Feature and Subfeature Selection for Classification Using Correlation Coefficient and Fuzzy Model

Hemanta Kumar Bhuyan, Chinmay Chakraborty, Subhendu Kumar Pani, and Vinayakumar Ravi

***Abstract*—This article presents an analysis of data extraction for classification using correlation coefficient and fuzzy model. Several traditional methods of data extraction are used for classification that could not provide sufficient information for further step of data analysis on class. It needs refinement of features data to distinguish a class that differs from a traditional class. Thus, it proposes the feature tiny data (subfeature data) to find distinguish class from a traditional class using two methods such as correlation coefficient and fuzzy model to select features as well as subfeature for distinguishing class. In the first approach, the correlation coefficient methods with gradient descent technique are used to select features from the dataset and in the second approach, the fuzzy model with supreme of minimum value is considered to get subfeature data. As per the proposed model, some features (i.e., three features from the acoustic dataset, two features from the QCM dataset, and eight features from the audit dataset, etc.) and subfeatures (as per threshold value like 20 for acoustic; 10 for QCM, and 20 for audit, etc.) are selected based on correlation coefficient as well as fuzzy methods, respectively. Further, the probability approach is used to find the association and availability of subfeature data from the dimensional reduced database. The experimental results show the proposed framework identifies and selects both feature and subfeature data with the effectiveness of the new class. The comparison results of several classifiers on several datasets are explained in the experimental section.**

***Index Terms*—Classification, correlation coefficient, data mining, feature selection, fuzzy model.**

I. Introduction

THE dimensional features in the dataset have been created a critical situation for the analysis of data in many cases. The lower dimension of the feature set with effective values has produced a valuable result for data analysis. For indifferent features, data analysis yields low-quality results, and classifiers produce wide results. Thus, data are always analyzed on compact effective values which produce well result forever.

Manuscript received 24 July 2020; revised 17 February 2021; accepted 25 February 2021. Date of publication 19 April 2021; date of current version 17 March 2023. This work was supported by the Laboratory of VFSTRU, AP, India. Review of this manuscript was arranged by Department Editor G. Kayakutlu. *(Corresponding author: Chinmay Chakraborty.)*

Hemanta Kumar Bhuyan is with the Department of Information Technology, Vignan's Foundation for Science, Technology and Research (Deemed to be University), Andhra Pradesh 522213, India (e-mail: hmb.bhuyan@gmail.com).

Chinmay Chakraborty is with the Electronics and Communication Engineering, Birla Institute of Technology Mesra, Jharhand 835215, India (e-mail: cchakrabarty@bitmesra.ac.in).

Subhendu Kumar Pani is with the Department of Computer Science and Engineering, Orissa Engineering College, Biju Patnaik University of Technology, Odisha 769004, India (e-mail: skpani.india@gmail.com).

Vinayakumar Ravi is with the Center for Artificial Intelligence, Prince Mohammad Bin Fahd University, Khobar 34754, Saudi Arabia (e-mail: vravi@pmu.edu.sa).

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TEM.2021.3065699.

Digital Object Identifier 10.1109/TEM.2021.3065699

In this article, data are analyzed on features as well as subfeatures data based on the proposed model, i.e., feature and subfeature selection model in data mining. Feature selection from pertinent features is an important task for data mining. It is done by the understanding of the underlying structure of features data and the discarding of disparaging features data. The multiple classifiers have been considered with the different features data to improve the performance. Thus, the designing of selected features data for evaluation has taken into account in this article. Although, the applications on feature selection have been developed by different researchers based on a biological dataset for diagnostic classification systems [1]–[3]. They showed a way to find the solution to the problem on the "dimensionality of feature set." It is concerned with dimensional reduction methods such as supervised, unsupervised, and semisupervised methods. But to solve a particular problem, these methods are not always necessary. The correlation methods may take power to select features data for classification, but most of the feature selection scheme does not comprise the dimensional reduction of features data whereas few tried among selected features [4]. Chen *et al.* [31] have developed a novel semisupervised embedded feature selection method using rescaling the regression coefficients in the least square regression. Ali *et al.* [32] have also made deep private-feature extractor using the sensitive selective exchange of information among users and service providers. Further, Xu *et al.* [33] proposed multiview classification and feature selection using support vector machines and iterative optimization method with better performance than other state-of-the-art learning algorithms.

Thus, this article aims to select features from the dataset for classification, where features are a very useful and independent manner for observation as well as evaluation. The desirable objectives on selected features have been taken into consideration for further processing of classification and preserved such selected features are used for subfeature data selection. Based on the above concerning situation, the article divides two parts for description such as (a) feature selection and (b) subfeature selection.

1) *Feature selection*: In this part, the correlation coefficient methods proposed to elaborate on the feature selection process based on the learning process with derivative methods. The gradient descent technique is used to

determine the selected features for classification through objective function. Thus, the problem formulation is considered by Sammon's error (SE) to determine Euclidian distance between two variables and it is one part of the objective function. Thus, the objective function is defined on SE and correlation coefficient which is mentioned in Section IV.

2) *Subfeature selection*: In this part, there is very little research paper available to collect the related information. The idea of subfeature is initiated by Bhuyan *et al.* [5], [6]. The fuzzy methods are used to find subfeature data. After selecting features from the dataset, two important tasks have been taken into consideration such as (a) association of subfeature values of different features and (b) the % of availability of subfeature values concerning features of the corresponding dataset.

Taking the above methodologies, the experimental results are analyzed by using three different datasets which are very effective to determine both feature and subfeature selection for classification that is mentioned in Section VI.

The rest of this article is organized as follows. The survey of related different research work is reviewed in Section II. The problem statement is formulated in Section III. Section IV explained the correlation framework for feature selection and the gradient descent technique. The fuzzy model for subfeature data is described in Section V. The experimental results are analyzed in Section VI. Finally, Section VII concludes this article.

II. Related Work

Since this article involves feature and subfeature selection, the background of this research work is considered based on two parts. In the first part, it describes the related feature selection approaches whereas subfeature selection based on respective selection approaches is elaborated in the second part which is very less due to lack of inadequate information regarding the corresponding related research work.

A. Several Approaches to Feature Selection

Several feature selection approaches have been developed by different researchers as in [7]–[9], [27]. Most of the researchers apply the feature selection method based on unsupervised approaches. The partition of feature set into several clusters is proposed by Mitra *et al.* [9] using an unsupervised feature selection scheme for reducing the feature set. The feature similarity is determined by using the maximum information compression index approach over various data sizes of real-life data set. Sometimes, the graph Laplacian for feature selection has described in [10] for multicluster data. Further, feature selection is an important step in machine learning. In an incremental feature selection approach, the past approaches have attempted to increase class relevance while simultaneously minimizing redundancy with previously selected features [11]. It proposes a general domain-specific feature transfer framework, which can link up different domains using common features and simultaneously reduce domain divergences [12]. Subsequently, it proposes a ranking procedure in the feature pool for balancing the information gain and the complexity to avoid overfitting [13].

Few studies have been conducted on test-cost-sensitive feature selection. It considers a new term to the evaluation function of a wrapper feature selection method so that the test cost of measuring features is taken into account [14]. An effective feature selection method is introduced to select useful features hierarchically [15], [35]. It proposes a novel multivariate feature selection method in which both search strategy and classifiers are based on multiobjective evolutionary computation (EC) [16]. It presents a comprehensive survey of the state-of-the-artwork on EC for feature selection, which identifies the contributions of these different algorithms [17]. Further, Bhuyan *et al*. describe the feature and subfeature selection in [18]–[20]. Similarly, many researchers have developed their techniques and implemented existing methods as in [24]–[26]. Further few authors have developed several fuzzy logic models for classification using different datasets [28]–[30].

B. Entropy-Based Feature Selection

The description of the support vector machine (SVM) method has been taken based on recursive feature elimination to remove the features. The significant features are selected by this method that yield better classification performance [20]. Alter *et al.* [21] have used singular value decomposition (SVD) for gene data processing based on Shannon entropy as

$$E(X_{|n\times p|}) = -\frac{1}{\log(N)}\sum_{j=1}^{N} V_j \log V_j \tag{1}$$

where the dataset $X \in R^{nxp}$ $N = \min\{n, p\}$, and v_j is the singular value of X. the relevance of each feature is determined by SVD based entropy in [16]. The contribution of each feature to the entropy (CE) is defined by

$$CE_i = \mathrm{E}(X_{|n\times p|}) - \mathrm{E}(X_{|n\times \bar{p}|}) \tag{2}$$

where $\bar{p} = \mathrm{p}\backslash(i\text{th feature})$. The selection of features has developed using simple ranking, forward selection, and backward elimination-based CE_i. Readers can follow the details in [22]. Further, the modified definition of CE_i as mCE_i is developed by Banerjee and Pal [8] as follows:

$$mCE_i = \mathrm{E}(X_{|n\times \bar{p}|}) - \mathrm{E}(X_{|n\times p|}). \tag{3}$$

This equation reduces redundancy than CE_i.

C. Subfeature Selection

In this part, very few papers are available including Bhuyan *et al.* [5], [6] for subfeature selection. There is no sufficient information on this part. This has motivated us to explore the research work in the direction of the subfeature selection method.

III. Problem Statement

The feature selection has been focused based on the removal and controls of the irrelevant features in a database for better classification performance. Banerjee *et al.* [22] have shown that how the irrelevant features are discarded and controlled the redundancy from any dataset. But when this model is considered for classification, the corresponding method may not be performed for predicted class when fewer features are available in datasets because the above method selects very few features in some cases. But for a large number of features in datasets, this method may consider in certain cases. In some cases, the feature selection does not reflect the performance of classification, but it is performed by feature's tiny values for the different identification of class (i.e., the several classes can be identified differently based on their feature's tiny data). Thus, it needs to select such feature's tiny data for better classification. The feature's tiny data called subfeature data have been described in [5] and [6]. Thus, it motivates to select subfeature data from each feature for individual dataset. The role of subfeature is very important in each dataset for classification especially to find a novel class or unique class. Thus, it considers the collection of similar feature data from a different database to make large dataset otherwise the novel class (unique class) may not get from a few datasets. Since the novel class is unknown in the existing class, it considers finding it by applying unsupervised methods. It mainly tries to find the subfeature set that can generate a new class that is different from the existing class based on their corresponding subfeature data. The construction of problems comprises two ways (a) problem formulation is defined through correlation coefficient for the feature set and (b) set a fuzzy logic approach are used for the subfeature set as follows.

