JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2280 - 2288
https://publishoa.com
ISSN: 1309-3452

Pattern Classification for Random Forest by using Fuzzy Logic

Archana P. Kale
Associate Professor

ABSTRACT
Due to digital advancement, the whole world contains the huge amount of data (irrelevant and redundant data). If we use such type of data then the performance of the system degrades. So, Prominent Feature Selection problem is one of the challenging problems in the Machine Learning Area. Principal component analysis (PCA) is the dimensionality reduction (DR) technique in which the original features are transformed from higher dimensional space into lower dimensional space. Though the PCA space has orthogonal principal components (PC), it does not provide a real reduction of dimensionality in terms of the original features (variables), as all features including irrelevant and redundant features are required to define a single PC. It is necessary to remove such type of features by using feature subset selection (FSS) for better generalization performance. The key objective of this paper is to introduce a PCA based Prominent Feature Selection for Random Forest with Fuzzy Logic algorithm (PF-FRF) approach which is able to handle uncertainty classification problem. The PF-FRF is divided into five subparts : Input, PCA, FSS, fuzzification and classification. FSS selects an Prominent feature order which is based on maximum occurrences within the various filter based ranking algorithms. The simulation results are computed and compared by using sequential forward search strategy for clinical datasets. With the results, it is inferred that PF-FRF provides 5.5% improved generalization performance as compared to P-FRF (without feature subset selection).

Keywords: Principal Component Analysis, Random Forest, Feature Subset Selection Problem, Classification Problem, Supervised.

1. Introduction

Machine learning (ML) algorithms are basically used to map input data to target output by using various potential functions. However, the presence of irrelevant and redundant features in the input data may negatively affect the generalization potential of ML methods. Thus, feature subset selection (FSS), feature transformation and dimension reduction (DR) are usually required to pre-process input data in order to attain better overall performance.

Principal Component Analysis (PCA) is one of the important feature extraction and feature transformation techniques where the features are transformed into lower dimensional space. PCA has many objectives like :

1. Finding relationships between observations

2. Extracting the most important information from the data

3. Detecting Outlier and it's removal

4. Reducing the dimension of the data by maintaining important information

These goals can be achieved by finding the PCA space, which represents the direction of the maximum variance of the given data [1]. The PCA space consists of orthogonal principal components, i.e. axes or vectors. Although, it may contains the irrelevant and redundant features.

FSS is one of the critical problems in the field of ML, Pattern Recognition and Data Mining. The key challenge of FSS is to provide the same or improved Classification accuracy with a minimum number of relevant and non-redundant features instead of using all features. Prominent features are those which are less redundant but are highly relevant [2]. Irrelevant features are excess and can be omitted from classification process. For example, in medical field, in the human heart disease classification problem, age and blood pressure are mostly required (relevant) features whereas skin and hair color features are not required (irrelevant) at all.

Classification algorithm is one of the supervised learning techniques which is used to detect the category of new observations on basis of trained data set. It is very critical task to categorize the large data by considering the all features. Therefore, various classification algorithms are used to provide accurate results. Thus, in various real time applications of data science and machine learning, different classifiers like decision tree, logistic regression, random forest, SVM, etc. are used to classify the provided dataset having many attributes. Research efforts affirmed that Random Forest (RF) is one of the best classifiers. Existing RF algorithms are unable to solve the classification problem with fuzzy logic. Random forest is one the famous and popular machine learning algorithms which provides easy interpretability, good generalization performance, better predictive performance, low over-fitting. but due to the presence of irrelevant and

redundant features the performance is decrease. So, one of the alternative solutions for this is to introduce a new integrated approach of PCA, Fuzzy Logic and Random Forest which provides a good generalization performance.

The main contribution of this paper is to design a PCA based Prominent Feature Selection for Random Forest with Fuzzy Logic algorithm (PF-FRF) approach which is able to handle the three problems like classification, FSS and DR problems. PF-FRF first transforms the original feature space into PCA space. It further selects the Prominent (relevant and non-redundant) features from PCA space by using Prominent feature rank which is based on various filter based FSS methods. The literature survey lacks to provide solution to the classification problem by using Random forest with fuzzy logic.

