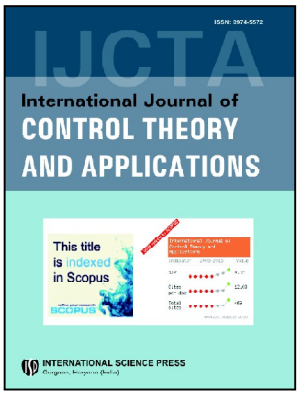

International Journal of Control Theory and Applications

ISSN : 0974-5572

Volume 10 • Number 11 • 2017

Hybridization of Fuzzy and Maximum Relevance Minimum Redundancy Techniques for Effective Stroke and Heart Disease Diagnosis

K. S. Thirunavukkarsu[1] and S. Sukumaran[2]

[1]Ph.D Research Scholar, Dept of Computer Science, Research and Development Centre, Manonmaniam Sundaranar University and Assistant Professor, Sri Amman Arts and Science College, Erode, Tamilnadu, India, E-mail: thirukst@gmail.com
[2]Associate Professor, Dept of Computer Science, Erode Arts and Science College, Erode, Tamilnadu, India
E-mail:prof_sukumar@yahoo.co.in

Abstract: The availability of clinical datasets encourages medical practitioners and researchers to pursue research in disease diagnosis. Different data mining methods have been used for feature selection, classification and mathematical models designed to assist the medical practitioners in decision making. The objective of this work is to build a classifier model for effective diagnosis of stroke and heart disease by learning from the minimal set of attributes that has been extracted from the clinical dataset. In this work Maximum Relevance Minimum Redundancy method with Trapezoidal Fuzzy Classifier (MRMR-TFC) is used. This work has two stages. The first stage is extracting more relevant attributes and reducing the redundant attributes using MRMR algorithm, resulting in high-level features. The second stage classifies the high-level features using Trapezoidal Fuzzy-based Classifier, improving classification accuracy. The classifier has been tested with stroke and heart disease dataset to evaluate the classification time and error rate. The experimental results show that MRMR-TFC has an effective performance for diagnosing stroke and heart disease with 15% and 14% accuracy and superior performance compared to other methods in the literature.

Keywords: Data Mining, Feature Selection, Maximum Relevance, Minimum Redundancy, Fuzzy Classifier

1. INTRODUCTION

One of the most recent studies in the field of medical data conducted by several researchers is to help the surgeons make effective decision making for disease diagnosis. Though many researchers have contributed in the area of medical data classification, still now, several problems remained unaddressed.

Rough Fuzzy Classifier [1] was designed to predict the heart disease that combined rough set theory with the fuzzy set. However, analysis to find relevant attributes for prediction of heart disease remained unsolved. Kernel Penalized Support Vector Machine (KP-SVM) [2] employed an explicit stopping condition that helped in eliminating the features that had adverse effects on classifier performance.

Various classification and regression methods have been used by several researchers in identifying the heart disease. In [3], to improve the accuracy of disease diagnosis rate being detected, a number of computational

intelligence techniques were used. A multistage classification algorithm was designed in [4] using two level classification that refined the prediction done by the first stage and later reduced the number of false positives. Classification algorithms like Naive Bayes, CART [5] were used to predict the heart disease. However, the time for disease diagnosis remained unsolved.

In [6], a cross validation approach using brain-computer interface was designed resulting in the improvement of true positive and reducing the false positive results, detecting the disease at an early stage. In [7], a group incremental approach for feature selection using rough set technique was investigated. This in turn found the new feature subset in a much shorter time. In [8], event categorization and prediction based on temporal patterns was designed in a dynamic data system. However, the design did not involved early disease diagnosis.

With the objective of detecting and diagnosis of disease at an early stage, in [9], an hybrid approach combining random forest, Naive Bayes and Support Vector Machine utilized a combination of attributes was designed. In [10], weighted fuzzy rules were applied for the diagnosis of heart disease improving sensitivity and specificity.

In this paper, we propose a novel method for stroke and heart disease diagnosis from medical datasets. This proposed method provides an efficient means of disease diagnosis through Maximum Relevance Minimum Redundancy algorithm and Trapezoidal Fuzzy Classifier, and thus achieves higher classification accuracy reducing the classification time.

The structure of paper is as follows. In Section 1, medical data classification for disease diagnosis with existing works is described. In Section 2, literatures related to disease diagnosis and classification obtained with respect to medical data is elaborated. Section 3 explains about the proposed work Maximum Relevance Minimum Redundancy with Trapezoidal Fuzzy Classifier (MRMR-TFC) with neat architecture diagram and algorithmic steps to increase the classification accuracy and reduce the classification time. Section 4 analyzes the experimental results and Section 5 provides the result analysis using pictorial representation in graphical form. Finally, the concluding remarks are included in Section 6.

2. RELATED WORKS

By minimizing the attributes in medical data, a number of aspects such as speed, storage and accuracy get improved. Upper and lower approximation reductions was investigated in [11] in ordered complicated decision tables with fuzzy decision resulting in the improvement of computational time. A case control study was analyzed in [12] for non alcoholic fatty liver disease. In [13], a systematic review of studies comparing prediction rules with clinical judgement was presented. In [14], prediction and personalized treatment for stroke prevention was presented.

Heart rate variability analysis was made in [15] using automatic prediction of cardiovascular and cerebrovascular events, aiming at improving the prediction rate. Another model was presented in [16] to predict the cardiovascular disease risk using real life data. In [17], a systematic review of studies was presented with the aid of diagnostic clinical rules resulting in the improvement of true positive rate. Multivariate pattern analysis was performed in [18] using support vector regression to improve the diagnostic classification rate. In [19], another method to enhance the prediction of cardiovascular disease was presented.

Based on the aforementioned techniques and methods, a design of medical data classification based on trapezoidal fuzzy classifier to improve the classification accuracy on medical data is explained in the forthcoming sections.