A. Problem Formulation Feature Set

Several approaches can be used to measure the appropriate feature selection based on class labels. If class labels are not known, SE can be used for formulating the feature extraction. Let $X \subseteq R^{nxp}$ be the dataset, i.e., X has n data samples and p dimensions (or features) X can be represented by $X = \{x_1, x_2, \ldots. x_n\}$, where $x_i = (x_{i1}, x_{i2}, \ldots. x_{ip})^T$, i = 1…..n. The Euclidean distance between x_i, x_j is represented as

$$d_{ij}^{0}(X_i, X_j) = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2}. \tag{4}$$

Again let the dataset in the reduced dimension be $Y = \{y_1, y_2, \ldots.y_n\}$ and its corresponding Euclidian distance $d_{ij}^{r}(y_i, y_j)$ can be defined between y_i & y_j, where Y can be determined by using gradient descent manner based on associating weight (w_i) which can use a monotonic different differentiable function with range [0, 1]. Thus, it defines

$$Y = (w_1 x_1, w_2 x_2, \ldots w_n x_n)^T.$$

Hence, the Euclidian distance between y_i and y_j i.e.,

$$d_{ij}^{r}(y_i, y_j) = \sqrt{\sum_{k=1}^{p} w_k (x_{ik} - x_{jk})^2}. \tag{5}$$

Then, the SE can be formulated based on the above Euclidian distance as follows:

$$\mathrm{SE} = \frac{1}{\sum_{i<j} d_{ij}^{0}} \sum \frac{\left(d_{ij}^{0} - d_{ij}^{r}\right)^2}{d_{ij}^{0}}. \tag{6}$$

To avoid dependent features, it considers the correlation coefficient between the *i*th and *j*th feature and controls the redundancy. The nonlinear correlation is considered for avoiding constant ratio to the amount of change in the other features values. Since the values of the features are varied for classification, it cannot be determined which features values will be appropriate for a novel class or distinguished class from the existing class. For correlation coefficients, the authors have taken the steps to control redundancy among features in [22].

B. Fuzzy Logic Approaches for the Subfeature Set

Although feature selection is an important task in most classification problems, albeit, it needs a new approach to determine the subfeature selection to improve the classification accuracy. Traditional classification cannot determine the exactness of class toward advanced knowledge for a different purpose. Different kind of subfeature data of corresponding features is being generated day-by-day. For example, a new disease is generated at a different location based on its environment. But initially, physicians are unable to find such a disease based on their knowing traditional class of disease. It can be recognized by verifying the features of tiny data (subfeatures data) based on the latest technology. Accordingly, the physician can treat the patient properly otherwise patient falls in death. Thus, to find sensitive such subfeature data is a difficult task. An effective subfeature selection method is always necessary for classification. Although a large number of classification and feature selection techniques have been used in the past by different researchers, still the subfeature selection remains a major and persistent issue for unique classes. Thus, it aims to identify the subfeature data from each feature to get a unique class for different purposes of the application. The next section is considered to get a partial solution to this problem. In this section, fuzzy logic has been taken into consideration with multivalued logic and binary operation for the association of feature set based on subfeature data. It helps to generate a distinguished class from the existing class. The details are explained in Section V.

IV. Framework for Feature Selection

A. Correlation Framework for the Feature Set

To avoid irrelevant dependent features and control redundancy, it generates a method to assess the redundancy among the selected features based on the correlation coefficient between the *i*th and *j*th features. Thus, the equation can be written as

$$\mathrm{CR} = \frac{1}{(p-r)} \sum_{i=1}^{p} w_i \sum_{j \neq i} \rho_{ij}\, w_j \tag{7}$$

where CR—controlling redundancy, with $p \neq r$, w_i and w_j weight factor of the *i*th and *j*th feature, ρ_{ij}—correlation coefficient between the *i*th and *j*th features, p—total no. of features, r—selected features. If $p = $ r, there is no meaning of the selection of features.

For selecting a set of features, it considers an associated weight to maintain the selection process based on the gradient descent method. Generally, it consider $w_i \epsilon$ [0, 1]. It generates different values of w_i for different features. It considers the activation function to learn w_i by gradient descent, especially it considers the sigmoid function of activation function for proposed work. Although the activation function has been defined as definition 1, but based on that activation function, the sigmoid function can be defined accordingly which is defined in definition 2 as in Appendix A.

These functions are smooth and bounded sigmoidal functions. Further, Benjamin Gompertz introduced a smooth sigmoidal function that can be generated as follows:

$$\sigma_{\alpha,\beta}(x) = e^{-\alpha\, e^{-\beta x}} \;\; x\, \epsilon R \tag{10}$$

where α, $\beta > 0$ represent an effective translation and scaling term, respectively. Readers can follow the details from [23], where (10) is defined by Benjamin Gompertz. For nonsmooth, continuous, the sigmoidal function can be written by following:

$$\alpha_R(x) = \begin{cases} 0 \left(x < \frac{1}{2}\right) \\ x + \frac{1}{2} \left(-\frac{1}{2} \le x \le \frac{1}{2}\right) \\ 1 \left(x > \frac{1}{2}\right) \end{cases} \tag{11}$$

The above function is called a ramp function. It considers the smooth and bounded function, especially, Gompertz function for the proposed model. Thus, (7) can be rewritten as

$$CR = \frac{1}{(p-r)} \sum_{i=1}^{p} \sigma_{\alpha,\beta}(x_i) \sum_{j \ne i} \rho_{ij}\, \sigma_{\alpha,\beta}(x_j) \tag{12}$$

where $w_i = \sigma_{\alpha,\beta}(x_i)$ and $w_j = \sigma_{\alpha,\beta}(x_j)$ and p≠r

This equation is equivalent to

$$CR = \frac{1}{(p-r)} \sum_{i=1}^{p} e^{-\alpha\, e^{-\beta\, x_i}} \sum_{j \ne i} \rho_{ij}\, e^{-\alpha\, e^{-\beta\, x_j}} \text{ with } p \ne r. \tag{13}$$

For different values of x such as x_1, x_2 ....x_n, and for simplicity, x_i can be written as simple "I," it is considered again $\beta_i = \beta x_i$ for a natural number of x. Thus, the equation can be rewritten as

$$CR = \frac{1}{(p-r)} \sum_{i=1}^{p} e^{-\alpha_i\, e^{-\beta_i}} \sum_{j \ne i} \rho_{ij}\, e^{-\alpha_j\, e^{-\beta\, x_j}}, \text{ with } p \ne r. \tag{14}$$

Further, the correlation coefficient ρ_{ij} finds the relation between two features. But to find the better feature, it needs to define the rank of features, which is determined by the rank of the correlation coefficient as follows.

Let (X_i, Y_i), i = 12,....*n* be the ranks of the individuals in two characteristics A & B. Then, the rank of the correlation coefficient is given by

$$R_{cc} = 1 - \frac{6}{n(n^2-1)} \sum_i d_i^2 \;\; i = 1, 2, 3 \ldots .n \tag{15}$$

where d_i is the difference between the ranks. This formula is called Spearman's formula for the rank of the correlation coefficient. Readers may refer to any statistical mathematics book to understand this formula or equation. The complexity may occur for a few databases to find a rank correlation coefficient among features. In this case, the applied method in [22] can help to select features from the above databases. Since it aims to find a subfeature selection for a unique class, it considers the following two cases.

Case 1: To select a feature from the feature set.
Case 2: To select a subfeature for generating a unique class.

Case 1 is considered a little method from [22] to control redundancy based on reduced features whereas it is more comfortable to select subfeature for the unique class as Case-2. Both cases are explained in the following two sections separately.

B. Gradient Descent Technique for Feature Selection

The gradient descent technique is derived on objective function with variable through needed data. This technique can be considered either a linear or nonlinear dependence of the dataset. Since the needed dataset is designed by the linear or nonlinear dependence of features data, it can be created the objective function by dependence. But the needed framework is designed through the correlation coefficient to control the above dependence. It is performed by the correlation coefficient ρ_{ij} between the *i*th and *j*th feature. Mutual information has considered measuring the dependence between the *i*th and *j*th features. Now the objective function (E) can be generated for controlling redundancy and also the approximate number of selected features based on SE as

$$E = \text{SE} + w * \text{CR}. \tag{16}$$

Here *E* depends on SE, nonnegative weight w, and CR value is in [0, 1]. Equation (16) will be optimized by the variation of the parameter (w) value. The role of α & β is important to determine CR and *E*. We can use random initialization for β in [0, 1] & α can be considered as a real constant value.

The several values of β help to select appropriate useful features depending on their corresponding interval such as [0, 1], [−3,+3], etc. sometimes, the recommended interval or random interval may consider for the selection procedure to predict useful features. It can repeat the experiment several times based on different parameters that are used in (14) and (16), etc. Accordingly, different performances will be generated that are described in the experiment section.