This paper is divided into five sections. In section II, the related work of PCA, Random Forest and FSS is overviewed. Section III illustrates the proposed methodology with the help of conceptual framework. Experimental results are shown in section IV. Finally, concluding remarks are given in section V.

2. Related Work

Random forest is one of the supervised learning algorithm which is used for classification purpose [3]. Random forest consist of various number of decision trees and various subsets of given dataset which takes the average of all to increase the generalization performance in terms of classification accuracy, precision , recall etc. Random forest considers the output from each tree present in it. Random Forest classify the data based on the maximum votes on predictions. As each and every attribute is considered in classification, it provides more the number of trees, so that more is the accuracy.

Additionally, Random forest is able to handle missing data problem witch takes less training time as compared to other algorithms. For large datasets also it is able to provide the good generalization performance and predicts the output with higher accuracy. One of the advantageous of it is that it maintain the accuracy even though the large proportion of data is missing. Thus it handle the missing value problem very efficiently. Rung Ching chen et. al. adopted Random Forest to select the important feature in classification [4]. X-Li et. al. used Random Forest models to calculate the importance of the features [5]. Shafizadeh-Moghadam et.al developed an integrated approach of forward feature selection algorithm, random forest, and cellular automata for the application of Iran [6]. Rong Zhu et. al. developed a model of incremental principal component analysis feature selection and a random forest classifier for predicting lncRNA-disease (IPCARF) [7]. Jaime Lynn Speiser developed a random forest method with feature selection for medical prediction models with the help of clustered and longitudinal data [8]. Yaoxin Wang et. al. built a prediction model Using Recursive Feature Selection with Random Forest which was helpful in the improvement of Protein Structural Class Prediction for Low-Similarity Sequences [9]. Prasetiyowati et. al. used information gain based feature selection for random forest which provided the increase speed and prediction accuracy [10].

The goal of the PCA technique is to _nd a lower dimensional space or PCA space (W) that is used to transform the data (X = x1; x2; :::; xN) from a higher dimensional space (RM) to a lower dimensional space (Rk), where N represents the total number of samples or observations and xi represents ith sample, pattern or observation. All samples have the same dimension. In other words, each sample is represented by M variables, i.e. each sample is represented as a point in M - dimensional space [11]. The PCA space consists of a number of PCs where each principal component has a different robustness according to the amount of variance in its direction.

Guo et. al. used genetic algorithm to select a subset of variables in PCA [12] which optimizes the consensus between the subset and the complete data set. Feature selection workflow for high-dimensional omits data by using PCA designed by Perez-Riverol et. al. [13]. Pacheco et. al. described the methods for variable selection in PCA [14]. Principal feature analysis is used in face tracking and content-based image retrieval (CBIR) problems [15].

3. Methods

The proposed PCA based Feature Selection for Random Forest with Fuzzy Logic (PF-FRF) approach is as shown in Figure 1. The PF-FRF approach consists of five subsystems - Input subsystem, PCA subsystem, FSS subsystem, fuzzification subsystem and classification subsystem. In PCA subsystem, all original features are transformed into PCA space by using covariance matrix method. Various filter algorithms are used to rank the feature order and finalize one prominent feature rank order in FSS subsystem. Fuzzification subsystem is used to convert the original features into linguistic /fuzzified features. Finally, forward selection is achieved by wrapping the ML method, such as ELM in classification subsystem. The detailed descriptions of all these subsystems are elaborated in detail.

3.1 Input subsystem

The clinical datasets from UCI Repository [16] like Pima Indian Diabetes (PID) and Heart-Statlog (SHD) are used in which uncertainty is present [17]. PID dataset contains 768 instance with 8 attributes and SHD dataset contains 270

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2280 - 2288
https://publishoa.com
ISSN: 1309-3452

instance with 13 attributes. These datasets are assigned to two classes; either +1 or -1 and are used for binary classification problem.

