Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 4, August 2025, pp. 2793~2806
ISSN: 2302-9285, DOI: 10.11591/eei.v14i4.9140

Incremental learning based fuzzy reasoning approach for diagnosis of thyroid disease

Thirumalaimuthu Thirumalaiappan Ramanathan, Md. Jakir Hossen
Centre for Advanced Analytics, CoE for Artificial Intelligence, Faculty of Engineering and Technology, Multimedia University, Melaka, Malaysia

Article Info

Article history:

Received Aug 13, 2024
Revised Jan 25, 2025
Accepted Mar 9, 2025

Keywords:

Fuzzy logic
Hybrid intelligent system
Incremental learning
Machine learning
Medical diagnosis

ABSTRACT

This paper presents a novel hybrid fuzzy logic approach for the classification of thyroid disease. Hybrid fuzzy logic approaches have brought many benefits to the medical data classification problems such as reasoning on uncertain or incomplete data. The machine learning algorithms had been used with the fuzzy expert systems to define the fuzzy rule base. The optimization techniques had been used in the fuzzy expert systems for optimizing the fuzzy membership functions and fuzzy rules. Enhancing the machine learning algorithms and optimization techniques that are integrated with the fuzzy logic method can improve the overall performance of the fuzzy expert system. To deal with the curse of dimensionality problem and to enhance the integration of machine learning algorithm and fuzzy logic method, this paper presents an incremental learning based parallel fuzzy reasoning system (IL-PFRS) for medical diagnosis. In this research work, the decision tree classifier is used to extract the features from dataset. IL-PFRS is applied to classify the thyroid disease which is serious disease that needs attention and earlier detection. The thyroid disease dataset obtained from the UCI machine learning repository is used in this research work where the IL-PFRS showed the classification accuracy of 99% when testing using this dataset.

Corresponding Author:

Md. Jakir Hossen
Centre for Advanced Analytics, CoE for Artificial Intelligence, Faculty of Engineering and Technology
Multimedia University
St. Ayer Keroh Lama, Bukit Beruang, 75450 Melaka, Malaysia
Email: jakir.hossen@mmu.edu.my

1. INTRODUCTION

Being an endocrine gland, the thyroid is located in the inferior, anterior neck and it regulates the body's iodine balance in addition to the production and secretion of thyroid hormones. A medical disorder known as thyroid illness impacts the action of thyroid gland by preventing the thyroid from producing the appropriate quantity of hormones [1]. There are different types of thyroid diseases such as hyperthyroidism, hypothyroidism, thyroid cancer [2]. Thyroid disease may result to number of health issues when not treated properly at the earlier stage. There are evidences were some of the serious thyroid diseases such as thyroid cancer are the main cause of many deaths in the world [3]. Hybrid artificial intelligence (AI) techniques based on the machine learning and soft computing approaches are proved to be efficient in various problems of medical data classification. This research paper focus on the enhancement of hybrid fuzzy expert system for the classification of thyroid disease. Fuzzy logic is a reasoning method that addresses the idea of partial truth. In fuzzy logic, a variable's truth value might fall anywhere between 0 (totally false) and 1 (absolutely false). In one of our previous research works [4], we presented a parallel fuzzy reasoning approach based on the naïve bayes (NB) algorithm for the classification of diabetes where the NB classifier [5] is used to set the fuzzy rules in the fuzzy expert system. In our another research work [6], we presented a blockchain technology based multi-agent fuzzy system for the prediction of breast cancer where each fuzzy agent system is integrated with different AI classifiers such as decision tree [7], K-nearest neighbor (KNN) [8], BN, random forest (RF) [9], and support vector machine (SVM) [10] classifiers. In this research paper, we enhance the parallel fuzzy expert systems to get integrated with the incremental learning method and investigate its performance on classification of thyroid disease. Some of the other recent research works done on application of hybrid fuzzy method for the medical diagnosis are reviewed below.

Genetic algorithm-based type-3 fuzzy inference system is used in the research work of Melin *et al*. [11] for medical diagnosis. Here, the genetic algorithm is used to identify the parameters of membership functions and fuzzy rules. Their proposed system showed the classification accuracy of 75.30, 87.13, 82.04, 77.76, 71.86, and 71.06 for Haberman's survival dataset [12], Cryotherapy dataset [13], Immunotherapy dataset [14], Pima Indian Diabetes (PID) dataset [15], Indian Liver patient (ILP) dataset [16], and Breast Cancer Coimbra dataset [17], respectively.

A hybrid of adaptive backpropagation neural network and fuzzy logic approach is used for diagnosis kidney disease in the research work of Vineetha *et al*. [18]. In their research work, Chronic Renal Disease dataset from UCI machine learning repository [19] is used for testing their proposed system. Their proposed system showed a classification accuracy of 98% for classifying the kidney disease.

The adaptive neuro-fuzzy inference system (ANFIS) is applied for classification of liver disease in the research work of Babatunde *et al*. [20]. The ILP dataset is used for testing their research work. Based on their experiments, the ANFIS showed the classification accuracy of 97% for diagnosing the liver disease.

A hybrid approach based on the fuzzy SVM and biased RF is used for the classification of diabetes in the research work of Hao *et al*. [21]. Here, fuzzy SVM is applied to enhance the SVM classifier and biased RF is used to extract the rules for diabetes patients based on the support vectors. In their research work, diabetes medical examination data collected by Beijing Hospital (DMED-BH) [21], PID dataset, and China health and nutrition survey dataset [22] were used for testing. Their proposed approach showed the classification accuracy of 96.92%, 96.84%, and 94.92% for the DMED-BH dataset, PID dataset and CHNS dataset, respectively.

A hybrid approach of fuzzy logic, principal component analysis (PCA), and SVM is used in the research work of Gupta *et al*. [23] for classification of medical datasets. Here, in their proposed approach, the fuzzy logic is used to obtain membership function from the dataset, PCA is used to reduce the number of input features, and SVM is used for classification. Their proposed method showed the classification accuracy of 79%, 87%, 85%, 83%, and 92% for the PID dataset, heart stroke prediction dataset [24], heart failure prediction dataset [25], body fat prediction dataset [26], and heart disease dataset [27], respectively.

The main common research drawback found in the reviewed hybrid fuzzy approaches is the curse of dimensionality problem. This problem occurs when a greater number of input variables are used in a single fuzzy expert system. When many input variables are used is in the fuzzy expert system, it will lead to a complicated fuzzy rile base. Although fuzzy expert system is improved in the reviewed hybrid fuzzy approaches by integrating it with the machine learning and AI optimization techniques, there is a research challenge to investigate the performance of fuzzy expert system in a parallel architecture when it gets optimized by an incremental learning method.

An incremental learning algorithm which can be also called as a continual learning approach is an extension of machine learning methodology where the incremental learning classifier extracts the patterns from different datasets in an incremental manner. There are different kinds of incremental learning techniques such as task based incremental learning, domain based incremental learning, and class based incremental learning techniques [28]. Domain based incremental learning method can be effective for the data mining applications in medical diagnosis. Some of the examples of incremental learning classifiers are iterative Dichotomize 5 revision [29], learn++ [30], and tree-convolution neural network [31].

The reviewed hybrid fuzzy approaches investigated the classification of medical data using the integration of single fuzzy expert system and machine learning classifiers. But this study investigates the performance of the parallel fuzzy systems which are integrated by an incremental learning classifier for the classification of medical data. To deal with the curse of dimensionality problem and to investigate the integration of fuzzy logic method and incremental learning method, this paper proposes an incremental learning based parallel fuzzy reasoning system (IL-PFRS) for the diagnosis of thyroid disease. The SVM classifier is enhanced and used as an incremental classifier in IL-PFRS where the incremental SVM classifier is based on the stochastic gradient descent (SGD) [32] optimization technique.

