performing feature selection first using m-GA and then FeExt using wgt-PCA. In stage-3, prediction of agricultural information was done. For the "real-world" farming data, the information for manures and crop were acquired from "IndiaStats" [12]. Information associated with rainfall was taken from "IndiaStats" [12] and "India Water Portal" [13]. The data for soil was acquired from the "Karnataka Atlas of Soil Fertility" [14]. The information about manures, yield of crops, rainfall, and soil were merged to create separate data files for each of the Indian crops namely "wheat, maize, bajra, jowar, cotton, groundnut, sugarcane and ragi". The standard datasets used were "Forest Fires" (FF) [15], "weather Ankara" (wAnk) [16] and "weather Izmir" (wIzm) [17].

Key contributions of this effort are as follows:

- Designed a modified Genetic Algorithm (m-GA) with an improved "fitness function" centred on mutual information and Root Mean Square Error.
- Designed a hybrid approach (m-GA+wgt-PCA) by integrating Feature Selection (m-GA) and Feature Extraction (wgt-PCA) techniques.
- Compared the designed hybrid approach (m-GA+wgt-PCA) with other feature selection and extraction methods such as Correlation Analysis, Singular Valued Decomposition, Genetic Algorithm, and weighted-Principal Component Analysis on 3 benchmark and 8 real-world agricultural datasets with respect to multiple performance metrics such as R-squared, root mean square error, and Mean Absolute Error.

The contents of this paper are outlined below. Segment 2 explains the work related to the hybrid feature selection and extraction methods. Segment 3 presents the designed approach. The results of the trials are demonstrated in Segment 4. The paper ends with conclusion and future work.

2. Related work

This segment presents feature selection and feature extraction techniques used in the field of agriculture and other domains. Following paragraphs present the works on FS and FeExt techniques in the domain of agriculture (with a focus on "crop yield prediction"):

In [18], authors used random search, best search, and genetic methods for determining the relevant features for land classification. Results revealed that electrical conductivity, exchangeable sodium percentage, soil texture and wetness were the most effective parameters for land classification. In [19], researchers employed several algorithms for feature selection. The selected features were fed to the Multiple Linear Regression, ArtNN, and M5Prime algorithms. Results exhibited that the MLR algorithm performed better on features selected using the Sequential Forward FS. Khaki et al. [20] performed feature selection using Deep Neural Network (DNN). They used the embedded approach of feature selection in which the ML algorithm itself performs the feature selection. Here, the authors utilized the DNN for feature selection and for crop yield prediction. Dhivya et al. [21] developed a hybrid feature extraction technique. The framework determined optimal features for crop yield prediction ML model. They employed decision tree, random forest, and gradient boosting as the ML models for crop yield prediction. However, FS, and FeExt techniques were not integrated. In [22], authors extracted important features for wheat yield from satellite data. They used convolutional neural network for extracting significant features. Authors were able to establish the interconnection between the

vegetation satellite data and the meteorological conditions. Comparison of the "Convolutional Neural Network" (CNN) with other FeExt or FS techniques was not done. Experiments were limited to one specific dataset (wheat). Klompenburg et al. [23], extensively reviewed the different ML methods applied for crop yield prediction. They observed that temperature, rainfall, and soil type were among the most used parameters for crop yield prediction. ArtNN was the most applied ML algorithm. Authors surveyed different Deep Learning techniques for crop yield prediction. They observed that CNN, Long Term Short Memory (LSTM), and DNN were the popularly used Deep Learning (DL) methods for crop yield prediction. They concluded that DL techniques could be employed for FeExt but did not propose any new or hybrid FS or FeExt technique. In [24], authors developed a crop recommendation system by employing FeExt and ML techniques. Authors employed PrCA for extracting features from the crop dataset. These features were then fed to the ML models such as random forest, logistic regression, and ArtNN. They observed that ArtNN performed better on the PrCA extracted features. However, the authors did not compare PrCA with other FeExt techniques. They did not propose any new FS or FeExt technique. Sharma et al. [25] utilized CNN for feature extraction from image data. The extracted features were fed to the LSTM for crop yield prediction.

Following works present the works on FS and FeExt techniques in other domains such as malware classification, text classification, etc.

In [26], malware classification was done by first selecting features using TF-IDF algorithm. The reduced feature set was transformed into principal components using PrCA and kernel PrCA. Land use classification was performed in [27] using hybrid FS method which combined filter and wrapper methods. Filter approach made use of the Relief F algorithm for selecting relevant features affecting the utilization of land. Particle Swarm Optimization was employed by the wrapper technique for searching features and the classification accuracy of Support Vector Machine was utilized to select the optimal features. In [28], a hybrid feature selection scheme, comprising of filter and wrapper selection stages, was proposed for text classification problems. In the first stage, features were selected using filter methods. Next, the features selected by the filters were combined and fed into genetic algorithm. But feature selection was not combined with feature extraction. In [29], authors present a feature selection technique by combining filter methods. Subsequently, exhaustive search was utilized to obtain optimal attributes. Although, the authors tested hybrid feature combination methods they did not attempts to integrate FS and FeExt techniques. In [30], authors proposed a new feature extraction method by combining PrCA with GA. Principal components were encoded as chromosomes of GA and later improved by fitness function. As per our survey, no work has been done to combine GA for feature selection and wgt-PCA for feature extraction.

Following points summarize certain drawbacks of the above works and benefits of our proposed work:

- As per our survey, very less work was done in the domain of combining feature selection and feature extraction techniques on agricultural data. In this work, the FS and FeExt methods were combined to acquire the benefits of both these techniques.
- Most of the previous works have employed existing FS or FeExt techniques. In this work, a modified "fitness function" was devised for the GA algorithm and the FS technique has been integrated with FeExt technique (wgt-PCA).
- Most of the works have not compared their methods with other FS or FeExt methods. In this work, the designed technique has been extensively matched with other techniques using different metrics.
- Many works have concentrated on specific crop dataset. In this work, experiments have been performed on multiple crop datasets to show the effectiveness of the proposed technique.

3. Proposed work

This segment presents the approach and algorithm of the devised hybrid FS and FeExt strategy for agricultural prediction. Fig. 1 depicts the flow of the devised technique comprising of 3 stages.

The designed approach constituted of 3 stages. In the 1st stage, pre-processing of data was performed by predicting the missing values and normalizing the information. The missing values that were in the "real-world" agricultural datasets were predicted using the "mean imputation" technique. The agricultural information was subjected to normalization in the range of 0 to 1 for reducing the dominance of features with higher values over features with lower values. Prediction of missing values was performed since they lead to inaccurate predictions as the data will not be complete [31,32].

Modified GA (m-GA) was employed in stage-2 for selecting relevant features. The GA was chosen as the base algorithm for FS since it can perform search in an exhaustive manner [33]. The m-GA was developed by designing a fitness function based on MutInf and RtMSE to choose the optimal feature subset that significantly affected the yield. The wgt-PCA was then employed to extract features from these selected features. The principal components (PC) were generated by wgt-PCA from the m-GA selected features. The wgt-PCA was selected over the PrCA for FeExt as it can achieve higher accuracy [32]. By combining GA and wgt-PCA, we were able to obtain the benefits of both feature selection and feature extraction. m-GA as feature selector eliminated the noisy and irrelevant data. wgt-PCA as feature extractor caused no loss of information as the extracted features represented combination of original features. During stage 3, the extracted features (principal components) were provided as input to the different ML models namely EnT, SuVR, ANFIS, ArtNN.

3.1. "Data pre-processing" (stage-1)

Prior to doing any ML task such as agricultural prediction there is a necessity to transform the original raw data into an appropriate structure so that the prediction techniques perform with higher accuracy. This technique is known as "data pre-processing" [34]. In this effort, the "data pre-processing" comprised of 2 actions. The 1st action included the prediction of the values that were absent by utilizing the "mean-imputation" technique. The normalization of information in the range of 0 to 1 was performed.

3.1.1. Predicting the values that were absent

Missing data values make the data incomplete which causes the ML models to predict inaccurately. In this paper, the 8 "real-world datasets" [12–14] contained "missing values". These values were predicted using the "mean imputation" technique in which the omitted values were replaced by the mean/average of the known values of that feature [35]. "Mean imputation" technique was chosen since its performance was found to be better than other techniques like "complete case analysis" and "multiple imputation" [35].

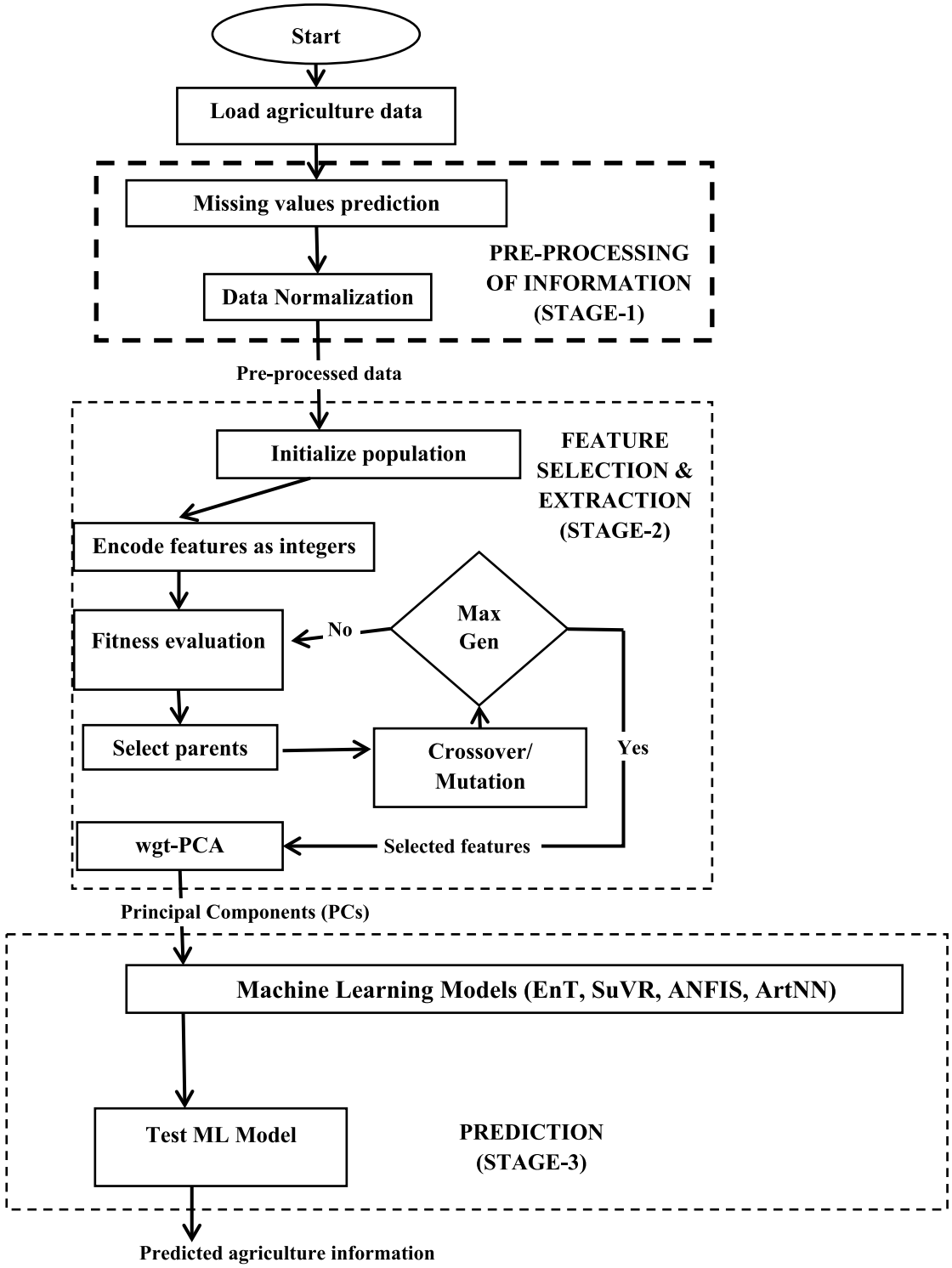

Fig. 1. Proposed Hybrid Feature Selection and Extraction Strategy.