3.2 Principal Component Analysis subsystem

By using original dataset, PCA transforms the features or finds the eigenvector by using covariance matrix method.

3.2.1 Principal Components (PCs)

The PCA space consists of k principal components. The principal components are orthonormal, uncorrelated and it represents the direction of the maximum variance. Orthonormal vectors have a unit length and are orthogonal, calculated by using Equation :

$$v_i^T v_j = \begin{bmatrix} 1, i = j \\ 0, i \neq j \end{bmatrix}$$

vi and vj are uncorrelated if Cov(vi; vj) = 0, where Cov(vi; vj) represents the covariance between ith and jth vectors. The first principal component ((PC1 or v1)2 RM_1) of the PCA space represents the direction of the maximum variance of the data. The second principal component has the second largest variance, and so on. Figure 2 shows how original data is transformed from the original space X and Y (RM) to the PCA space PCA1 and PCA2 (Rk). Thus, the PCA technique is considered as an orthogonal transformation due to its orthogonal principal components. It also considered an axes rotation due to the rotation of the original axes [11] [18].

3.2.2 Covariance Matrix Method

Two main steps are included in covariance matrix method -

- The data matrix (F) of covariance matrix
- The eigenvalues (γ) and eigenvectors (V)

Calculating Covariance Matrix (P) The variance measures the deviation of the variable from its mean value. The covariance matrix is calculated as:

$$\sum = DD^T$$

Covariance matrix is positive semi-definite and symmetric matrix (i.e. F = FT). The diagonal values represent the variance of the variable fi, for i = 1,2,..,M of the covariance matrix. The covariance between two different variables is represented by the off-diagonal entries as shown in Equation.

The positive and negative correlation or statistical independence in terms of un-correlation can be obtained [18].

$$\begin{bmatrix} Var(f_1, f_1) & Cov(f_1, f_2) & \dots & Cov(f_1, f_M) \\ Cov(f_2, f_1) & Var(f_2, f_2) & \dots & Cov(f_2, f_M) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(f_M, f_1) & Cov(f_M, f_2) & \dots & Var(f_M, f_M) \end{bmatrix}$$

Calculating Eigenvalues (ʎ) and Eigenvectors (V):

The covariance matrix is solved by calculating the eigenvalues (ʎ) and eigenvectors (V) as follows:

$$V\sum = ʎV$$

Where, V and ʎ represent the eigenvectors and eigenvalues of the covariance matrix, respectively.

The eigenvalues are scalar values, while the eigenvectors are non-zero vectors, representing the principal components i.e. each eigenvector illustrating one principal component. The eigenvectors represent the directions of the PCA space and the corresponding eigenvalues represent the scaling factor, length, magnitude or the robustness of the eigenvectors [20] [21]. The eigenvector with the highest eigenvalue represents the first principal component and it has the maximum variance [20]. However, PCA does not produce a real reduction of dimensionality in terms of the original variables, as all original variables are required to represent a single Principal Component (PC) [14].

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2280 - 2288
https://publishoa.com
ISSN: 1309-3452

3.3 Feature Subset Selection Subsystem

PCA space is the robust database as it contains various irrelevant and redundant features also. To remove these features or to select the Prominent features, various filter based FSS algorithms like relief [22], chi-square test (CHI) [23], t-test [24], information gain (IG) [22], correlation-based feature selection (CFS) [22] and gain ratio (GR) [23] are used to rank the features. Four evaluation measures like information, distance, consistency and dependency are used to rank the features as shown in Table 1. The feature order set is sorted and listed as per the mentioned importance priority. Relief algorithm is an example of the distance measure which calculates the relevance. For information measure, GR and IG are used to calculate the redundancy. As name indicates, dependency measure is used to calculate the dependency between variables. Student's T-test, SU and CFS attribute evaluation are some of the examples of this measure. CFS can only detect linear dependencies of feature and the target.

