

Research Article

A Hybrid Neuro-Fuzzy and Feature Reduction Model for Classification

Himansu Das ¹, Bighnaraj Naik ², and H. S. Behera ¹

¹Department of Information Technology, Veer Surendra Sai University of Technology, Burla, Sambalpur 768018, Odisha, India

²Department of Computer Application, Veer Surendra Sai University of Technology, Burla, Sambalpur 768018, Odisha, India

Correspondence should be addressed to Himansu Das; das.himansu2007@gmail.com

Received 15 September 2019; Revised 18 November 2019; Accepted 16 January 2020; Published 1 March 2020

Academic Editor: Ibrahim Ozkan

Copyright © 2020 Himansu Das et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The evolvement of the fuzzy system has shown influential and successful in many universal approximation capabilities and applications. This paper proposes a hybrid Neuro-Fuzzy and Feature Reduction (NF-FR) model for data analysis. This proposed NF-FR model uses a feature-based class belongingness fuzzification process for all the patterns. During the fuzzification process, all the features are expanded based on the number of classes available in the dataset. It helps to deal with the uncertainty issues and assists the Artificial Neural Network- (ANN-) based model to achieve better performance. However, the complexity of the problem increases due to this expansion of input features in the fuzzification process. These expanded features may not always contribute significantly to the model. To overcome this problem, feature reduction (FR) is used to filter out the insignificant features, resulting the network less computational cost. These reduced significant features are used in the ANN-based model to classify the data. The effectiveness of this proposed model is tested and validated with ten benchmark datasets (both balanced and unbalanced) to demonstrate the performance of the proposed NF-FR model. The performance comparison of the NF-FR model with other counterparts has been carried out based on various performance measures such as classification accuracy, root means square error, precision, recall, and f-measure for quantitative analysis of the results. The obtained simulated results have been tested using the Friedman, Holm, and ANOVA tests under the null hypothesis for statistical validity and correctness proof of the results. The result analysis and statistical analysis show that the NF-FR model has achieved a considerable improvement in accuracy and is found to be efficient in eliminating redundant and noisy information.

1. Introduction

In the last few decades, machine learning [1] is a key research area among the researchers due to the dynamic generation and availability of large volume data. Converting this large volume of data into knowledge is one of the biggest challenges. There are various machine learning techniques such as classification [2–5], clustering [6], prediction [7, 8], and system control [9, 10] used for the aforesaid problems. Classification is one of the important machine learning techniques for constructing a model that classifies data into different class labels. It gives the detailed knowledge [11] of the domain that are being classified. To address the real-world problems, ANN [12] is used as a tool for the classification of different tasks such as classification, clustering,

and regression similar to the human brain. ANN is well-known for massive parallel in structure that process a large amount of data simultaneously. It has also high precision and high learning ability even in the presence of very little amount of information. ANN is already applied successfully in different problem domains such as time series prediction [13], clustering, and system control [14]. The major drawback of ANN is to handle the imprecise or uncertain data [15]. Due to the presence of imprecise and ambiguous input information, some uncertainties [16, 17] may arise at any stage of the data classification process. Fuzzy Set (FS) is a most suitable technique to handle these uncertainty issues. The degree of membership values of each feature with respect to several class labels are determined using the FS. These membership values can easily deal with the uncertain

and imprecise information. The distinct feature of FS is that it can work efficiently even with incomplete or imprecise data as compared to other mathematical models. But this fuzzy expansion is a widely used concept in most of the complex problems in order to handle uncertainty. However, this leads to computationally heavy tasks while processing large datasets. But, on the contrary, the use of fuzzy expansion is also essential too. So, we cannot completely eliminate the fuzzification process. Rather, we may avoid the processing of less significant fuzzified values. Hence, it is essential to integrate different individual techniques to form various hybrid techniques. These hybridization techniques provide an intelligent system that performs better than individual techniques to deal with the real-world problems. The motivation behind the hybrid model is to eliminate the drawbacks of individual techniques and build hybrid models that are more efficient and transparent than individual models. The hybridization of various techniques is successfully applied in several applications such as biomedical signal processing [18], cloud optimization [19, 20], forecasting [21], and healthcare [22–24]. The mostly used hybridized model called Adaptive Neuro-Fuzzy Inference System (ANFIS) [25] is successfully applied to different problem domains such as classification, prediction, and pattern recognition. But the major drawback of this model is that it is governed by fuzzy rule sets. These rule sets take more time to train the model and makes the network more complex. It provides better performance only when the fuzzy rule sets are properly designed. Another variant of the NF model that hybridizes both the NN and FS models is described in [26, 27]. In this model, the capability of the fuzzy system is to adapt the problems in a way the human perceives along with the learning ability of NN. It is hybridized to enhance different characteristics such as adaptability, speed, and flexibility of the model. Ghosh et al. [26] used a Π -type membership function [28] for the fuzzification process that expands the input features based on the number of classes available in the class label attribute. Due to this fuzzy expansion of these input features, the complexity of the network increases and takes a longer time for training and testing the model. It increases the dimensionality [29] of the problem due to the fuzzification process and leads to a major obstruction that improvises the performance of the model [30–32]. This NF model gives a considerable improvement in accuracy than ANN, but its computational time increases. Apart from this, all the expanded features may not always contribute significantly to the network. Hence, it is essential to remove these features that are not significantly contributing much to the network. In this regard, dimensionality reduction algorithms play an important role to eliminate the features that are not significantly contributing to the network. Therefore, a feature reduction algorithm called Principal Component Analysis (PCA) [33] is used to eliminate the irrelevant features to improve the performance and reduce the computational cost of the model. The NF models have been found successful in many real-life applications [34] in science and engineering. It is suitable to process the imprecise and uncertainty nature of problems [35] by

interpreting the input features before processing. It makes the model more robust and transparent in the decision-making process. This attracted many researchers to adopt the NF model for data analysis. Motivated by the above facts, this work projects a scheme for the improvement of the NF model.

The objective of the work is to develop a hybrid model called NF-FR for data classification. In the first step, Π -type membership function is employed for fuzzification of input patterns. Then, the feature reduction algorithm is applied on fuzzified variables after the fuzzification process (postfeature reduction) of the NF model to develop neuro-fuzzy with the feature reduction (NF-FR) model. In this model, postfeature reduction has been employed on fuzzified patterns to filter out the irrelevant, redundant, and noisy features. Unlike prefeature reduction, this allows all the features to participate in the fuzzification process and then identify irrelevant features from the fuzzified patterns. This approach allows exploring potential fuzzified features from the weak feature set. The NF-FR model extracts the fuzzified information that is truly contributing to the network for speeding up the classification process by eliminating the irrelevant fuzzified features. The major observation is that the overall time required for running the algorithm decreases considerably by using the FR algorithm called PCA. Thus, the NF-FR model not only provides more accurate results but also reduces the execution time. We have compared four models such as ANN, ANN with FR using PCA (ANN-FR), NF, and NF-FR using ten benchmark datasets from the UCI machine learning repository. Each dataset is then evaluated on various performance measuring parameters such as root mean square error (RMSE), f-measure, precision, and recall from the confusion matrix.

The remainder of this paper is organized as follows: Section 2 provides the review of related works. Section 3 describes the basic preliminaries of NF model, PCA, and the detailed proposed NF-FR model of this research work. Section 4 presents the detailed experimental setup along with simulating environment and result analysis of the proposed model. Section 5 describes the statistical analysis of all the models, and finally, Section 6 concludes the work with future scope of the article.

2. Literature Survey

Several soft computing techniques such as FS, NN, NF and dimensionality reduction play a critical role in the development of hybrid models in the last few last decades. The hybridization of these techniques is also considered to be one of the benchmark works in the field of data mining, machine learning, and pattern recognition. This literature review indicates the most recent development of aforesaid models and their applications in various fields. FS proposed by Zadeh represents the way of representation of human perception specially in diverse fields such as language communication, pattern recognition, and information abstraction that solves uncertainty issues. These uncertainty problems can be resolved through different fuzzification techniques that are used to convert the input features into

its corresponding fuzzified feature sets. This fuzzification process can be represented in two ways such as class belongingness fuzzification and class nonbelongingness fuzzification. Ghosh et al. [27] proposed a NF classification model in which features are fuzzified based on the bell-shaped membership function. The fuzzified matrix formed from the input features was associated with a degree of belongingness to the different classes. The class labels determined the value of the degree of belongingness towards that class. Pal and Mitra [28] used a membership function that converts the crisp values into its linguistic values, and these linguistic values are used as input patterns to the network instead of numeric values. They have used Π -type membership functions for fuzzification and ANN-based MLP with backpropagation model. But they do not consider any prone or addition to the network structure. Meher [34] proposed NF classification using the rough set approach that utilizes the best possible extracted features. This is obtained through feature-wise belongingness of patterns using fuzzy set to deal with impreciseness and rough set for uncertainty. Kar et al. [35] provided a recent survey on the NF classification model development during the period of 2002 to 2012 in different application fields such as traffic control, economic system, medical system, and image processing. Viharos and Kis [36] conducted a detailed survey on different NF models such as ANFIS, FALCON, GARIC, NEFCON, and SONFIN along with their architecture. It also gives the detailed survey of the use of these models in the technical diagnostics and measurement field. The detailed survey of NF models from 2000 to 2017 for classification is described in [37]. Das and Pratihari [38] used neuro-fuzzy with multiobjective optimization techniques to inherent fuzziness in the manufacturing process. Škrjanc et al. [39] addressed a review on evolving neuro-fuzzy and fuzzy rule-based models used in real-world environments for classification, clustering, regression, and system identification. In the data analysis process, dimensionality reduction techniques such as feature selection and feature reduction are used in the preprocessing [40] stage in which the original features are transformed into either original feature or transformed features. Chattopadhyay [41] addressed a NF model for the diagnosis of human depression based on certain symptoms. This model used PCA for feature reduction from fourteen features to seven features of those that are relevant and significantly contributing to the decision-making process of disease identification. Ibrahim et al. [42] used a data-adaptive NF inference model for early detection and classification of diabetes disease based on symptoms. Alvanitopoulos et al. [43] proposed a NF classification technique for the identification of damages produced by an earthquake on construction. After the manifestation of the earthquake, the evaluation of the safety of existing structure and measure to be taken for automatic damage classification of buildings is considered. Chen [44] proposed an online NF model for the deadline constraint message scheduling system. It adapts the network structure and parameters to explore the dynamic behaviour of the message scheduling system. Azhari and Kumar [45]