C. Learning Process-Based Derivative for Feature Selection

The learning starts with zero selection of features (i.e., the selected feature is zero). The initial value of β is considered in such a way that $e^{-\alpha\, e^{-\beta}}$ is nearly zero. For a fixed value of β, the evaluation of proposed work can be done, but the randomness of β value deals with correlated features. In some cases, the strongly correlated features generate ambiguity situations such as either one or both features maintain an important role in the feature selection procedure. We cannot discard either one or both from it. Thus, it needs the randomness initialized of β, the learning process is generated by gradient descent technique as per the following equation:

$$\beta_k^{new} = \beta_k^{old} - \rho \frac{D_1}{\sqrt{\sum_{i=1}^{p} D_{1i}^2}} \quad \forall k = 1, 2, \ldots.p \tag{17}$$

where $D_1 = \frac{\partial E}{\partial \beta^{old}}$ and

$$\alpha_k^{new} = \alpha_k^{old} - \rho \frac{D_2}{\sqrt{\sum_{i=1}^{p} D_{2i}^2}} \quad \forall k = 1, 2, \ldots.p \tag{18}$$

where $D_2 = \frac{\partial E}{\partial \alpha^{old}}$ and β_k^{new}, β_k^{old} , α_k^{new}, α_k^{new} are new and old α and β values of the *k*th feature, respectively, $\rho > 0$ is the learning rate. Since E contains two variables (α_k and β_k) for derivation, the second-order partial derivation is considered concerning α_k and β_k. The corresponding derivatives are derived as follows:

$$\frac{\partial^2 E}{\partial \alpha_k \, \partial \beta_k} = \frac{\partial^2 SE}{\partial \alpha_k \, \partial \beta_k} + \frac{w \partial^2 CR}{\partial \alpha_k \, \partial \beta_k} \; \forall k = 1, 2, 3 \ldots.p \tag{19}$$

where

$$\frac{\partial^2 SE}{\partial \alpha_k \, \partial \beta_k} = \frac{e^{-\alpha \, e^{-\beta}} + 2\beta(e^{-\beta} + 2\beta + 1)}{\sum_{i=1}^{n} \sum_{j>1} d_{ij}^0} \sum_{i=1}^{n} \sum_{j>i} \frac{(d_{ij}^0 - d_{ij})\,(x_{ik} - x_{jk})^2}{d_{ij}^0 \, d_{ij}^r} \tag{20}$$

$$\frac{\partial^2 CR}{\partial \alpha_k \, \partial \beta_k} = e^{-2\beta_j - \alpha_j \, e^{-\beta_j}}. \tag{21}$$

The objective function is defined in (16) is maintained by the effectiveness of several gradients based on their corresponding positions. Especially the influence of feature selection is differentiated by different β values. The selection of features can be controlled and managed by adjustment of α and β values iterativewise. The process of the feature selection procedure is stopped based on the changes of α, β values with no effective significant result, i.e., the difference between old and new of α, β values is very less value of η and Γ ($\eta > 0$ and $\Gamma > 0$ are two threshold values that are defined before implementation) when the selection process stops, it determines the number of values of selected features from the procedure based on α and β in $e^{-\alpha \, e^{-\beta}} > \delta$ (where δ is predefined value). In some cases, the δ can be considered with certain predefined range values. Its importance is determined during evaluation. Based on the above concept, the algorithm 1 is developed as follows.

In algorithm 1, δ values are considered with a certain real constant for selecting the required features experimentally. The δ values are tested during the experiment. Because the values of each feature in the dataset vary as compared to another same name dataset. The above method is considered for feature selection, but the role of subfeature is important. Thus, the next section is mentioned for determining subfeature data for classification.

Algorithm 1: Gradient Descent Based Feature Selection.

Input: Dataset X_{nxp} with n data samples and p dimensions non-negative weight w, learning rate ρ, two threshold values η, δ, Γ.
Output: Get selected features.
Initially determine data samples n and p dimensions of X_{nxp}.
  Evaluate dataset using Pearson's Coefficient of correlation between each pair of features
    Compute d_{ij}^0 using (4)
    Initialize α and β values randomly between predefined certain ranges individually.
      Evaluate d_{ij}^r using (5)
    Evaluate the error function E based on the constants w and (α, β) values using (16)
Compute new values of (α_k and β_k) using (17) to (21)
  Recomputed new E based on new d_{nxp}, new (α_k and β_k)
If ($\| \beta^{new}$ - $\beta^{old} \| > \eta$ and $\| \alpha^{new}$ - $\alpha^{old} \| > \Gamma$) then
  $\beta^{old} \leftarrow \beta^{new}$, $\alpha^{old} \leftarrow \alpha^{new}$, and $d_{n\times p}^{old} \leftarrow d_{n\times p}^{new}$.
    Prev E $\leftarrow$ New E
  End
    Else $\rho \leftarrow$ c x ρ (where c is real constant)
End
If ($e^{-\alpha \, e^{-\beta i}} > \delta$) then i[th] feature will be selected $\forall$ i = 12, …..p
End

V. Fuzzy Model for SubFeature Data

Several variables can consider for determining the usefulness and goodness of both selected features and subfeatures. Now it discusses the subfeature selection based on random variables. Here randomness of subfeature data occurs because nobody can say where the exactness of data at a particular place is. It depends on location, environment, opinion, etc., from any given individual. It may or may not occurr in a particular situation. Once it occurred, there is uncertainty about the exactness of data. For this, it considers the uncertainty data that can be determined by fuzzy logic (i.e., it will be represented by a set of logical data). It considers random variables to find out subfeature data from a huge database by fuzzy values only, not real numbers since fuzzy values are vaguely defined and assume with a degree of acceptability. These are considered truth values and are handled by the rules of fuzzy logic. After feature selection, the ranking of features is considered by the ranker's method. The ranking of features is assumed by three categories that recognize the importance of features for classification. The category of the ranking of features may be characterized as fuzzy values for classification, which is summarized in Table I.

Table I consists of two columns that describe set fuzzy values and classification. These are assumed by the generating classification. Generally, each dataset consists of an individual class, which should be categorized by Table I. The fuzzy values may be varied but categories of classification can filter the classes for further investigation. Here three categories of classifications are considered for any huge dataset. Novel or distinguished class

TABLE I
FUZZY SET CLASSIFICATION

fuzzy values	Classification
0.2	Novel/Distinguished class
0.7	Traditional Class
0.1	No Class/meaningless class

always affects society or database, which is difficult to find out. People are habituated with a traditional class, where anybody can easily identify the required class. The last row of Table I is no class or meaningless class means it is not required at all for further investigation because it is a useless class for the experiment.

A. Fuzzy Subfeature Set

In this section, fuzzy logic is considered with multivalued logic and binary operation for the association of feature set based on subfeature data. It helps to generate a distinguished class from the existing class. It is considered binary operations {∧(‘and’) and ∨ (‘or’)} to connect the usual properties of fuzzy logic. Let Γ is a subset of features and α is the single feature of F. Then, the statement can be generated $\wedge a_\alpha$ or $\vee a_\beta$, where $\alpha \in \Gamma$ and $a_\alpha \in$ F (F is the set of features). The truth valued function T is defined on F as T: F → [0, 1]. Thus, T(a) can receive one of two values 0 (false) or 1 (true).

Let functions R and S are considered with mapping [01] x [01] into [01] such that

T (a∧b) = R (T(a), T(b))

T (a∨b) = S (T(a), T(b)) ∀ a ϵ F, b ϵ F.

Again R(x,y) = min(x,y) and S(x,y) = max(x,y), x ϵ [01] and y ϵ [01].

For complete distributive lattice and the subset of statements $a_\alpha \in$ F, $\alpha \in \Gamma$, it follows:

$$T\left(\wedge_{\alpha\epsilon\Gamma}a_\alpha\right) = \in f_{\alpha\epsilon\Gamma}T\left(a_\alpha\right) \text{ and } T\left(\vee_{\alpha\epsilon\Gamma}a_\alpha\right) = \sup_{\alpha\epsilon\Gamma}T\left(a_\alpha\right). \quad (22)$$

B. Functional Mapping on the Feature Set

Let a set G = (A, μ, g) be defined by the required element of set G as

1) A is the ordinary set, called the basic space.
2) μ is a membership function, mapping as μ: X →[01].
3) g is mapping as g: A→F, where F is a set of features and is also called the universe of discourse.

For each element a ϵ A, g(a) can be generated and the corresponding membership function μ(a) can also be generated from a ϵ A. it is considered that the truth value of g(a) is a membership functional value, i.e.,

$$\mu\left(a\right) = T\left(g\left(a\right)\right). \quad (23)$$

Thus, the fuzzy logic is assigned or considered based on the truth value of each member of the feature set and ordinary set. Further, let B be another ordinary set and γ be a function mapping A → B. Let the elements a ϵ A and b ϵ B holds g(a) and γ(a) = b, where a≠ b. Here it clarifies that some elements are out of range (i.e., above or below the range of feature’s data) included in the B set. Since it aims to find a new class from the existing class then it considers both set based on fuzzy logic as [g(a)∧(γ(a) = b))] and assigned to m(b)

$$m\left(b\right) = \wedge_{a\epsilon A}\left[g\left(a\right)\vee\gamma\left(a\right) = b\right]. \quad (24)$$

[Note that g(a) is set of all feature’s value and γ(a) = b is set off out of feature’s range value]

The truth value of (24) is

$$T\left(m\left(b\right)\right) = T\left(\vee_{a\epsilon A}\left[g\left(a\right)\wedge\gamma\left(a\right) = b\right]\right) = \sup_{a\epsilon A}\min\left[T\left(g\left(a\right)\right), T\left(\gamma\left(a\right) = b\right)\right]. \quad (25)$$

The new class can be generated based on following criteria:

1) T(γ(a) = b) = 1 if γ(a) = b,
2) T(γ(a) = b) = 0 if γ(a) ≠b.

From the above two criteria, only criteria (a) are considered for the proposed work, otherwise, it cannot be determined new class from traditional class based on criteria (b). Because if criteria (b) are used, it will again consider the same features data that cannot be distinguished from others.

C. Fuzzy Expected Subfeature Set

Let the probability triple is (Ω, F, P) and U is a random variable defined on this triple. This random variable is perceived by a set of windows W_i, i ϵ J with J a finite or countable set. Thus, it can be established U(w) ϵ W_i for i ϵ J. Further, let S be the space of discrete function mapping R → [01]. The ordinary set mapping A: Ω → S by $w \xrightarrow{A} A_w$ based on ordinary random variable U. The characteristic function A_w is an element of S, where w ϵ Ω. It considers the random variable U as the original of the fuzzy random variable (FRV) through perception. Based on this point, it defines an FRV as a map ξ: Ω → F_N, where F_N is the set of fuzzy numbers. Let the set A′ of all possible originals of ξ is defined as the set of all random variables that are defined on (Ω′,F′,P′), where (Ω′,F′,P′) is more details than (Ω, F, P). Thus, U′ ϵ A′ can also be generated based on (Ω′,F′,P′). Thus, the original is accepted by the truth value as m(U′) where

$$m(U') = (U' \text{ is an original of } \xi).$$

Applying lattice,

$$m\left(U'\right) = \wedge_{w\epsilon\Omega, w\epsilon\Omega'}g_w\left(U'\left(w, w'\right)\right) \quad (26)$$

i.e., original assumes the value U′(w,w′) at the point (w,w′). Applying truth value into (26), then

$$T\left(m\left(U'\right)\right) = \inf_{w\epsilon\Omega, w\epsilon\Omega'}T(g_w\left(\left(U'\left(w, w'\right)\right)\right) = \inf_{w\epsilon\Omega, w\epsilon\Omega'}A_w\left(U'\left(w, w'\right)\right) \quad (27)$$

where A_w is the truth value of g_w.