3.1.2. Standardization of information

The original farming data was pre-processed between values 0 and 1. This disallows the higher valued features from dominating the lower valued features during prediction. Eq. (1) shows how value (t_i) is normalized for the feature G in row '*i*' [36].

$$Normalized\,(t_i) = \frac{t_i - G_{min}}{G_{max} - G_{min}} \quad (1)$$

Where, G_{min} and G_{max} are the minimum and maximum values of feature *G*, respectively.

3.2. Feature selection and extraction (stage 2)

After standardizing the farming information, the hybrid strategy of FS and FeExt to pick out the relevant attributes from this standardized data was applied. FS is a method for removing irrelevant and identical traits [37]. It is primarily done to enhance the prediction "accuracy" of ML models. In feature extraction all the original data features are utilized to create new features [38] that denote the combinations of the original features. Thus, in

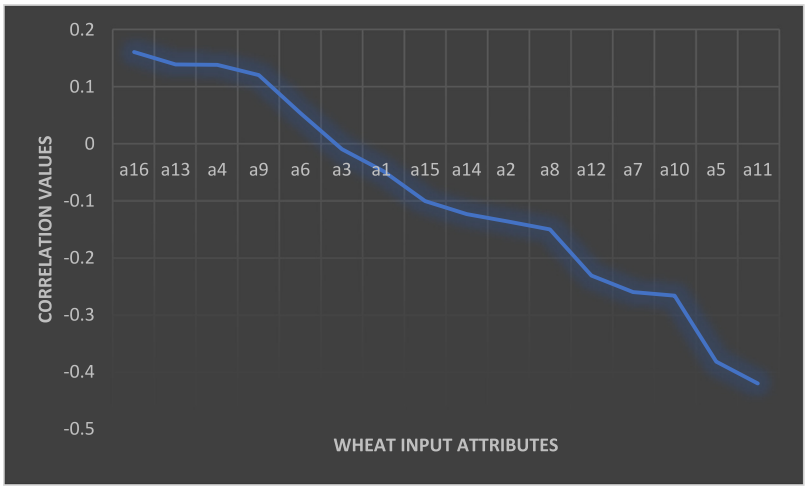

Fig. 2. Correlation between Wheat input features with the Wheat yield.

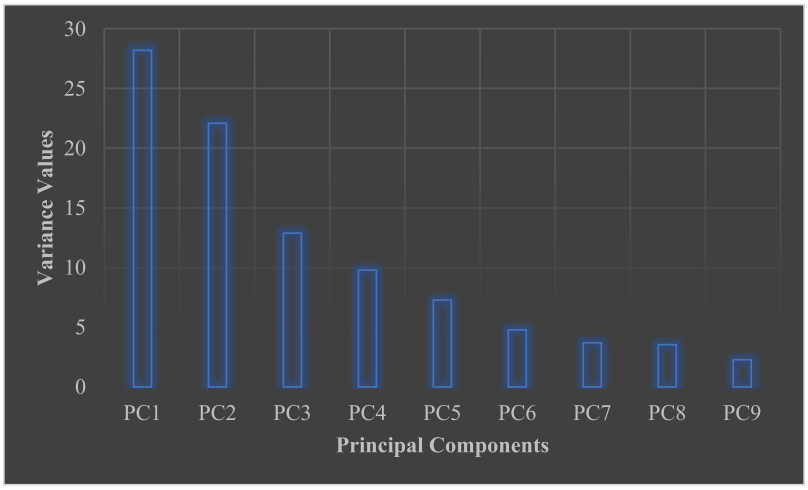

Fig. 3. wgt-PCA variance values on the m-GA selected features for the Wheat Data.

feature selection loss of information may occur, while in feature extraction the extracted features may contain all the useful information present in the original data. Retaining these points in mind, we propose a novel hybrid method for selecting features by combining feature selector (m-GA) with feature extractor (wgt-PCA). In stage-2 of the proposed work, the features are first selected using m-GA. Then, the selected features are fed to wgt-PCA to perform feature extraction.

The combination of m-GA and wgt-PCA is done for the following reasons:

- Firstly, the feature selection method (m-GA) gives the essential features which affect the crop yield (target).
- These features are then subjected to feature extraction method using wgt-PCA. This is done to capture the relevant information in the selected features.
- The combined benefits of FS and FeExt methods are achieved. Noisy and irrelevant data will be removed by feature selection. The benefits of feature extraction are no loss of information since the extracted features represent combinations of the original features.
- The prediction accuracy of the ML model improves due to this combination.

The devised modified genetic algorithm (m-GA) for feature selection is explained in Section 3.2.3. The m-GA was compared with Correlation Analysis and Singular Valued Decomposition (SiVD) described in segments 3.2.1 and 3.2.2 correspondingly.

3.2.1. "Correlation Analysis" (CA)

CA is the procedure of finding how tightly the attributes are associated to one another by utilizing "correlation coefficient" ($Corr - Coeff$). The $Corr - Coeff$ provides the measure of the linear dependency amongst 2 features. It varies from "−1 to 1". The "positive correlation" is shown by "1", whilst "0" implies "no correlation". Features possessing "negative values" imply that the "correlation" between them is negative. In this effort, the $Corr - Coeff$ was calculated by employing the "covariance" (Cov) technique since it is appropriate for features that are scalar in nature [39]. The $Corr$ - $Coeff$ with respect to Cov of A and B features is presented in Eqs. (2) and (3) [39]:

$$Corr - Coeff = \frac{Cov\,(A, B)}{\sigma_A \sigma_B} \tag{2}$$

$$Cov\,(A, B) = (A - \mu_A)\,(B - \mu_B) \tag{3}$$

Here,

- μ_A, μ_B refer to the average of A and B.
- σ_A, σ_B signify the "standard deviations" of A and B.

In this paper, the FS using "correlation" was performed by picking out the input traits having "high correlation" with the dependent feature ("crop yield").

3.2.2. "Singular Valued Decomposition" (SiVD)

In SiVD, a "rectangular matrix" is made to undergo factorization. For a universal rectangular c× d "matrix" L, its SiVD is L = PQR^T where P is a c × f "matrix", R^T is the transpose of order f × d and Q is a diagonal matrix f × f. The diagonal components of 'Q' denote "singular values" like "$\delta_1 \geq \delta_2 \geq ... \geq \delta_r > 0$" where "r" signifies "matrix rank" of L. With regards to decomposition, SiVD utilizes the blend of columns and rows of L in a direct manner [40]. In this paper, SiVD was employed as the FeExt technique using Eq. (4):

$$L \approx L_k = P_{c\times k} Q_{k\times k} R^T_{k\times d} \tag{4}$$

Where, "k" is smaller than the rank "r". The "singular values property $\delta_1 \geq \delta_2 \geq ... \geq \delta_r > 0$" ensures that the first "k values" are better than the discarded ones. This work considered the attributes having high "singular values".

3.2.3. Genetic Algorithm (GA) based on Mutual Information (MutInf) and Root Mean Square Error (RtMSE) for feature selection

GA is generally utilized in optimization & search problems. It is normally comprised of 4 stages. During the first stage, preliminary "population" of chromosomes is generated. Each "chromosome" is coded as a sequence of values by employing "binary" or real valued coding. 2nd stage forms the selection stage, in which a part of the "population" is chosen by utilizing a problem related fitness "function". The 3rd stage encompasses the formation of 2nd "generation" of "population" by applying chromosomal operators ("crossover" & "mutation"). The GA process iterates until maximum number of generations are reached or fitness reaches a satisfactory level [41,42].

In this work, we propose a modified fitness "function" for GA to achieve FS. The creation of the preliminary population was performed randomly. Subsequently, each "chromosome "in the "population" was produced using the fitness "function" in Eq. (5). The proposed fitness "function" comprises of 2 components: one related to the MutInf between the dependent & independent features while the other component was related to the RtMSE value given in Eq. (5):

$$Ftns(e) = c * MutInf(e,f) + d * (1/RtMSE) \tag{5}$$

Where:

- *Ftns(e) → "fitness" of the attribute 'e',*
- *RtMSE → "Root Mean Square Error",*
- *MutInf(e, (f) → "Mutual Information" between independent attribute 'e' and the dependent attribute 'f'*
- *c=0.2 and d=0.7 → weights*

'c' and 'd' values were selected by trial-and-error method. Their values were varied in the range of 0.1 to 1.0. "c=d=0.1" proved exceedingly small, whilst "c=d=0.5" allocated identical weights to MutInf and RtMSE components and "c=d=0.9" proved very high. Consequently, modest values of "0.2" & "0.7" were chosen for the research.

'c' was allocated a smaller value than "d"' to assign RtMSE with a higher significance than MutInf in Eq. (5). This was done since improving the "accuracy" of the prediction technique was more important than the MutInf between 'e' and 'f'.

The first element of Eq. (5) includes MutInf(e,f) as defined in Eqs. (6), (7), and (8).

$$MutInf(e,f) = Ent(e) + Ent(y) - Jnt - Ent(e,f); \tag{6}$$

$$Ent(e) = -\sum pr(e) log\ pr(e) \tag{7}$$

Table 1
GA parameters.

GA Parameter	Value
"Population Size"	"100"
"Chromosome size"	"9"
"Population Category"	Floating point
"Fitness Function"	MutInf and RtMSE with weights 'c' and 'd'
"Number of Generations"	"80"
"Crossover"	"Arithmetic Crossover"
"Crossover Likelihood"	"0.8"
"Mutation"	"Uniform Mutation"
"Mutation Likelihood"	"0.1"
"Selection Scheme"	"Rank based selection"
EliteCount	"1" (i.e., 1 chromosome was chosen)

$$Jnt - Ent(e,f) = -\sum pr(e,f) log\ pr(e|f) \tag{8}$$

Here, computation of MutInf between input attributes ('e') and the target attribute ('f') was performed and not among the input attributes because these selected features would be subjected to wgt-PCA which would remove any correlation between the input attributes 'e'. Here, more significance was given to the correlation between input features 'e' and target feature 'f' and not among the input features.

The likely ambiguity in 'e' was measured by entropy (Entr) in Eq. (7) [43]. Ambiguity in 'e' and 'f' variables was assessed by Joint-Entropy (Jnt-Ent) in Eq. (8). The relationship between 2 random features is calculated by MutInf. MutInf among 2 features is 0 in case of statistically autonomous features [43]. Consequently, large MutInf among 2 random features implies that they share high MutInf. So, Entr must be low and MutInf must be high. The modified GA for feature selection is described in Algorithm-3:

Algorithm-3 Genetic Algorithm for feature selection

```
1: procedure FS-GA (Feat_Num,TrningInput, Trget)
2:     k←0
3:     AF(k) ← Init-Population (AF(k)) //Initialize population of all features
4:     F(k) ← Compute-Fitness (AF(k)) //Select initial features
5:     while (k <= Max-Gen) //loop over generations
6:         k<-k+1
7:         FS(k)<-crsover (FS(k-1))
8:         FS(k)<-mut (FS(k))
9:         F(k)<-Comp-Fitns (FS(k))
10:    end while
11:    FS (k)<-Rank-Features(F(k))
12:    return FS (k)
```

"Algorithm-3" operates by considering the number of attributes to be selected (*Feat_Num*), "training input" (*TrningInput*) & the "training class" (*Trget*) as the parameters in line 1. "Population" involving the entire attribute set was initialized at line 3. At line 4, our modified fitness function given in Eq. (5) was employed for computing the fitness of the features. At line 5, the "while" loop iterates until the maximum number of generations (*Max-Gen*) is reached. At lines 7 and 8, genetic operations i.e., "crossover" (*crsover*) and "mutation" (*mut*) were performed on the attributes. At line 9, the fitness of the new attributes was computed. Finally, at line 11, the features were ranked based on their fitness values. For the "real-world" farming datasets the *Feat_num* was changed from "4" to "15" since the total independent attributes were "16". For the standard farming datasets, the *Feat_num* was changed from "4" to "9" since total input traits were "10".