addressed a NF approach for text summarization. It filters the high-quality summary sentences on the document understanding conference data corpus. Singh et al. [46] proposed an enhanced NF model used for clustering that reduces the number of linguistic variables as compared to the NF model. Nilashi et al. [47] used ensembles ANFIS model, clustering along with dimensionality reduction for prediction of hepatitis disease diagnosis. Shihabudheen et al. [48] addressed a PSO-based ELM-ANFIS model for regression and classification to reduce the computational cost, randomness, and better generalization.

PCA also plays an important role to eliminate redundant features from the input pattern which improves system performance along with accuracy. It extracts the important information from the datasets and represents it as a new set of orthogonal variables called principal components. It is a statistical method used to reduce the number of variables by collecting highly correlated variables. Polat and Güneş [49] used PCA and ANFIS techniques. They have used the feature reduction algorithm to reduce the number of input features of the diabetes dataset from eight features to four features and conducted the predictive diagnosis by passing the inputs through the ANFIS model. Wang and Paliwal [50] proposed dimensionality-based feature extraction algorithms such as linear discriminant analysis and PCA for vowel recognition. It transforms the input parameter into the feature vector and reduces its dimension to make the classification process more efficient. Azar [51] addressed a feature selection method based on linguistic hedges in the adaptive NF model for medical diagnosis. It reduces the dimensions of the problem and also enhances the performance of the classification by eliminating redundant and noisy features. It also speeds up the computational time of the learning algorithm and simplifies the classification task. Keles et al. [52] proposed a NF tool for prostate cancer classification. This model diagnosis to find a set of rules that can be interpreted linguistically. Gabrys [53] addressed a general fuzzy max-min network model for uncertain information processing in the industry. It analyses and identifies whether to combine or not combine different techniques to form hybridization. Übeyli [54] applied the ANFIS model for the classification of ECG signals. They used Lyapunov exponents for feature extraction, and ANFIS is used for classification. Kolodyazhnyi et al. [55] used PCA for dimension reduction and NF Kolmogorov's network for classification for waste water treatment plant data. Schclar et al. [56] ensemble various models based on the dimensionality reduction. Due to an increase in the dimension of input features in the NF model, the computational cost is increased. To address this issue, various feature reduction techniques are employed in the preprocessing stage. However, our present investigation of the proposed NF-FR model is justified with class belongingness fuzzification of input features. These fuzzified features are filtered out by the PCA to produce the reduced features. These reduced features are passed to the ANN-BPN based model for training and testing. This experimentation is done with ten numbers of both balanced and unbalanced datasets.

3. System Model

Last few decades, researchers are trying to design the hybrid systems by using the fuzzy system and neural network for pattern classification. The basic concepts of NF models, PCA, and the proposed hybrid model of NF-FR models are presented in the following sections.

3.1. Neuro-Fuzzy Model. In real-world problems, uncertainty is one of the major challenges which leads to incomplete and imprecise information about the input data in pattern classification problems. Therefore, it is necessary to provide ample provision to handle uncertainty. In the NF model, instead of a normal crisp input value, fuzzy values are being inputted to the neural network. The fuzzified matrix is a result of the fuzzification process that generates a membership matrix in which the total number of element present in this matrix is equal to the product of the number of features and number of classes present in the dataset, which is input to the neural network. The fuzzified input matrix is associated with the degree of belongingness with respect to the classes, which extracts the feature-wise information of the input pattern. Each feature value of a pattern represents membership values for each class, where membership values are measured by using Π -type membership functions as represented in Figure 1. This fuzzification matrix is passed to the ANN model to train the network.

3.2. Principal Component Analysis. The fuzzification process may result in high dimensional data, where all the features may not carry significant information for discrimination of the pattern. Furthermore, this increase in dimension affects the complexity of machine learning algorithms. This section describes the working principle of a feature reduction algorithm called PCA to extract the relevant features from the original feature set to reduce the dimensions of the data. This can be achieved by transforming the high dimensional features into new small transformed features without losing the essential information of the original datasets. These new sets of features are called principal components in which the data varies as the linear combination of original features. PCA considers only those components that have a larger variance of the data. The major objective of PCA is to identify the hidden patterns of the data, determine the correlation among the features, and decrease the dimensionality of the features by eliminating the redundant and noisy features.

3.3. Proposed Hybrid Neuro-Fuzzy and FR Model. In this model, feature-wise information of input patterns is extracted from the original data with respect to different classes. Since all features are not equally important in discriminating the instances, the feature-wise belonging is expected to help in the classification process. In this section, a detailed schematic diagram of the novel NF-FR model has been proposed for the classification of nonlinear data. This proposed model is moving parts into three major steps such

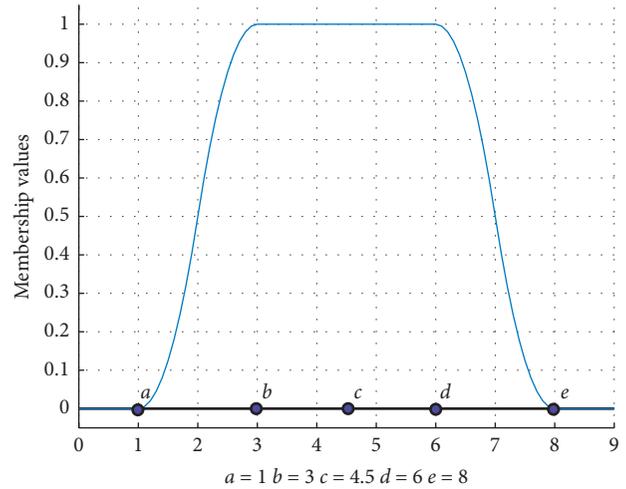


FIGURE 1: Π -type membership function.

as input feature fuzzification process, feature reduction using PCA, and classification using ANN with back-propagation learning.

Initially, this NF-FR model is used to extract the feature-wise information from the input pattern to its corresponding fuzzified matrix by using the class belongingness fuzzification technique. In this present study, we have used a popular Π -type membership function for fuzzification of the input pattern. Since all the features may not be significantly contributing to the classification process, it is essential to find the class belongingness of each attribute. In order to achieve this, here Π -type membership functions have been used for the fuzzification process which provides the degree of the belongingness of individual features with respect to the class labels. As a result, each feature value of the input patterns has been expanded to C number of values, where C is the number of class labels. Furthermore, such expansion of input patterns may include some insignificant features, and thus, PCA has been used for the pruning of irrelevant features. Finally, the ANN model has been used for the classification process with backpropagation learning, and the output of the ANN model is defuzzified to get the final result. The block diagram of the proposed NF-FR classification model is shown in Figure 2, and the detailed working model is shown in Figure 3. This proposed model has been discussed in detail below.

Step 1 (Fuzzification process). In this step, the n -dimensional input pattern $P_i = (F_{i,1}, F_{i,2}, \dots, F_{i,n})$ is considered as an input pattern, where n is the number of features available in the dataset. Here, the membership value of each feature of the dataset is computed by using Π -type membership function, which is represented in Figure 1. The membership value of the j^{th} instance of the i^{th} feature with respect to class labels C is denoted as $\mu_C(F_{j,i})$, where $F_{j,i}$ is the j^{th} instance of i^{th} feature of the dataset, and $C = 1, 2, \dots, k$ represents the number of classes available in the dataset. This fuzzification process $\mu_C(F_{j,i})$ provides the degree of membership of individual features with respect to the different class labels.