3. MAXIMUM RELEVANCE MINIMUM REDUNDANCY WITH TRAPEZOIDAL FUZZY CLASSIFIER

One of the most commonly used methods for diagnosis of stroke and heart disease is fuzzy logic model [1]. Fuzzy logic model is considered an effective classifier model if it has relevant training data, reducing the redundant attributes present in it. So, in this work, a combination of a Maximum Relevance Minimum Redundancy algorithm

and Trapezoidal Fuzzy Classifier in one system is designed called Maximum Relevance Minimum Redundancy with Trapezoidal Fuzzy Classifier (MRMR-TFC), which combines the advantages of both approaches.

3.1. Maximum Relevance Minimum Redundancy algorithm

The MRMR-TFC method extracts the features from stroke and heart dataset for disease diagnosis serves as inputs for classification. Feature selection is one of the predominant steps to be followed for reducing classification time for disease diagnosis. The extracted features when stored in an attribute matrix involve a time consuming process.

In the MRMR-TFC, the features extracted are stored in an attribute matrix and instead of using all the features present in the dataset high-level features that are important in terms of the performance are extracted. This in turn reduces classification time and therefore the processing time for disease diagnosis. For this purpose, the MRMR-TFC uses Maximum Relevance Minimum Redundancy (MRMR) algorithm. Table 1 describes the attributes of heart disease and table 2 describes the attributes of stroke disease respectively.

Table 1
Description of heart disease dataset (Statlog heart disease dataset)

Attribute	*Description*	*Domain value*
Age	Age in years	29 – 77
Sex	Sex	[1, 0]
CP	Chest pain type	[1, 2, 3, 4]
Trestbps	Resting blood sugar	94 to 200mm Hg
Chol	Serum Cholesterol	126 to 564mg/dl
Fbs	Fasting blood sugar	>120mg/dl, True (1),False (0)
Restecg	Resting ECG result	[0, 1, 2]
Thalach	Maximum heart rate achieved	71 to 202
Exang	Exercise induced angina	[1, 0]
Oldpeak	ST depression induced by exercise relative to rest	0 to 6.2
Slope	Slope of peak exercise	[1, 2, 3]
Ca	Number of major vessels coloured by fluoroscopy	[0 to 3]
Thal	Defect type	[3, 6, 7]
Num	Heart disease	0 – 4

Table 2
Description of stroke disease database (Baseline variables from International Stroke Trial Database)

Attribute	*Domain value*
Age	29 – 77
Sex	[1, 0]
Arthritis	[1, 0]
Hypertension	[1, 0]
Heart disease	[1, 0]
Diabetes	[1, 0]
Alcohol	Regular or occasional drinker, No drinks past 12 months
BMI class	(Normal, Overweight, Obsese)
Smoking status	(Never, Smokes Daily, former smoker)
Physical activity	(Inactive, Active, Moderate)

The attributes provided in table 1 and table 2 servers as input (i.e. patient's data). Figure 1 shows the structure of Maximum Relevance Minimum Redundancy model.

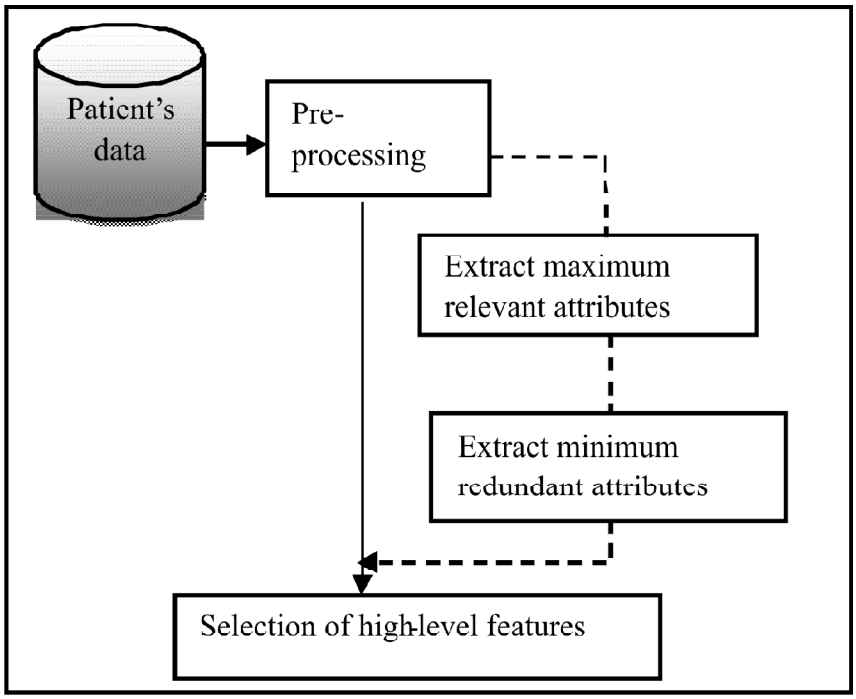

Figure 1: Maximum Relevance Minimum Redundancy model

As shown in the figure 1, the Maximum Relevance Minimum Redundancy model obtains the patients data (from heart disease and stroke disease dataset) and performs pre-processing. To reduce the classification time, high-level features comprising of maximum relevant and minimum redundant attributes are extracted using MRMR algorithm.

Let us consider an information system table represented in the form of two dimensions and applied to the MRMR algorithm. The MRMR algorithm in the MRMR-TFC shows the number of objects in row whereas the column shows the values of the attributes and class label of the objects. It is expressed as given below.

$$I = (O, A \cup C) \tag{1}$$

From (1), information system table '*I*' is expressed with finite set of object '*O*' (i.e. stroke and heart dataset), '*A*' representing the set of attributes (present in stroke and heart disease dataset) and class label (stored in the form of class vector) denoted as '*C*' respectively.