To find out a set of minimum number of features that define the classes, consistency measure is used, for example F-score (FS) [24]. The algorithm gives the ranking of feature dependency on feature's significance. The features of the clinical data set are organized in a descending order for feature selection methods.

From all these ranking lists, one order list is finalized as an optimal feature rank which has maximum occurrences as shown in the last column of Table 1.

Sequential forward search (SFS) is used to insert one by one features and classify with F-ELM classifier. For example, the feature order for PID and SHD data set is finalized as <2; 6; 8; 1; 7; 5; 4; 3> and <13; 12; 3; 9; 10; 11; 2; 1; 7; 4; 5; 6; 8> respectively. The ordered features are incrementally added one at a time for each subset by using SFS, illustrating PID data set which has feature order as: <2; 6; 8; 1; 7; 5; 4; 3>. First, the classification accuracy is calculated by using only the first feature with high rank, feature 2 (feature subset 1). Then feature 6 is inserted with feature 2 (feature subset 2). Similarly, total 8 feature subsets are created. Out of all feature subsets, one feature subset is selected as efficient, which gives the maximum classification accuracy as indicated in the last column of Table 2.

Table 1: Ordered List of Feature Selection Methods Distance Information Dependency Consistency Prominent Feature Order

Relief	IG	GR	t-test	CR	SU	FSore	Prominent Feature Order
2	2	2	2	2	2	2	2
6	6	6	6	6	6	6	6
4	8	8	8	8	8	8	8
8	5	1	1	1	5	1	1
1	1	5	7	7	1	7	7
7	4	7	5	5	4	5	5
5	7	4	4	4	7	4	4
3	3	3	3	3	3	3	3

Table 2: Classification accuracy with the Feature order < 2; 6; 8; 1; 7; 5; 4; 3> for PID

Classifier	Classification Accuracy with Number of features-(Feature Subset (FS)) Max Efficient FS								Max	Efficient FS
PCA RF	74.78	69.56	70	66.08	68.26	72.6	69.13	67.82	74.78	2
PCA FRF	76.52	78.26	79.13	81.73	82.6	80.43	99.56	81.3	99.56	1,2,4-8

3.4 Fuzzification Subsystem

The feature F is converted into fuzzified features (Mq j) by using the trapezoidal membership function given in Equation 8 [17].

$$f(M,a,b,c,d) = \begin{cases} 0 & M < a, M > d \\ \frac{M-a}{b-a} & a \leq M \leq b \\ 1 & d \leq M \leq c \\ \frac{d-M}{d-c} & c \leq M \leq d \end{cases}$$

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2280 - 2288
https://publishoa.com
ISSN: 1309-3452

The parameters a; b; c; d are used to control the membership values of the M feature. These parameters are decided by considering the clinician's suggestion [25]. In fuzzification, eight features of PID data set are converted into twenty- five linguistic variables [17]. All features of SHD data set are also transformed into the linguistic variable (fuzzified features).

3.5 Classification Subsystem

Random forest is one of the supervised learning algorithm which is used for classification purpose. Random forest consist of various number of decision trees and various subsets of given dataset which takes the average of all to increase the generalization performance in terms of classification accuracy, precision, recall etc. Random forest considers the output from each tree present in it. Random Forest classify the data based on the maximum votes on predictions. As each and every attribute is considered in classification, it provides more the number of trees, so that more is the accuracy. Additionally, Random forest is able to handle missing data problem which takes less training time as compared to other algorithms. For large datasets also it is able to provide the good generalization performance and predicts the output with higher accuracy. One of the advantageous of it is that it maintain the accuracy even though the large proportion of data is missing. Thus it handles the missing value problem very efficiently.