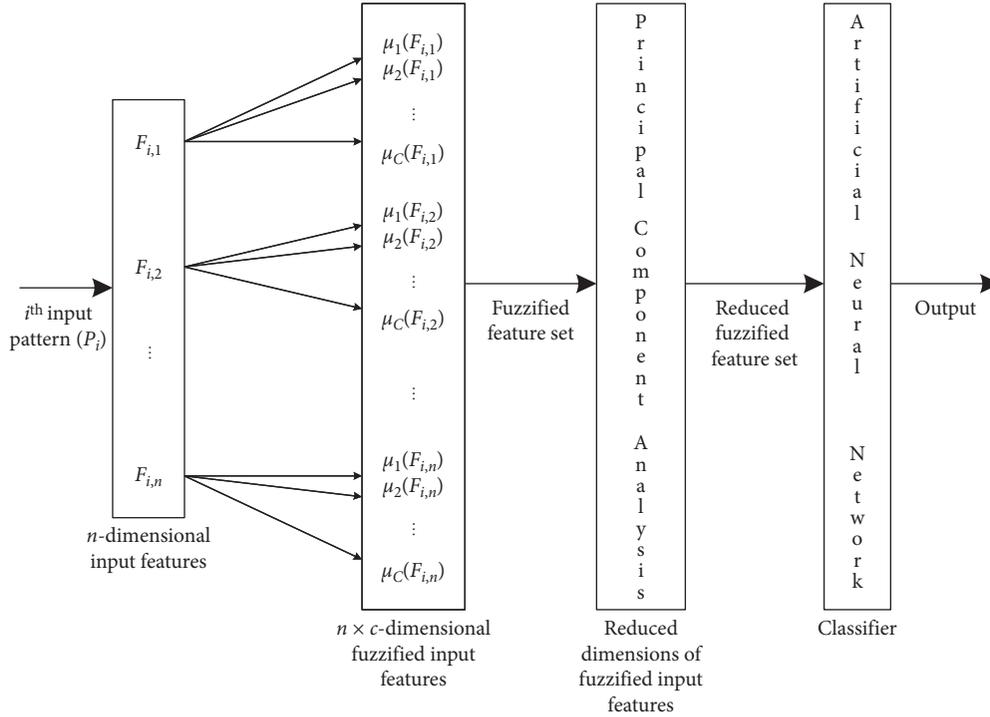


FIGURE 2: Framework of the NF-FR model.

Here, Π -type membership function has been used for fuzzification and controlling the steepness of the model, which may be realized in the following equation:

$$\mu(F_{ji}; a, b, d, e) = \begin{cases} 0, & F_{ji} \leq a, \\ 2\left(\frac{F_{ji} - a}{b - a}\right)^2, & a \leq F_{ji} \leq f, \\ 1 - 2\left(\frac{F_{ji} - b}{b - a}\right)^2, & f \leq F_{ji} \leq b, \\ 1, & b \leq F_{ji} \leq d, \\ 1 - 2\left(\frac{F_{ji} - d}{e - d}\right)^2, & d \leq F_{ji} \leq g, \\ 2\left(\frac{F_{ji} - e}{e - d}\right)^2, & g \leq F_{ji} \leq e, \\ 0, & F_{ji} \geq e. \end{cases} \quad (1)$$

In equation (1), the membership value is minimum at points a and e . Here, the membership value gradually increases from the points a and b , retains the maximum value between the points b and d , and afterward gradually decreases from the points d to e . The center c is computed as

$c = \sum_{j=1}^m F_{j,i}/m$ of the training dataset. The computation of crossover point at f and g are represented in the following equations, respectively:

$$f = c - \frac{(\text{Max}_{j \in m}(F_{j,i}) - \text{Min}_{j \in m}(F_{j,i}))}{2}, \quad (2)$$

$$g = c + \frac{(\text{Max}_{j \in m}(F_{j,i}) - \text{Min}_{j \in m}(F_{j,i}))}{2}. \quad (3)$$

In equations (2) and (3), Max and Min are two mathematical functions used to calculate the minimum and maximum value of the j^{th} instance of the i^{th} feature ($F_{j,i}$) of the dataset. The membership value of the crossover points f and g is 0.5.

After the fuzzification process, the fuzzified matrix $\mu(P)$ of the complete dataset is computed by using Π -type membership function, which is expressed in as follows:

$$\mu(P) = \begin{bmatrix} \mu(F_{1,1}) & \mu(F_{1,2}) & \cdots & \mu(F_{1,n}) \\ \mu(F_{2,1}) & \mu(F_{2,2}) & \cdots & \mu(F_{2,n}) \\ \vdots & \vdots & \ddots & \vdots \\ \mu(F_{m,1}) & \mu(F_{m,2}) & \cdots & \mu(F_{m,n}) \end{bmatrix}, \quad (4)$$

where $\mu(F_{j,i})$ represents the membership value of the j^{th} input pattern of the i^{th} feature of the dataset, and it can be represented by the following equation:

$$\mu(F_{j,i}) = [\mu_{C1}(f_{j,i}), \mu_{C2}(f_{j,i}), \dots, \mu_{Ck}(f_{j,i})], \quad (5)$$

where $\mu_C(f_{j,i})$ represents the membership value of the j^{th} input pattern of the i^{th} feature of the dataset with respect to

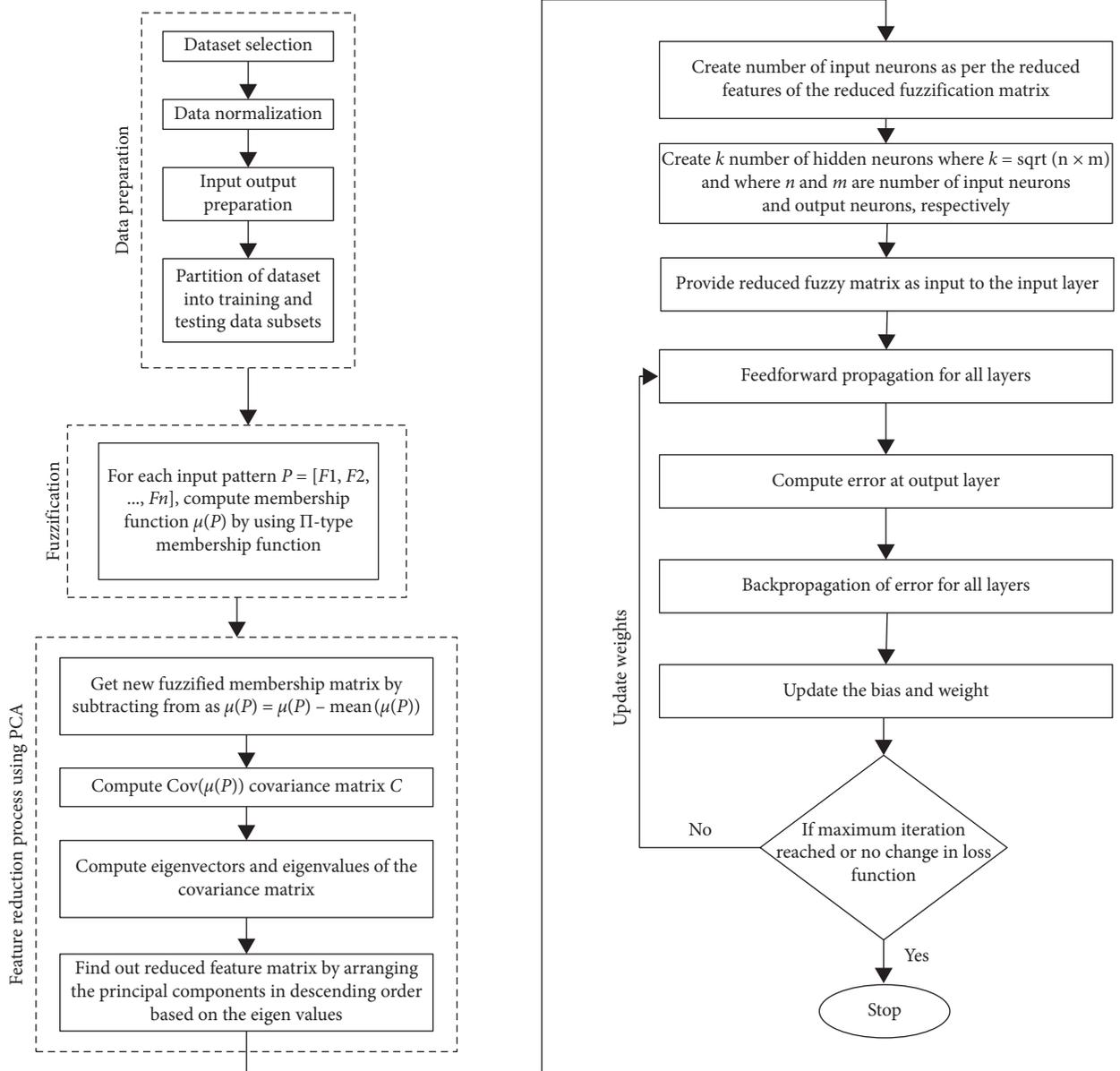


FIGURE 3: The detailed working model of proposed models.

the C class level. Here, $C = C_1, C_2, \dots, C_k$ is the number of classes available in the dataset. The output of this process is a fuzzified matrix that contains the expanded fuzzified features of the input pattern. The example of fuzzification results of one feature (petal width) of the IRIS dataset is represented in Figure 4. Each membership value of all the features of the dataset will be transformed within the range of [0-1] as shown in Figure 4.

Step 2 (Feature reduction process). Due to expansion of input feature, the complexity of the model increases. To make the classification process more effective and efficient, PCA is used to reduce the features of the fuzzified membership matrix. In this step, this fuzzified membership matrix $\mu(P)$ is used as an input to the PCA algorithm to reduce the dimensions of the features.