Table 1 reveals the "parameters" of m-GA employed in this research.

Table 2 below gives the example for computation of fitness based on the proposed fitness equation (5):

Table 2
Fitness Computation example-1.

Chromosome	MutInf(e, f)	RtMSE	Fitness
1	5	0.75	1.93
2	9	0.85	2.62

Table 3
Fitness Computation example-2.

Chromosome	MutInf(e, f)	RtMSE	Fitness
5	4	0.39	4.2
6	7	0.39	5.7

Here, chromosome-1 is chosen as it has the highest fitness value, among others. As can be noted, though the chromosome-2 has a very high value of MutInf, it is not selected as its RtMSE is high. This is due to the fact that we have given more prominence to RtMSE than MutInf. Table 3 provides another illustration of fitness computation using the proposed fitness equation (5).

It can be inferred by observing Table 3 that although "chromosomes" 5 & 6 have equal RtMSE values, the "chromosome-6" was picked because of its higher fitness value. The higher fitness for "chromosome-6" was obtained since its MutInf was higher than "chromosome-5". This is the result of including the MutInf component in our fitness function along with RtMSE. Hence, the proposed fitness "function" for FS took care of both MutInf & "accuracy/error".

The 2nd component of Eq. (1) includes RtMSE [44] which is the difference between the actual and predicted values as presented in Eq. (9):

$$\mathbf{RtMSE} = \frac{\mathbf{1}}{\mathbf{Nu}} \sum_{\mathbf{i=1}}^{\mathbf{Nu}} (\mathbf{Act_i} - \mathbf{Pred_i})^{\mathbf{2}} \tag{9}$$

Where:

- *Nu* – Number of instances
- Act_i- Real output
- $Pred_i$-Predicted output

In this effort, the RtMSE value of each chromosome from ArtNN was taken as its fitness value for m-GA. i.e., the attributes possessing higher MutInfo & lower RtMSE values were chosen. The "chromosome" possessing the greatest value in Eq. (5) would be taken as the fittest and would be sent to the subsequent generation.

3.2.4. "Weighted-Principal Component Analysis" (wgt-PCA)

"Correlated" features are transformed into linearly un-associated features known as "principal components" by PrCA which employs orthogonal conversion [45]. The wgt-PCA was employed for transforming the attributes chosen by m-GA onto the "Principal Component" (PC) space in which the components are not linked to each other [10]. In PrCA, the chosen m-GA features minus the dependent attribute containing "d-dimensions" are mapped into a "low" dimensional space [46]. The mapping signifies the "invertible" linear conversion (information rotation). The altered information is more appropriate for ArtNN than the raw information being fed to ArtNN, as the information interpretation impacts the prediction ability of ArtNN more than the category of learning rule. Algorithm-4 describes the actions for extracting attributes by employing wgt-PCA.

***Algorithm-4* w-PCA for Feature Extraction**

```
Procedure w-PCA(TrngInput)
    1.numDim ← SetNoDim();
    2.Compute_Weighted_Mean();
    3.Compute_Weighted_Covariance();
    4.[wcoeff, ~, latent,~,explained] =
      pca(input,'VariableWeights','variance');
    5.coeff = inv(diag(std(input))) * wcoeff;
    6.reducedDimension = coeff(:1: numDim);
    7.input = input * reducedDimension;
    8.[COEFF, SCORE] = pca(input);
    9.input=SCORE (:1: numDim);
end procedure
```

In Algorithm-4, TrngInput is the information comprising of "d" input features. For "real-world" information, TrngInput comprised of "15" attributes minus the dependent attribute. The "benchmark" information comprised of "9" features. Utilizing wgt-PCA, "9" PCs were extracted out of "15" for the "real-world" dataset, & "5" PCs were extracted out of "9" attributes for the "benchmark" dataset. This was since these PCs described maximum variance. At line 2, the "*Compute_Weighted_Mean ()*" module determined the average of each dimension of the TrngInput. "*Compute_Weighted_Covariance ()*" module at line-3 computed the "weighted covariance matrix" of the entire dataset by considering the inverted attribute "variances" as "weights". Later, the "eigenvectors" (e1, e2, …, ed) and corresponding eigenvalues (λ1, λ2..., λd) were estimated. The examples were converted into different "subspace" by employing the "eigenvector matrix" [46]. This is demonstrated by Eq. (10):

$$y = W^T x \tag{10}$$

Here,

- "x" → "d×1"-dimensional vector indicating 1 record,
- "y" → converted "$k \times 1$"-dimensional record in the "new space".

The wgt-PCA was utilized to obtain attributes from m-GA selected features as it removes the correlated data. The related attributes are combined to produce improved features by wgt-PCA. Moreover, many "real-world" datasets that were utilized to train ArtNN possessed correlated data where input records overlapped with each other [47]. ArtNN suffers from poor "generalization" ability if redundant information is given as input [47].

4. Experimental setup and findings

This segment discusses the trial setup used for conducting the trials along with the results obtained from feature selection methods. "Matlab R2016a" was employed as the coding language on "windows 7 environment".

In this work, experiments were conducted on Indian crop datasets, and benchmark datasets. On each data set we compared our hybrid feature selection method (m-GA+wgt-PCA) with other feature selection methods (m-GA, GA and Correlation), and Feature Extraction methods (SiVD, wgt-PCA) with respect to Rsq^2, RtMSE, and MAE on different prediction models (EnT, SuVR, ANFIS, and ArtNN). The details of the experimental results are described in the subsequent sections. In all the cases, the datasets were split into 80% as training set and remaining 20% as test set.

4.1. Trial findings on Indian crop datasets

For the "real-world" data sets the crops "wheat, maize, bajra, jowar, cotton, groundnut, sugarcane and ragi" from [12–14] were used for experimentation. The attribute to be predicted was the "crop yield" in "tonnes/hectare". Table 4 describes the attributes of these datasets.

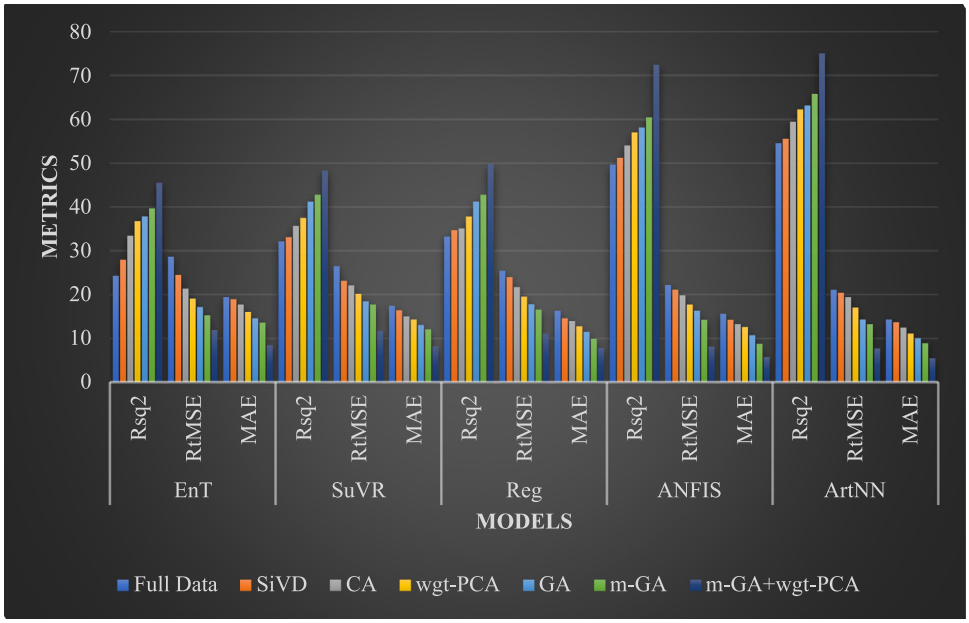

Fig. 4. Comparison of proposed hybrid feature selection method with other methods on Indian Wheat Dataset.

Table 4
Attribute description of real-world datasets.

Attributes	Description
a1	"Nitrogen" content in manure (Nit)
a2	"Phosphorus" content in manure (Phos)
a3	"Potassium" employed in manure (Pot)
a4	Mean "Rainfall" in regions (Rain)
a5	Mean "temperature" in areas (Temp)
a6	"Precipitation" in regions (ppt)
a7	"Soil pH" value (Pow-H)
a8	"Electrical conductivity" of "soil" in "dS/m" (Ele-Cond)
a9	Volume of "Organic carbon" present in "soil" in "percentage" (Org-C)
a10	Average "phosphorus" amount in "soil" in "ppm" (Aver-P)
a11	Average "potassium" present in "soil" in "ppm" (Aver-K)
a12	Average "sulphur" existing in "soil" in "ppm" (Aver-S)
a13	Average "zinc" prevailing in "soil" in "ppm" (Aver-Zn)
a14	Average "boron" prevalent in "soil" in "ppm" (Aver-B)
a15	"Area" in "hectares" (Ar)
a16	"Production" in "tonnes" (Pr)

In the case of wheat data, the number of records were 149 with 16 features. The target attribute was the wheat yield. The dataset was split into training set with 119 records (80%) and test set with 30 records (20%). The values that were absent were estimated by utilizing the "mean imputation" technique. The correlation between the input features (a1 to a16) and the Wheat yield (a17) is shown in Fig. 2.

As can be inferred from Fig. 2, positive correlations were found between a16(Pr), a13(Aver-Zn), a4(Rain), a9(Org-C), a6(ppt) with the wheat yield (a17). Negative correlations were observed between a3(Pot), a1 (Nit), a15(Ar), a14(Aver-B), a2 (Phos), a8(Ele-Cond), a12(Aver-S), a7(pow-H), a10(Aver-P), a5(Temp), and a11(Pot) with the Wheat Yield (a17).

For the proposed method (m-GA+wgt-PCA), the m-GA selected 14 features (a6, a1, a16, a15, a3, a5, a7, a4, a11, a12, a9, a8, a14) which were subjected to feature extraction by wgt-PCA which in turn generated 9 Principal components shown in Fig. 3.

From Fig. 3 it can be observed that PC1 to PC9 together captured 94.61% of the variances in the data.

Fig. 4 compares the proposed hybrid feature selection strategy (m-GA+wgt-PCA) with other feature selection methods (SiVD, CA, wgt-PCA, GA, and m-GA) on the Indian wheat dataset.

The performance evaluation was done by applying the features selected by the feature selection methods (SiVD, CA, wgt-PCA, GA, m-GA, m-GA + wgt-PCA) on the predictive models (EnT, SiVR, Regr, ANFIS, ArtNN) using the performance metrics Rsq^2, RtMSE, and MAE. The proposed hybrid feature selection method (m-GA +wgt-PCA) performed better than the original data with 16 input features, correlation with 9 features (a16, a13, a4, a9, a6, a3, a1, a15, a14), SiVD with 9 features (a3, a1, a5, a8, a7, a2, a10, a15, a9), GA with 11 features (a16, a1, a12, a6, a3, a15, a10, a4, a2, a7, a5), and wgt-PCA with 9 extracted features. The performance was measured in terms of Rsq^2, RtMSE, and MAE. In case of the proposed feature selection method, the m-GA selected 14 features (a6, a1, a16, a15, a3, a5, a7, a4, a11, a12, a9, a8, a14) which were subjected to feature extraction by wgt-PCA which in turn extracted 9 features. These features were then fed to the ML models and the performance of the feature selection methods were evaluated. The proposed hybrid feature selection strategy performed with Rsq^2 improvements ranging from 5.85% to 21.22% on the EnT model, 7.05% to 16.16% on the SiVR model, 8.48% to 16.48% on the regression model, 14.32% to 22.78% on the ANFIS model, 11.91% to 20.48% on the ArtNN model with respect to the Rsq^2 statistic when compared to the original data, SiVD, Correlation, wgt-PCA, GA, and m-GA feature selection methods.