The MRMR algorithm selects the attributes that are most relevant to the class vectors and filter out the rest of the attributes. While identifying the most relevant attributes, the MRMR algorithm in the MRMR-TFC minimizes the redundancy among the selected attributes. To measure the similarity between two attributes and the class vector, the MRMR uses mutual information factor '*I*(*a*, *b*)' and is expressed as given below.

$$I(P,Q) = \Sigma_{i,j}\, prob(p_i,q_j) \log \frac{prob(p_i q_j)}{prob(p_i) prob(q_i)}, p,q \in A \tag{2}$$

For '*p*' and '*q*' attributes from (2), '*prob* (p_i)' and '*prob* (q_j)' represents probability functions whereas '*prob*(p_i q_j)' forms the combined probability distribution factor. Let '' represents the attribute to be selected,

while '$|A|$' denotes the number of elements present in the stroke and heart disease dataset. In order to select the attribute, the MRMR algorithm in the MRMR-TFC evaluates the maximum relevant '*XREL*' and minimum redundant '*NRED*' attributes and is expressed as given below.

$$XREL = \frac{1}{|A|}\Sigma_{i\in A} I(C,i) \tag{3}$$

$$NRED = \frac{1}{|A|^2}\Sigma_{i,j\in A} I(i,j) \tag{4}$$

From (3), '$C = \{c_1, c_2, ..., c_n\}$' represent the class variable of stroke and heart disease dataset with 'n' possibilities. With the maximum relevant and minimum redundant attributes, high-level features '*HF*' are evaluated as expressed below.

$$HF = I(XREL, NRED) = \Sigma_{i,j\in A} prob(XREL_i, NRED_j) \log \frac{prob(XREL_i, NRED_j)}{prob(XREL_i) prob(NRED_j)} \tag{5}$$

Figure 2 shows the Maximum Relevance Minimum Redundancy (MRMR) algorithm.

Input: set of attributes 'A', class label 'C', objects 'O'
Output: Extracted high-level features
1: Begin 2: For attributes 'A' in objects 'O' 3: Measure mutual information factor using (2) 4: Measure maximum relevant attributes using (3) 5: Measure minimum redundant attributes using (4) 6: Measure high-level features using (5) 7: End for 8: End

Figure 2. Maximum Relevance Minimum Redundancy (MRMR) algorithm

Based on the given steps in the MRMR algorithm, the feature selected of table 3 is computed as shown in figure 2. The information system table shown in table 3 has two high-level features as shown in the MRMR algorithm. They are '{*Chap, Vessel*}' and '{*ECG, Vessel*}' respectively, based on three attributes '*Cp, ECG, Vessel*'. This attributes are taken from the heart disease dataset information.

Table 3
Sample of information table of heart disease

O	*A*			*C*
	Cp	*ECG (Restecg)*	*Vessel, (ca)*	
P_1	2	0	3	Yes
P_2	1	2	1	No
P_3	3	2	3	Yes
P_4	2	0	0	No
P_5	3	0	0	No

Table 3 shows that the attribute information about the heart disease. Cp denotes the chest pain types which has the four values. The value 1 represent the typical angina, value 2 is the atypical angina; value 3 is the non-anginal pain, value 4 is the asymptomatic.

In second column, ECG is the electro cardio graphic results. This has 3 values such as 0, 1, and 2. The value 0 is the normal. Value 1 is the ST-T wave abnormality. Value 2 showing particular left ventricular hypertrophy by Estes' criteria.

In third column, number of major vessels which is varied from 0 to 3 and it is coloured by flourosopy. Based on these three attribute information values, the class (C) variables results are obtained.

3.2. Trapezoidal Fuzzy Rule-based Classifier model

With the high-level features (i.e. attributes) obtained using the MRMR algorithm, the MRMR-TFC uses a Trapezoidal Fuzzy Rule-based Classifier model using stroke and heart disease dataset. The MRMR-TFC uses trapezoidal membership functions for medical variables and improves the classification performance by adjusting the certainty factor of each rule. Figure 3 shows the structure of classifier model used in the MRMR-TFC.

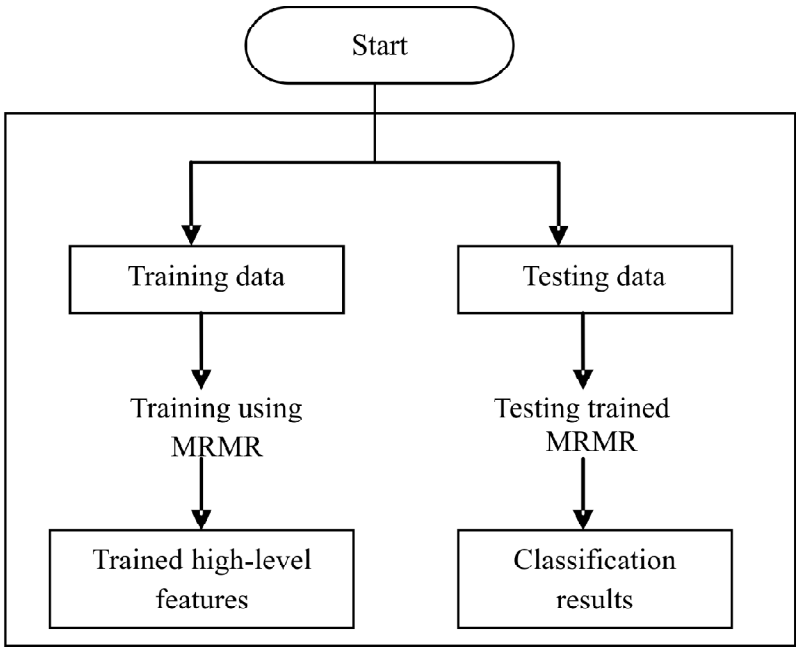

Figure 3: Structure of Classifier

As shown in the figure 3, the raw stroke and heart disease dataset is trained using MRMR algorithm to obtained high-level features. With the obtained high-level features, the training data is tested using the trapezoidal fuzzy rule-based classifier. To start with, the number of attributes (i.e. high-level features) is equal to the number of input variables.