Let, the aforesaid fuzzified membership matrix $\mu(P)$ having q number of fuzzified features that can be expressed as $P_i = (F_{i,1}, F_{i,2}, \dots, F_{i,q})$, $q = n \times C$, n is the number of features, and C is the number of class labels. The covariance matrix of the fuzzified membership matrix is computed by using the following equation:

$$CM = Cov(\mu(P)) = \frac{1}{m-1} \sum_{i=1}^m \left(F_{j,i} - \overline{\mu(F_{j,i})} \right) \times \left(F_{j,i} - \overline{\mu(F_{j,i})} \right)^T, \quad (6)$$

where $\overline{\mu(F_{j,i})}$ is the sample mean of the feature $F_{j,i}$ and m represents the number of samples to be considered. The components of the covariance matrix ($CM_{j,i}$) represents the variances of the features F_i and F_j .

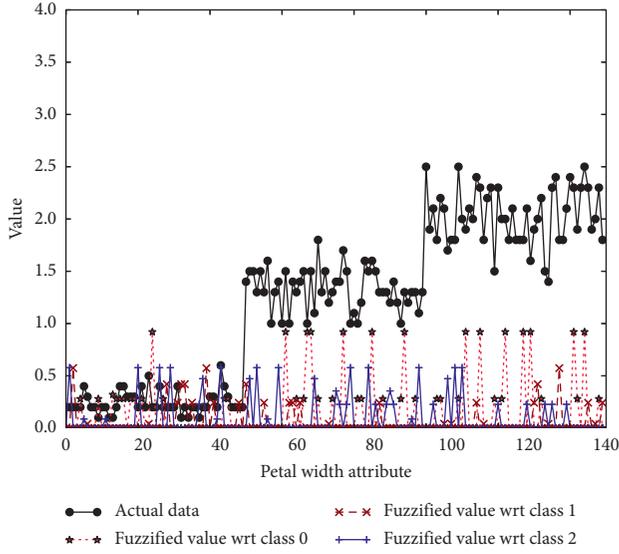


FIGURE 4: Fuzzification process of IRIS dataset of feature (petal width).

Let r be the number of principal axes A_1, A_2, \dots, A_r , which represents the eigenvectors of the covariance matrix, where $1 \leq r \leq q$, in which the variance is maximum in the projected space. The mean value of each feature of the fuzzified membership matrix is computed as $\mu(F_{j,i})$ where $i = 1, 2, \dots, q$ and $j = 1, 2, \dots, m$. The mean value of each feature is subtracted from each of the data dimensions to produce a dataset whose mean is zero.

The eigen values (α_i) and eigenvectors (A_i) of the covariance matrix are computed easily as it is a symmetric matrix. The eigenvectors A_i and its corresponding eigen values α_i are computed by using the following equation:

$$CM \times A_i = \alpha_i \times A_i, \quad (7)$$

where $i = 1, 2, \dots, r$ and r is the number of principal components that can be derived by using equation (8) and can be represented by the descending order of the eigenvalues of the corresponding eigen vectors.

$$X_i = [x_{i,1}, x_{i,2}, \dots, x_{i,r}] = A^T \times P_i. \quad (8)$$

The output of this step is the reduced matrix (X) that contains the relevant information of the input features that are necessary in decision-making process of the classification. In the third step, this reduced matrix is passed to the ANN as input.

Step 3 (Building ANN-BPN model). In this step, the artificial neural network with backpropagation (ANN-BPN) model is used for the classification process. The ANN-BPN model uses backpropagation as a supervised learning algorithm to train the artificial neural network. It updates the weights of the model to minimize the loss by efficiently computing the gradients. This network uses the reduced fuzzified matrix as input which is generated from the Step 2 to this model. In this model, all the weights of the input layer are fully connected to the hidden layers. All the

hidden layers are also fully connected to themselves. The last hidden layer is also connected to the output layer. Initially, all weights are assigned with random weights in between the range of 0 and 1. The number of nodes available in the input layer is equal to the number of features available in the reduced fuzzified matrix. The number of nodes in the output layer is equal to the number of class labels available in the dataset. The number of nodes available in the hidden layer is computed by using equation (9), in which input_nodes, hidden_nodes, and output_nodes represent the number of input nodes, hidden nodes, and output nodes, respectively. The number of hidden layers and the number of neurons in each hidden layer depends on the complexity of the problem. There is no any standard method available to compute it in the literature, but some authors use the following equation to compute the number of neurons available in the hidden layer:

$$\text{hidden_nodes} = \sqrt{\text{input_nodes} \times \text{output_nodes}}. \quad (9)$$

In the feedforward step, the model is trained based on the reduced fuzzified matrix input information. The net input is computed by making the sum of product of input patterns and the assigned weights and add the bias. Mathematically, the performance of the net input of the N^{th} neuron can be expressed in the following equation:

$$U_N = B_N + \sum_{j=1}^r W_{i,j} \times X_{i,j}, \quad (10)$$

where B_N is the bias of the N^{th} neuron, $X_i = [x_{i,1}, x_{i,2}, \dots, x_{i,r}]$ is the input pattern of the reduced fuzzified matrix, $W_i = [w_{i,1}, w_{i,2}, \dots, w_{i,r}]$ are connection weights of the N^{th} neuron, and U_N is the net input of the model. Similarly, the net input of each layer is computed, and apply the activation function for output, which is determined between different connecting layers. The output of the output layer is computed by using the sigmoid activation function described in the following equation, where ϕ is the activation functions and O_N is the output of the neuron:

$$O_N = \phi(U_N). \quad (11)$$

In the backpropagation step, the error is computed by subtracting the actual output with the target output, and the error is expressed in as follows:

$$\text{error} = \sum_{j=1}^m \text{target_output} - \text{actual_output}, \quad (12)$$

where $j = 1, 2, \dots, m$ indicates the number of output neurons. Here, the root mean square error (RMSE) [5–10] can be computed by using the following equation:

$$\text{RMSE} = \sqrt{\text{error}^2}. \quad (13)$$

Similarly, the errors are computed, and the weights and bias are updated in the learning process. The weights of the connecting path between the different layers are adjusted by

computing the change in weights in the model to reduce the overall error of the model, which is realized in equation (14), where α is the learning rate between the range of [0, 1]:

$$\Delta \text{weight} = -\alpha \times \frac{\partial \text{error}}{\partial \text{output}} \times \frac{\partial \text{output}}{\partial \text{net}} \times \frac{\partial \text{net}}{\partial \text{weight}} \quad (14)$$

The new weights and bias of the model can be computed by using the following equations, respectively:

$$\text{new_weight} = \text{old_weight} + \Delta \text{weight}, \quad (15)$$

$$\text{new_bias} = \text{old_bias} + \Delta \text{bias}. \quad (16)$$

This process is repeated multiple times to minimize the root-mean-square error of the model or till the stopping criteria is reached. This proposed approach is different from ANN (feedforward with backpropagation), ANN-FR, and NF as follows: (i) in ANN (feedforward with backpropagation), all the input features are processed in parallel without extracting insignificant features which take more time to train the model and also leads to uncertainty problem. (ii) In the ANN-FR model, PCA is used in the preprocessing stage to eliminate insignificant features, but this model is unable to address the uncertainty issues. (iii) The NF model solves uncertainty issues by the using the class belongingness fuzzification process, but it is unable to eliminate the redundant or noisy features. (iv) By considering aforesaid issues, the proposed model solves the uncertainty issue by using the class belongingness fuzzification process and also eliminates the insignificant fuzzified features by using PCA instead of eliminating the complete features which seem insignificant.

4. Result Analysis

In this section, the simulation environment and the dataset used for the training and testing phases for the analysis of the proposed model are presented. Here, four models (ANN, NF, ANN-FR, and NF-FR) are implemented using Matlab (version R2015a) with the Windows 7 operating system. The benchmark datasets from the UCI machine learning repository [57] are collected and tested in different classification models. The detail descriptions about all these datasets [57, 58] can be found at “<http://archive.ics.uci.edu/ml/>” and “<http://keel.es/>.”

Several performance comparison techniques such as classification accuracy, root mean square error (RMSE), precision, recall, and f-measure are obtained from the confusion matrix for all the benchmark datasets of each model, and the comparison result is presented. The details of these performance measures are outlined below. The comparison of RMSE of ANN, NF, ANN-FR, and NF-FR models is shown in Table 1. Error plots of four datasets (Titanic, Mammographic, Breast Cancer, and Wine) with four models (ANN, NF, ANN-FR, and NF-FR) are shown in Figure 5. There are some configuration variations in these four models as follows: ANN is a simple model in which no fuzzification and no PCA are used for analysis, the ANN-FR is a no fuzzification with PCA-based model, NF is a

TABLE 1: Comparison of RMS errors of ANN, NF, ANN-FR, and NF-FR models.

Datasets/models	ANN	NF	ANN-FR	NF-FR
Iris	0.18721	0.10258	0.15471	0.11721
Mammographic	0.30163	0.19111	0.26163	0.18031
Breast Cancer	0.39734	0.17052	0.21637	0.02546
Pima Indian	0.41819	0.29887	0.38248	0.30353
Hayes-Roth	0.32242	0.14770	0.35525	0.16222
Thyroid	0.15335	0.05315	0.14085	0.08838
Titanic	0.19373	0.28599	0.07685	0.34237
Wine	0.05249	0.03854	0.08942	0.10774
Haberman	0.25329	0.33007	0.35470	0.35749
Blood Transfusion Service	0.31176	0.31284	0.13597	0.29339

fuzzification with the non-PCA based model, and similarly, NF-FR is a fuzzification with the PCA-based model.