Figs. 5 to 11 shows the correlation between the input attributes of the remaining Indian crop datasets with their respective yield as given by the correlation technique.

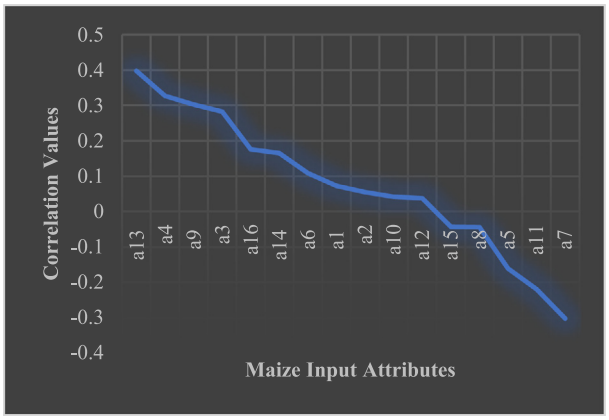

Fig. 5. Correlation between Maize input features with its yield.

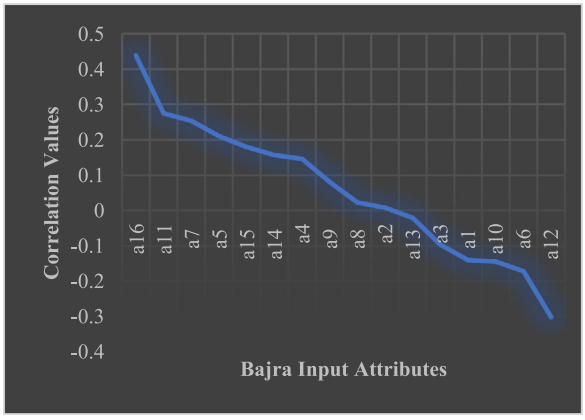

Fig. 6. Correlation between Bajra input features with its yield.

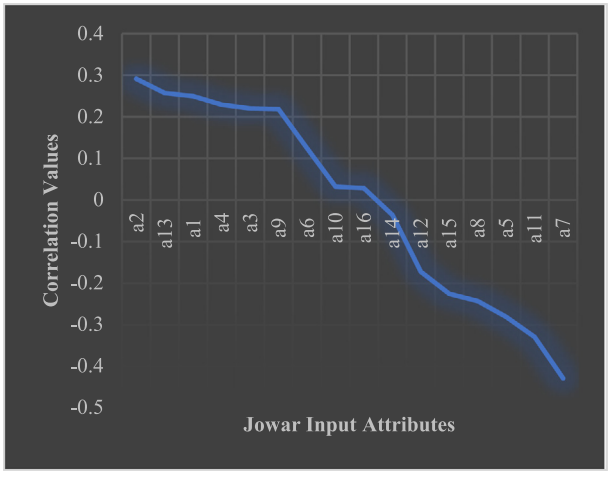

Fig. 7. Correlation between Jowar input features with its yield.

Following are the observations from Figs. 5 to 11.

- For the Indian maize dataset, positive correlation was found between a13, a4, a9, a3, a16, a14, a6, a1, a2, a10, a12 (in that order from the highest to lowest correlation) and the maize yield. Negative correlation was found between a15, a8, a5, a11, and a7 (in that order from the highest to lowest correlation).

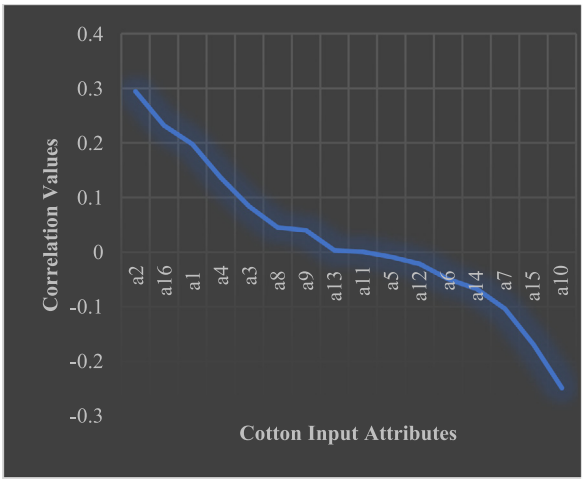

Fig. 8. Correlation between Jowar input features with its yield.

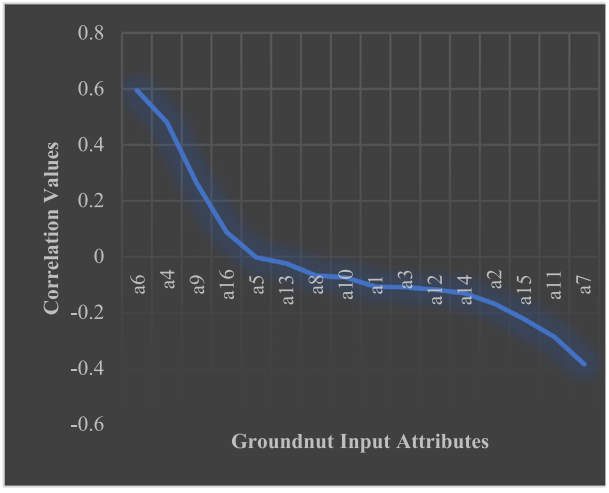

Fig. 9. Correlation between Groundnut input features with its yield.

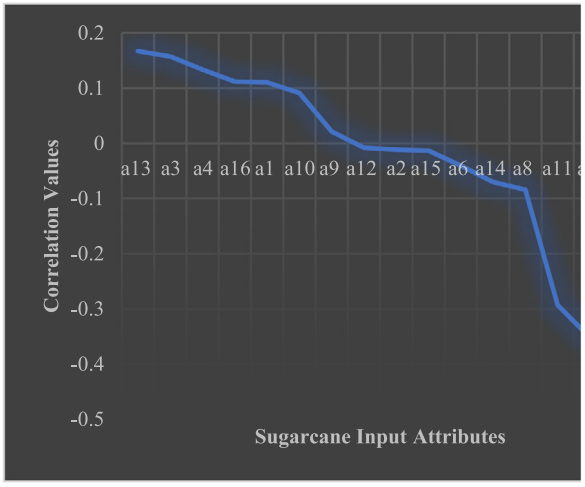

Fig. 10. Correlation between Sugarcane input features with its yield.

- For the Indian Bajra dataset, positive correlation was found between a16, a11, a7, a5, a15, a14, a4, a9, a8, a2 (in that order from the highest to lowest correlation) and the Bajra

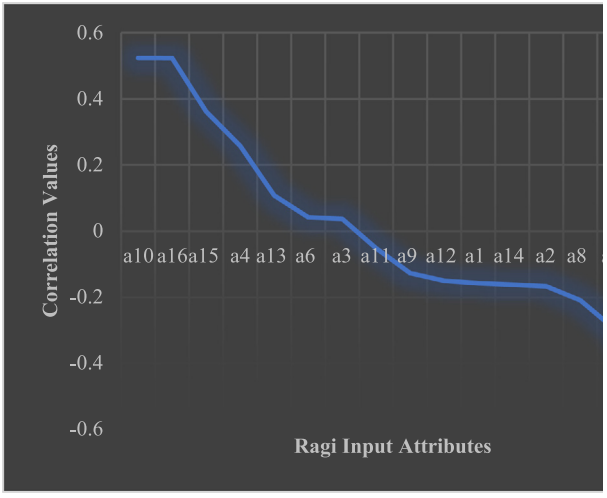

Fig. 11. Correlation between Ragi input features and its yield.

yield. Negative correlation was found between a13, a3, a1, a10, a6, and a12 (in that order from the highest to lowest correlation).

- With regards to the Indian Jowar dataset, a2, a13, a1, a4, a3, a9, a6, a10, and a16 exhibited positive correlation (from highest to lowest in that order) with the jowar yield (a17). Negative correlation was observed between a14, a12, a15, a8, a5, a11, a7 and the jowar yield.
- In case of the Indian Cotton dataset, positive correlation was demonstrated between a2, a16, a1, a4, a3, a8, a9, a13, a11 with the cotton yield. However, negative correlation was exhibited between a11, a5, a12, a6, a14, a7, a15, a10 and the cotton yield.
- For the Indian Groundnut dataset, positive correlation was observed between a6, a4, a9, a16, and the groundnut yield. For the other input attributes negative correlation with the yield was observed.
- With respect to the Indian Sugarcane dataset, a13, a3, a4, a16, a1, a10, a9 showed positive correlation with the yield of sugarcane while the other attributes exhibited negative correlation.
- Finally, for the Indian Ragi dataset, positive correlation between a10, a16, a15, a4, a13, a6, a3 and the Ragi yield was observed. The remaining input attributes exhibited negative correlation with the Ragi yield.

Figs. 12 to 18 illustrate the variance captured by the wgt-PCA when applied on the other Indian crop dataset for feature extraction on the features selected by m-GA.

Following are the observations from Figs. 12 to 18:

- For Indian Maize data set, m-GA selected 9 features (a2, a3, a4, a5, a6, a12, a13, a15, a16). The wgt-PCA extracted 7 PCs from these selected features. Together PC1 to PC7 captured 92.3% variance as shown in Fig. 12.
- For the Indian Bajra data set, m-GA selected 12 features (a1, a2, a3, a4, a5, a6, a7, a12, a13, a14, a15, a16). The wgt-PCA extracted 9 PCs from these selected features capturing a variance of 92.3% as depicted in Fig. 13.
- With regards to the Indian Jowar dataset, m-GA selected 13 features (a14, a11, a2, a9, a12, a5, a3, a4, a15, a10, a16, a13). The wgt-PCA extracted 10 PCs from these chosen features capturing a variance of 94.84% as demonstrated in Fig. 14.
- With respect to the Indian cotton dataset, m-GA selected 9 features (a12, a7, a13, a3, a16, a1, a11, a6, a4) out of which wgt-PCA extracted 8 PCs capturing a variance of 92.88% as illustrated in Fig. 15.
- For the Indian Groundnut dataset, m-GA selected 8 features (a15, a14, a8, a7, a1, a12, a5, a3) out of which wgt-PCA extracted 6 PCs capturing a variance of 84.21% as shown in Fig. 16.
- For the Indian Sugarcane dataset, m-GA selected 9 features (a9, a2, a14, a3, a8, a12, a15, a16, a5) out of which wgt-PCA extracted 8 PCs that together captured a variance of 90.41% as shown in Fig. 17.

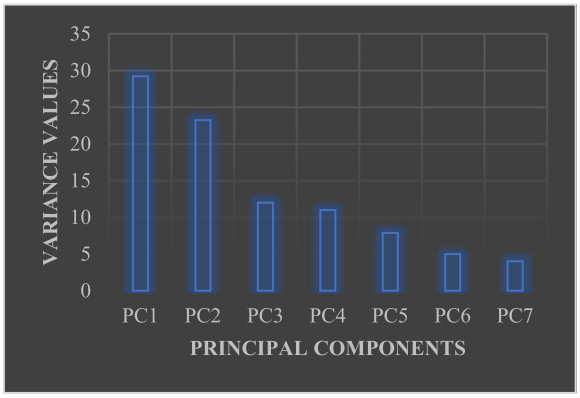

Fig. 12. wgt-PCA variance values on the m-GA selected features for the Maize Data.

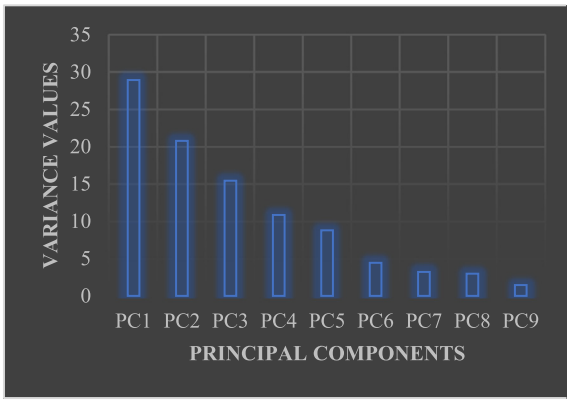

Fig. 13. wgt-PCA variance values on the m-GA selected features for the Bajra data.