Basically, Random forest task is divided into two phases -

- To create forest containing N decision trees
- To predict the accuracy for each tree and afterward at last step calculate the average of all outputs

First, Random Forest selects random `k' data points by using the provided training dataset. It builds the decision tree associated with that selected data points. It chooses the number `N' for decision trees according to the used applications for build. These things are repeated. For any new data point and for each decision tree, first and the predictions and then assign the new data point.

4. Experimental Results

Experiments have been conducted using MATLAB R2014a. The given dataset is divided into training and testing set with 70% and 30% respectively. The evaluation metrics like accuracy, precision, recall, f-measure and g-measure are used with sigmoidal activity function to compare the performances of P-FRF and PF-FRF. The observations are obtained by considering true negatives (TN), true positives (TP), false positives (FP) and false negatives (FN) [17].

The accuracy of the classification model on a given test is the percentage of test set that is correctly classified by the classifier [26]. Precision is the measure of correctness of positive labeled examples. Recall is the measure of completeness or accuracy of positive samples that how many objects of the positive class are labeled correctly [27]. In statistical analysis of binary classification, the f-score or f-measure is a measure of a test's accuracy. F-measure is the harmonic mean of recall and precision while the g-measure is the geometric mean [28].

As F-ELM is the next version of ELM, the experimental results are also calculated by using ELM classifier for comparison. To verify that the F-ELM classifier provides an improved generalization performance as compared to ELM, the evaluation performance is compared by using three strategies like:

RF with All features and transformed PCA Features

RF with All fuzzified features and transformed fuzzified PCA Features

RF with All main features and All fuzzified Features (RF and F-RF)

PF-FRF and PF-RF (Transformed Prominent features)

PF-FRF and Existing Methods

4.1 Results of Random Forest with All features and transformed PCA Features

The Performance of the proposed approach is compared by using all features and transformed features with various existing classifiers like Decision Tree Classifier, Logistic Regression, Naive Bays, K-Nearest Neighbourhood as shown in Figure 3. With the results it is observed that Transformed features provides 6.07% improved accuracy. RF with PCA Transformed features provides 1.99% and 4.66% improved accuracy as compared to other existing algorithms by using all features and transformed features respectively.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2280 - 2288
https://publishoa.com
ISSN: 1309-3452

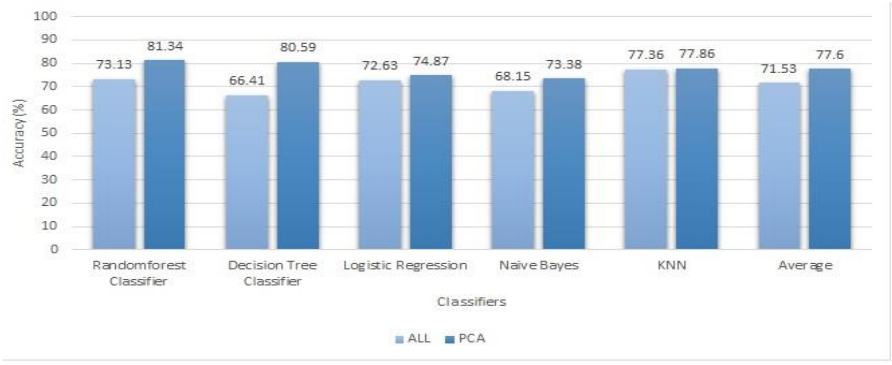

Figure 3: Comparative Performance of all features and transformed features with existing classifiers

4.2 Results of Random Forest with All fuzzified features and transformed fuzzified PCA features

The Performance of the proposed approach is compared by using all fuzzified features and transformed fuzzified features with various existing classifiers like Decision Tree Classifier, Logistic Regression, Naive Bays, K-Nearest Neighbourhood as shown in Figure 4. With the results it is observed that Transformed features provides 8.61% improved accuracy. RF with PCA Transformed features provides 18.71% improved accuracy as compared to RF by using all features.