The results presented here are exclusively based on the experiments that are observed. Table 2 describes the comparison of classification accuracy of ANN, NF, ANN-FR, and NF-FR models for worst, average, and best cases. Every model is executed ten times with random weights, and observations are recorded. Based on the ten times execution history of all the models, the worst, average, and best case classification accuracies are obtained and shown in Table 2. Apart from classification accuracy, some additional measures such as precision, recall, and f-measure are also considered here to measure the performance of ANN, NF, ANN-FR, and NF-FR models. Table 3 describes the comparison of precision, recall, and f-measure of ANN, NF, ANN-FR, and NF-FR models.

In Tables 2 and 3, classification accuracy, precision, recall, and f-measure (worst case, average case, and best case) of four models such as ANN, NF, ANN-FR, and NF-FR are presented. There are very few results in Tables 2 and 3 indicating the result of the proposed hybrid method (NF-FR) is lower than other techniques. Generally, this happens in any machine learning model. A single machine learning technique may not be suitable for all the benchmark datasets or problems (as per the principle of “no-free-lunch theorem”). Hence, in order to draw a generalized performance measure, we have conducted various statistical tests such as ANOVA test, Tukey and Dunnett test, Friedman test, and Holm procedure. To validate the performance of the proposed model with other models, several statistical analyses such as ANOVA test, Tukey and Dunnett test, Friedman test, and Holm procedure are made in the next section. For example, in Friedman test, the average rank of four models such as ANN, ANN-FR, NF, and NF-FR is computed based on assigned rank (in Table 4), which is represented in equation (17). This average rank of four models: ANN, ANN-FR, NF, and NF-FR, can be computed and assigned as $\{R_4 = 4, R_3 = 3, R_2 = 1.9, R_1 = 1.1\}$, respectively. Based on these ranks, null-hypothesis has been tested. The results of the aforesaid statistical analysis also show that the overall performance of the proposed NF-FR model is statically significantly different and better than other models. It means that our proposed model may not be suitable for few cases but works well for many datasets.

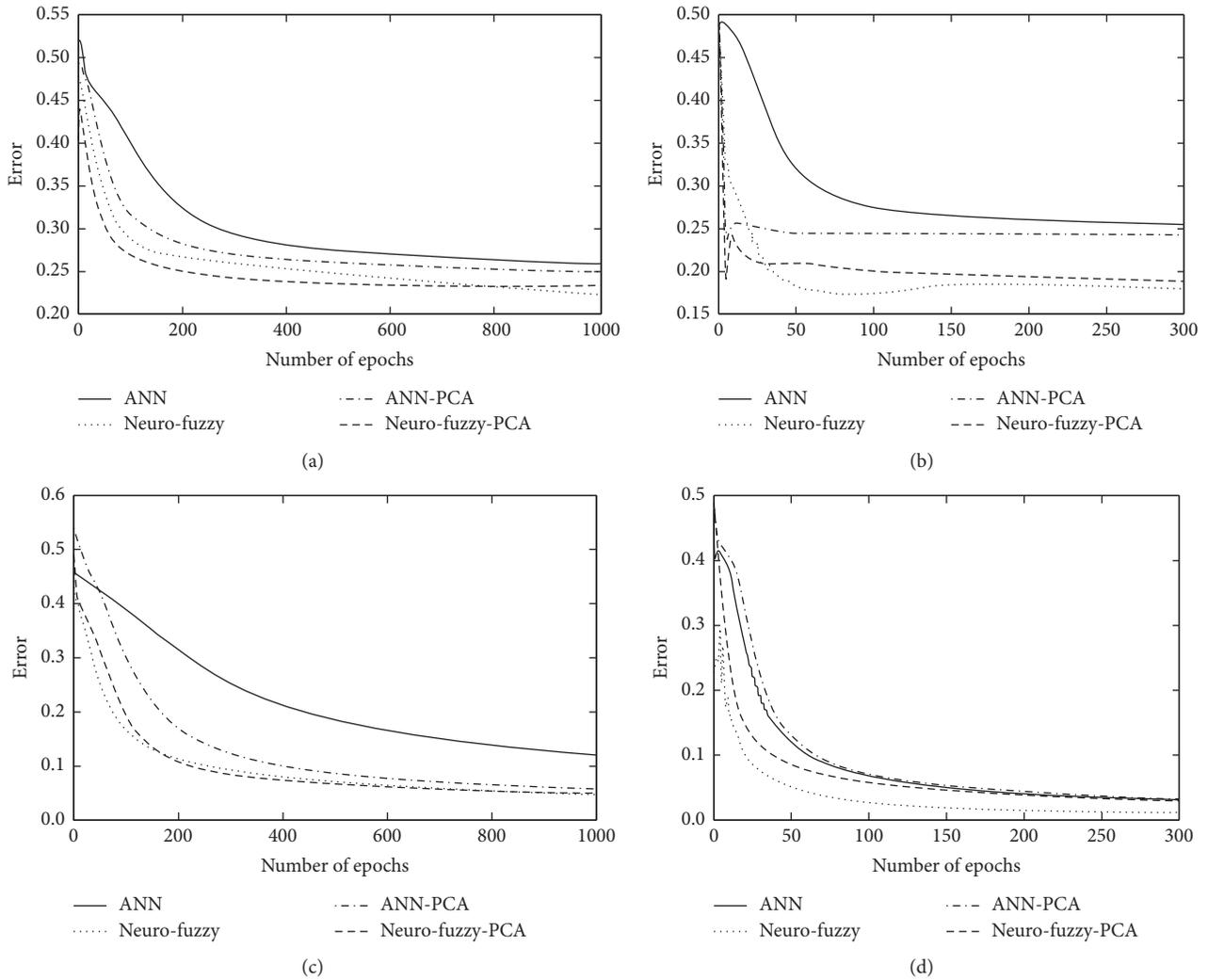


FIGURE 5: Error plots of four datasets: (a) Titanic; (b) Mammographic; (c) Breast Cancer; (d) Wine.

TABLE 2: Comparison of classification accuracy of ANN, NF, ANN-FR, and NF-FR.

Datasets/models	ANN			NF			ANN-FR (studied)			NF-FR (proposed)		
	WC	AC	BC	WC	AC	BC	WC	AC	BC	WC	AC	BC
Iris	85.71	91.56	94.44	90.00	92.63	97.06	90.63	94.02	98.02	91.43	95.15	99.26
Mammographic Mass	73.91	79.21	84.62	79.99	84.19	87.35	82.53	84.50	86.36	83.05	85.31	87.72
Breast Cancer Wisconsin	85.42	91.07	95.00	91.18	93.98	96.48	93.66	95.39	97.74	93.86	95.44	97.04
Pima Indian diabetes	65.52	72.96	78.41	71.76	78.78	86.84	76.00	81.66	86.49	76.74	82.55	85.94
Hayes-Roth	56.25	66.34	76.00	60.00	73.85	83.33	69.23	75.20	83.87	70.59	77.15	86.95
Thyroid	83.72	86.76	92.85	86.04	92.01	96.07	88.46	92.53	97.29	92.98	95.60	99.55
Titanic	75.13	77.64	81.87	76.00	79.25	83.74	77.00	80.54	85.00	77.10	81.47	85.94
Wine	85.29	91.24	94.12	88.88	93.1	97.22	90.90	93.94	98.07	90.62	95.59	99.88
Haberman	70.21	75.42	78.79	75.00	78.05	82.81	76.60	80.29	83.93	77.97	81.02	86.89
Blood Transfusion Service Center	73.37	75.86	80.00	73.61	76.79	80.41	76.00	80.14	84.50	76.58	80.22	84.91

WC: worst case; AC: average case; BC: best case.

On the similar way, we have tested under ANOVA, Tukey, and Dunnett tests, and the performance of our proposed model was found to be significantly better as compared to other models.

In this experiment, few parameters were considered in the design process of the models. The fuzzy expansion, number of input neurons, and number of output neurons for ten datasets are presented in Table 5. The number of hidden layer used is

TABLE 3: Comparison of different performance parameters (precision, recall, and f-measure) of ANN, NF, ANN-FR, and NF-FR models.

Datasets/Models	Precision, recall, and f-measure of models											
	ANN			NF			ANN-FR			NF-FR		
	P	R	FM	P	R	FM	P	R	FM	P	R	FM
Iris	0.918	0.910	0.911	0.926	0.928	0.923	0.947	0.942	0.940	0.950	0.953	0.948
Mammographic Breast Cancer	0.733	0.523	0.610	0.846	0.545	0.663	0.808	0.491	0.610	0.852	0.532	0.653
Pima Indian	0.972	0.359	0.501	0.914	0.343	0.499	0.942	0.327	0.485	0.925	0.324	0.480
Hayes-Roth	0.393	0.181	0.246	0.903	0.750	0.819	0.633	0.263	0.368	0.936	0.723	0.815
Thyroid	0.683	0.706	0.673	0.781	0.754	0.751	0.750	0.751	0.734	0.801	0.758	0.756
Titanic	0.953	0.751	0.803	0.915	0.868	0.875	0.933	0.831	0.863	0.932	0.944	0.933
Wine	0.762	0.389	0.512	0.781	0.681	0.728	0.840	0.324	0.465	0.816	0.686	0.740
Haberman	0.9155	0.9289	0.9186	0.9361	0.9419	0.9359	0.8512	0.9437	0.9363	0.9629	0.9544	0.9569
Blood Transfusion	0.6233	0.0807	0.1417	0.635	0.0699	0.1245	0.7014	0.074	0.132	0.5233	0.0308	0.0576
	0.5015	0.0839	0.0574	0.7283	0.9923	0.8703	0.129	0.1015	0.1011	0.8105	0.9741	0.8843

P: precision; R: recall; FM: f-measure.