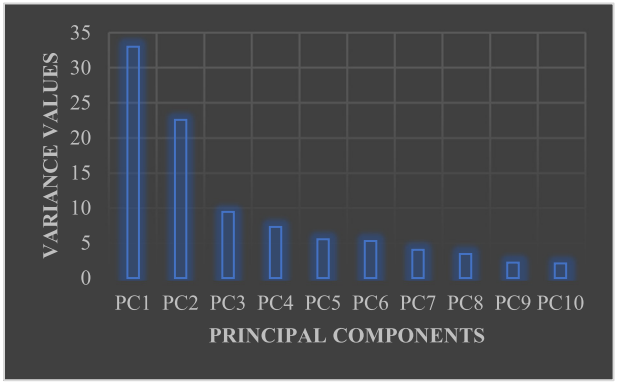

Fig. 14. wgt-PCA variance values on the m-GA chosen features for the Jowar dataset.

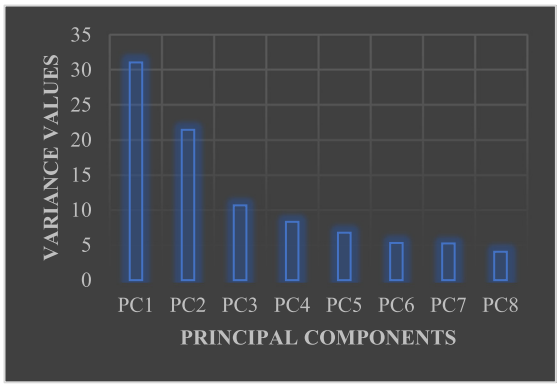

Fig. 15. wgt-PCA variance values on the m-GA selected features for the Cotton dataset.

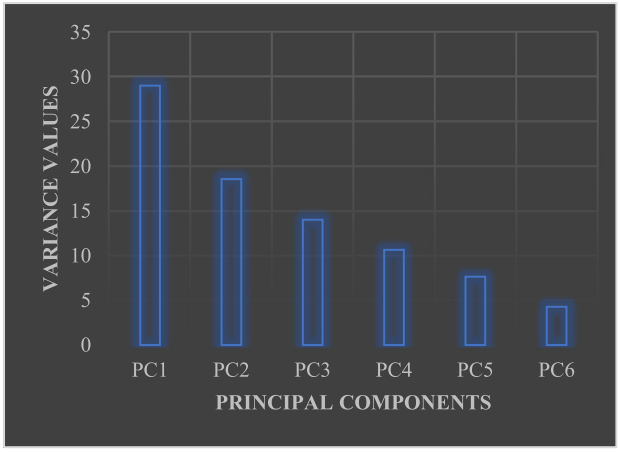

Fig. 16. wgt-PCA variance values on the m-GA selected features for the Groundnut dataset.

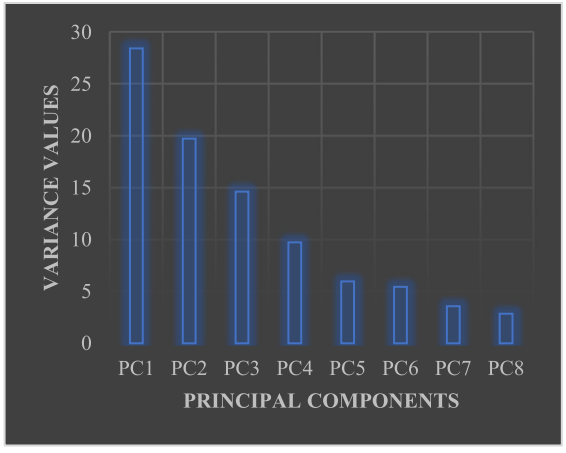

Fig. 17. wgt-PCA variance values on the m-GA selected features for the Sugarcane dataset.

- For the Indian Ragi dataset, m-GA selected 11 features (a13, a1, a4, a7, a10, a3, a9, a15, a12, a8, a6) out of which wgt-PCA extracted 9 PCs that together captured a variance of 93.74% as depicted in Fig. 18.

Figs. 19 to 25 show the consolidated experimental results on the India Crop datasets "Maize, Bajra, Jowar, Cotton, Groundnut, Sugarcane, and Ragi".

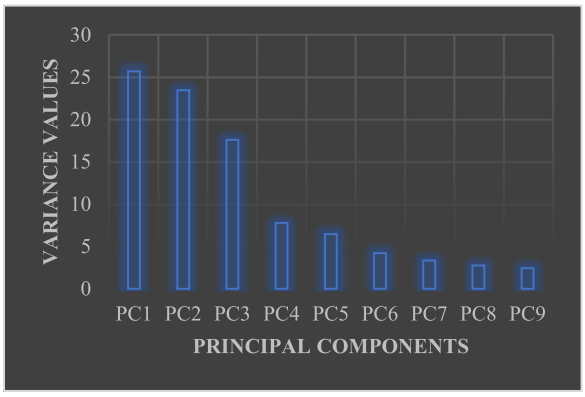

Fig. 18. wgt-PCA variance values on the m-GA selected features for the Ragi dataset.

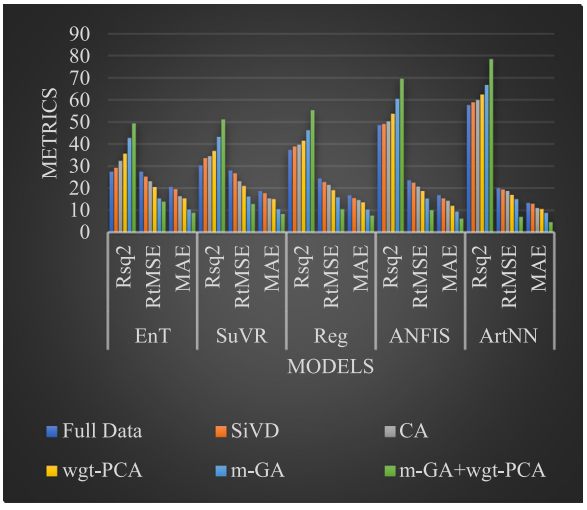

Fig. 19. Comparison on Maize Dataset.

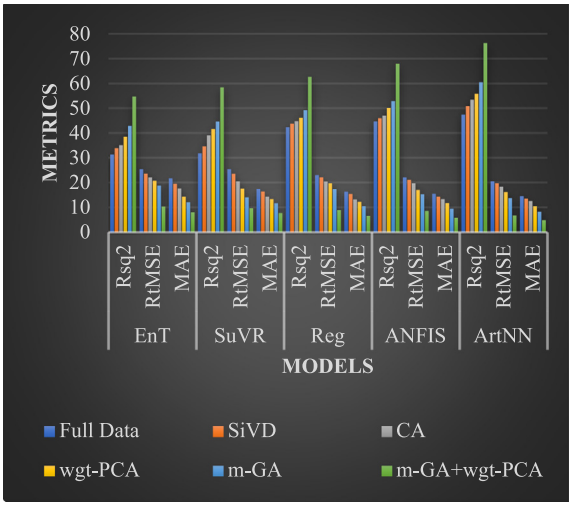

Fig. 20. Comparison on Bajra Dataset.

As explained previously, the features selected by the different feature selection methods were fed to the ML models and their performance was evaluated with our proposed method. Fig. 26 presents the "performance" improvements of the proposed method in comparison to the other techniques.

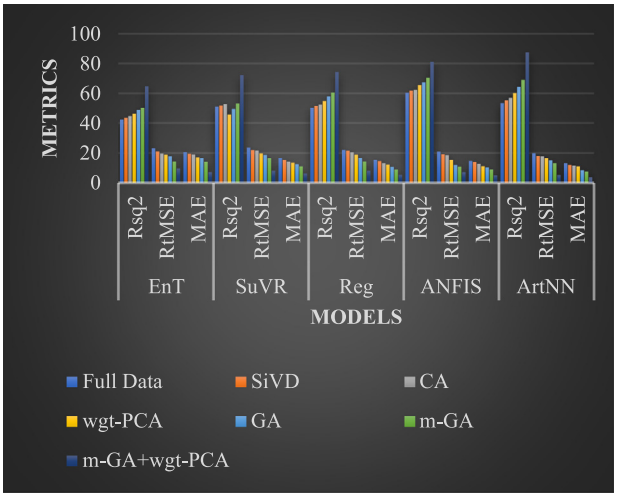

Fig. 21. Comparison on Jowar Dataset.

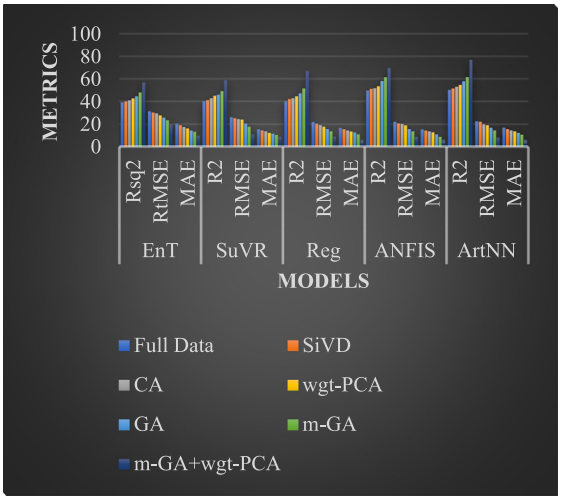

Fig. 22. Comparison on Cotton Dataset.

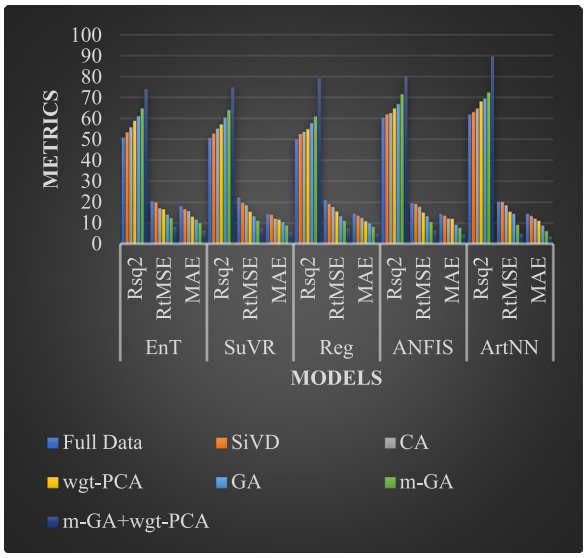

Fig. 23. Comparison on Groundnut Dataset.

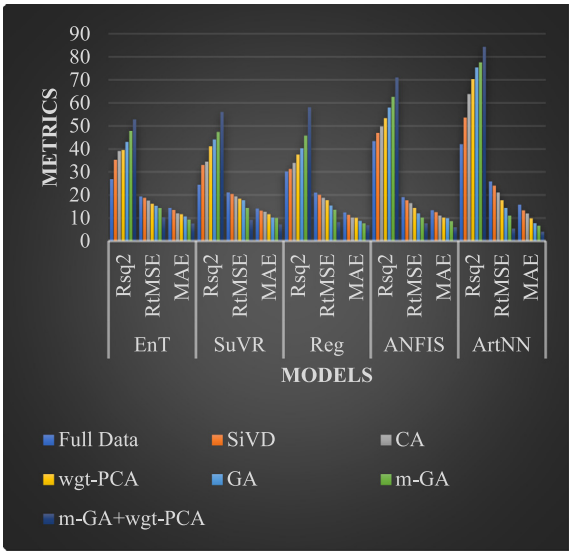

Fig. 24. Comparison on Sugarcane Dataset.