Each attribute represents one fuzzy set, in which the measure for the trapezoid membership function of the input variable (i.e. high-level attributes) to the fuzzy set is performed. Each attribute represents one fuzzy rule. Therefore, the number of attributes is related to the number of fuzzy rules and is expressed as given below.

$$Rule\ R_i : If\ \ a_1\ is\ A_{i1}\ and, ..., and\ a_n\ is\ A_{in}\ then\ class\ C_i\ with\ CF_i,\ i = 1,2,..,n \tag{6}$$

$$\mu_i(Y_p) = \mu_{i1}(y_{p1}) * ... * \mu_{in}(y_{pn}) \tag{7}$$

From (7), '$\mu_i(Y_p)$ is the degree of membership function of fuzzy set 'A_{ij}' where the parameter of j th antecedent member ship function in ith rule. A membership function for a fuzzy set A_{ij} is defined as μA: $(Y_p) \rightarrow [0,1]$, where each element of Y_p is mapped to a value between 0 and 1. This value is called as degree of membership. The value of 0 is said to be Y_p is not a member in fuzzy set whereas 1 represents the fully member in fuzzy set.

For each class, 'C' is calculated where '$\alpha_c(R_i)$' is expressed as given below.

$$\alpha_c(R_i) = w_i * \Sigma_{i=1}^{n} \mu_i(Y_p) \tag{8}$$

From (8), $\alpha_c(R_i)$ represents the membership function for each class C, 'w_i', forms the weighting coefficient which is multiplied with degree of membership function of fuzzy set 'A_{ij}. The certainty factor 'CF_i' of each rule is obtained as given below.

$$CF_i = \mu_i(Y_p) * \alpha_c(R_i) \tag{9}$$

From (9), the certainty factor 'CF_i' is measured using the product of the fuzzy rule formation for each class '$\mu_i(Y_p)$' and membership function for each class '$\alpha_c(R_i)$' respectively. Finally, to improve the classification performance (i.e. classification accuracy), the certainty factor of each rule is adjusted.

When a training pattern with stroke and heart disease dataset is not successfully classified using the fuzzy if-then rule (1), the MRMR-TFC decreases the grade of certainty 'NCF_i^{new}' and is expressed as given below.

$$NCF_i^{new} = (CF_i - \beta) * (w_i * CF_i) \tag{10}$$

On the other hand, if the training pattern with stroke and heart disease is successfully classified, the MRMR-TFC increases the grade of certainty 'NCF_i^{new}' expressed as given below.

$$NCF_i^{new} = (CF_i + \beta) * (w_i * CF_i) \tag{11}$$

From (10) and (11) 'β' represents the constant factor with 'w_i' representing the weight coefficient. Figure 4 shows the Trapezoidal Fuzzy Rule-based Classifier (TFRC) algorithm.

As shown in figure 4, for each high-level features extracted using MRMR algorithm, the Trapezoidal Fuzzy Rule-based Classifier algorithm uses certainty factor to improve the classification performance or

Input: High-level features '***HF***',
Output: Improved disease diagnosis rate
1: Begin 2: For each High-level features '***HF***' 3: Measure membership function using (7) 4: Evaluate the rule for each class using (8) 5: Measure the certainty factor using (9) 6: If training pattern successfully classified 7: Measure the new certainty factor using (10) 8: End if 9: If training pattern not successfully classified 10: Measure the new certainty factor using (11) 11: End if 12: End for 13: End

Figure 4: Trapezoidal Fuzzy Rule-based Classifier algorithm

classification accuracy. With the high-level features obtained, the TFRC algorithm measures the membership function and evaluates the rule for each class.

Next, certainty factor is obtained and the training pattern is used to measure whether the patterns are successfully classified or not. According to the results of the classification, the certainty factor is either incremented or decremented by 1. This in turn improves the classification performance rate for stroke and heart disease diagnosis.

4. EXPERIMENTAL SETUP

In this section, the method Maximum Relevance Minimum Redundancy with Trapezoidal Fuzzy Classifier (MRMR-TFC) is tested with a comprehensive set of experiments. In order to allow reproducibility, all the Heart disease dataset and Stroke disease database freely available in UCI repository are used for experimentation. Maximum Relevance Minimum Redundancy with Trapezoidal Fuzzy Classifier (MRMR-TFC) uses JAVA language with WEKA tool to perform the experimental work. The dataset details for conducting experiments in MRMR-TFC method is provided in table 1 and table 2.

Maximum Relevance Minimum Redundancy with Trapezoidal Fuzzy Classifier (MRMR-TFC) is compared against the existing methods Rough Fuzzy Classifier (RFC) [1] and Kernel Penalized Support Vector Machine (KP-SVM) [2]. The experiment is conducted on the factors such as classification time and classification accuracy with respect to patient's data using the heart disease and stroke database.

5. RESULTS ANALYSIS OF MRMR-TFC

The performance of Maximum Relevance Minimum Redundancy with Trapezoidal Fuzzy Classifier (MRMR-TFC) method is compared with the existing Rough Fuzzy Classifier (RFC) [1] and Kernel Penalized Support Vector Machine (KP-SVM) [2]. The performance is evaluated according to the following metrics.

5.1. Impact of classification time

The classification time for stroke and heart disease diagnosis measures the time required to classify the attributes for disease diagnosis with respect to number of patients. The classification time for disease diagnosis is formulated as given below.

$$CT = (\Sigma_{i=1}^{n} Pat_i * Time(Disease\,diagnosis)) \tag{12}$$

From (12), the classification time 'CT' is measured using the input data size 'Pat_i' in terms of milliseconds (ms). Lower classification time ensures the effectiveness of the method. The classification time is tabulated in table 4 for heart disease dataset and stroke disease database with respect to the number of features (i.e. patient's)

Table 4
Tabulation for computation time

No. of patient	*Classification time (ms) – for heart disease dataset*			*Classification time (ms) – for stroke disease database*		
	MRMR-TFC	*RFC*	*KP-SVM*	*MRMR-TFC*	*RFC*	*KP-SVM*
5	6.16	6.41	6.95	7.05	7.89	8.50
10	8.37	8.92	9.04	12.35	13.52	14.25
15	10.95	11.15	11.89	18.82	20.15	22.05
20	13.53	13.98	14.78	25.98	28.32	29.89
25	15.89	16.78	17.29	32.14	35.91	36.18
30	18.15	19.05	20.14	39.13	44.21	45.23

given as input and comparison of our method MRMR-TFC is made with RFC and KP-SVM. The figurative representation is shown in figure 5 and 6 respectively.