This paper is organized as follows. The section 2 describes about the proposed method. The section 3 describes the performance of proposed system for the classification of thyroid disease. The section 4 gives conclusion about the research work presented in this paper.

2. THE PROPOSED METHOD

This section describes about the dataset used for the IL-PFRS, the data preprocessing methods applied on the dataset, and the proposed IL-PFRS.

2.1. Dataset and data preprocessing

Before we discussed about the proposed classification system (IL-PFRS), this sub section describes about the dataset used in this research work and the data preprocessing technique applied on it. The thyroid disease dataset obtained from UCI machine learning repository [33] is used for testing the IL-PFRS. The thyroid disease dataset contains 3772 instances with 29 input attributes and one output attribute. The input attributes of thyroid disease dataset are tumor, hypopituitary, thyroxine utilization rates (T4U) measured, thyroid stimulating hormone (TSH) measured, age, thyroxine-binding globulin (TBG) measured, sex, I131 treatment, query hypothyroid, on thyroxine, triiodothyronine (T3) measured, TSH, query on thyroxine, free thyroxine index (FTI) measured, on antithyroid medication, sick, TBG, pregnant, thyroid surgery, thyroxine (TT4) measured, TT4, referral source, FTI, query hyperthyroid, lithium, T3, goitre, psych, and T4U. The output attribute contains two categorical values: 0 (negative) and 1 (positive). The thyroid disease dataset has many missing values which are imputed using the KNN imputation method [34].

The decision tree classifier is used to select the best features from the thyroid disease dataset. A decision tree model with a hierarchical tree-like structure, comprising a root node, internal nodes, and terminal (leaf) nodes is used for classification and feature selection. The directed edges are used to connect the decision tree's nodes. The input features of the training dataset are represented by the internal and root nodes. The output categories linked to the training dataset are represented by the terminal nodes. In order for the internal and root nodes to be divided according to their respective types, there will be specific test criteria. Until the decision tree locates every category of the output variable provided in the training dataset, the splitting process is repeated. Depending on the kind of attributes, the decision tree algorithm uses the attribute test condition. When testing the binary qualities, there are only two possible outcomes. When it comes to nominal attributes, the number of unique values connected to the appropriate characteristics determines the outputs generated by the test condition. Ordinal attributes allow for the grouping of the number of unique values associated with the relevant attributes without violating their property, which could result in numerous splits or binary outcomes. When dealing with continuous attributes, the test condition can provide numerous splits using distinct ranges of values, or a binary split using a comparison test.

Several metrics, including entropy [35], Gini impurity [36], and classification error, can be used to identify the node that divides the training dataset's samples in the most efficient way. When using the Gini impurity measure, the splitting technique selects the node with the lowest value, and when using the information gain [37] measure, the splitting procedure selects the node with the highest value.

Assume that the subset of samples at node x that belong to category j is denoted by $s(j|x)$ and there are t number of output categories. With using (1) to (3), respectively, the entropy $E(x)$, Gini impurity $G(x)$, and classification error $CE(x)$ measures are determined.

$$E(x)=-\sum_{i=0}^{t-1} s(j|x)\log_2 s(j|x) \quad (1)$$

$$G(x)=1-\sum_{i=0}^{t-1}[s(j|x)]^2 \quad (2)$$

$$CE(x)=1-\max_{j} s(j|x) \quad (3)$$

The gain, Δ formula, which takes into account the impurity value of the parent node and its child nodes, can be used to determine the quality of the attribute test condition. In this case, the parent node's impurity value is obtained before to splitting, and the child node's impurity value is obtained following splitting. When Δ values are large, the attribute test condition quality is better. Suppose u be the parent node, S be the number of samples at the parent node, S (v_l) be the number of samples at the child node (v) that correspond to the category l, and w be the number of feature values. In (4) describes the information gain, $\Delta_{\text{info.}}$ which calculates the quality of the attribute test condition by contrasting the entropy measure values of the parent node and its child nodes.

$$\Delta_{\text{info.}}=E(p)-\sum_{l=1}^{w}\frac{S(\mathrm{v_j})}{S}E(v_{\mathrm{j}}) \quad (4)$$

Based on the value of Gini impurity measure, the decision tree classifier will be able to select the best features from thyroid disease dataset for the IL-PFRS where the features with lower value of Gini impurity measure are consider as more important. The features importance scores for the thyroid disease dataset estimated by the decision tree classifier are shown in Figure 1. Five best features were selected from the thyroid disease dataset based on the feature importance scores shown in Figure 1. The selected input features of the thyroid disease dataset are shown in Table 1.

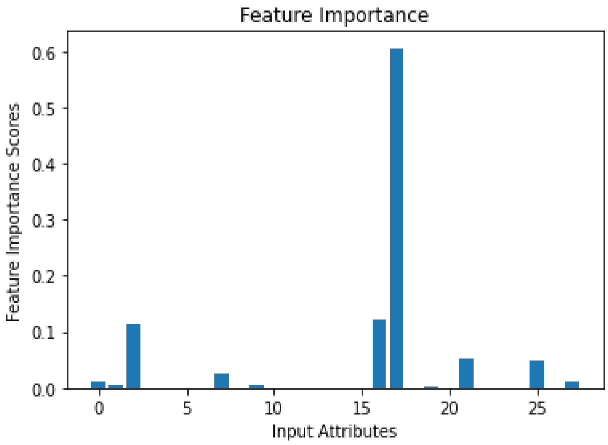

Figure 1. Feature importance scores for the thyroid disease dataset

Table 1. Selected input attributes of the thyroid disease dataset

Attributes	Data type
On thyroxine	Categorical
TSH measured	Categorical
TSH	Continuous
TT4	Continuous
FTI	Continuous

2.2. The proposed classification system

The obtained reduced dataset based on the KNN-imputation method and decision tree classifier is used to model the proposed classification system: IL-PFRS. The overall architecture of the proposed data classification method is shown in Figure 2. The architecture of IL-PFRS is shown in Figure 3.

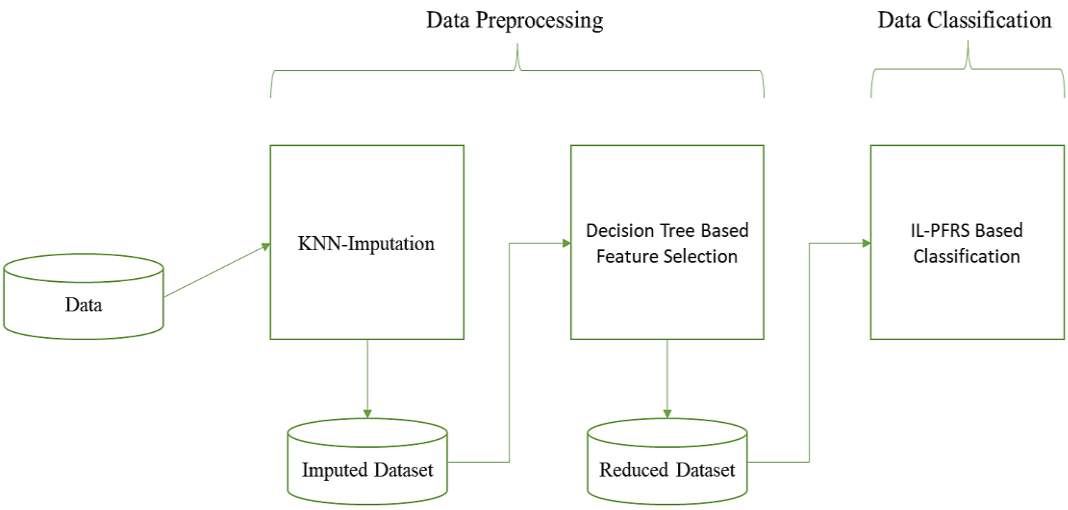

Figure 2. The overall architecture of the proposed method

As shown in Figure 3, IL-PFRS is designed by using two hybrid sub fuzzy systems (HSFSs) in a parallel architecture. The two HSFSs: HSFS1 and HSFS2 are implemented based on the Mamdani type fuzzy inference approach [38] which involves the following steps.