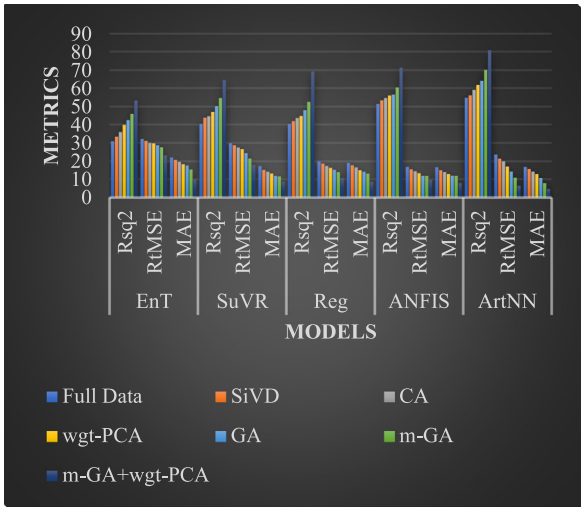

Fig. 25. Comparison on Ragi Dataset.

As shown in Fig. 26, significant improvements of the proposed method were observed on the Indian crop datasets over other feature selection methods. The performance improvements ranged from 2.39 to 5.7% on the maize dataset, 3.88% to 8.53% on Bajra data, 8.28% to 13.04% on Jowar data, 7.34% to 8.94% on Cotton data, 7.68% to 9.28% on Groundnut data, 7.42% to 16.21% on Sugarcane dataset, and 8.95% to 11.14% on ragi dataset.

4.2. Experimental results on Benchmark Datasets

This section describes the experiments conducted on the Benchmark Datasets viz. Forest Fires (FF) [15], Weather Ankara (wAnk) [16], and Weather Izmir (wIzm) [17].

4.2.1. Experimental results on Forest Fire Dataset

With regards to the FF Dataset, the purpose was to predict the burnt region of "forest fires", in the "northeast region of Portugal", by utilizing atmospheric & other associated information. There were 517 records, with 12 input features. Dataset was divided into training set which comprised of 414 records and

Table 5
Attribute description of FF dataset.

Attributes	Description
X	"x-axis spatial coordinate within the Montesinho park map: 1 to 9"
Y	"y-axis spatial coordinate within the Montesinho park map: 2 to 9"
Month	"Month of the year: 'jan' to 'dec'"
Day	"Day of the week: 'mon' to 'sun'"
FFMC	"FFMC index from the FWI system: 18.7 to 96.20"
DMC	"DMC index from the FWI system: 1.1 to 291.3"
DC	"DC index from the FWI system: 7.9 to 860.6"
ISI	"ISI index from the FWI system: 0.0 to 56.10"
Temp	"Temperature in Celsius degrees: 2.2 to 33.30"
RH	"Relative humidity in %: 15.0 to 100"
Wind	"Wind speed in km/h: 0.40 to 9.40"
Rain	"Outside rain in mm/m^2: 0.0 to 6.4"
Burned Area	"The burned area of the forest (in ha): 0.00 to 1090.84"

test set constituting of 103 records. Table 5 shows the attribute description for the FF dataset.

Missing values were predicted using the mean imputation method. The correlation between the input features ("X, Y, Month, Day, FFMC, DMC, DC, ISI, Temp, RH, Wind, and Rain") and the Burned-Area is shown in Fig. 27.

As can be inferred from Fig. 27, positive correlations were found Temp, DMC, X, Month, DC, Y, FFMC, Day, ISI, Wind with the target attribute (Burned-Area) in that order. Negative correlations were observed between Rain, RH with the target attribute (Burned-Area) in that order.

Best performance was obtained when ML models were run on 8 PCs extracted from 9 m-GA selected features (Temp, Y, Wind, X, RH, Day, Month, FFMC, Rain). Fig. 28 shows the variance values of each of the principal components PC1 to PC8. Together they captured 90.16% of the total variance.

From Fig. 28 it can be observed that PC1 to PC8 together captured more than 90% of the variance in the data. In the proposed method when m-GA selected features were subjected to wgt-PCA, the PC1 to PC8 together explained 90.16% of the data variance and performed better than the other methods on most of the datasets.

Fig. 29 compares the proposed hybrid feature selection strategy (m-GA+wgt-PCA) with other feature selection methods (SiVD, Correlation, wgt-PCA, GA, and m-GA) on the FF dataset.

The performance evaluation was done by applying the features selected by the feature selection methods (SiVD, Correlation, wgt-PCA, GA, m-GA, m-GA + wgt-PCA) on the predictive models (EnT, SuVR, Regression, ANFIS, ArtNN) using the performance metrics Rsq^2, RtMSE, and MAE. The proposed method with 9 m-GA selected features (Temp, Y, Wind, X, RH, Day, Month, FFMC, Rain) and 8 wgt-PCA extracted features performed the best with performance improvements ranging from 15.33% to 27.56% over the original data with 12 features, 13.92% to 26.66% over SiVD with 9 features (Rain, Month, DC, FFMC, Temp, DMC, RH, Y, Wind), 13.43% to 25.88% over Correlation with 9 features (Temp, DMC, X, Month, DC, Y, FFMC, Day, ISI), 12.52% to 24.34% over wgt-PCA with 9 extracted features, 9.1% to 21.88% over GA with 9 features (Temp, Y Wind, X, RH, Day, Month, FFMC, Rain), and 6.65% to 19.22% over m-GA with 8 selected features (Temp, Y Wind, X, RH, Day, Month, FFMC).

4.2.2. Experimental results on Weather Datasets

The wAnk [16] and wIzm [17] files were taken for experimentation as the other benchmark datasets. Entire attributes were "real-valued" in nature. The target variable to be predicted was the "mean temperature". The weather Ankara contained the climate information of "Ankara" from "01/01/1994 to 28/05/1998". The aim was to predict the average temperature. The training information constituted of 1287 records while the test data comprised of 322 records. The weather Izmir dataset contained the weather data of "Izmir" from "01/01/1994 to 31/12/1997". The purpose was to predict the average temperature. The total records of 1461 were divided into 1168 as training records and the remaining 293 records as test data. In both cases, 80%–20% rule was followed for the train test data split. The attribute descriptions of both the weather datasets are shown in Table 6.

Table 6
Attribute description of wAnk and wIzm datasets.

Features	Description
MaxT	"Maximum temperature"
MinT	"Minimum temperature"
Dew	"Dewpoint"
Ppt	"Precipitation"
SLP	"Sea_level_pressure"
SP	"Standard_pressure"
V	"Visibility"
WS	"Wind speed"
MWS	"Maximum wind speed"

Fig. 30 shows the correlation graph for the weather datasets of wAnk and wIzm.

In Fig. 10, AirT-WA and AirT-Wiz denote the Air Temperature (target attribute) of wAnk and wIzm datasets, respectively. For the wAnk dataset, positive correlation between MaxT, MinT, Dew, V, MWS, WS with AirT-WA (target attribute) was found, while negative correlation was observed between PPT, SP, SLP with AirT-WA. For the wIzm dataset, positive correlation was obtained between MaxT, MinT, Dew, V, WS, with the AirT-Wiz, while negative correlation was observed between Ppt, SLP, MWS with the target attribute (AirT-Wiz).

Superior performance on the wAnk dataset was obtained on 5 PCs extracted by wgt-PCA on 8 m-GA selected features (WS, Dew, SLP, MinT, V, MaxT, MWS, Ppt). Together PC1 to PC5 captured a variance of 93.44%. In case of wIzm dataset the best results were obtained on 5 PCs extracted by wgt-PCA on 8 m-GA selected features (WS, Dew, SLP, MinT, SP, MaxT, MWS, Ppt). Together PC1 to PC5 captured 93.15% variance. Fig. 31 depicts the wgt-PCA variance values captured by PC1 to PC5 for wAnk and wIzm datasets after applying m-GA.

As depicted in Fig. 31, the principal components (PC1 to PC5) together captured more than 90% variance of the original data on both the wAnk and wIzm datasets.

Fig. 32 compares the proposed hybrid feature selection strategy (m-GA+wgt-PCA) with other feature selection methods (SiVD, Correlation, wgt-PCA, GA, and m-GA) on the wAnk dataset.

The performance evaluation was done by applying the features selected by the feature selection methods (SiVD, Correlation, wgt-PCA, GA, m-GA, m-GA + wgt-PCA) on the predictive models (EnT, SuVR, Regression, ANFIS, ArtNN) using the performance metrics

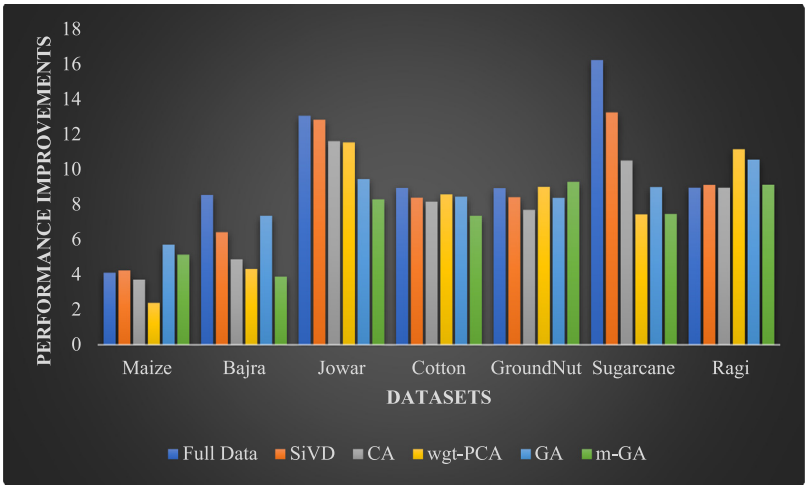

Fig. 26. Performance improvements of the proposed feature selection technique with other feature selection methods including original data (Full Data).

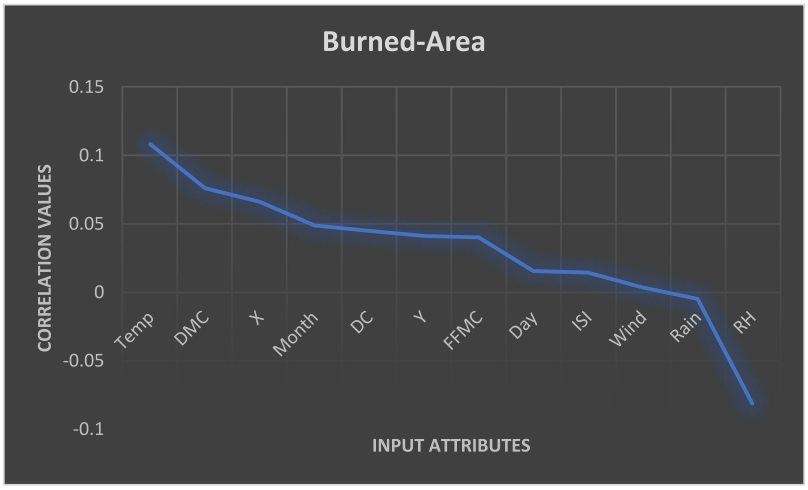

Fig. 27. Correlation between Forest Fire input features with the Target attribute.

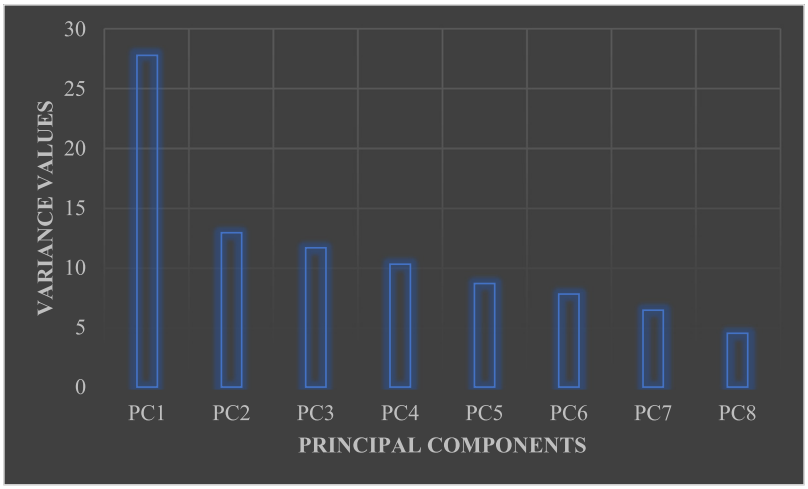

Fig. 28. wgt-PCA variance values on the m-GA selected features of FF Dataset.