Now the expectation of A is defined as the fuzzy number EA and its image of the fuzzy set $A = (A', A)$ is considered by the mapping

E: A $\rightarrow$ R such that $U' \xrightarrow{E} E\,U'$. Let the membership function of the fuzzy number EA is denoted as (EA), it may explicitly write EA = (R, (EA)) where

$$(\mathrm{EA})\,(a) = \sup_{U'\epsilon A':EU=a} \inf_{w\epsilon\Omega, w\epsilon\Omega'} A_w\,(U'\,(w, w'))\,, a\epsilon R. \quad (28)$$

Although (28) is very difficult to find subfeature data for the novel class through a fuzzy random variable, but it has tried to find such data through experimentally using the above difficult techniques for the database. Here two sample spaces are considered for subfeature data such as (a) within range for Ω and (b) out of range for Ω'. When the subfeature data evaluated for classification, some data may or may not be involved in within range or out of range as per instance. The data either within range or out of range are easily identified to a specific class, but some data of different features involved in both ranges are very difficult to identify the distinguished class. In such a situation, (28) is very much helpful to identify such a new class. In $\mathrm{inf}_{\mathrm{w}\epsilon\Omega,\mathrm{w}\epsilon\Omega'}\mathrm{A_w}(\mathrm{U'(w,w')})$, the random subfeature data can be determined which are involved in both range (i.e., within and also out of range) instancewise. Supreme of such subfeature data can be determined in (28). After getting subfeatures from each dataset, the subfeatures data are to test for generating a new class using the subfeature selection process. The selection process is considered based on a simple probability of subfeature data. Again the selection of subfeatures is determined by two ways (a) association of subfeature values of different features from dimensional reduction dataset and (b) the percentage of availability of subfeature values w.r.t. features from dimensional reduction dataset. The detailed evaluation and experimental analysis of the proposed model based on the different datasets are mentioned in the next section for better understanding.

VI. Experiments and Analysis

In this section, the different datasets have been taken for experiments as per the proposed model and analyze accordingly. Three different kinds of datasets such as Audit dataset, Alcohol QCM Sensor Dataset, and Parkinson dataset with replicated acoustic features dataset from UCI machine learning repository are considered to test the model [34]. Readers can refer to the above dataset for more information from the UCI dataset. The details of the above three datasets are as follows:

1) Audit dataset: This dataset helps the auditors by building a classification model that can predict the fraudulent firm based on the present and historical risk factors. The information about the sectors and the counts of firms are listed respectively as irrigation, public health, buildings and roads, forest, corporate, animal husbandry, communication, electrical, land, science and technology, tourism, fisheries, industries, agriculture. The description of the feature information is as follows: Many risk factors are examined from various areas like past records of audit office, audit-paras, environmental conditions reports, firm reputation summary, on-going issues report, profit-value records, loss-value records, follow-up reports, etc. After auditing, important risk factors are evaluated and their probability of existence is calculated from the present and past records.
2) Alcohol QCM Sensor Dataset Dataset: The gas sample is passed through the sensor in five different concentrations. These concentrations are (a) Concentration Air ratio (ml) (b) Gas ratio (ml). There are two channels in the sensor. One of these circles forms channel 1, and the other forms channel 2. MIP and MP ratios are used in the QCM sensors. In the dataset, there are five types of the dataset such as QCM3, QCM6, QCM7, QCM10, and QCM12. It considers only four QCM sensors such as QCM3, QCM6, QCM7, and QCM10. In each dataset, there is alcohol classification of five types, 1-octanol, 1-propanol, 2-butanol, 2-propanol, and 1-isobutanol.
3) Acoustic features: In this dataset, there are 240 instances but for only 80 subjects, so they are not independent. The nature of data is dependent for each subject but independent from one to another subject. So, a traditional technique from machine learning cannot be applied to this dataset, because those techniques are based on the independent nature of the instances. The concept of replication considered here does not match the classical concept of statistical repeated measurements. The term "replications" refers to the collection of features extracted from voice recordings belonging to the same subject. Since, in this context, features are extracted from multiple consecutive voice recordings from the same subject, in principle, the features should be identical. The biological variability result in nonidentical replicated features is more similar to one another than features from different subjects. The status of this dataset is either healthy or PD. It considers three kinds of features as pitch local perturbation measures, amplitude perturbation measures, and harmonic-to-noise ratio measures.

Although different theoretical methods have been explained in this article, the learning process-based derivative and fuzzy logic-based data are very important for feature selection as well as subfeature selection. It experiments with the proposed model in two ways (a) feature selection and (b) subfeature selection, which are explained below.

A. Feature Selection

Initially, the given dataset is taken into account for different experiments by proposed methods. The correlation coefficient among features is determined from each dataset. Several classifiers with correlation coefficient values are given in Table II. Many classifiers and their correlation coefficient can be taken for experiments, but here it considered only five classifiers with correlation coefficient values for needed experiments.

The correlation coefficient values have been generated by the abovementioned classifier from a healthy and PD dataset. Here both healthy and PD dataset is the status of Parkinson dataset with replicated acoustic features dataset. As per the proposed

TABLE II
CLASSIFIER WITH CORRELATION COEFFICIENT VALUES

S.No.	Classifier	Correlation coefficient values of Healthy dataset	Correlation coefficient values of PD dataset
1	Gaussian Processes	0.5723	0.8307
2	Linear Regression	0.6729	0.8591
3	Multilayer Perceptron	0.4732	0.7312
4	Random tree	0.3578	0.7183
5	Decision Tree	0.415	0.7934

TABLE III
CLASSIFIER WITH CORRELATION COEFFICIENT VALUES OF SELECTED FEATURES

S.No.	Classifier	Correlation coefficient values of selected features of the Healthy dataset	Correlation coefficient values of selected features of PD dataset
1	Gaussian Processes	0.7834	0.6554
2	Linear Regression Model	0.9989	0.9985
3	Multipeceptron	0.9974	0.9948
4	Random tree	0.8661	0.9608
5	Decision Table	0.9355	0.9457

model, the features are selected based on the correlation coefficient (CC). The CC is strong in selected features from the dataset that are used for further experiments like subfeature selection. The CC of selected features is as shown in Table III from the healthy and PD dataset.

The selected features of a healthy dataset are effective to each other as shown in the following figures. In Fig. 1, healthy dataset are taken into account, where the features are very close to each other in (a), (b), and (c), especially in Fig. 1(a), the features in pitch local perturbation measures are very close to each other excluding the feature "relative Jitter." Similarly, in Fig. 1(b) the feature "5-point amplitude perturbation quotient (Shim_APQ5)" is different from other features. The mapping from instances versus features; the result is satisfied with the closeness of features excluding few features, but the features are much closed when mapping is generated among them as shown in Fig. 1(d). Similarly, it has generated the approximate result by PD dataset to compare to a healthy dataset as shown in Fig. 2. As per the proposed model, the correlation coefficient is very effective in both datasets. Since the features from the Healthy and PD dataset are very close to each other as per the correlation coefficient method, these features are selected for classification.

Further, the QCM dataset has taken as per similar approaches, but few features data are discarded from others as per alcohol classification. Here five types of alcohol classification have been taken into consideration for experiments based on four types of QCM datasets. Particularly, the class "isobutanol" consists of the supreme features data, which are separated from other features data. In Fig. 3(a), (b), and (d) are approximately in similar state values, whereas in (c) the QCM7 dataset consists of distinguished features value from others.

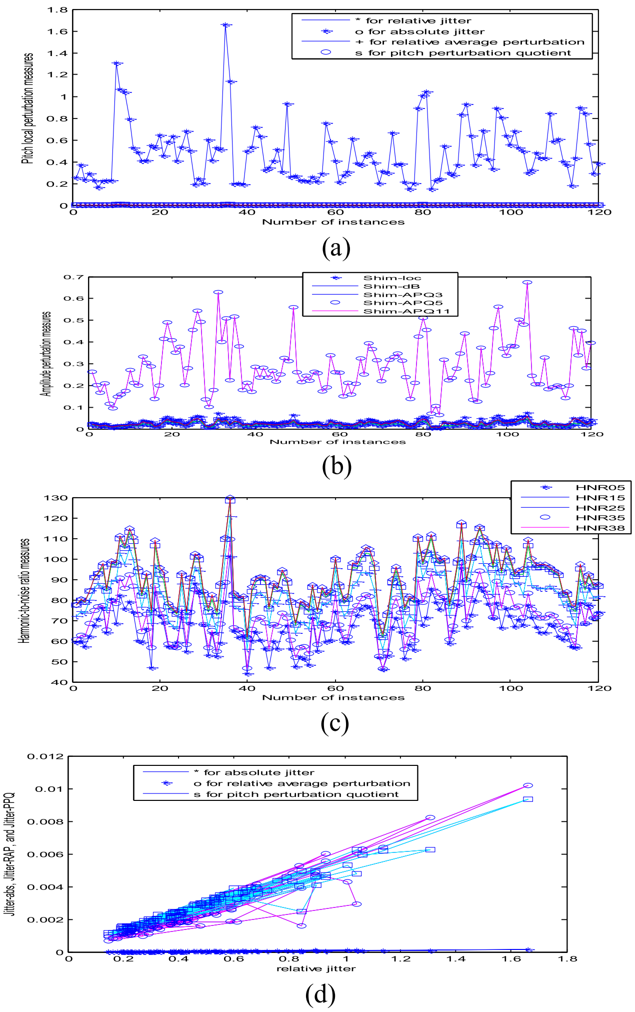


Fig. 1. Healthy dataset on (a) number of instances versus pitch local perturbation measures, (b) number of instances versus amplitude perturbation measures, (c) number of instances versus harmonic-to-noise ratio measures, and (d) relative jitter versus other jitters.

Thus, the concentration air ratio (ml) and gas ratio (ml) should be measured by sensors carefully; otherwise, it will be harmful to the environment. Since the correlation coefficient of the above features is very strong, for which these features are selected for consideration.

Equation (14) helps to find several features from a dataset. It controls the redundancy among features. Since few features have irrelevant data, it is discarded such features by techniques [1/(p-r)] from (14). Although the ranking of features is considered, it has given the importance of features for selection. It was a very difficult task to evaluate (16). The role of α_k and β_k is very important. The first derivative is not sufficient to find the needed result. So it has taken the second derivative to find the needed solution. When old and new values of α and β are determined, then the objective function value will be updated by the new E value assigned to the previous value E.