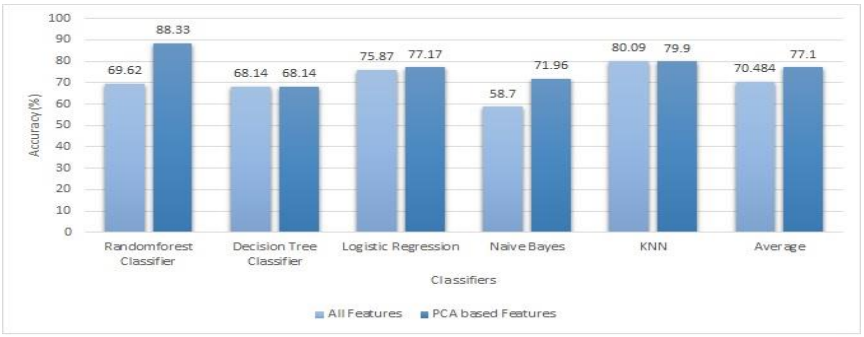

Figure 4: Comparative Performance of all features and transformed features with existing classifiers

4.3 Results of Random Forest with All main features and All fuzzified Features (RF and F-RF)

The Performance of the proposed approach is compared by using all main features and all fuzzified features with various existing classifiers like Decision Tree Classifier, Logistic Regression, Naive Bays, K-Nearest Neighbourhood as shown in Figure 5. With the results it is observed that F-RF provides 7.1% improved accuracy as compared to RF. RF with PCA Transformed features provides 1.08% and 4.66% improved accuracy as compared to other existing algorithms by using all main features and all fuzzified features respectively.

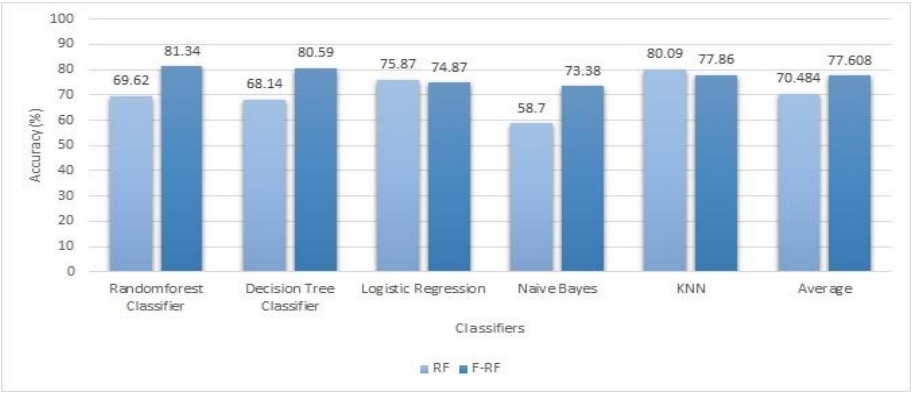

Figure 5: Comparative Performance of Random Forest and Fuzzy based Random Forest with all features

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2280 - 2288
https://publishoa.com
ISSN: 1309-3452

4.4 PF-FRF and PF-RF (Transformed Prominent features)

The introduced PF-FRF is compared with PF-RF by using optimal feature subset. Figure 6 shows that the PF-FRF provides 5.5% improved performance as compared to PF-RF for PID dataset. RF with PCA Transformed features provides 1.09% and 14.03% improved accuracy as compared to other existing algorithms by using optimal pca based main features and optimal pca based fuzzified features respectively.

Figure 6: Comparative Performance of PF-FRF and PF-RF by using Optimal features for PID

4.5 PF-FRF and Existing Methods

The performance of the proposed PF-FRF is compared with six other [29] machine learning algorithms like FRF, RF, ELM, SVM, BN, RBF and J4.8 in terms of accuracy as shown in Table 3. It is observed that PF-FRF provides an improved 25.96 % and 10.15 % average accuracy with 12.5 % and 15.38 % feature reduction rate for PID and SHD datasets. Some of the comparison between P-FRF with already existing methods like F-ELM [30], enhanced generalized adaptive resonance theory FIS (EGART-FIS) [31] and modified fuzzy min-max NN-FIS (MFMM-FIS) [32]. The feature reduction rate for P-FRF is 12.5% with 6.2 % improvement in accuracy as compared to already available methods.