a. The implementation of the fuzzy reasoning system starts with the fuzzification process of creating fuzzy sets using different kinds of membership functions where the membership degree is assigned between 0 and 1 for all crisp inputs and outputs. Triangular, trapezoidal, gaussian, and sigmoid MFs are few of the frequently utilized membership functions for the fuzzy sets [39].
b. In order to facilitate fuzzy reasoning, a fuzzy rule base is created. The fuzzy rule base is made up of several rules, each of which has antecedents and consequents that are, respectively, the membership functions of the input and output fuzzy sets. The AND, OR operators are used to carry out the fuzzy operation between the membership functions of various input fuzzy sets. When the fuzzy rule base is executed, the outputs of all declared rules combine to generate a single aggregate output fuzzy set.
c. The last stage in the fuzzy reasoning system involves computing the crisp output from final aggregate output fuzzy set. The centre of gravity method (COG) [40] is one of the effective defuzzification techniques that uses (5) to generate the crisp output from the final aggregate output fuzzy set. In this case, $\mu_J(x)$ is the membership degree on fuzzy set J.

$$\mathrm{COG}=\frac{\int \mu_J(x)x\,dx}{\int \mu_J(x)\,dx} \tag{5}$$

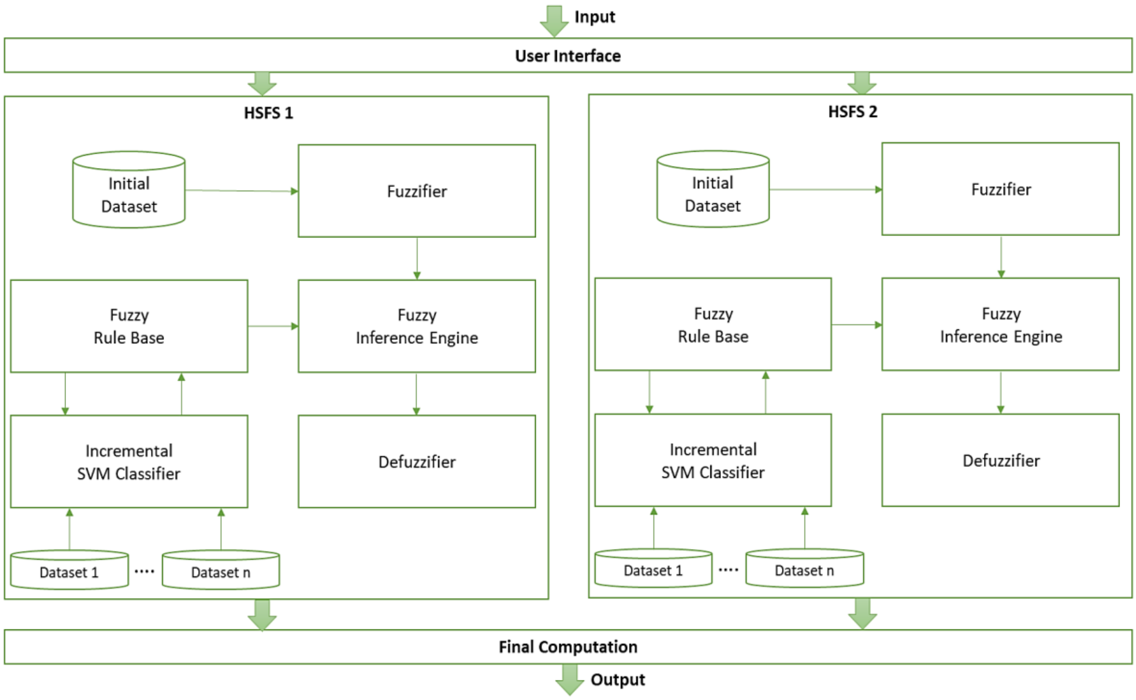

Figure 3. The architecture of IL-PFRS

The reduced dataset obtained from the decision tree classifier is divided into two groups. The first group of reduced thyroid disease dataset contains the input features: on thyroxine, TSH measured, and TSH. The second group of reduced thyroid disease dataset contains the input features: TT4 and FTI. Two dataset groups are further divided into training and testing dataset groups, respectively, based on 80:20 ratio. The first training dataset group is used to model the HSFS1 and the second training dataset group is used to model the HSFS2. The two testing dataset groups are joined as one testing dataset to test the IL-PFRS. Figure 4 shows the flowchart of HSFS.

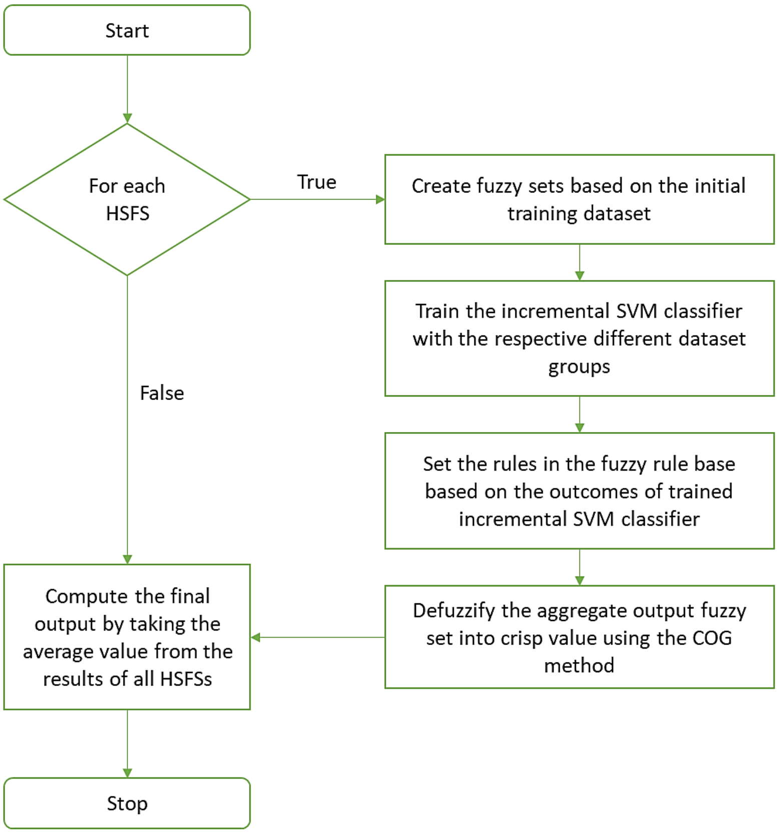

Figure 4. Flowchart of HSFS

2.2.1. Fuzzification in incremental learning based parallel fuzzy reasoning system

In the each HSFS of IL-PFRS, the gaussian membership functions are used to assign the membership degrees in the fuzzy sets which is based on the quartile concept. In (6) illustrates how the mean *m* and standard deviation σ of the fuzzy set *J* can be used to describe the gaussian membership function for any two points on the x-axis.