The MRMR-TFC method is analyzed against RFC [1] and KP-SVM [2]. Each method has its own respective classification time. The existing and proposed result is analyzed by providing several features in JAVA using the values provided in the table and graph points. The classification time for disease diagnosis is reduced using MRMR-TFC method than the state-of-art methods.

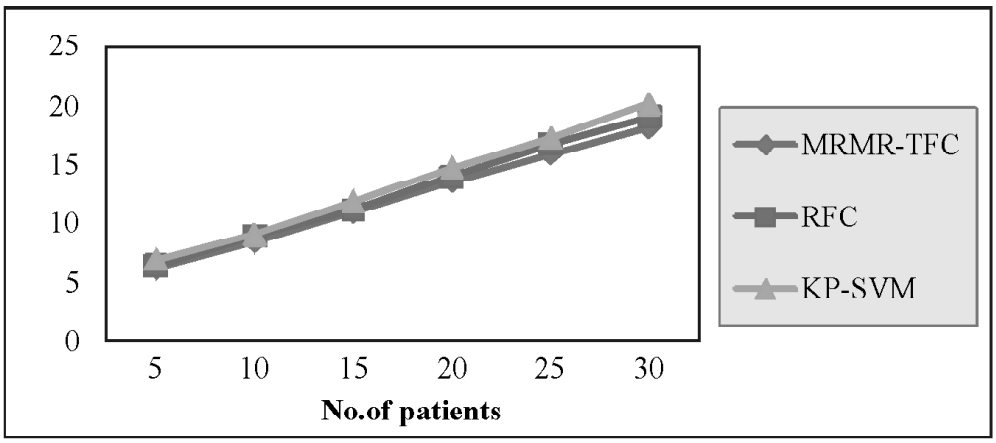

Figure 5: Measure of classification time (using heart disease dataset)

Figure 5 shows the time taken to perform classification on medical data extracted from UCI repository (heart disease dataset) with respect to 30 patient's for experimental purposes. As depicted in the figure with the increase in the number of patient's, the classification time is also increased. But when compared to the state-of-the-art works, the classification time is reduced in the MRMR-TFC method. The classification time on medical data is improved owing to the fact that the MRMR-TFC method uses Maximum Relevance Minimum Redundancy algorithm.

By applying MRMR algorithm, relevant attributes with minimum redundancy factor is considered for classification in the MRMR-TFC method. This in turn reduces the classification time using MRMR-TFC by 4% compared to RFC. Moreover, with the application of Mutual Information factor (5), features are not classified on a whole, but are classified based on the similarity of features helping the surgeons to reduce the classification time by 9% compared to KP-SVM.

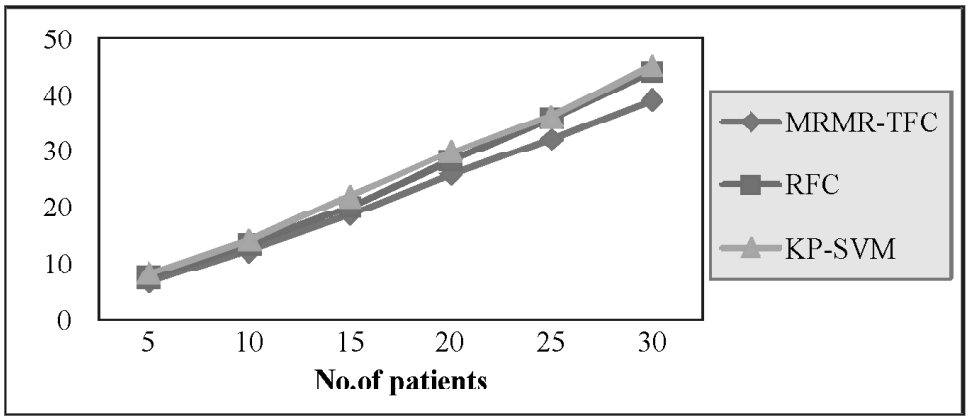

Figure 6: Measure of classification time (using stroke disease database)

Figure 6 shows the measure of classification time using MRMR-TFC, RFC [1] and KP-SVM [2] respectively using stroke disease database. The classification time is observed to be more using stroke disease dataset than compared to heart disease dataset. This is because the high-level features obtained using stroke disease database is observed to be higher than when applied using heart disease dataset.

However, as shown in the figure, the MRMR-TFC method minimizes the classification time for disease diagnosis when compared to RFC [1] method and KP-SVM [2] method. This is because of diagnosis of disease performed using high-level features extracted from MRMR algorithm, rather than with the attributes without pre-processing that removes the redundant attributes and therefore obtains minimum subset of attributes. Therefore, the classification time for disease diagnosis is reduced compared to the state-of-the-art methods. By applying the MRMR algorithm, unwanted attributes are removed reducing the classification time for disease diagnosis by 10% compared to RFC and 16% compared to KP-SVM respectively.

5.2. Impact of classification accuracy

The classification accuracy is one of the important metrics to ascertain the effectiveness of the method. The classification accuracy helps in determining the number of patients correctly diagnosed with stroke or heart disease. The classification accuracy is formulated as given below.

$$A = \left(\frac{Total\ number\ of\ patients\ correctly\ diagnosed}{Pat_i} \right) * 100 \tag{13}$$

Where 'A', corresponds to the classification accuracy measure made with number of patient's correctly diagnosed with stroke and heart disease, to the total patients 'Pat_i' considered for experimentation. It is measured in terms of percentage with higher quality measure ensuring the effectiveness of the method.