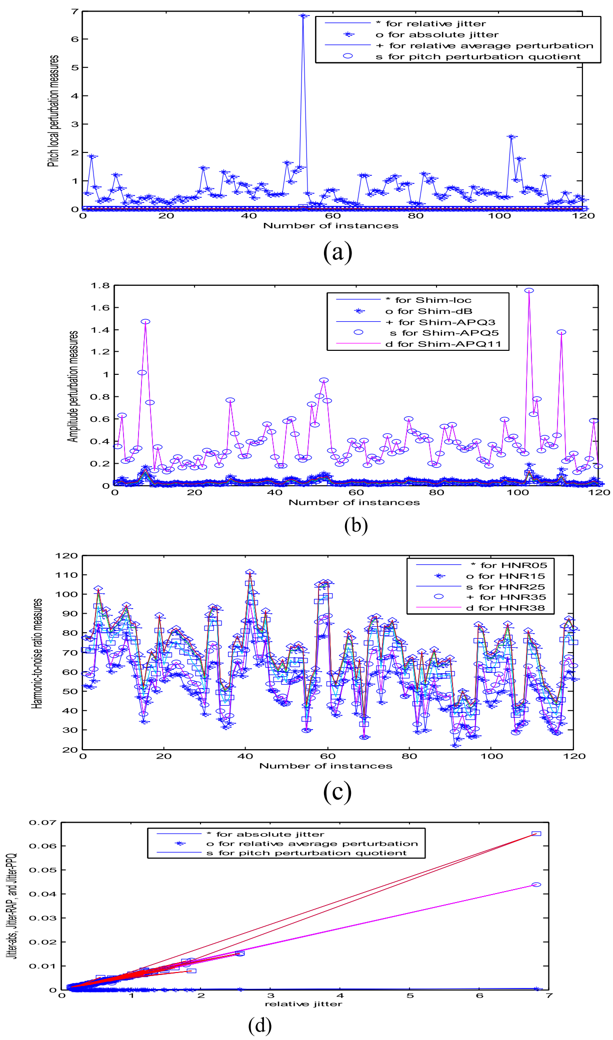


Fig. 2. PD dataset on (a) number of instances versus pitch local perturbation measures, (b) number of instances versus amplitude perturbation measures, (c) number of instances versus harmonic-to-noise ratio measures, and (d) relative Jitter versus other Jitters.

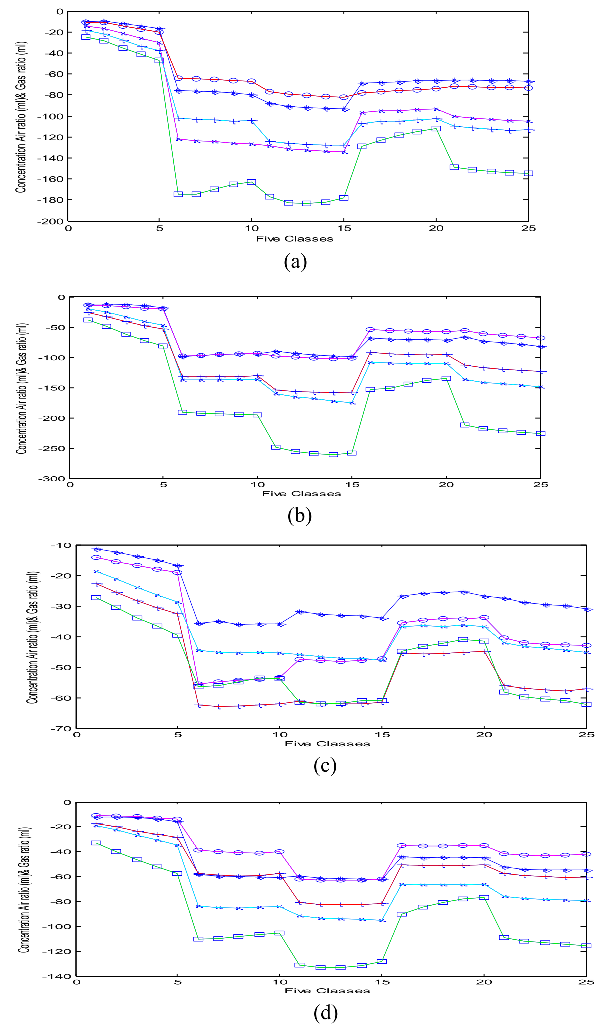


Fig. 3. QCM dataset on four types such as (a) QCM3 dataset, (b) QCM6 dataset, (c) QCM7 dataset, and (d) QCM10 dataset.

Here, correlation coefficient ρ_{ij} is considered to determine the relationship between two features i and j. Since it is needed to select the number of features from the dataset, it is considered a correlation coefficient ρ_{ij} through different classifiers. Five classifiers are taken into account for experiment with evaluation correlation coefficient, which helps to select the features from the dataset. When the proposed model is applied for experiments on different dataset, then the dataset divides into a sub dataset based on a similar feature community of data. For example: from the acoustic dataset, two datasets are generated such as (a) healthy dataset and (b) PD dataset. Further from both the above dataset, similar feature community datasets have been taken for experiments. For example, the number of similar feature communities under Pitch local perturbation measures is one set. Similarly, the same feature community under amplitude perturbation measures in another feature set, etc. Accordingly, the results are shown in Figs. 1–4 with corresponding datasets. Again audit dataset are used in the proposed model. In this data, it is very easy to select the needed features by threshold values. It has taken eight risk factors for experiments for corresponding instances. In Fig. 4(c), (e), and (f) are easily selected and filtered the features from the audit dataset based on resultant performance, but it is less effective due to similar data. Another feature set is determined by the proposed model. Since the above features have a strong correlation coefficient, these features are taken into consideration.

The importance of correlation coefficient and gradient descent methods is very effective for feature selection. These methods help to select the above features from the given dataset. A detailed description of these methods for the experiment is not mentioned in this article to avoid the length of this article. Figs. 4 to 13 are mentioned in Appendix C and Table IV in Appendix B as supplementary data.

B. Subfeature Selection

This part, it considers the subfeature data from each feature for experiments toward new classification. The subfeature data are generated by the second part of the proposed model using a fuzzy logic model. This part is a very typical part to select subfeature data. Getting subfeature data is generated by choosing the threshold value of each feature data from the dataset using a learning supervised process. It will vary from one dataset to another. After getting subfeature values, such values are used for selecting the process of subfeature based on probability and associating among subfeatures. The selected features and subfeatures from each dataset are shown in the following Table IV based on their strong correlation coefficient.

Different dataset have several features under the main feature set as per feature values. For example: in acoustic dataset, a particular feature has some subset features like the main feature "Pitch local perturbation measures" has subset features such as relative jitter (Jitter_rel), absolute jitter (Jitter_abs), relative average perturbation (Jitter_RAP), and pitch perturbation quotient (Jitter_PPQ). Thus, four features are under the main feature "Pitch local perturbation measures." Similarly, five features from each main feature "Amplitude perturbation measures" and "Harmonic-to-noise ratio measures" are selected. Further 14 features are selected from the acoustic dataset. The Healthy and PD dataset under the acoustic dataset contains 14 selected features from the main dataset for further experiments, i.e., subfeature selection. Similarly, ten features from the QCM dataset and eight features from the audit dataset are selected for further experiments. Based on the above description, the association of subfeature data from a subset of the feature of individual data is shown in Figs. 5–7, where association and availability of subfeatures are mentioned.

Further in this part, the subfeature data are analyzed for experiments based on a fuzzy logic model. The subfeatures prefer to test in the category of classification as per the fuzzy logic model. When the subfeature generates new classes or classes, it will be tested for the category of classification by the fuzzy logic value as Table I. It is very interesting to associate the subfeatures from the instance which are determined by (22).

The subfeature data are identified by a fuzzy logic model based on true value and false value as follows. If subfeatures are selected or determined for a new class, then it will be recognized by true value otherwise false values, i.e., not selected subfeature data are false values. Thus, (25) is determined by the true value of subfeature data, which helps to determine a new class. The subfeature data are selected through the membership function based on the association of subfeatures. It motivates such subfeatures toward the new class. Accordingly, Figs. 5–13 has been generated as per the proposed models. From each dataset and feature set, it has taken different subfeature data for further experiments. The association and percentage of availability of subfeature data are most important rather than getting subfeature data. The above two concepts are very important to create distinguish class. Otherwise, classes will fall in the category of classification under traditional or meaningless classes as per Table I. The aim is to find a new class from the existing class. This association and percentage of availability of subfeature data are considered based on the probability of these data. Accordingly, the different subfeature data are selected and fall into distinguished classes.

TABLE V
GETTING SUBEATURE DATA BASED ON THE THRESHOLD VALUE.

	Acoustic Dataset	QCM Dataset	Audit Dataset
Total number of data	240	125	776
Threshold value	20	10	20

These selected subfeatures are shown in different figures as per the corresponding dataset. For example, the percentage of availability of selected features in both the Healthy dataset as well as PD dataset from the acoustic dataset is shown in Figs. 8 and 9. Similarly, the association of subfeatures and percent of availability of subfeatures from different QCM datasets are shown in Figs. 10–13. Here, it considers four kinds of QCM datasets such as QCM3 dataset, QCM6 dataset, QCM7 dataset, and QCM10 dataset. Although these sets have the same features name, their feature values are different as per the dataset. Getting the subfeatures data by choosing the threshold value depends on the corresponding dataset. The task of association and percentage of availability of subfeatures is very important for subfeature data. Lack of these tasks is meaningless of subfeature selection. Thus, it considers the threshold value for getting subfeature data as in Table V. The subfeatures can be selected from the dimensional reduction database where choose subfeature data are available.