$$\mu_J\ (x)=\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) \tag{6}$$

For the fuzzification process, initially, the lowest (x_{min}) and highest (x_{max}) data values of the relevant input feature from the respective training dataset are used to determine the bounds of the x-axis in input fuzzy sets. Four points (L_1, L_2, L_3, L_4) are chosen on the x-axis from the input fuzzy sets: L_1 which is the median of x_{min} and earliest quartile value; L_2 which is the median of the earliest quartile value and second quartile value; L_3 which is the median of the second and third quartile values; and L_4 which is the median of the third quartile value and x_{max}. Four gaussian membership functions are used to set the membership degrees based on the four defined points identified on the input fuzzy sets. L_1 is the mean value for the first gaussian membership function, and the standard deviation is estimated between x_{min} and L_2. L_2 is the mean value for the second gaussian membership function, and the standard deviation is estimated between L_1 and L_3. L_3 is the mean value for the third gaussian membership function, and the standard deviation is estimated between L_2 and L_4. L_4 is the mean value for the fourth gaussian membership function, and the standard deviation is estimated between L_3 and x_{max}.

For the input fuzzy set that needs to model the input feature containing categorical data, its x-axis is divided into equal number of sections to depict the number of triangle membership functions as per the respective categorical data where the x-axis bounds are selected at random from 0 to 100. Each triangle membership function on the input fuzzy set represents each categorical data of the respective input feature. Similarly, for the output fuzzy set, where the x-axis is divided into two equal sections to depict two triangle membership functions, the x-axis bounds are selected at random from 0 to 100. Plotting two triangle membership functions on the output fuzzy set represents the thyroid disease dataset's output classes. Figure 5 shows the input fuzzy sets and output fuzzy set obtained from the thyroid disease dataset. The output fuzzy set is same for both HSFS1 and HSFS2. As seen by (7), the triangle membership function can be represented by the three parameters {a, b, c}, which stand for the fuzzy set J's lower limit, peak value, and upper limit, respectively, on the x-axis.

$$\mu_J\ (x)=\begin{cases} 0, & \text{if } x \le a \\ \frac{x-a}{b-a}, & \text{if } a\ \le\ x\ \le\ b \\ \frac{c-x}{c-b}, & \text{if } b\ \le\ x\ \le\ c \\ 0, & \text{if } c\ \le\ x \end{cases} \tag{7}$$

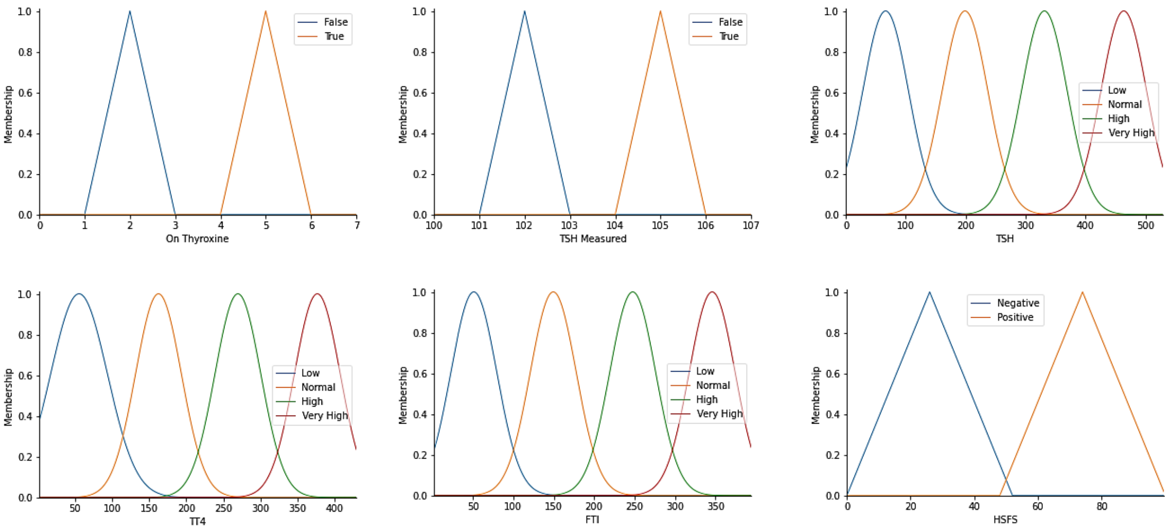

Figure 5. Fuzzy sets of IL-PFRS

2.2.2. Incremental learning based fuzzy rule base

An incremental SGD based SVM classifier (SGD-SVM) is used to set the output of each rule in the fuzzy rule base of HSFS. SVM is one of the prediction techniques that has been extensively accepted for many applications. The curse of dimensionality problem can be handled by using SVM when using the dataset with large number of features. Based on the maximal margin hyperplane approach, SVM classifier performs classification that can be utilized for both linear and nonlinear data. The training samples of datasets are divided into groups using the hyperplanes in accordance with the related class labels as per the maximal margin hyperplane technique. While a large number of hyperplanes can be plotted to divide the samples, not all of them are effective at categorizing the test samples. When compared to hyperplanes with lesser margins, the hyperplane with a higher margin will perform better in classifying the test samples. The data points that lie on the boundaries of various data categories are used in the linear kernel based SVM classifier to identify a hyperplane that divides the training data points displayed on the feature space into distinct categories. These data points which are used to spot the hyperplane are called as support vectors. Given the training data, consider the data points (x_i, y_i). In this case, where i=1, 2, ..., n, x_i is the n-dimensional vector and y_i is the target variable connected to x_i. In (8) illustrates how the decision boundary works, which is used to segregate the training data points.

$$w \cdot x + b = 0 \tag{8}$$

In this case, b is the scalar and w is the n-dimensional weight vector. The training phase involves computation of the parameters w and b. When a certain type of linear model for the data that are linearly separable is identified, the SVM classifier increases the margin of hyperplanes.

Gradient descent is one of popular optimization techniques that is used to estimate the optimum value of cost function in machine learning classifiers. The prediction accuracy of the machine learning classifiers can be increased by using proper appropriate parameters which can be identified by using the gradient descent method. SGD is a simple yet very efficient approach to optimize the parameters of linear kernel based SVM classifier. The SGD method optimizes the parameter of linear kernel based SVM classifier based on the following steps.

a. At first, the parameters of the SVM model are initialized.
b. The parameters of the model are tuned based on the learning rate and number of iterations.
c. The following steps are repeated until the maximum level of iterations is reached.
   – The randomness process is executed by shuffling the dataset.
   – Each batch of dataset are iterated to train the SVM in a shuffled order.
   – The cost function’s gradient is measured based on the weight of the trained SVM classifier.
   – The weights of the SVM classifier are updated.
   – The difference of the cost function between the iteration levels is evaluated.
   – The optimized parameters of SVM classifier are found when reaching the maximum level of iterations.
d. At first, the parameters of the SVM model are initialized.
e. The parameters of the model are tuned based on the learning rate and number of iterations.
f. The following steps are repeated until the maximum level of iterations is reached.
   – The randomness process is executed by shuffling the dataset.
   – Each batch of dataset are iterated to train the SVM in a shuffled order.
   – The cost function’s gradient is measured based on the weight of the trained SVM classifier.
   – The weights of the SVM classifier are updated.
   – The difference of the cost function between the iteration levels is evaluated.
   – The optimized parameters of SVM classifier are found when reaching the maximum level of iterations.