TABLE 4: Friedman's rank to all the models based on avg. classification accuracy.

Datasets/models	ANN	NF	ANN-FR	NF-FR
Iris	91.56 (4)	92.63 (3)	94.02 (2)	95.15 (1)
Mammographic Mass Breast Cancer Wisconsin	79.21 (4)	84.19 (3)	84.50 (2)	85.31 (1)
Pima Indian Diabetes	91.07 (4)	93.98 (3)	95.39 (1)	95.04 (2)
Hayes-Roth	72.96 (4)	78.78 (3)	81.66 (2)	82.55 (1)
Thyroid	66.34 (4)	73.85 (3)	75.20 (2)	77.15 (1)
Titanic	86.76 (4)	92.01 (3)	92.53 (2)	95.60 (1)
Wine	77.64 (4)	79.25 (3)	80.54 (2)	81.47 (1)
Haberman	91.24 (4)	93.1 (3)	93.94 (2)	95.59 (1)
Blood Transfusion Service	75.42 (4)	78.05 (3)	80.29 (2)	81.02 (1)
Friedman's Rank	4	3	1.9	1.1

one, and number of neurons in the hidden layer is computed by using equation (9) for all the models. The learning rate of all the models is 0.76. In the FR process, reduction of the dimensions of the principal components is 5% of the original data.

The complexity of the model describes how efficient the proposed NF-FR model is. This model comprises three components such as fuzzification, feature reduction, and ANN classification. Fuzzification step requires constant amount of time for initialization of initial parameters that take $O(1)$ time, and for each feature, the fuzzification process expands the feature space into its corresponding fuzzified feature space based on the class labels that are available in the datasets will take $O(n \times C)$ times. Here, n is the number of features and C is the number of class labels in the dataset. The total time required for all the instances of the dataset to do fuzzification requires $O(n \times C \times m)$ times. Here, m is the number of instances in the dataset. In the feature reduction step, the computation of eigen values and eigen vector requires $O(1)$ time and covariance matrix requires $O(q \times q)$ times, where q is the fuzzified feature set. So, the feature reduction step requires $O(q^2)$ time. Finally, ANN-BPN step consists of both the feedforward and backpropagation steps that take $O(n^2)$ and $O(n^4)$, respectively. So, the total complexity of the model is $O(n \times C \times m + q^2 + n^4)$.

5. Statistical Analysis of Results

Statistical analysis is a well-known method to analyze the performance of various models with several datasets. Generally, different statistical tools are used to analyze the nature of data and algorithms. In this section, statistical analysis [59] along with the comparison of all the models over multiple datasets is presented. Several statistical tests such as analysis of variance (ANOVA) test [60], Tukey test [61], Dunnett test [62], Friedman test [63, 64], and post hoc test [65, 66] have been used to prove the proposed classification algorithm is more efficient than other existing classification algorithms. It will find out the best classification algorithm among a set of classification algorithms based on certain measuring parameters.

5.1. ANOVA Test. ANOVA [60] is a parametric statistical technique used to be compared among the different models. It generally compares the mean and relative variance in the performance of different models. This method is suitable when more than two models are compared with different datasets. ANOVA uses a null hypothesis and an alternative hypothesis. The null hypothesis is valid only when the performances of all the models are equal or there is no significant difference among these models. Alternatively, the alternative hypothesis is valid only when any one of the models is different from the rest of the models. The one-way ANOVA test has been carried out in SPSS (Version 16.0) with 95% confident interval, and the result has been presented in Tables 6 and 7.

5.2. Tukey and Dunnett Tests. To reject the null hypothesis, Tukey test [61] and Dunnett test [62] have been conducted. In the Tukey test, the comparison of the performance of every model is compared with every other model, but the Dunnett test compares the performance of every model with the proposed model. The control group for this test is NF-FR is compared with different models such as ANN, ANN-FR, and NF. The result of Tukey and

TABLE 5: Parameters of the models.

Datasets	Fuzzy expansion	No. of input neurons	No. of output neurons
Iris	3	4	3
Mammographic	2	5	2
Breast Cancer	2	9	2
Pima Indian	2	8	2
Hayes-Roth	3	6	3
Thyroid	3	6	3
Titanic	2	14	2
Wine	3	14	3
Haberman	2	4	2
Blood Transfusion Service	2	5	2

TABLE 6: ANOVA test with descriptive statistics.

Models	N	Mean	Std. deviation	Std. error	95% confidence interval for mean		Minimum	Maximum
					Lower bound	Upper bound		
ANN	10	80.806	8.83292	2.79321	74.4873	87.1247	66.34	91.56
NF	10	84.263	7.89241	2.4958	78.6171	89.9089	73.85	93.98
ANN-FR	10	85.821	7.39622	2.33889	80.5301	91.1119	75.2	95.39
NF-FR	10	86.91	7.53305	2.38216	81.5212	92.2988	77.15	95.6
Total	40	84.45	7.97159	1.26042	81.9006	86.9994	66.34	95.6

TABLE 7: ANOVA test with mean square and sum of squares.

		Sum of squares	df	Mean square	F	Sig.	
Between groups	(Combined)	212.449	3	70.816	1.125	0.352	
	Linear term	Contrast	197.408	1	197.408	3.136	0.085
		Deviation	15.041	2	7.521	0.119	0.888
	Quadratic term	Contrast	14.019	1	14.019	0.223	0.64
		Deviation	1.022	1	1.022	0.016	0.899
Within groups		2265.856	36	62.94			
Total		2478.305	39				

TABLE 8: Tukey test and Dunnett test.

	(I) Algo	(J) Algo	Mean difference (I-J)	Std. error	Sig.	95% confidence interval		
						Lower bound	Upper bound	
Tukey HSD	ANN	NF	-3.457	3.54797	0.765	-13.0125	6.0985	
		ANN-FR	-5.015	3.54797	0.499	-14.5705	4.5405	
		NF-FR	-6.104	3.54797	0.328	-15.6595	3.4515	
	NF	ANN	3.457	3.54797	0.765	-6.0985	13.0125	
		ANN-FR	-1.558	3.54797	0.971	-11.1135	7.9975	
		NF-FR	-2.647	3.54797	0.878	-12.2025	6.9085	
	ANN-FR	ANN	5.015	3.54797	0.499	-4.5405	14.5705	
		NF	1.558	3.54797	0.971	-7.9975	11.1135	
		NF-FR	-1.089	3.54797	0.99	-10.6445	8.4665	
	NF-FR	ANN	6.104	3.54797	0.328	-3.4515	15.6595	
		NF	2.647	3.54797	0.878	-6.9085	12.2025	
		ANN-FR	1.089	3.54797	0.99	-8.4665	10.6445	
	Dunnett <i>t</i> (2-sided) ^a	ANN	NF-FR	-6.104	3.54797	0.221	-14.8041	2.5961
		NF	NF-FR	-2.647	3.54797	0.798	-11.3471	6.0531
		ANN-FR	NF-FR	-1.089	3.54797	0.98	-9.7891	7.6111

Means for groups in homogeneous subsets are displayed. ^aDunnett *t*-tests treat one group as a control and compare all other groups against it.

Dunnett tests is presented in Table 8. A homogeneous group of models based on their level of significance is presented in Table 9.

5.3. *Friedman Test.* Friedman test [63, 64] is a nonparametric statistical technique developed by M. Friedman. It is used to find out the differences among different models by

TABLE 9: Homogeneous group of models based on their level of significance.

	Algorithms	N	Subset for alpha = 0.05 1
Tukey HSD ^a	ANN	10	80.806
	NF	10	84.263
	ANN-FR	10	85.821
	NF-FR	10	86.91
	Sig.		0.328
Duncan ^a	ANN	10	80.806
	NF	10	84.263
	ANN-FR	10	85.821
	NF-FR	10	86.91
	Sig.		0.124

^aUses harmonic mean sample size = 10.000.

assigning certain ranks to the resultant values represented in Table 4. The average rank of algorithms can be computed by using the following equation, where r_i^j is the rank of the j^{th} model on the i^{th} dataset and N is the number of dataset:

$$R_j = \frac{1}{N} \sum_i r_i^j. \quad (17)$$

The average rank of four models: ANN, ANN-FR, NF, and NF-FR is computed based on assigned rank, which is represented in equation (17). This average rank of four models: ANN, ANN-FR, NF, NF-FR can be represented as $\{R_4 = 4, R_3 = 3, R_2 = 1.9, R_1 = 1.1\}$, respectively. The value of X_F^2 is computed from average rank R_i and can be realized by the following equation (18), where N is the number of datasets and m is the number of models. In this case, the value of X_F^2 is 48.2:

$$X_F^2 = \frac{12N}{m(m-1)} \left[\sum_j R_j^2 - \frac{m(m+1)^2}{4} \right]. \quad (18)$$