From Table V, it can consider a manual threshold value as per the learning process of testing. If the threshold value is small than the mentioned value then, it will lose some subfeature data. Accordingly, the association among subfeature will be less, which is meaningless of subfeature selection. In contrast, the threshold value is large, then irrelevant subfeature data will come to a picture that is not necessary as per the selection process. Thus, it cannot consider for experiments. The subfeature data are taken from a dataset based on fuzzy logic subfeature data. Equation (24) is very important to find subfeature data through a membership function since it is a membership function. But it is tried to use that membership function for needed work. The subfeature set is taken through the membership function of the fuzzy logic values. Initially, it is created the subfeature set by using a threshold value, and this set use for the fuzzy model through the membership function. The membership function is made as per association and percentage of availability of subfeature data. If subfeature data comes under the membership function, it will be selected otherwise it will be discarded from the dataset. Thus, the selected subfeature dataset are shown above different figures. Based on the association among subfeatures, select the number of a new class from the existing class and can recognize it as the preferred name. The detailed description avoids in this section due to the length of pages of paper. Finally, it concluded all parts of this article in the following section.

VII. CONCLUSION

This article investigated the important and challenging problem of subfeature selection based on the proposed model. Both feature and subfeature selection had an important role in every dataset for distinguished class. The selection of a minimum of subfeatures from maximum values is a very challenging classification task. The correlation coefficient technique by several classifiers helps to select the features from each dataset. The learning process of testing helps to generate different experimental results. The fuzzy model also helps to select subfeatures based on association and percentage of availability of subfeature data, which is very important for classification. The probability distribution and gradient descent method in a different manner can be applied to select subfeatures for classification in the future.

APPENDIX A

Definition 1: A measurable function S: R $\rightarrow$ R is called "activation function" where

$$\lim_{x\to-\infty} S(x) = a \text{ and } \lim_{x\to\infty} S(x) = b \text{ with } a \neq b. \tag{8}$$

Definition 2: A measurable function σ: R→R is called " a sigmoidal function" whenever

$$\lim_{x\to-\infty} S(x) = a \text{ and } \lim_{x\to\infty} S(x) = b. \tag{9}$$

APPENDIX B

TABLE IV
SELECTED FEATURES FROM THREE MENTIONED DATASET

Dataset name	Feature's name	Dataset name	Feature's name	Dataset name	Feature's name
Acoustic	Pitch local perturbation measures	QCM	Concentration Air ratio (ml)	Audit	Risk-A
	Amplitude perturbation measures		Gas ratio (ml)		Risk-B
	Harmonic-to-noise ratio measures				Risk-C
					Risk-D
					Risk-E
					Risk-F
					Inherent-Risk
					Audit-Risk

APPENDIX C

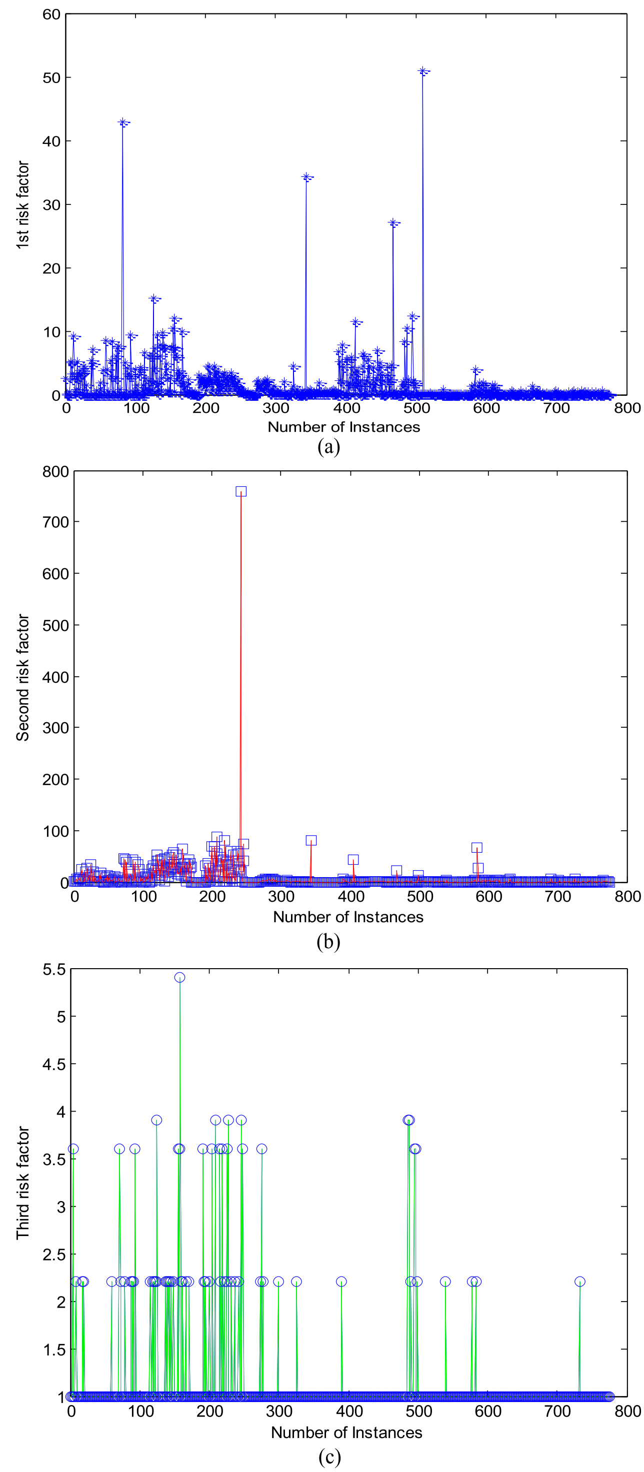
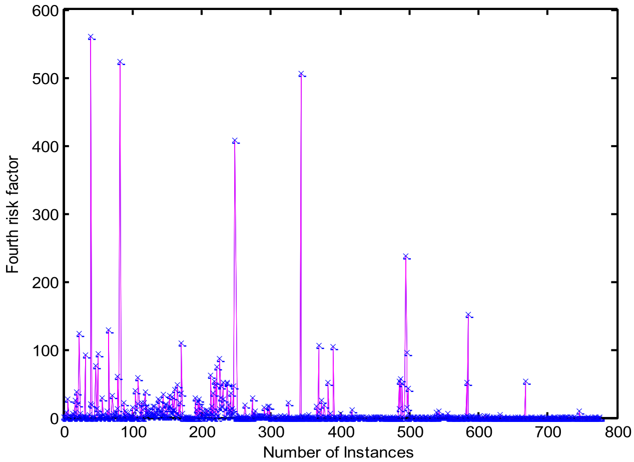
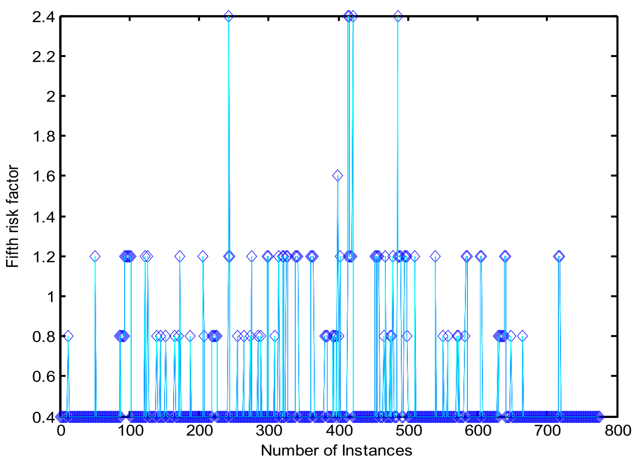


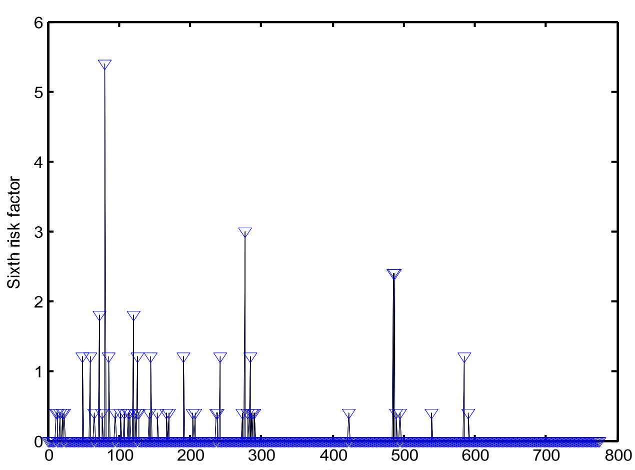
Fig. 4. Different risk factors w.r.t. instances on audit dataset as (a) number of instances versus first risk factor (b) number of instances versus second risk factor (c) number of instances versus third risk factor (d) number of instances versus fourth risk factor (e) number of instances versus fifth risk (f) number of instances versus sixth risk factor (g) number of instances versus audit risk factor (h) number of instances versus inherent risk factor and (i) number of instances versus several risk factor.



(d)

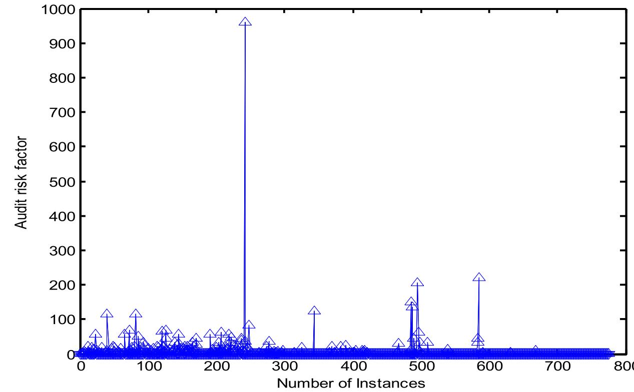


(e)

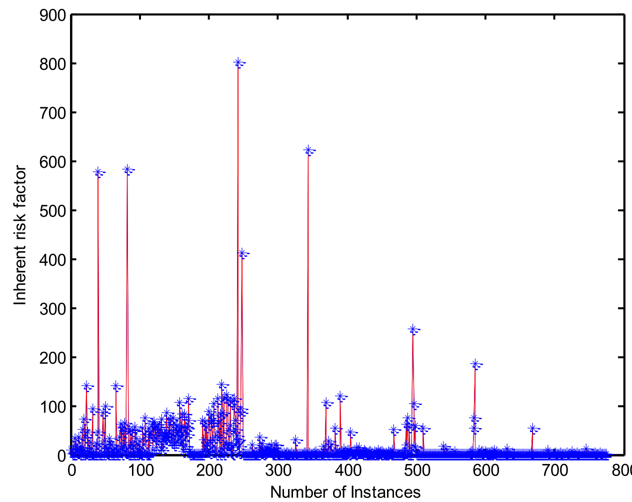


(f)