Suppose v be the instance with a pair (x, y) where x is the arbitrary input and y is the scalar output. The weights of the SVM classifier are updated after t level of iteration using (9):

$$wt + 1 = wt - \alpha t \nabla w \, L\,(vt, wt) \tag{9}$$

Here, the cost function L which is estimated based on the predicted output y' and actual output y.

The incremental SGD-SVM incremental learning follows the domain based incremental learning approach. SGD-SVM incremental learning is based on the following steps [41].

a. Initially, the SGD-SVM incremental learning classifier is trained with one dataset.
b. The SGD-SVM incremental learning classifier can process with only one dataset at a time.
c. The trained SGD-SVM incremental learning classifier can be used to train with another dataset without modifying or eliminating the previously extracted patterns.
d. Based on the domain-incremental learning approach, the trained SGD-SVM incremental learning classifier can be used to train with different number of datasets where the newly extracted patterns only get added to the existing patterns.

With help of the respective training dataset group from the respective HSFS, the incremental SGD-SVM classifier is trained. Every rule in the fuzzy rule base has its consequent membership function set by the trained incremental SGD-SVM classifiers. Each HSFS employs the COG defuzzification technique. The average value of the outputs from the HSFSs is the final output of IL-PFRS.

The implementation of IL-PFRS is done using python programming language. Python machine learning library called SciPy toolkit [42] is used to implement the fuzzy logic method and incremental SGD-SVM classifier. The incremental SGD-SVM classifier is implemented based on the SGD classifier by using the ‘partial_fit’ method. When using the SGD classifier with ‘partial_fit’ method, the training dataset is split into number of batches and each batch are used to train the SGD classifier incrementally. Here, an incremental SGD-SVM classifier is obtained by using the hinge loss function along with the ‘partial_fit’ method. To demonstrate the incremental learning method, each training dataset group are split into ten number of batches and each batch of the training dataset group is used to train the SGD-SVM classifier incrementally.

3. RESULTS AND DISCUSSION

The testing dataset group discussed in subsection 2.2 is used to test the IL-PFRS. The effectiveness of IL-PFRS is assessed using performance metrics that are based on the confusion matrix, including accuracy, precision, sensitivity, specificity, and F-measure. Using the test outcomes: true positive (TP), true negative (TN), false positive (FP), and false negative (FN), the confusion matrix is used to compare the classification system's final output with the actual result provided in the dataset. In this case, TP denotes the number of samples that were correctly classified with the positive class label, TN denotes the number of samples that were correctly classified with the negative class label, FP denotes the number of samples that were incorrectly classified with the positive class label, and FN denotes the number of samples that were incorrectly classified with the negative class label. The performance measures: accuracy, sensitivity, and specificity are estimated using as (10) to (12), respectively [43]. The performance measures: precision and F-measure are estimated using as (13) and (14), respectively [44]. The performance of IL-PFRS is also examined using receiver operating characteristics (ROC) curve [45].

$$\text{Accuracy}=\frac{\text{TP+TN}}{\text{TP+TN+FP+ FN}} \tag{10}$$

$$\text{Precision}=\frac{\text{TP}}{\text{TP+FP}} \tag{11}$$

$$\text{Sensitivity}=\frac{\text{TP}}{\text{TP+FN}} \tag{12}$$

$$\text{Specificity}=\frac{\text{TN}}{\text{TN+FP}} \tag{13}$$

$$\text{F-measure}=\frac{2 \times \text{precision} \times \text{sensitivity}}{\text{precision+sensitivity}} \tag{14}$$

Figure 6 shows the accuracy, precision, sensitivity, specificity, and F-measure of incremental SGD-SVM classifier for the first and second training dataset groups which are discussed in section 2.2. Figure 7 shows the accuracy, precision, sensitivity, specificity, F-measure and the area under the ROC curve (AUC) of IL-PFRS for the optimized thyroid disease dataset. As shown in Figure 7, the IL-PFRS shows its best performance by showing the classification accuracy of 99% for the thyroid disease dataset. The experiments were also done to investigate the performance of feature selection process done based on the decision tree classifier. Various machine learning classifiers such as KNN, multi-layer perceptron (MLP) [46], SVM, RF, and NB classifiers were also tested on the thyroid disease dataset. When experimenting with the KNN classifier, the nearest neighbors value was set to three. When experimenting with the MLP classifier, the neural network architecture was set to have one hidden layer with five hidden neurons and SGD technique was used for weight optimization. When experimenting with the NB classifier, the Bernoulli NB model was used. When experimenting with the RF classifier, the value of tree depth was set to two. When experimenting with the SVM classifier, the radial basis function (RBF) kernel was used.

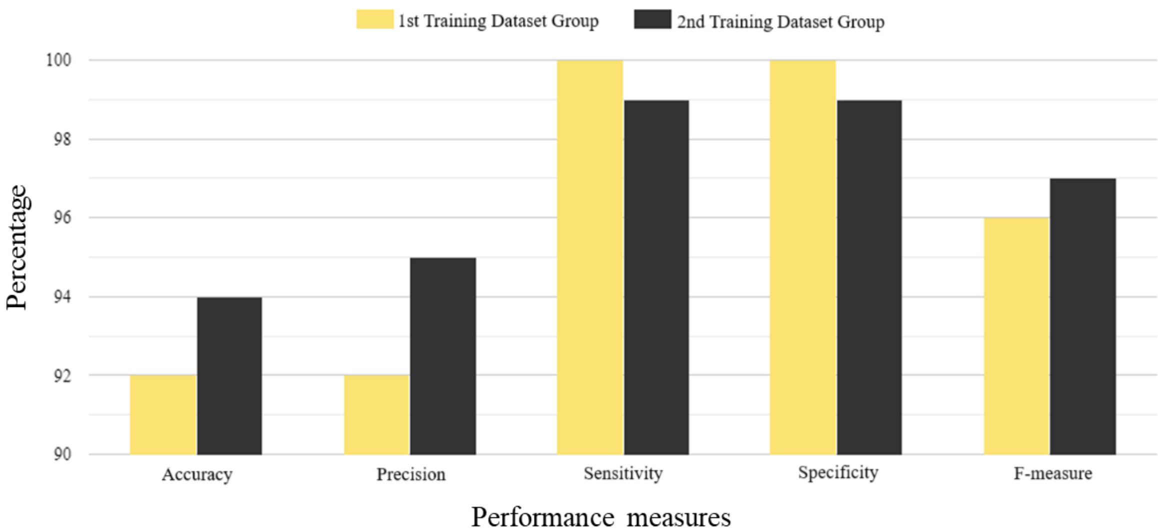

Figure 6. Performance of incremental SGD-SVM classifier for the training dataset groups

Figure 8 shows the accuracy, precision, sensitivity, specificity, F-measure and AUC of KNN, MLP, SVM, RF, and NB classifiers for the whole thyroid disease dataset before the feature selection process. Figure 9 shows the accuracy, precision, sensitivity, specificity, F-measure and AUC of KNN, MLP, SVM, RF, and NB classifiers for the reduced thyroid disease dataset after the feature selection process. When examining Figures 8 and 9, it can be seen that the performance of traditional machine learning classifiers that were evaluated before and after the feature selection process have only little difference between the values of performance measures for the thyroid disease dataset.