The Friedman statistics F_F is measured by using X_F^2 with $(m-1)$ degree of freedom and can be realized in equation (19). The critical value [64] can be obtained from Friedman statistics F_F with $(m-1)$ and $(m-1) \times (N-1)$ degree of freedom. In this approach, four numbers of models m and ten numbers of datasets N are used. In this case, the value of Friedman statistics F_F is 241 with four numbers of models and ten numbers of datasets:

$$F_F = \frac{(N-1)X_F^2}{N(m-1) - X_F^2}. \quad (19)$$

The performances of models are different, if the corresponding average rank differs by at least the critical difference. The critical value is computed as 4.6 with $(4-1=3)$ and $(4-1=3) \times (10-1=9)$ degree of freedom and significance level $\alpha = 0.01$. The density plot is obtained and shown in Figure 6 with a degree of freedom (3, 27). It is noted that the null hypothesis is rejected as the critical value (4.6) is less than the Friedman statistics ($F_F = 241$). The post hoc test experiment has been conducted by using

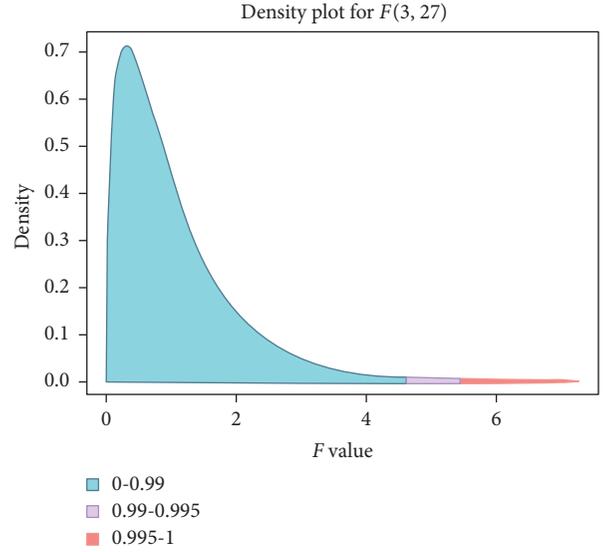


FIGURE 6: Density plot.

TABLE 10: Results of the Holm statistical test.

Models	Models	z -values	p values	$\alpha/(m-i)$
1	NF-FR: ANN	5.02	2.584871e-7	0.0033
2	NF-FR: ANN-FR	3.29	0.000501	0.005
3	NF-FR: NF	1.38	0.083793	0.01

the Holm procedure, after the rejection of the null hypothesis.

5.4. Holm Procedure. This Holm procedure [65–67] computes the performance of every individual model with the rest of the models by using the z -value and p value. The z -value is computed by using equation (20), and the p value is computed from the z -value and the normal distribution table accordingly:

$$z = \frac{(R_i - R_j)}{\sqrt{m(m+1)/6N}}, \quad (20)$$

where m is the number of models, z is the z -score value, and N is the number of datasets. The average rank of the i^{th} and j^{th} models is denoted by R_i and R_j , respectively. All the three models are compared with the proposed model based on z -value, p value, and $\alpha/(m-i)$, and the result is represented in Table 10. Here, we noticed that, in almost all the cases, the p values are less than $\alpha/(m-i)$ by using the Holm test. Hence, it is concluded that the null hypothesis is rejected. Thus, it indicates that the proposed model NF-FR is statistically significantly different and better than other classification models.

6. Conclusion and Future Scope

In this paper, the proposed NF-FR model is demonstrated successfully for solving data classification problems in data mining. Initially, this model uses the fuzzification process for expansion of the input features class-wise belongingness of the features to various classes which provide to handle imprecise and uncertainty problems. Due to the expansion

of features, the model structure becomes massively parallel and also found that all the features may not contribute significantly to the model. In the next step, PCA is applied to reduce the dimension of the expanded features by selecting the best suitable relevant and nonredundant features. As a result, the learning time of the proposed model was also reduced with the selected relevant features. However, a particular Π -type membership function considered for the fuzzification process may not always be suitable for the entire datasets. In such cases, the selection of suitable membership functions may be taken into consideration for data analysis. As per the experimental analysis, the proposed method is able to classify the datasets with superior classification performance as compared to ANN, NF, and ANN-FR models. After statistical analysis, it is found that the proposed NF-FR model is valid and efficient as compared to ANN, NF, and ANN-FR models. In the future, this proposed model can be used in various real-life problems such as gene expression classification, document classification, and satellite image classification.

Data Availability

The data used to support the findings of this study are included in Section 4 within the article. We have used ten benchmark datasets from the UCI machine learning repository. The detail descriptions about all these datasets can be found at “<http://archive.ics.uci.edu/ml/>” and “<http://keel.es/>.”

Disclosure

This study was not funded by any research organization.

Conflicts of Interest

All authors declare that there are no conflicts of interest.

Acknowledgments

This research work was supported by the Science and Engineering Research Board (SERB), Department of Science and Technology (DST), New Delhi, Govt. of India, under the research project grant Sanction Order No. EEQ/2017/000355.

References

- [1] D. Michie, D. J. Spiegelhalter, and C. C. Taylor, *Machine Learning. Neural and Statistical Classification*, vol. 13, Ellis Horwood, Hemel Hempstead, UK, 1994.
- [2] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley & Sons, Hoboken, NJ, USA, 2012.
- [3] H. Das, B. Naik, and H. S. Behera, “An experimental analysis of machine learning classification algorithms on biomedical data,” in *Proceedings of the 2nd International Conference on Communication, Devices and Computing*, pp. 525–539, Springer, Singapore, 2020.
- [4] A. K. Sahoo, C. Pradhan, and H. Das, “Performance evaluation of different machine learning methods and deep-learning based convolutional neural network for health decision making,” in *Nature Inspired Computing for Data Science*, pp. 201–212, Springer, Cham, Switzerland, 2020.
- [5] H. Das, B. Naik, and H. S. Behera, “Medical disease analysis using neuro-fuzzy with feature extraction model for classification,” *Informatomics in Medicine Unlocked*, vol. 18, p. 100288, 2020.
- [6] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data clustering: a review,” *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.
- [7] J. Makhoul, “Linear prediction: a tutorial review,” *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.
- [8] H. Das, N. Dey, and V. E. Balas, *Real-Time Data Analytics for Large Scale Sensor Data*, Academic Press, Cambridge, MA, USA, 2019.
- [9] F. Khoshbin, H. Bonakdari, S. H. Ashraf Talesh, I. Ebtehaj, A. H. Zaji, and H. Azimi, “Adaptive neuro-fuzzy inference system multi-objective optimization using the genetic algorithm/singular value decomposition method for modelling the discharge coefficient in rectangular sharp-crested side weirs,” *Engineering Optimization*, vol. 48, no. 6, pp. 933–948, 2016.
- [10] V. T. Yen, Y. N. Wang, and P. V. Cuong, “Recurrent fuzzy wavelet neural networks based on robust adaptive sliding mode control for industrial robot manipulators,” *Neural Computing and Applications*, vol. 31, no. 11, pp. 6925–6958, 2019.
- [11] B. H. Kwasnik, “The role of classification in knowledge representation and discovery,” *Library Trends*, vol. 48, no. 1, pp. 22–47, 1999.
- [12] S. Haykin, *Neural Networks*, vol. 2, Prentice-Hall, New York, NY, USA, 1994.
- [13] H. Yoon, J. Lim, and J. S. Lim, “Reconstructing time series GRN using a neuro-fuzzy system,” *Journal of Intelligent & Fuzzy Systems*, vol. 29, no. 6, pp. 2751–2757, 2015.
- [14] M. Lee, S.-Y. Lee, and C. H. Park, “Neuro-fuzzy identifiers and controllers,” *Journal of Intelligent and Fuzzy Systems*, vol. 2, no. 1, pp. 1–14, 1994.
- [15] L. A. Zadeh, “Fuzzy sets,” *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.
- [16] L. A. Zadeh, “Fuzzy sets and information granularity,” in *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers*, L. A. Zadeh, Ed., pp. 433–448, World Scientific, Singapore, 1996.
- [17] L. A. Zadeh, “Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic,” *Fuzzy Sets and Systems*, vol. 90, no. 2, pp. 111–127, 1997.
- [18] C. Pradhan, H. Das, B. Naik, and N. Dey, *Handbook of Research on Information Security in Biomedical Signal Processing*, pp. 1–414, IGI Global, Hershey, PA, USA, 2018.
- [19] J. Nayak, B. Naik, A. K. Jena, R. K. Barik, and H. Das, “Nature inspired optimizations in cloud computing: applications and challenges,” in *Cloud Computing for Optimization: Foundations, Applications, and Challenges*, pp. 1–26, Springer, Cham, Switzerland, 2018.
- [20] B. S. P. Mishra, H. Das, S. Dehuri, and A. K. Jagadev, *Cloud Computing for Optimization: Foundations, Applications, and Challenges*, vol. 39, Springer, Berlin, Germany, 2018.
- [21] M. Rout, A. K. Jena, J. K. Rout, and H. Das, “Teaching-learning optimization based cascaded low-complexity neural network model for exchange rates forecasting,” in *Smart Intelligent Computing and Applications*, pp. 635–645, Springer, Singapore, 2020.