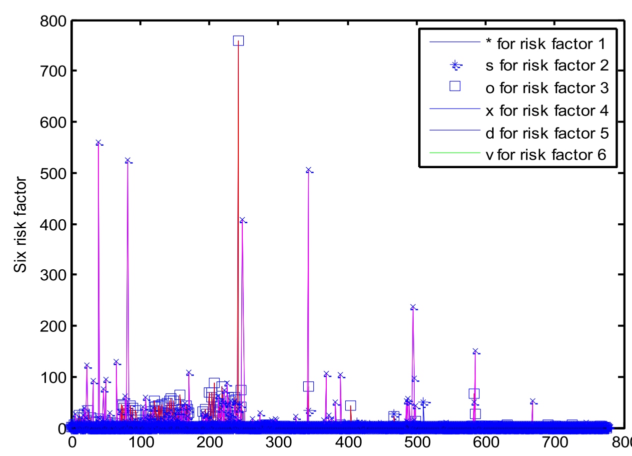
Fig. 4. *(Continued).*



(g)



(h)



(i)

Fig. 4. *(Continued).*

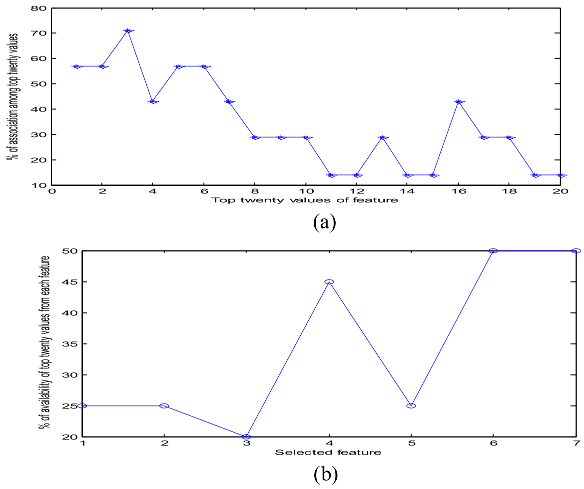


Fig. 5. Audit subfeature data on (a) percentage of association among top twenty values of features and (b) percentage of availability of top 20 values from seven selected features.

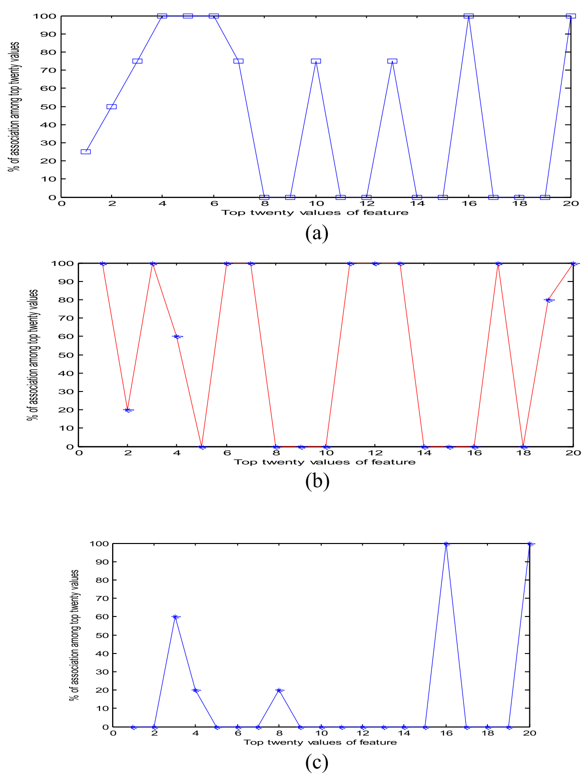


Fig. 6. Association of sub-feature values of different selected sub-feature data of health data set as (a) Pitch local perturbation measures of health data. (b) Amplitude perturbation measures of health data and (c) harmonic-to-noise ratio measures of health data.

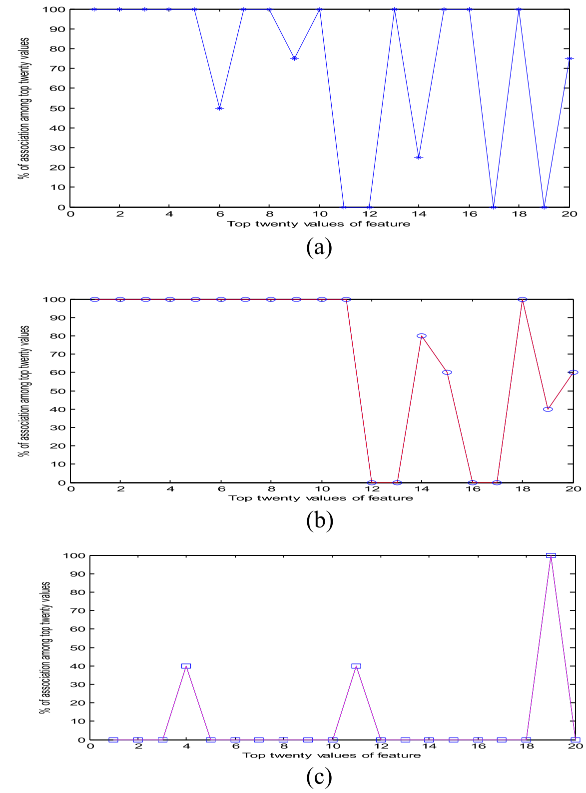


Fig. 7. Association of sub-feature values of different selected sub-feature data of PD data set as (a) Pitch local perturbation measures of PD data. (b) Amplitude perturbation measures of PD data and (c) harmonic-to-noise ratio measures of PD data.

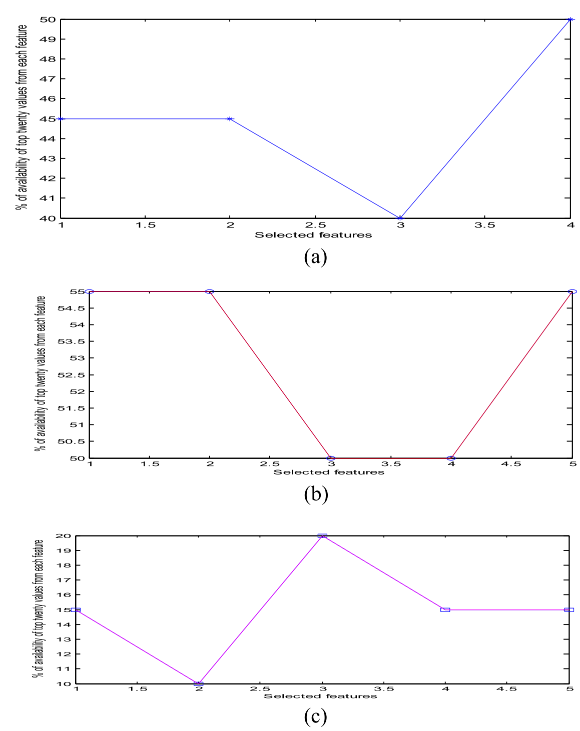
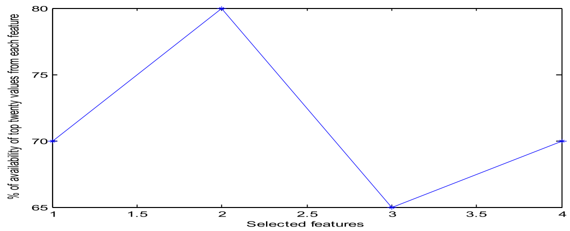
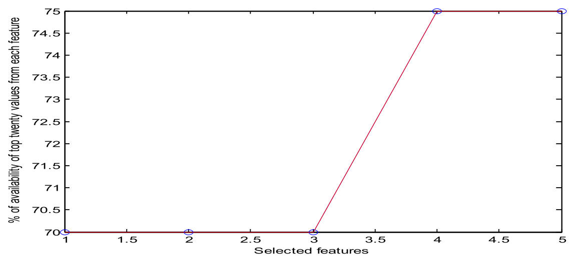


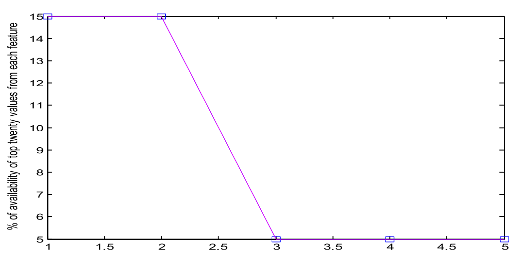
Fig. 8. (a) % of availability of top twenty values w.r.t. features of Pitch local perturbation measures of Healthy dataset (b) % of availability of top twenty values w.r.t. features of amplitude perturbation measures of healthy dataset and (c) % of availability of top twenty values w.r.t. features of harmonic-to-noise ratio measures of healthy dataset.



(a)

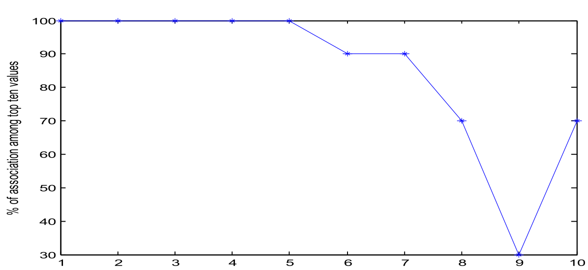


(b)

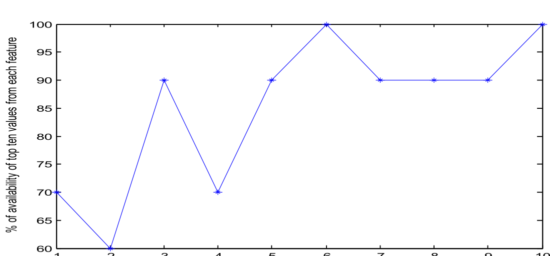


(c)

Fig. 9. (a) % of availability of top twenty values w.r.t. features of pitch local perturbation measures of PD dataset, (b) % of availability of top twenty values w.r.t. features of amplitude perturbation measures of PD dataset, and (c) % of availability of top twenty values w.r.t. features of Harmonic-to-noise ratio measures of PD dataset.