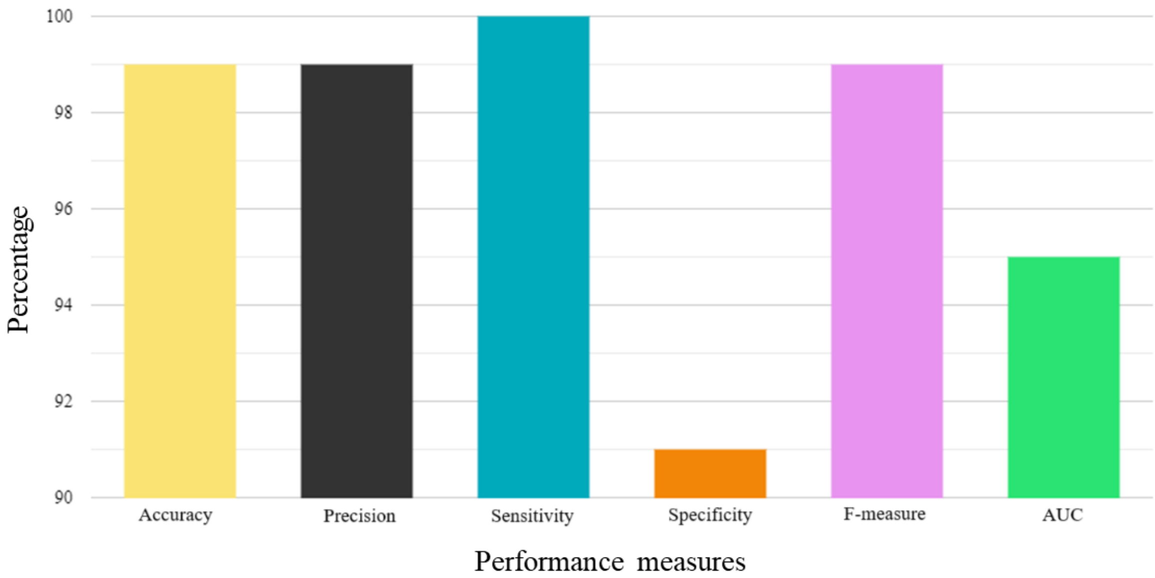

Figure 7. Performance of IL-PFRS

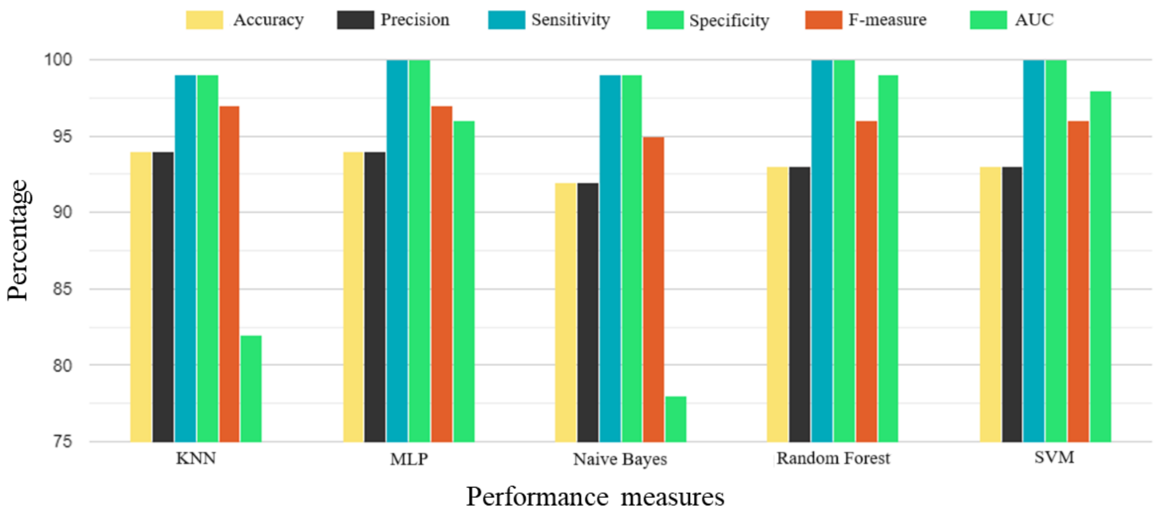

Figure 8. Performance of machine learning classifiers before the feature selection process

The accuracy values of MLP and RF classifiers are relatively high for the reduced dataset after the feature selection process when compared to the dataset before involving the feature selection process. The accuracy of KNN classifier is relatively low for the reduced dataset after the feature selection process when compared to the dataset before involving the feature selection process. The accuracy values of KNN and SVM classifiers are same when testing with and without the feature selection process.

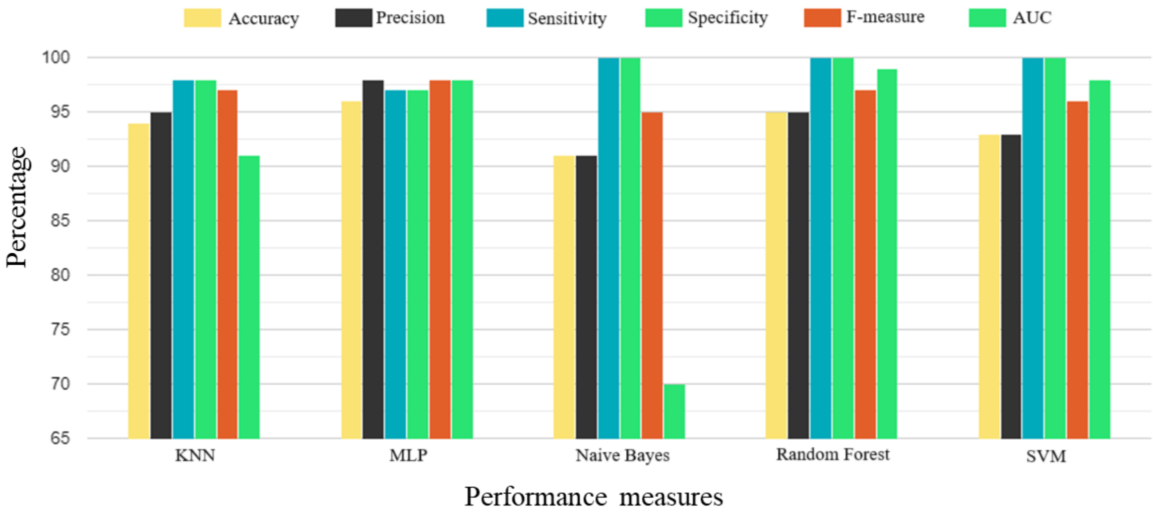

Figure 9. Performance of machine learning classifiers after the feature selection process

The precision values of KNN, MLP, and RF classifiers are relatively high for the reduced dataset after the feature selection process when compared to the dataset before involving the feature selection process. The precision of NB classifier is relatively low for the reduced dataset after the feature selection process when compared to the dataset before involving the feature selection process. The precision of SVM classifier is same when testing with and without the feature selection process.

The sensitivity and specificity of NB classifier are relatively high for the reduced dataset after the feature selection process when compared to the dataset before involving the feature selection process. The sensitivity and specificity values of KNN and MLP classifiers are relatively low for the reduced dataset after the feature selection process when compared to the dataset before involving the feature selection process. The sensitivity and specificity values of RF and SVM classifiers are same when testing with and without the feature selection process.