Table 5
Tabulation for classification accuracy

No. of patients	*Classification accuracy (%) – for heart disease dataset*			*Classification accuracy (%) – for stroke disease database*		
	MRMR-TFC	*RFC*	*KP-SVM*	*MRMR-TFC*	*RFC*	*KP-SVM*
5	92.14	86.14	73.28	90.25	85.14	71.34
10	90.29	84.23	71.29	89.35	78.45	69.24
15	87.35	81.25	68.45	85.53	80.98	66.78
20	85.28	79.14	66.23	83.78	77.14	64.45
25	83.14	76.32	63.14	81.42	74.63	61.32
30	80.26	74.28	61.23	78.32	72.35	60.32

Table 5 summarizes the three methods that we experimented for disease diagnosis using Heart disease Data Set and stroke disease database with respect to 30 different patients. With respect to the increasing number of patients, though the classification accuracy of disease diagnosis is reduced, but shows gradual improvement by applying MRMR-TFC method when compared to RFC and KP-SVM.

A comparative analysis for classification accuracy with respect to different number of patient's with the existing RFC and KP-SVM is shown in Figure 7. The increasing number of patient's in the range of 5 to 30 is considered for experimental purpose for disease diagnosis using heart disease dataset. As illustrated in figure, comparatively while considering higher number of patient's, the classification accuracy also decreases, though

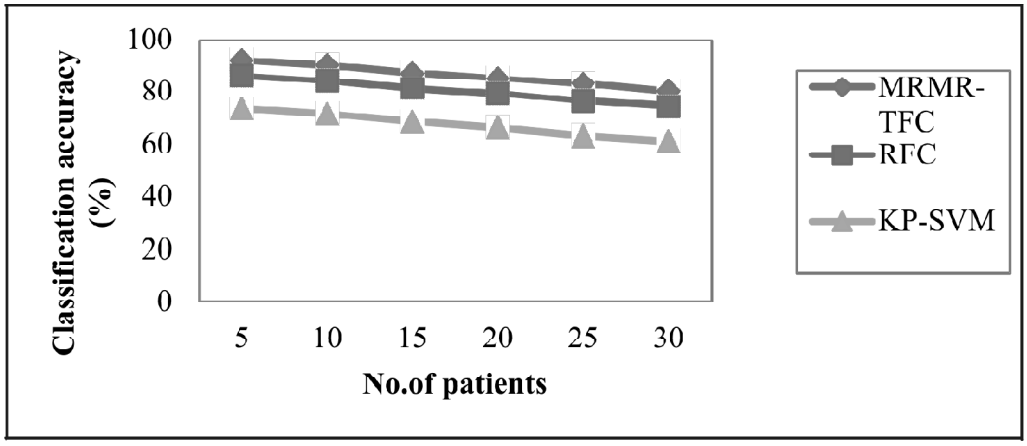

Figure 7: Measure of classification accuracy (using heart disease dataset)

betterment achieved using the MRMR-TFC method. We consider the experimental process for a subset of 3 attributes for experimental purpose, with the resultant values shown in table 3.

The significant improvement in classification accuracy is because of the application of Trapezoidal Fuzzy Rule-based Classifier model that distributes the function into correct and wrongly classified data into two improving the classification accuracy by 7% than RFC [1]. Also by adjusting the certainty factor on the basis of correct and wrong classification made, the classification accuracy in MRMR-TFC is improved by 22% compared to KP-SVM.

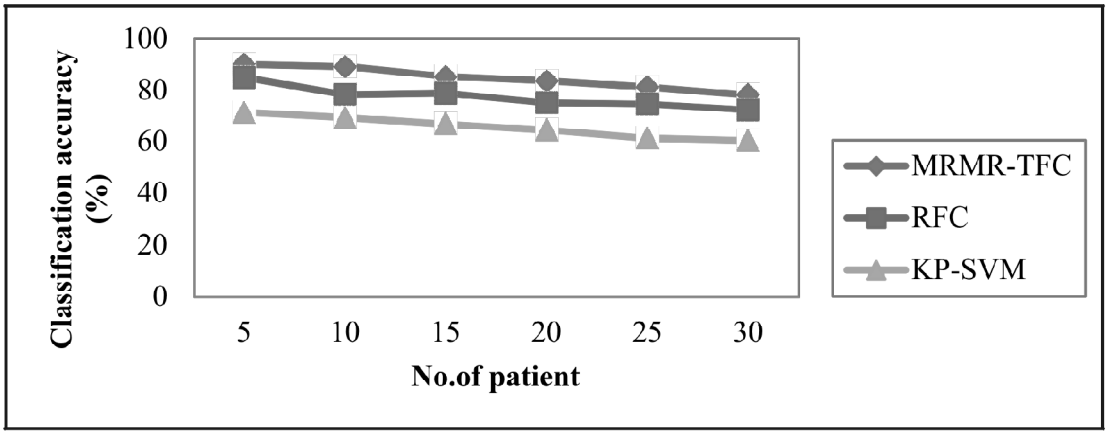

Figure 8: Measure of classification accuracy (using stroke disease database)

To ascertain the performance of the classification accuracy of disease diagnosis, comparison is made with two other existing methods RFC and KP-SVM. In figure 8, the number of patient is varied between 5 and 30. From the figure it is illustrative that the classification accuracy of disease diagnosis for different number of patient is improved using the MRMR-TFC method when compared to the two other existing methods. This is because using certainty factor for each rule in MRMR-TFC method, the fuzzy rule formation clearly differentiates between classified and non-classified rules. This in turn improves the classification accuracy of disease diagnosis using stroke disease database by 8% compared to RFC [1]. Furthermore, based on the high-level features extracted using MRMR algorithm, trapezoid membership function is used for each class. This symbolizes the improved quality of MRMR-TFC method by 22% when compared to KP-SVM [2] method.