Rsq^2, RtMSE, and MAE. The proposed method with 8 m-GA selected features (WS, Dew, SLP, MinT, V, MaxT, MWS, Ppt), and 5 wgt-PCA extracted features performed superior with improvements ranging from 16.8% to 21.21% over the original data, 23.36% to 28.06% over SiVD with 7 extracted features, 21.64% to 24.56% over Corr with 8 selected features, 16.8% to 20.33% over wgt-PCA with 8 extracted features, 12.28% to 18.64% over GA with 8 selected features, and 9.51% to 14.56% over m-GA with 7 selected features.

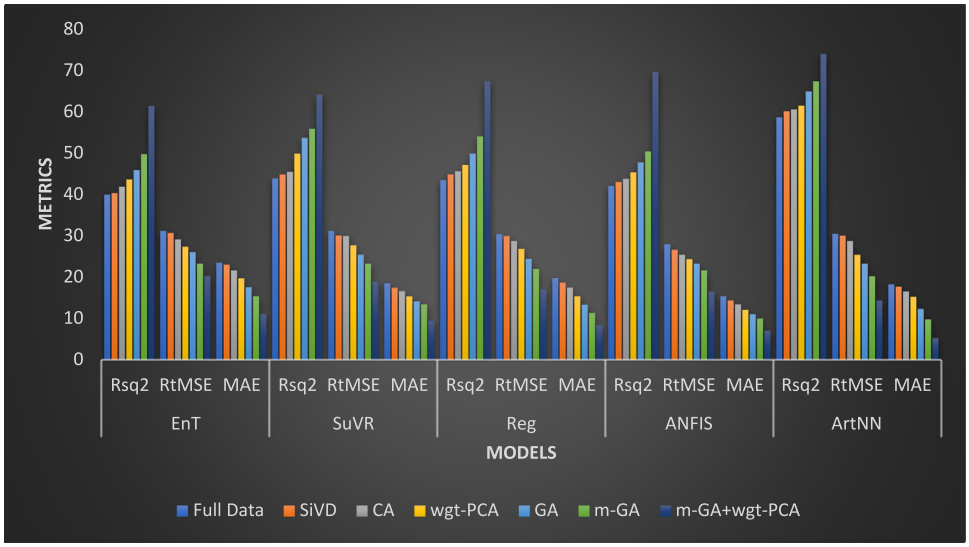

Fig. 29. Comparison of proposed hybrid feature selection method with other methods on Forest Fire Dataset.

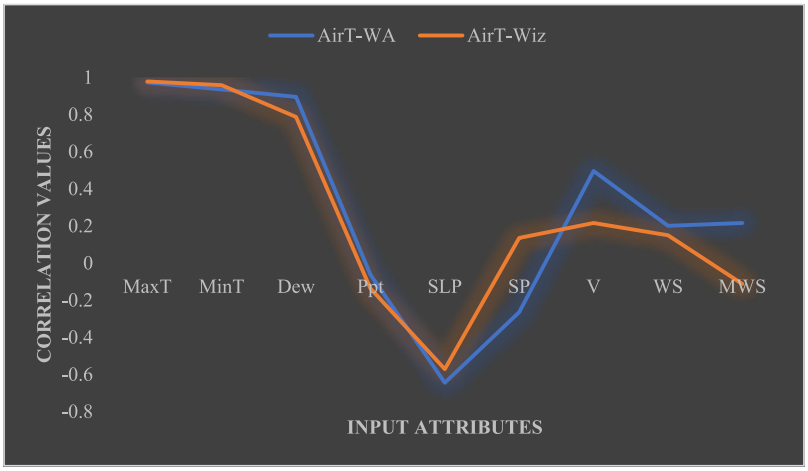

Fig. 30. Correlation between input features of wAnk and wIzm with the Target attribute (Air Temperature).

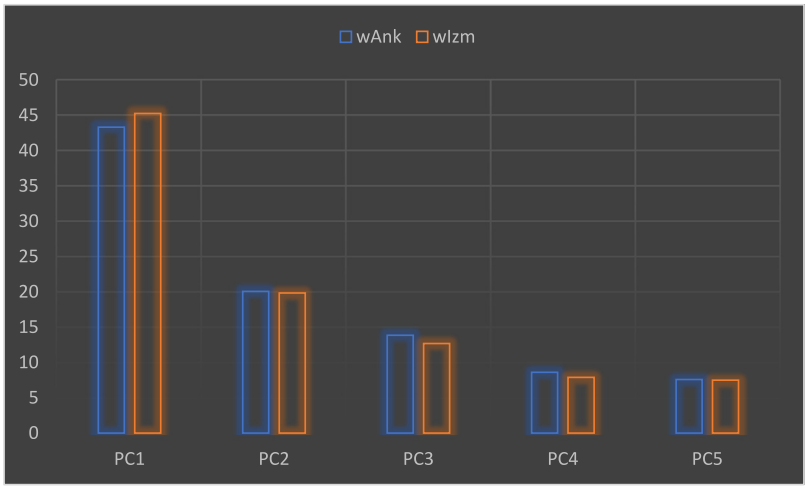

Fig. 31. wgt-PCA variance values on the m-GA selected features of wAnk and wIzm Datasets.

Fig. 33 compares the proposed hybrid feature selection strategy (m-GA+wgt-PCA) with other feature selection methods (SiVD, Correlation, wgt-PCA, GA, and m-GA) on the wIzm dataset.

The performance evaluation was done by applying the features selected by the feature selection methods (SiVD, Correlation, wgt-PCA, GA, m-GA, m-GA+wgt-PCA) on the predictive models (EnT,

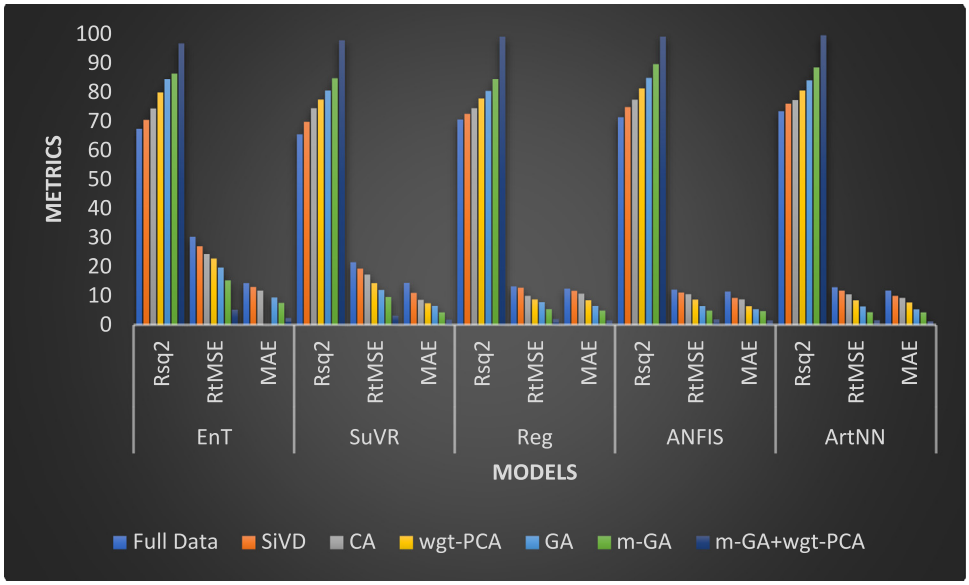

Fig. 32. Comparison of proposed hybrid feature selection method with other methods on Weather Ankara Dataset.

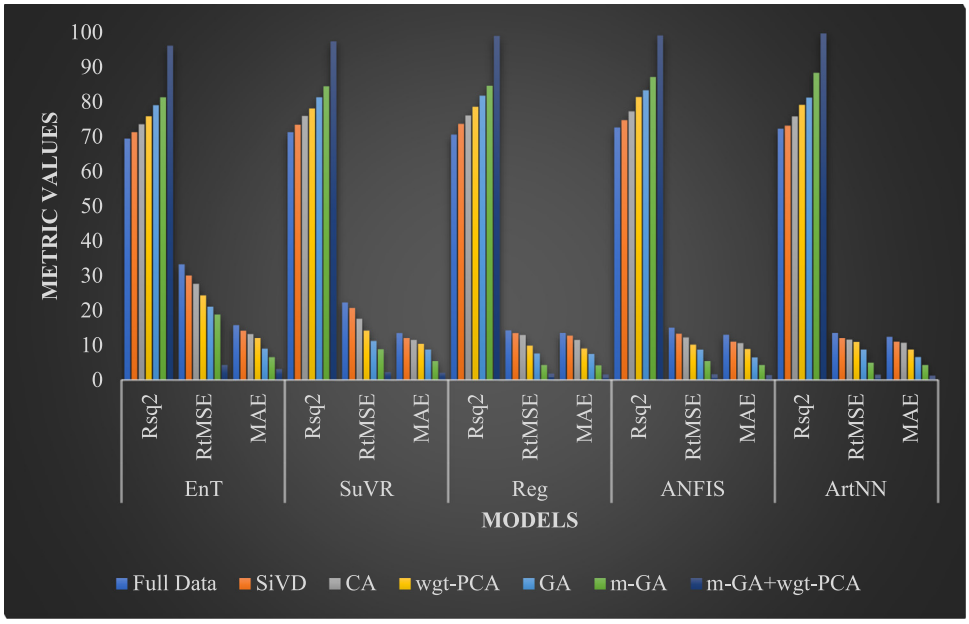

Fig. 33. Comparison of proposed hybrid feature selection method with other methods on Weather Izmir Dataset.

SiVR, Regression, ANFIS, ArtNN) using the performance metrics Rsq^2, RtMSE, and MAE. The proposed method with 8 m-GA selected features (WS, Dew, SLP, MinT, SP, MaxT, MWS, Ppt), and 5 wgt-PCA extracted features performed superior with improvements ranging from 26% to 28% over the original data, 23.92% to 26.54% over SiVD with 7 extracted features, 21.45% to 23.93% over Correlation with 8 selected features, 17.57% to 20.32% over wgt-PCA with 5 extracted features, 15.69% to 18.55% over GA with 8 selected features, and 11.36% to 14.69% over m-GA with 8 selected features.

5. Discussion

In this section, justification of why the proposed hybrid feature selection method performed better than the other feature selection methods are discussed.

- The proposed method performed better than the original data (data comprising of all the features) since the original data consisted of features that did not contribute to the target attribute. Removing the unimportant attributes from the original data made the predictive models perform with higher accuracy. For instance, with regards to Indian wheat data m-GA selected 14 features (a6, a1, a16, a15, a3, a5, a7, a4, a11, a12, a9, a8, a14) which were subjected to feature extraction by wgt-PCA which in turn extracted 9 features. This extraction happened during stage-2 of the proposed work.
- The proposed method performed better than SiVD since SiVD suffers on non-linear data. The agricultural data was non-linear in nature. SiVD also may discard useful information since it focuses more on data variance [48].
- Devised technique performed better than correlation since correlation does not determine what features impact the

target variable. It can induce bias in the analysis as an external feature may impact the correlation [49].

- The primary reason for the superior performance of the hybrid feature selection technique was that it combined the benefits of both FS and FeExt methods. The wgt-PCA was able to obtain important traits from the records without information loss and the m-GA was able to remove the noisy and irrelevant information [50].