The F-measure values of MLP and RF classifiers are relatively high for the reduced dataset after the feature selection process when compared to the dataset before involving the feature selection process. The F-measure values of KNN, NB, and SVM classifiers are same when testing with and without the feature selection process.

The AUC values of KNN and MLP classifiers are relatively high for the reduced dataset after the feature selection process when compared to the dataset before involving the feature selection process. The AUC of NB classifier is relatively low for the reduced dataset after the feature selection process when compared to the dataset before involving the feature selection process. The AUC values of RF and SVM classifiers are same when testing with and without the feature selection process.

When analyzing the efficiency of the decision tree algorithm based on the performance metrics for the feature selection process in thyroid disease dataset, it can be seen that the performance of some of the classifiers are improved and the performance of some of the classifiers are slightly reduced after the feature selection process. But the feature selection process by the decision tree algorithm has an additional advantage that it reduces the input variables of the fuzzy system which supports in modelling the parallel fuzzy systems in IL-PFRS as the curse of dimensionality is the main problem when applying the fuzzy logic method in IL-PFRS.

Table 2 compares the performance of IL-PFRS with the traditional machine learning classifiers. As shown in Table 2, the accuracy, precision, and F-measure of IL-PFRS is relatively high when compared to the traditional machine learning classifiers. The sensitivity of the IL-PFRS is relatively high when compared to the KNN and MLP classifiers. The sensitivity is the same for the IL-PFRS, SVM, RF, and NB classifiers. The specificity of the IL-PFRS is relatively low when compared to the traditional machine learning classifiers. The AUC of IL-PFRS is relatively low when compared to the MLP, RF, and SVM classifiers. The AUC of IL-PFRS is relatively high when compared to the KNN and NB classifiers. As the reviewed hybrid fuzzy logic approaches are not applied for the classification of thyroid disease, the performance of IL-PFRS is compared with the performance of traditional machine learning classifiers. Based on the comparison shown in Table 2, it is evident that the IL-PFRS showed an improved accuracy for the classification of thyroid disease. However, in-depth research on IL-PFRS is required to study its performance on classification of other medical data as the medical datasets may involve many uncertain or incomplete data which can be handled by improving the hybrid fuzzy logic method.

Table 2. Performance comparison of IL-PFRS with machine learning classifiers

Classifier	Accuracy	Precision	Sensitivity	Specificity	F-measure	AUC
KNN	0.94	0.95	0.98	0.98	0.97	0.91
MLP	0.96	0.98	0.97	0.97	0.98	0.98
NB	0.91	0.91	1.0	1.0	0.95	0.7
RF	0.95	0.95	1.0	1.0	0.97	0.99
SVM	0.93	0.93	1.0	1.0	0.96	0.98
IL-PFRS	0.99	0.99	1.0	0.91	0.99	0.95

4. CONCLUSION

Although curse of dimensionality is the main problem in applying hybrid fuzzy expert systems for medical data classification, the other main research challenges are the enhancement of machine learning and soft computing techniques that are integrated with the fuzzy expert systems. The IL-PFRS proposed in this paper can overcome the curse of dimensionality problem by using a number of fuzzy reasoning systems in a parallel architecture. In the IL-PFRS, the enhancement of machine learning and soft computing techniques are done by using an incremental SGD-SVM classifier which are integrated to the parallel fuzzy systems. The incremental learning method has an advantage that it can add patterns from the new dataset to the already existing patterns without replacing it. So far, the integration of incremental learning method and fuzzy logic method has not been studied and applied for the classification of thyroid disease. In this paper, the integration of incremental learning method and fuzzy logic method is studied through the proposed IL-PFRS for the classification of thyroid disease. The IL-PFRS showed the effectiveness of the integration of incremental learning classifier and parallel fuzzy reasoning systems for the thyroid disease diagnosis where the IL-PFRS showed the accuracy of 99% for the classification of thyroid disease. The future work will be investigating the performance of IL-PFRS for the classification of multi-labelled medical datasets.

ACKNOWLEDGMENTS

We would like to thank the Research Centre for Advanced Analytics, Multimedia University for providing resources and supervision.

FUNDING INFORMATION

This work was supported by Multimedia University. We gratefully acknowledge the financial and administrative support provided by Multimedia University which contributed significantly to the successful completion of this research.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Thirumalaimuthu Thirumalaiappan Ramanathan	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓			
Md. Jakir Hossen	✓					✓				✓		✓	✓	✓

C : **C**onceptualization	I : **I**nvestigation	Vi : **Vi**sualization
M : **M**ethodology	R : **R**esources	Su : **Su**pervision
So : **So**ftware	D : **D**ata Curation	P : **P**roject administration
Va : **Va**lidation	O : Writing - **O**riginal Draft	Fu : **Fu**nding acquisition
Fo : **Fo**rmal analysis	E : Writing - Review & **E**diting	

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

The research related to human use has been complied with all the relevant national regulations and institutional policies in accordance with the tenets of the Helsinki Declaration and has been approved by the authors' institutional review board or equivalent committee.

DATA AVAILABILITY

The data that support the findings of this study are openly available in UCI machine learning repository at https://archive.ics.uci.edu/dataset/102/thyroid+disease/ [doi: 10.24432/C5D010], [33].