Table 3: Performance Comparison of PF-FRF and other classifiers

Method	Accuracy(%)	
	PID	SHD
PF-FRF	88.33	87.65
FRF	81.3	87.65
RF	69.62	68.77
ELM	71.3	77.77
SVM	65.10	55.92
BN	74.73	82.59
RBF	75.39	83.33
J4.8	71.22	77.77

5. Conclusion

The key objective of this paper is to develop a dimensionality reduction (DR) framework in combination with feature subset selection (FSS) i.e. PCA based optimal FSS for fuzzy based Random Forest (PF-FRF). The PF-FRF is broadly

categorized into five subsystems like Input, PCA, FSS, fuzzification and classification subsystems. The proposed PF-FRF first transforms original features or finds out the eigenvectors into PCA space by using covariance method which is robust database as it contains various irrelevant and redundant features also. To remove these features or to select the optimal features, FSS is used. Various FSS methods like relief, CHI, t-test, IG, CFS and GR are used to rank the features. Through all ranking lists, one order list is finalized as an optimal feature rank which has maximum occurrences. RF classifier is used for classification.

The performance comparative analysis in terms of learning algorithm shows that the proposed PF-FRF provides 18.37% and 6.98% improved performance as compared to PF-ELM for PID and SHD dataset respectively. And the performance comparative analysis in terms of number of features shows that the PF-FRF provides 5.5% improved performance as compared to PF-RF for PID dataset. PF-FRF is able to cope with the four issues like: 1) Binary classification 2) DR 3) FSS 4) Fuzzy Logic. The work can be extended for multi-class classification problem. The multi-objective non-dominated sorting genetic algorithm II (NSGA-II) can also be used for Feature Selection [33] which may support with clear insight and direction for further improvements.