5.3. Impact of error rate

Error rate in the disease diagnosis is defined as number of patients incorrectly diagnosed as stroke or heart disease. The error rate is measured in terms of percentage (%). Lower the error rate more efficient the method is said to be. The mathematical formula for error rate is expressed as follows,

$$Error\ rate = \frac{No.of\ patient\ incorrectly\ diagonized\ as\ disease}{No.of\ patients} * 100 \tag{14}$$

Table 6
Tabulation for error rate

No. of patients	*Error rate (%) – for heart disease dataset*			*Error rate (%) – for stroke disease dataset*		
	MRMR-TFC	*RFC*	*KP-SVM*	*MRMR-TFC*	*RFC*	*KP-SVM*
5	12.32	15.24	18.21	13.89	16.35	20.15
10	14.15	18.65	21.35	16.20	20.32	24.65
15	16.35	22.20	24.86	18.54	25.65	28.11
20	20.37	24.58	26.24	21.35	28.59	30.34
25	22.34	28.20	31.20	25.36	31.22	34.67
30	24.58	30.58	33.36	28.62	33.65	36.48

Table 6 shows that the Error rate measurement with three methods using Heart disease Data Set and stroke disease dataset based on 30 different patients. While increasing the number of patient, the error rate gets increased. But comparatively it is reduced in MRMR-TFC method using two different dataset when compared to existing RFC [1] and KP-SVM [2].

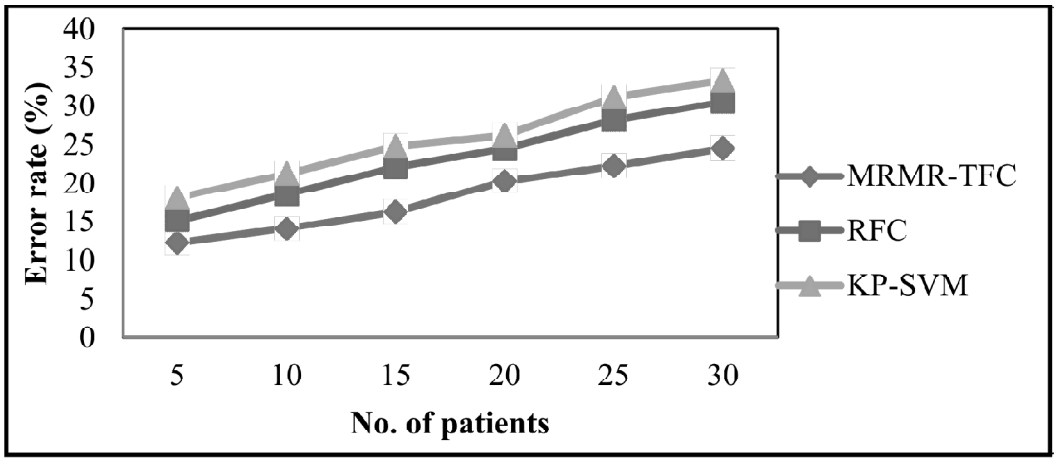

Figure 9: Measure of Error rate (using Heart disease dataset)

Figure 9 illustrates the performance analysis of error rate using heart disease dataset with respect different number of patients. From the figure it is clearly evident that the MRMR-TFC method reduces the incorrect classification of disease. This is because, the MRMR-TFC method extracts the relevant attributes and mimeses the irrelevant attributes in dataset using maximum Relevance Minimum Redundancy algorithm. In addition, the Trapezoidal Fuzzy Rule-based Classifier algorithm is applied in MRMR-TFC method for efficiently diagnosing

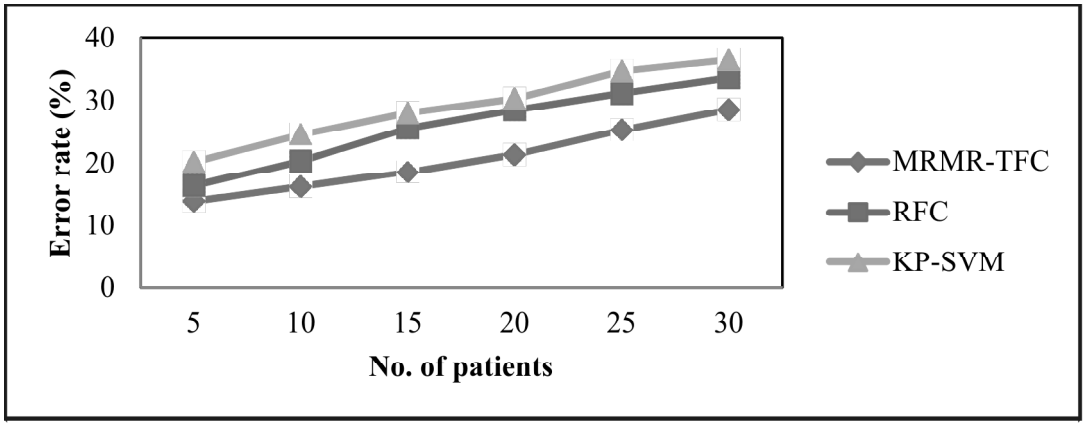

Figure 10: Measure of Error rate (using stroke disease dataset)

the diseases. This helps to diagnosis the stroke and heart disease based on the classified and non-classified rule reducing the error rate. The error rate is reduced by 27% and 42% than the existing RFC [1] and KP-SVM [2] respectively.

Figure 10 describes the error rate measurement using stroke disease dataset with respect to number of patients. From the figure, the error rate is reduced in proposed MRMR-TFC method using stroke disease dataset compared to state-of-the art methods [1] [2]. This is because; the MRMR-TFC uses a Trapezoidal Fuzzy Rule-based Classifier model using stroke disease dataset. The MRMR-TFC method also uses trapezoidal membership functions for medical variables and classifies the diseases effectively. This in turn diagnosis the stroke disease based on the classified and non-classified rule and reducing the error rate and therefore improving the classification accuracy. Therefore, the error rate obtained during the disease diagnosis is reduced by 26% and 43% compared to existing RFC [1] and KP-SVM [2] respectively.