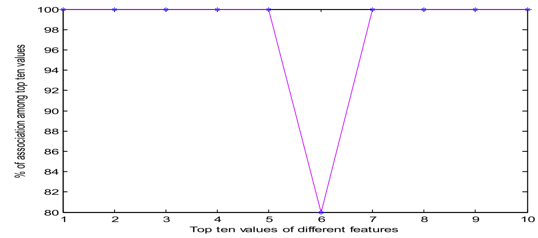


(a)

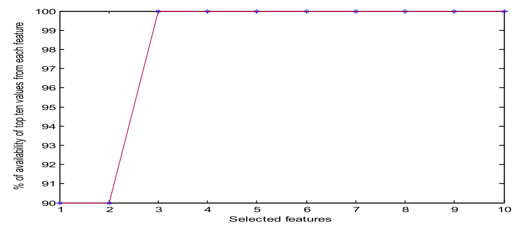


(b)

Fig. 10. (a) Association of subfeature values of different features in QCM3 data, and (b) % of availability of top ten values w.r.t. features of QCM3 dataset.

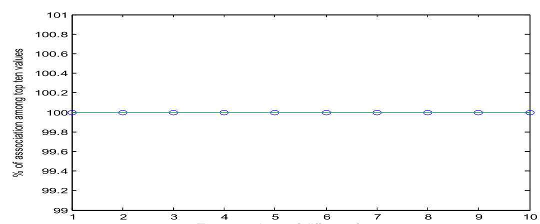


(a)

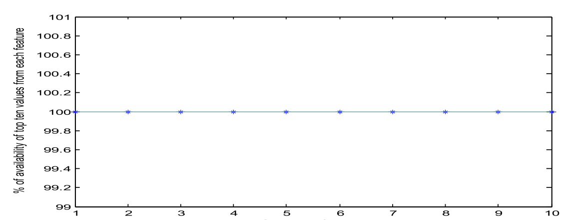


(b)

Fig. 11. (a) Association of subfeature values of different features in QCM6 data, and (b) % of availability of top ten values w.r.t. features of QCM6 dataset.

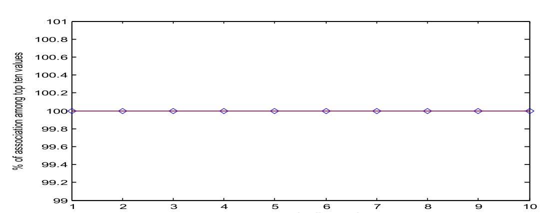


(a)

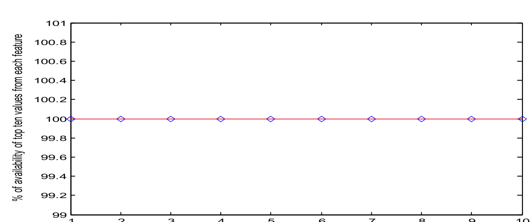


(b)

Fig. 12. (a) Association of subfeature values of different features in QCM7 data, and (b) % of availability of top ten values w.r.t. features of QCM7 dataset.



(a)



(b)

Fig. 13. (a) Association of subfeature values of different features in QCM10 data, and (b) % of availability of top ten values w.r.t. features of QCM10 dataset.

ACKNOWLEDGMENT

The author would like to thank Dr. K. V. Kishore Kumar for helping with experimental evaluation in the department of Information Technology at Vignan University. Dr. M.Y. S. Prasad was supported by Vignan University to complete this research work.

REFERENCES

[1] N. R. Pal, K. Aguan, A. Sharma, and S. Amari, "Discovering biomarkers from gene expression data for predicting cancer subgroups using neural networks and relational fuzzy clustering," *BMC Bioinf.*, vol. 8, 2007, Art. no. 5.

[2] N. R. Pal, "A fuzzy rule-based approach to identify biomarkers for diagnostic classification of cancers," in *Proc. IEEE Int. Fuzzy Syst. Conf.*, 2007, pp. 1–6.

[3] Y.-S. Tsai, C.-T. Lin, G. C. Tseng, I.-F. Chung, and N. R. Pal, "Discovery of dominant and dormant genes from expression data using a novel generalization of SNR for multi-class problems," *BMC Bioinf.*, vol. 9, 2008, Art. no. 425.

[4] L. Zhou, L. Wang, and C. Shen, "Feature selection with redundancy-constrained class separability," *IEEE Trans. Neural Netw.*, vol. 21, no. 5, pp. 853–858, May 2010.

[5] H. K. Bhuyan and N. K. Kamila, "Privacy preserving Sub-feature selection based on fuzzy probabilities," *Cluster Comput.*, vol. 17, no. 4, pp. 1383–1399, 2014.

[6] H. K. Bhuyan and N. K. Kamila, "Privacy preserving sub-feature selection in distributed data mining," *Appl. Soft Comput.*, vol. 36, pp. 552–569, 2015.

[7] Z. Li, J. Liu, Y. Yang, X. Zhou, and H. Lu, "Clustering-guided sparse structural learning for unsupervised feature selection," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 9, pp. 2138–2150, Sep. 2013.

[8] M. Banerjee and N. R. Pal, "Feature selection with SVD entropy: Some modification and extension," *Inf. Sci.*, vol. 264, pp. 118–134, 2014.

[9] P. Mitra, C. A. Murthy, and S. K. Pal, "Unsupervised feature selection using feature similarity," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 301–312, Mar. 2002.

[10] D. Cai, C. Zhang, and X. He, "Unsupervised feature selection for multi-cluster data," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2010, pp. 333–342.

[11] Z. Zhang and K. K. Parhi, "MUSE: Minimum uncertainty and sample elimination based binary feature selection," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 9, pp. 1750–1764, Sep. 2019.

[12] P. Wei, Y. Ke, and C. K. Goh, "A general domain specific feature transfer framework for hybrid domain adaptation," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 8, pp. 1440–1451, Aug. 2019.

[13] Q. Cui, K. El-Arroudi, and Y. Weng, "A feature selection method for high impedance fault detection," *IEEE Trans. Power Del.*, vol. 34, no. 3, pp. 1203–1215, Jun. 2019.

[14] L. Jiang, G. Kong, and C. Li, "Wrapper framework for test-cost-sensitive feature selection," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 3, pp. 1747–1756, Mar. 2021.

[15] Q. Wang, J. Wan, F. Nie, B. Liu, C. Yan, and X. Li, "Hierarchical feature selection for random projection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 5, pp. 1581–1586, May 2019.

[16] F. Jiménez, C. Martínez, E. Marzano, J. T. Palma, G. Sánchez, and S. Guido, "Multiobjective evolutionary feature selection for fuzzy classification," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 5, pp. 1085–1099, May 2019.

[17] B. Xue, M. Zhang, W. N. Browne, and X. Yao, "A survey on evolutionary computation approaches to feature selection," *IEEE Trans. Evol. Comput.*, vol. 20, no. 4, pp. 606–626, Aug. 2016.

[18] N. K. Kamila, L. D. Jena, and H. K. Bhuyan, "Pareto-based multi-objective optimization for classification in data mining," *Cluster Comput.*, vol. 19, no. 4, pp. 1723–1745, Dec 2016.

[19] H. K. Bhuyan and C. V. M. Reddy, "Sub-feature selection for novel classification," in *Proc. 2nd Int. Conf. Inventive Commun. Comput. Technol.*, Apr. 2018, pp. 477–482.

[20] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, vol. 46, pp. 389–422, Mar. 2002.

[21] O. Alter, P. O. Brown, and D. Botstein, "Singular value decomposition for genome-wide expression data processing and modeling," *Proc. Nat. Acad. Sci. USA*, vol. 97, no. 18, pp. 10 101–10 106, 2000.

[22] M. Banerjee and N. R. Pal, "Unsupervised feature selection with controlled redundancy (UFeSCoR)," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 12, pp. 3390–3403, Dec. 2015.

[23] D. Costarelli, "Sigmoidal functions approximation and applications," Ph.D. dissertation, Dept. Math. Comput. Sci., University of Milan, Milan, Italy, 2014.

[24] H. Liu, M. Shao, and Y. Fu, "Feature selection with unsupervised consensus guidance," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 12, pp. 2319–2331, Dec. 2019.

[25] Y. Zhang *et al.*, "Unsupervised nonnegative adaptive feature extraction for data representation," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 12, pp. 2423–2440, Dec. 2019.

[26] D. Wang, F. Nie, and H. Huang, "Feature selection via global redundancy minimization," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 10, pp. 2743–2755, Oct. 2015.

[27] M. Krishnan, C. Chakraborty, S. Banerjee, C. Chakraborty, and A. K. Ray, "Statistical analysis of mammographic features and its classification using support vector machine," *Expert Syst. Appl.*, vol. 37, pp. 470–478, 2009.

[28] M. Nilashi *et al.*, "Coronary heart disease diagnosis through self-organizing map and fuzzy support vector machine with incremental updates," *Int. J. Fuzzy Syst.*, vol. 22, pp. 1376–1388, 2020.

[29] G. Arji Hossein Ahmadi, M. Nilashi, and A. Tarik, "Fuzzy logic approach for infectious disease diagnosis: A methodical evaluation, literature and classification," *Biocybernetics Biomed. Eng.*, vol. 39, no. 4, pp. 937–955, Oct.–Dec. 2019.

[30] M. Nilashi *et al.*, "Factors influencing medical tourism adoption in Malaysia: A DEMATEL-Fuzzy TOPSIS approach," *Comput. Ind. Eng.*, vol. 137, Nov. 2019, Art. no. 106005.

[31] X. Chen, G. Yuan, F. Nie, and Z. Ming, "Semi-supervised feature selection via sparse rescaled linear square regression," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 1, pp. 165–176, Jan. 2020.

[32] S. A. Osia, A. Taheri, A. S. Shamsabadi, K. Katevas, H. Haddadi, and H. R. Rabiee, "Deep private-feature extraction," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 1, pp. 54–66, Jan. 2020.

[33] J. Xu, J. Han, F. Nie, and X. Li, "Multi-view scaling support vector machines for classification and feature selection," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 1, pp. 54–66, Jan. 2020.

[34] UCI. *Machine Learning Repository*. 2007. [Online]. Available: https://archive.ics.uci.edu/ml/datasets.php

[35] C. Chakraborty, B. Gupta, S. K. Ghosh, D. Das, and C. Chakraborty, "Telemedicine supported chronic wound tissue prediction using different classification approach," *J. Med. Syst.*, vol. 40, no. 3, pp. 1–12, 2016.