6. Conclusion and future scope

In this work, a hybrid feature selection technique was developed by combining feature selector m-GA with the feature extractor wgt-PCA. The m-GA was developed by designing a fitness function based on MutInf, and RtMSE to select the best features which significantly affected the yield of crops. The selected features were then subjected to feature extraction using wgt-PCA. By integrating m-GA and wgt-PCA the strengths of FS and FeExt were achieved. Extensive experiments were conducted on 8 "real-world" farming information & 2 "benchmark" farming data sets. Significant performance improvements of the proposed method over the original data (containing all the features), SiVD, Correlation, wgt-PCA, GA were observed. Superior results of the designed technique can be attributed to various facts. First and foremost, features were chosen by utilizing the blend of FS (m-GA) and FeExt (wgt-PCA) techniques. This combined the advantages of both methods. The feature selection method selected the most important features that affected the target ("crop yield" & "temperature") while the feature extraction technique (wgt-PCA) extracted a linear combination of the selected features without information loss. This resulted in a well-defined input for the prediction model. In future, the methods may be tested on data sets other than agriculture. Also, various combinations of feature selectors and extractors may be tested. Satellite images can be considered for extracting important features.

CRediT authorship contribution statement

K. Aditya Shastry: Conceptualization, Methodology, Software, Data curation, Writing – original draft, Visualization, Investigation, Writing – review & editing. **Sanjay H.A.:** Conceptualization, Supervision, Validation, Reviewing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Z. Guan, T. Ji, X. Qian, Y. Ma, X. Hong, A survey on big data preprocessing, in: 2017 5th Intl Conf on Applied Computing and Information Technology/4th Intl Conf on Computational Science/Intelligence and Applied Informatics/2nd Intl Conf on Big Data, Cloud Computing, Data Science, ACIT-CSII-BCD, 2017, pp. 241–247, http://dx.doi.org/10.1109/ACIT-CSII-BCD.2017.49.

[2] C. Bauckhage, K. Kersting, Data mining and pattern recognition in agriculture, Künstl Intell. 27 (2013) 313–324, https://doi.org/10.1007/s13218-013-0273-0.

[3] S. Visalakshi, V. Radha, A literature review of feature selection techniques and applications: Review of feature selection in data mining, in: 2014 IEEE International Conference on Computational Intelligence and Computing Research, 2014, pp. 1–6, http://dx.doi.org/10.1109/ICCIC.2014.7238499.

[4] L.J. Herrera, V. Lafuente, R. Ghinea, M.M. Perez, I. Negueruela, H. Pomares, I. Rojas, A. Guillén, Mutual information-based feature selection in spectrometric data for agriculture applications, in: Proceedings of the International Multiconference of Engineers and Computer Scientists, IMECS, Mar 18–20, 2015, Vol I Hong Kong.

[5] M. Cherrington, F. Thabtah, J. Lu, Q. Xu, Feature selection: Filter methods performance challenges, in: 2019 International Conference on Computer and Information Sciences, ICCIS, 2019, pp. 1–4, http://dx.doi.org/10.1109/ICCISci.2019.8716478.

[6] N. El Aboudi, L. Benhlima, Review on wrapper feature selection approaches, in: 2016 International Conference on Engineering & MIS, ICEMIS, 2016, pp. 1–5, http://dx.doi.org/10.1109/ICEMIS.2016.7745366.

[7] Girish Chandrashekar, Ferat Sahin, A survey on feature selection methods, Comput. Electr. Eng. 40 (1) (2014) 16–28, https://doi.org/10.1016/j.compeleceng.2013.11.024.

[8] H. Liu, Feature selection, in: C. Sammut, G.I. Webb (Eds.), Encyclopedia of Machine Learning, Springer, Boston, MA, 2011, https://doi.org/10.1007/978-0-387-30164-8_306.

[9] I. Guyon, A. Elisseeff, An introduction to feature extraction, in: I. Guyon, M. Nikravesh, S. Gunn, L.A. Zadeh (Eds.), Feature Extraction, in: Studies in Fuzziness and Soft Computing, vol. 207, Springer, Berlin, Heidelberg, 2006.

[10] Z. Fan, E. Liu, B. Xu, Weighted principal component analysis, in: H. Deng, D. Miao, J. Lei, F.L. Wang (Eds.), Artificial Intelligence and Computational Intelligence. AICI 2011, in: Lecture Notes in Computer Science, vol. 7004, Springer, Berlin, Heidelberg, 2011.

[11] Z. Shang, M. Li, Combined feature extraction and selection in texture analysis, in: 2016 9th International Symposium on Computational Intelligence and Design, ISCID, 2016, pp. 398–401, http://dx.doi.org/10.1109/ISCID.2016.1098.

[12] India Stats. Available at https://www.indiastat.com/data/agriculture. Accessed online in January 2016.

[13] India Water Portal. Available at http://www.indiawaterportal.org/articles/meteorological-datasets-download-entire-datasets-various-meteorological-indicators-1901. Accessed online in January 2016.

[14] Suhas P. Wani, Kanwar Lal Sahrawat, Sarvesh KV, Baburao Mudbi, Krishnappa K (Eds.), Soil Fertility Atlas for Karnataka, India, International Crops Research Institute for the Semi-Arid Tropics, Patancheru 502 324, Andhra Pradesh, India, 2011.

[15] Forest Fires Dataset. Available at: http://archive.ics.uci.edu/ml/datasets/Forest+Fires. Accessed online in March 2016.

[16] Weather Ankara Dataset. Available at: http://sci2s.ugr.es/keel/dataset.php?cod=41. Accessed online in April 2016.

[17] Weather Izmir data set. Available at: http://sci2s.ugr.es/keel/dataset.php?cod=78. Accessed online in April 2016.

[18] S. Hamzeh, M. Mokarram, A. Haratian, H. Bartholomeus, A. Ligtenberg, A.K. Bregt, Feature selection as a time and cost-saving approach for land suitability classification (case study of shavur plain, Iran), Agriculture 6 (2016) 52, https://doi.org/10.3390/agriculture6040052.

[19] S. Maya, R. Bhargavi, Selection of important features for optimizing crop yield prediction, Int. J. Agricult. Environ. Inf. Syst. 10 (2019) 54–71, http://dx.doi.org/10.4018/IJAEIS.2019070104.

[20] Khaki Saeed, Wang Lizhi, Crop yield prediction using deep neural networks, Front. Plant Sci. 10 (2019) 621, http://dx.doi.org/10.3389/fpls.2019.00621.

[21] D. Elavarasan, P.M.D.R. Vincent, K. Srinivasan, C.-Y. Chang, A hybrid CFS filter and RF-RFE wrapper-based feature extraction for enhanced agricultural crop yield prediction modeling, Agriculture 10 (9) (2020) 400, https://doi.org/10.3390/agriculture10090400.

[22] Aleksandra Wolanin, Luis Guanter, Gustau Camps-Valls, Grégory Duveiller, Extracting important features for crop yield prediction with convolutional neural networks on remote sensing and meteorological data, Geophys. Res. Abst. 21 (2019).

[23] Thomas van Klompenburg, Ayalew Kassahun, Cagatay Catal, Crop yield prediction using machine learning: A systematic literature review, Comput. Electron. Agric. 177 (2020) https://doi.org/10.1016/j.compag.2020.105709.

[24] Soumya Attaluri, Nowshath Batcha, Mafas Raheem, Crop plantation recommendation using feature extraction and machine learning techniques, 4 (2020) 1–4.

[25] Sagarika Sharma, Sujit Rai, Narayanan C. Krishnan, Wheat crop yield prediction using deep LSTM model, Comput. Vis. Pattern Recognit. (2020) arXiv - CS.

[26] C.-T. Lin, N.-J. Wang, Claudia Xiao, Feature selection and extraction for malware classification, J. Inf. Sci. Eng. 31 (2015) 965–992.

[27] Yan Xiao, Qigang Jiang, Bin Wang, Yuanhua Li, Shu Liu, Can Cui, Object based land-use classification based on hybrid feature selection method of combining Relief F and PSO, Trans. Chin. Soc. Agric. Eng. 32 (4) (2016) 211–21 (6).

[28] Serkan Gunal, Hybrid feature selection for text classification, Turk. J. Electr. Eng. Comput. Sci. 20 (2012) 1296–1311, http://dx.doi.org/10.3906/elk-1101-1064.

[29] Silvia Cateni, Valentina Colla, Marco Vannucci, A hybrid feature selection method for classification purposes, in: 2014 UKSim-AMSS 8th European Modelling Symposium, IEEE.

[30] D. Somvanshi, R. Yadava, Boosting principal component analysis by genetic algorithm, Def. Sci. J. 60 (4) (2010) 392–398, http://dx.doi.org/10.14429/dsj.60.495.

[31] PCA MathWorks. Available at: https://in.mathworks.com/help/stats/pca.html. Accessed online July 2017.
[32] K.A. Severson, M.C. Molaro, R.D. Braatz, Principal component analysis of process datasets with missing values, Processes 5 (2017) 38.
[33] M. Skurichina, R.P.W. Duin, Combining feature subsets in feature selection, in: N.C. Oza, R. Polikar, J. Kittler, F. Roli (Eds.), Multiple Classifier Systems. MCS 2005, in: Lecture Notes in Computer Science, vol. 3541, Springer, Berlin, Heidelberg, 2005.
[34] S.B. Kotsiantis, D. Kanellopoulos, P.E. Pintelas, Data pre-processing for supervised leaning, Int. J. Comput. Sci. 1 (2) (2006) 111–117.
[35] Nikolas Mittag, Imputations: Benefits, Risks and a Method for Missing Data, Technical report, Harris School Of Public Policy, University of Chicago, 2013.
[36] S. Gopal Patro, Kishore Kumar Sahu, Normalization: A pre-processing stage, IARJSET (2015).
[37] Marina Skurichina, Robert P.W. Duin, Combining feature subsets in feature selection, in: Multiple Classifier Systems, Springer Berlin Heidelberg, 2005, pp. 165–175.
[38] I. Guyon, A. Elisseeff, An introduction to feature extraction, in: Guyon, I. and Nikravesh, M. and Gunn, S. and Zadeh, L.A., in: Studies in Fuzziness and Soft Computing, vol. 207, Springer, Berlin, Heidelberg, 2006.
[39] Agustin Garcia Asuero, Ana Sayago, Gustavo Gonzalez, The correlation coefficient: An overview, Crit. Rev. Anal. Chem. 36 (2006) 41–59.
[40] Francesca Fallucchi, Fabio Massimo Zanzotto, Singular value decomposition for feature selection in taxonomy learning, in: International Conference RANLP, 2009, pp. 82–87.
[41] J.H. Holland, Adaptation in Natural and Artificial Systems, MIT Press, Cambridge MA, 1975.
[42] Oswaldo Ludwig, Urbano Nunes, Novel maximum-margin training algorithms for supervised neural networks, IEEE Trans. Neural Netw. 21 (6) (2010) 972–984.
[43] T. Cover, J. Thomas, Elements of Information Theory, John Wiley & Sons, New York, 2006.
[44] R.J. Hyndman, G. Athanasopoulos, Forecasting: Principles and Practice, OTexts, Melbourne, Australia, 2013, http://otexts.org/fpp/.
[45] I.T. Jolliffe, Principal Component Analysis, second ed., in: Springer Series in Statistics, 2002, http://dx.doi.org/10.1007/978-1-4757-1904-8.
[46] Sebastian Raschka, Implementing a principal component analysis (PCA)-in Python, step by step, Apr 13, 2014.
[47] Junita Mohamad-Saleh, Brian S. Hoyle, Improved neural network performance using principal component analysis on Matlab. N.o. 162, 0002, pp. 1–8.
[48] R. Mehta, K. Rana, An empirical analysis on SVD based recommendation techniques, in: 2017 Innovations in Power and Advanced Computing Technologies, I-PACT, 2017, pp. 1–7, http://dx.doi.org/10.1109/IPACT.2017.8245024.
[49] K.P. Shroff, H.H. Maheta, A comparative study of various feature selection techniques in high-dimensional data set to improve classification accuracy, in: 2015 International Conference on Computer Communication and Informatics, ICCCI, 2015, pp. 1–6, http://dx.doi.org/10.1109/ICCCI.2015.7218098.
[50] S. Jia, Y. Qian, J. Li, W. Liu, Z. Ji, Feature extraction and selection hybrid algorithm for hyperspectral imagery classification, in: 2010 IEEE International Geoscience and Remote Sensing Symposium, 2010, pp. 72–75, http://dx.doi.org/10.1109/IGARSS.2010.5652463.