REFERENCES

[1] J. R. Arthur and G. J. Beckett, "Thyroid function," *British Medical Bulletin*, vol. 55, no. 3, pp. 658–668, Jan. 1999, doi: 10.1258/0007142991902538.
[2] P. A. Farling, "Thyroid disease," *British Journal of Anaesthesia*, vol. 85, no. 1, pp. 15–28, Jul. 2000, doi: 10.1093/bja/85.1.15.
[3] Y. Deng *et al.*, "Global Burden of Thyroid Cancer from 1990 to 2017," *JAMA Network Open*, vol. 3, no. 6, p. e208759, Jun. 2020, doi: 10.1001/jamanetworkopen.2020.8759.
[4] T. T. Ramanathan, J. Hossen, and S. Sayeed, "Naïve Bayes based multiple parallel fuzzy reasoning method for medical diagnosis," *Journal of Engineering Science and Technology*, vol. 17, no. 1, pp. 472–490, 2022.
[5] G. I. Webb, "Naïve Bayes," *Encyclopedia of Machine Learning and Data Mining*, vol. 15, pp. 713–714, 2010, doi: 10.1007/978-1-4899-7687-1_581.
[6] T. T. Ramanathan, M. J. Hossen, and M. S. Sayeed, "Blockchain integrated multi-agent system for breast cancer diagnosis," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 2, pp. 998–1008, May 2022, doi: 10.11591/ijeecs.v26.i2.pp998-1008.
[7] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, "*Classification and regression trees*," *Routledge*, 2017, doi: 10.1201/9781315139470.
[8] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *American Statistician*, vol. 46, no. 3, pp. 175–185, Aug. 1992, doi: 10.1080/00031305.1992.10475879.
[9] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
[10] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/bf00994018.
[11] P. Melin, D. Sánchez, and O. Castillo, "Interval Type-3 Fuzzy Inference System Design for Medical Classification Using Genetic Algorithms," *Axioms*, vol. 13, no. 1, p. 5, Dec. 2023, doi: 10.3390/axioms13010005.
[12] S. Haberman, "Haberman's Survival," *UCI Machine Learning Repository*, 1999, doi: 10.24432/C5XK51.
[13] F. Khozeimeh R. Alizadehsani, M. Roshanzamir, and P. Layegh, "Cryotherapy Dataset," *UCI Machine Learning Repository*, 2018.
[14] F. Khozeimeh, R. Alizadehsani, M. Roshanzamir, and P. Layegh, "Immunotherapy Dataset," *UCI Machine Learning Repository*, 2018.
[15] D. P. Kumar, "Pima Indians Diabetes Database," 2016. Online. [Available]: https://www.kaggle.com/code/mragpavank/pima-indians-diabetes-database, (Date accessed: Aug 5, 2024).
[16] B. Ramana and N. Jeevannagaraj, "Indian Liver Patient Dataset," *UCI Machine Learning Repository*, 2012.
[17] F. Patrcio, J. Pereira, J. Crisstomo, P. Matafome, R. Seia, and F. Caramelo, "Breast Cancer Coimbra," *UCI Machine Learning*, p. UCI Machine Learning Repository, 2018, doi: 10.24432/C52P59.
[18] K. R. Vineetha, M. S. Maharajan, B. K, and N. Sivakumar, "Classification of adaptive back propagation neural network along with fuzzy logic in chronic kidney disease," *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, vol. 7, p. 100463, Mar. 2024, doi: 10.1016/j.prime.2024.100463.
[19] L. Rubini, P. Soundarapandian, and P. Eswaran, "Chronic kidney disease," *UCI Machine Learning Repository*, 2015.
[20] R. S. Babatunde *et al.*, "A Neuro-Fuzzy-based Approach to Detect Liver Diseases," *Pakistan Journal of Engineering and Technology*, vol. 7, no. 2, pp. 50–58, 2024, doi: 10.51846/vol7iss2pp50-58.
[21] J. Hao, S. Luo, and L. Pan, "Rule extraction from biased random forest and fuzzy support vector machine for early diagnosis of diabetes," *Scientific Reports*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-14143-8.
[22] University of North Carolina: Carolina Population Center, "China Health and Nutrition Survey," *China Health and Nutrition Survey*, 2013, [Online]. Available: http://www.cpc.unc.edu/projects/china. (Date accessed: Aug, 2024).
[23] K. Gupta, P. Kumar, S. Upadhyaya, M. Poriye, and S. Aggarwal, "Fuzzy Logic and Machine Learning Integration: Enhancing Healthcare Decision-Making," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 16, no. 3, pp. 515–534, 2024.
[24] F. Soriano, "Stroke Prediction Dataset," *Kaggle*, 2021, [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset. (Date accessed: Aug 5, 2024).
[25] F. Soriano, "Heart Failure Prediction Dataset," *Kaggle*, 2021, [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction. (Date accessed: Aug 5, 2024).
[26] F. Soriano, "Body Fat Prediction Dataset," *Kaggle*, 2021, [Online]. Available: https://www.kaggle.com/datasets/fedesoriano/body-fat-prediction-dataset. (Date accessed: Aug 5, 2024).
[27] M. Siddhartha, "Heart disease dataset (comprehensive)," *IEEE Dataport*, 2020.

[28] G. M. van de Ven, T. Tuytelaars, and A. S. Tolias, "Three types of incremental learning," *Nature Machine Intelligence*, vol. 4, no. 12, pp. 1185–1197, Dec. 2022, doi: 10.1038/s42256-022-00568-3.

[29] P. E. Utgoff, "Incremental Induction of Decision Trees," *Machine Learning*, vol. 4, no. 2, pp. 161–186, 1989, doi: 10.1023/A:1022699900025.

[30] R. Polikar, L. Udpa, S. S. Udpa, and V. Honavar, "Learn++: An incremental learning algorithm for supervised neural networks," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 31, no. 4, pp. 497–508, 2001, doi: 10.1109/5326.983933.

[31] D. Roy, P. Panda, and K. Roy, "Tree-CNN: A hierarchical Deep Convolutional Neural Network for incremental learning," *Neural Networks*, vol. 121, pp. 148–160, Jan. 2020, doi: 10.1016/j.neunet.2019.09.010.

[32] P. Netrapalli, "Stochastic Gradient Descent and Its Variants in Machine Learning," *Journal of the Indian Institute of Science*, vol. 99, no. 2, pp. 201–213, Jun. 2019, doi: 10.1007/s41745-019-0098-4.

[33] R. Quinlan, "Thyroid disease," *UCI Machine Learning Repository*, 1987, doi: 10.24432/C5D010.

[34] O. Troyanskaya *et al.*, "Missing value estimation methods for DNA microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, Jun. 2001, doi: 10.1093/bioinformatics/17.6.520.

[35] I. I. Hirschman, "A note on entropy," *American Journal of Mathematics*, vol. 79, no. 1, pp. 152-156, 1957, doi: 10.2307/2372390.

[36] T. Daniya, M. Geetha, and K. S. Kumar, "Classification and regression trees with gini index," *Advances in Mathematics: Scientific Journal*, vol. 9, no. 10, pp. 8237–8247, Sep. 2020, doi: 10.37418/amsj.9.10.53.

[37] Y. Wang, Y. Li, Y. Song, X. Rong, and S. Zhang, "Improvement of ID3 algorithm based on simplified information entropy and coordination degree," *Algorithms*, vol. 10, no. 4, 2017, doi: 10.3390/a10040124.

[38] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *International Journal of Man-Machine Studies*, vol. 7, no. 1, pp. 1–13, Jan. 1975, doi: 10.1016/S0020-7373(75)80002-2.

[39] M. Negnevitsky, "Artificial Intelligence: A Guide to Intelligent Systems," North York, Ontario: *Pearson Education Limited*, vol. 2, p. 414, 2005.

[40] W. Van Leekwijck and E. E. Kerre, "Defuzzification: Criteria and classification," *Fuzzy Sets and Systems*, vol. 108, no. 2, pp. 159–178, Dec. 1999, doi: 10.1016/s0165-0114(97)00337-0.

[41] D. Snell and Q. Wu, "An Introduction to Incremental Learning," *Predictive Analytics and Futurism*, no. 13, pp. 18–21, 2016.

[42] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[43] W. Zhu, N. Zeng, and N. Wang, "Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS® implementations," *Northeast SAS Users Group 2010: Health Care and Life Sciences*, pp. 1–9, 2010.

[44] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *International Journal of Machine Learning Technology*, vol. 2, no. 1, pp. 37–63, Oct. 2011.

[45] T. Fawcett, "An Introduction to ROC Analysis," *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861–874, 2006.

[46] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychological Review*, vol. 65, no. 6, pp. 386–408, 1958, doi: 10.1037/h0042519.

BIOGRAPHIES OF AUTHORS

Thirumalaimuthu Thirumalaiappan Ramanathan is currently working as a research fellow in the Faculty of Engineering and Technology, Multimedia University, Malaysia. He received the bachelor of engineering degree in computer science from Anna University, India in 2010. He received the master by research degree in information science from University of Canberra, Australia in 2016. He received the Ph.D. degree from Multimedia University, Malaysia in 2024. His research interests are application of artificial intelligence techniques in health informatics. He can be contacted at email: thirumalaimuthu@mmu.edu.my.

Md. Jakir Hossen is currently working as an Associate Professor in the Department of Robotics and Automation, Faculty of Engineering and Technology, Multimedia University, Malaysia. He received the master degree in communication and network engineering from Universiti Putra Malaysia, Malaysia in 2003. He received the Ph.D. degree in smart technology and robotic engineering from Universiti Putra Malaysia, Malaysia in 2012. His research interests are application of artificial intelligence techniques in data analytics, robotics control, data classifications, and predictions. He can be contacted at email: jakir.hossen@mmu.edu.my.