- [22] N. Dey, H. Das, B. Naik, and H. S. Behera, *Big Data Analytics for Intelligent Healthcare Management*, Academic Press, Cambridge, MA, USA, 2019.
- [23] A. K. Sahoo, S. Mallik, C. Pradhan, B. S. P. Mishra, R. K. Barik, and H. Das, "Intelligence-based health recommendation system using big data analytics," in *Big Data Analytics for Intelligent Healthcare Management*, pp. 227–246, Academic Press, Cambridge, MA, USA, 2019.
- [24] N. Dey, A. S. Ashour, H. Kalia, R. Goswami, and H. Das, *Histopathological Image Analysis in Medical Decision Making*, pp. 1–340, IGI Global, Hershey, PA, USA, 2019.
- [25] J.-S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 23, no. 3, pp. 665–685, 1993.
- [26] A. Ghosh, B. Uma Shankar, and S. K. Meher, "A novel approach to neuro-fuzzy classification," *Neural Networks*, vol. 22, no. 1, pp. 100–109, 2009.
- [27] S. Ghosh, S. Biswas, D. Sarkar, and P. P. Sarkar, "A novel neuro-fuzzy classification technique for data mining," *Egyptian Informatics Journal*, vol. 15, no. 3, pp. 129–147, 2014.
- [28] S. K. Pal and S. Mitra, "Multilayer perceptron, fuzzy sets, and classification," *IEEE Transactions on Neural Networks*, vol. 3, no. 5, pp. 683–697, 1992.
- [29] A. T. Azar and A. E. Hassanien, "Dimensionality reduction of medical big data using neural-fuzzy classifier," *Soft Computing*, vol. 19, no. 4, pp. 1115–1127, 2015.
- [30] H. Das, A. K. Jena, J. Nayak, B. Naik, and H. S. Behera, "A novel PSO based back propagation learning-MLP (PSO-BP-MLP) for classification," in *Computational Intelligence in Data Mining*, vol. 2, pp. 461–471, Springer, New Delhi, India, 2015.
- [31] R. Sahani, C. Rout, J. C. Badajena, A. K. Jena, and H. Das, "Classification of intrusion detection using data mining techniques," in *Progress in Computing, Analytics and Networking*, pp. 753–764, Springer, Singapore, 2018.
- [32] H. Das, B. Naik, and H. S. Behera, "Classification of diabetes mellitus disease (DMD): a data mining (DM) approach," in *Progress in Computing, Analytics and Networking*, pp. 539–549, Springer, Singapore, 2018.
- [33] L. I. Smith, "A tutorial on principal components analysis," 2002, http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf.
- [34] S. Kar, S. Das, and P. K. Ghosh, "Applications of neuro fuzzy systems: a brief review and future outline," *Applied Soft Computing*, vol. 15, pp. 243–259, 2014.
- [35] S. K. Meher, "Efficient pattern classification model with neuro-fuzzy networks," *Soft Computing*, vol. 21, no. 12, pp. 3317–3334, 2017.
- [36] Z. J. Viharos and K. B. Kis, "Survey on neuro-fuzzy systems and their applications in technical diagnostics and measurement," *Measurement*, vol. 67, pp. 126–136, 2015.
- [37] K. V. Shihabudheen and G. N. Pillai, "Recent advances in neuro-fuzzy system: a survey," *Knowledge-Based Systems*, vol. 152, pp. 136–162, 2018.
- [38] A. K. Das and D. K. Pratihari, "A novel approach for neuro-fuzzy system-based multi-objective optimization to capture inherent fuzziness in engineering processes," *Knowledge-Based Systems*, vol. 175, pp. 1–11, 2019.
- [39] I. Škrjanc, J. A. Iglesias, A. Sanchis, D. Leite, E. Lughofer, and F. Gomide, "Evolving fuzzy and neuro-fuzzy approaches in clustering, regression, identification, and classification: a survey," *Information Sciences*, vol. 490, pp. 344–368, 2019.
- [40] C. A. Murthy, "Bridging feature selection and extraction: compound feature generation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 4, pp. 757–770, 2017.
- [41] S. Chattopadhyay, "A neuro-fuzzy approach for the diagnosis of depression," *Applied Computing and Informatics*, vol. 13, no. 1, pp. 10–18, 2017.
- [42] S. Ibrahim, P. Chowriappa, S. Dua et al., "Classification of diabetes maculopathy images using data-adaptive neuro-fuzzy inference classifier," *Medical & Biological Engineering & Computing*, vol. 53, no. 12, pp. 1345–1360, 2015.
- [43] P. F. Alvanitopoulos, I. Andreadis, and A. Elenas, "Neuro-fuzzy techniques for the classification of earthquake damages in buildings," *Measurement*, vol. 43, no. 6, pp. 797–809, 2010.
- [44] M.-S. Chen, "Neuro-fuzzy approach for online message scheduling," *Engineering Applications of Artificial Intelligence*, vol. 38, pp. 59–69, 2015.
- [45] M. Azhari and Y. J. Kumar, "Improving text summarization using neuro-fuzzy approach," *Journal of Information and Telecommunication*, vol. 1, no. 4, pp. 367–379, 2017.
- [46] H. R. Singh, S. K. Biswas, and B. Purkayastha, "A neuro-fuzzy classification technique using dynamic clustering and GSS rule generation," *Journal of Computational and Applied Mathematics*, vol. 309, pp. 683–694, 2017.
- [47] M. Nilashi, H. Ahmadi, L. Shahmoradi, O. Ibrahim, and E. Akbari, "A predictive method for hepatitis disease diagnosis using ensembles of neuro-fuzzy technique," *Journal of Infection and Public Health*, vol. 12, no. 1, pp. 13–20, 2019.
- [48] K. V. Shihabudheen, M. Mahesh, and G. N. Pillai, "Particle swarm optimization based extreme learning neuro-fuzzy system for regression and classification," *Expert Systems with Applications*, vol. 92, pp. 474–484, 2018.
- [49] K. Polat and S. Güneş, "An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease," *Digital Signal Processing*, vol. 17, no. 4, pp. 702–710, 2007.
- [50] X. Wang and K. K. Paliwal, "Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition," *Pattern Recognition*, vol. 36, no. 10, pp. 2429–2439, 2003.
- [51] A. T. Azar, "Neuro-fuzzy feature selection approach based on linguistic hedges for medical diagnosis," *International Journal of Modelling, Identification and Control*, vol. 22, no. 3, pp. 195–206, 2014.
- [52] A. Keles, A. Samet Hasiloglu, A. Keles, and Y. Aksoy, "Neuro-fuzzy classification of prostate cancer using NEFCLASS-J," *Computers in Biology and Medicine*, vol. 37, no. 11, pp. 1617–1628, 2007.
- [53] B. Gabrys, "Learning hybrid neuro-fuzzy classifier models from data: to combine or not to combine?," *Fuzzy Sets and Systems*, vol. 147, no. 1, pp. 39–56, 2004.
- [54] E. D. Übeyli, "Adaptive neuro-fuzzy inference system for classification of ECG signals using Lyapunov exponents," *Computer Methods and Programs in Biomedicine*, vol. 93, no. 3, pp. 313–321, 2009.
- [55] V. Kolodyazhnyi, F. Klawonn, and K. Tschumitschew, "A neuro-fuzzy model for dimensionality reduction and its application," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 15, no. 5, pp. 571–593, 2007.
- [56] A. Schlar, L. Rokach, and A. Amit, "Ensembles of classifiers based on dimensionality reduction," *Intelligent Data Analysis*, vol. 21, no. 3, pp. 467–489, 2017.
- [57] K. Bache and M. Lichman, *UCI Machine Learning Repository*, University of California, School of Information and Computer Science, Irvine, CA, USA, 2013.

- [58] J. Alcalá-Fdez, A. Fernández, J. Luengo et al., “Keel data-mining software tool: data set repository. Integration of algorithms and experimental analysis framework,” *Journal of Multiple-Valued Logic & Soft Computing*, vol. 17, no. 2-3, pp. 255–287, 2011.
- [59] J. Demšar, “Statistical comparisons of classifiers over multiple data sets,” *Journal of Machine Learning Research*, vol. 7, pp. 1–30, 2006.
- [60] R. A. Fisher, *Statistical Methods and Scientific Inference*, Hafner Publishing Co., New York, NY, USA, 2nd edition, 1956.
- [61] J. W. Tukey, “Comparing individual means in the analysis of variance,” *Biometrics*, vol. 5, no. 2, pp. 99–114, 1949.
- [62] C. W. Dunnett, “A multiple comparison procedure for comparing several treatments with a control,” *Journal of the American Statistical Association*, vol. 50, no. 272, pp. 1096–1121, 1955.
- [63] M. Friedman, “The use of ranks to avoid the assumption of normality implicit in the analysis of variance,” *Journal of the American Statistical Association*, vol. 32, no. 200, pp. 675–701, 1937.
- [64] M. Friedman, “A comparison of alternative tests of significance for the problem of m rankings,” *The Annals of Mathematical Statistics*, vol. 11, no. 1, pp. 86–92, 1940.
- [65] R. L. Iman and J. M. Davenport, “Approximations of the critical region of the Friedman statistic,” *Communications in Statistics—Theory and Methods*, vol. 9, no. 6, pp. 571–595, 1980.
- [66] S. García, A. Fernández, J. Luengo, and F. Herrera, “Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: experimental analysis of power,” *Information Sciences*, vol. 180, no. 10, pp. 2044–2064, 2010.
- [67] J. Luengo, S. García, and F. Herrera, “A study on the use of statistical tests for experimentation with neural networks: analysis of parametric test conditions and non-parametric tests,” *Expert Systems with Applications*, vol. 36, no. 4, pp. 7798–7808, 2009.