R e f e r e n c e s

1. C a m b r i a, E r i k, G u a n g – B i n H u a n g, O t h e r. Extreme learning machines [trends & controversies]. IEEE Intelligent Systems 28(6). 2013, 30-59.
2. Y u, L e i, H u a n L i u, and I s a b e l l e G u y o n. Efficient feature selection via analysis of relevance and redundancy. Journal of Machine Learning Research 5, 2004, 1205-1224.
3. P a h u j a, G u n j a n and T. N. N a g a b h u s h a n. A novel ga-elm approach for Parkinson's disease detection using brain structural t1-weighted mri data. In IEEE Cognitive Computing and Information Processing (ccip), second international conference on 2016, 1-6.
4. M a o, L i, L i d o n g Z h a n g, O t h e r. Improved extreme learning machine and its application in image quality assessment. Mathematical Problems in Engineering. 2014.
5. L i a n g, N a n – Y i n g, G u a n g – B i n H u a n g, and O t h e r. A fast and accurate online sequential learning algorithm for feedforward networks. IEEE Transactions on neural networks, , 2006, 17 (6): 1411-1423.
6. Z h u, Q i n – Y u, A K a i Q i n, P o n n u t h u r a i and O t h e r. Evolutionary extreme learning machine. Pattern recognition, 2005, 38 (10): 1759-1763.
7. H a n, F e i, H a i – F e n Y a o and Q i n g – H u a L i n g. An improved evolutionary extreme learning machine based on particle swarm optimization. Neurocomputing, 2013, 116: 87-93.
8. C h e n, Z h i x i a n g X and O t h e r. A modified extreme learning machine with sigmoidal activation functions. Neural Computing and Applications, 2013, 22 (3-4): 541-550.
9. N g u y e n, D e r r i c k, and B e r n a r d W i d r o w. Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights. In IEEE IJCNN international joint conference on Neural networks, 1990, 21-26.
10. W o n g, S h e n Y u o n g, and O t h e r. A truly online learning algorithm using hybrid fuzzy artmap and online extreme learning machine for pattern classification. Neural Processing Letters, 2015 42 (3): 585-602.
11. L a n, Y u a n, Y e n g C h a i S o h, and G u a n g – B i n H u a n g. Ensemble of online sequential extreme learning machine. Neurocomputing, 2009, 72 (13): 3391-3395.
12. H u a n g, G u a n g – B i n, and L e i C h e n. Enhanced random search based incremental extreme learning machine. Neurocomputing, 2008, 71 (16): 3460-3468.
13. M, S e e r a , and L i m C P. A hybrid intelligent system for medical data classification extreme learning machine: theory and applications. Expert System Application, 2014, 41 (05): 2239-2249.
14. Z h a n g, R u i, Y u a n L a n, G u a n g – b i n H u a n g, and Z o n g – B e n X u. Universal approximation of extreme learning machine with adaptive growth of hidden nodes. IEEE Transactions on Neural Networks and Learning Systems, 2012, 23 (2): 365-371.
15. A s u n c i o n, A r t h u r, and D ssa v i d N e w m a n. UCI machine learning repository. 2007.
16. Li J, Liu H e n t r i d g e. Bio-medical data set repository. School of Computer Engineering, Nanyang Technological University, Singapore. 2004.
17. A k u s o k, A n t o n, K a j – M i k a e l B j o r k, Y o a n M i c h e and A m a u r y L e n d a s s. High-performance extreme learning machines: a complete toolbox for big data applications. In: IEEE Access 3, 2015, 1011-1025.
18. Z h u o, L i, J i n g Z h e n g, and O t h e r. A genetic algorithm based wrapper feature selection method for classification of hyper spectral images using support vector machine. In Geoinformatics 2008 and joint conference on built environment: Classification of remote sensing images, 2008, 71471-71471.
19. Y. F. C h e n g, W. S h a o, S.J. Z h a n g and Y.P. L i, An improved multi-objective genetic algorithm for large planar array thinning," IEEE Transactions on Magnetics, , 2016, vol. 52, no. 3, pp. 1-4.

20. F e i, Y e, and H a n M i n. Simultaneous feature with support vector selection and parameters optimization using ga-based svm solve the binary classification. IEEE international conference on Computer communication and the internet (iccci), 2016, 426-433.
21. N a h a t o, K i n d i e B i r e d a g n, and O t h e r. Hybrid approach using fuzzy sets and extreme learning machine for classifying clinical datasets. Informatics in Medicine Unlocked, 2016, 2: 1-11.
22. L e C u n, Y a n n A, L e o n B o t t o u, G e n e v i e v e B O r r, and K l a u s – R o b e r t M u l le r. Efficient backprop. Springer In Neural networks: Tricks of the trade, 2012, 9-48.
23. H u a n g, G u a n g – B i n, and O t h e r. An efficient sequential learning algorithm for growing and pruning rbf (gap-rbf) networks. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2004, 34 (6): 2284-2292.
24. H u a n g, G u a n g – B i n, and O t h e r. A generalized growing and pruning rbf (ggap-rbf) neural network for function approximation. IEEE Transactions on Neural Networks, 2005, 16 (1): 57-67.
25. H u, B i n, Y o n g q i a n g D a i, Y u n S u, and O t h e r. Feature selection for optimized high-dimensional biomedical data using the improved shuffled frog leaping algorithm. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2016.
26. D e b, K a l y a n m o y, A m r i t P r a t a p, and O t h e r. A fast and elitist multi objective genetic algorithm: Nsga-ii. IEEE transactions on evolutionary computation, 2002, 6 (2): 182-197.
27. K a r a k a y a, G u l s a h, S t e f a n o G a l e l l i, O t h e r. Identifying (quasi) equally informative subsets in feature selection problems for classification: a max-relevance min-redundancy approach. IEEE transactions on cybernetics, 2016, 46(6): 1424-1437.