6. CONCLUSION

In this work, an effective Maximum Relevance Minimum Redundancy method with Trapezoidal Fuzzy Classifier (MRMR-TFC) method is presented. The method reduced the classification time for disease diagnosis with reduced error rate and therefore provides improved quality of disease diagnosis for different patients and attributes. The goal of MRMR-TFC method is to extract more relevant attributes by reducing the redundant attributes that provides an insight into the medical practitioners for efficient stroke and heart disease diagnosis. To do this, we first designed a Maximum Relevance Minimum Redundancy algorithm to reduce the classification time resulting in high-level features and therefore improve the efficiency of the method. Then, based on this high-level feature Trapezoidal Fuzzy Rule-based Classifier algorithm is investigated by adjusting the certainty factor of each rule. This in turn diagnosis the stroke and heart disease based on the classified and non-classified rule reducing the error rate and therefore improving the classification accuracy. Through the experiments, we observed that our method of decision rule formation based on maximum relevance minimum redundancy model provided more accurate results compared to existing classification methods. The results show that MRMR-TFC method offers better performance with an improvement of classification accuracy and reduces the computational time and error rate compared to the state-of-the-art methods.

REFERENCES

[1] K. Srinivas, G. Raghavendra Rao, A. Govardhan," Rough-Fuzzy Classifier: A System to Predict the Heart Disease by Blending Two Different Set Theories", Arab Journal of Science and Engineering, Springer, Feb 2014.

[2] Sebastián Maldonado, Richard Weber, Jayanta Basak," Simultaneous feature selection and classification using kernel-penalized support vector machines", Information Sciences, Elsevier, July 2010.

[3] Jesmin Nahar, Tasadduq Imama, Kevin S. Tickle, Yi-Ping Phoebe Chen," Computational intelligence for heart disease diagnosis: A medical knowledge driven approach", Expert Systems with Applications, Elsevier, May 2013.

[4] Durga Toshniwal, Bharat Goel and Hina Sharma," Multistage Classification for Cardiovascular Disease Risk Prediction", Springer, June 2015.

[5] Hlaudi Daniel Masethe, Mosima Anna Masethe," Prediction of Heart Disease using Classification Algorithms", Proceedings of the World Congress on Engineering and Computer Science 2014 Vol II WCECS 2014, 22-24 October, 2014.

[6] Quentin Noirhomme, Damien Lesenfants, Francisco Gomez, Andrea Soddu , Jessica Schrouff, Ga e tan Garraux , Andr e Luxen , Christophe Phillips, Steven Laureys," Biased binomial assessment of cross-validated estimation of classification accuracies illustrated in diagnosis predictions", NeuroImage: Clinical, Elsevier, June 2014.

[7] Jiye Liang, Feng Wang, Chuangyin Dang, and Yuhua Qian," A Group Incremental Approach to Feature Selection Applying Rough Set Technique", IEEE Transactions on Knowledge And Data Engineering, Vol 26, No 2, February 2014.

[8] Wenjing Zhang, and Xin Feng," Event Characterization and Prediction Based on Temporal Patterns in Dynamic Data System", IEEE Transactions on Knowledge And Data Engineering, Vol. 26, No. 1, January 2014.

[9] Guan-Mau Huang, Kai-Yao Huang, Tzong-Yi Lee, Julia Tzu-Ya Weng," An interpretable rule-based diagnostic classification of diabetic nephropathy among type 2 diabetes patients", Bioinformatics, Elsevier, June 2015.

[10] P.K. Anooj, "Clinical decision support system: Risk level prediction of heart disease using weighted fuzzy rules", Journal of King Saud University – Computer and Information Sciences, Elsevier, Nov 2011.

[11] Xiaoyan Zhang, Shihu Liu, and Weihua Xu1," Rough Set Approach to Approximation Reduction in Ordered Decision Table with Fuzzy Decision", Hindawi Publishing Corporation. Mathematical Problems in Engineering Volume 2011.

[12] Surya Prakash Bhatt, Anoop Misra, Priyanka Nigam, Randeep Guleria, M. A. Qadar Pasha," Phenotype, Body Composition, and Prediction Equations (Indian Fatty Liver Index) for Non-Alcoholic Fatty Liver Disease in Non-Diabetic Asian Indians: A Case-Control Study", PLOS ONE | DOI:10.1371/journal.pone.0142260 November 24, 2015.

[13] Sharon Sanders, Jenny Doust, Paul Glasziou," A Systematic Review of Studies Comparing Diagnostic Clinical Prediction Rules with Clinical Judgment", PLOS ONE | DOI:10.1371/journal.pone.0128233 June 3, 2015.

[14] Thomas M Helms, Giang Duong, Bettina Zippel-Schultz, Roland Richard Tilz, Karl-Heinz Kuck and Christoph A Karle," Prediction and personalised treatment of atrial fibrillation—stroke prevention: consolidated position paper of CVD professionals", The EPMA Journal, May 2014.

[15] Paolo Melillo, Raffaele Izzo, Ada Orrico, Paolo Scala, Marcella Attanasio, Marco Mirra, Nicola De Luca, Leandro Pecchia," Automatic Prediction of Cardiovascular and Cerebrovascular Events Using Heart Rate Variability Analysis", PLOS ONE | DOI:10.1371/journal.pone.0118504 March 20, 2015.

[16] Thi Phuong Lan Nguyen, C.C.M. Schuiling Veninga,Thi BachYenNguyen,Vu Thi Thu Hang, E. Pamela Wright, M.J. Postma," Models to Predict the Burden of Cardiovascular Disease Risk in a Rural Mountainous Region of Vietnam", VALUE IN HEALTH REGIONAL ISSUES, Elsevier, June 2014.

[17] Sharon Sanders, Jenny Doust, Paul Glasziou," A Systematic Review of Studies Comparing Diagnostic Clinical Prediction Rules with Clinical Judgment", PLOS ONE | DOI:10.1371/journal.pone.0128233 June 3, 2015.

[18] Stefan P. Koch1, Claudia Hägele, John-Dylan Haynes, Andreas Heinz, Florian Schlagenhauf, Philipp Sterzer," Diagnostic Classification of Schizophrenia Patients on the Basis of Regional Reward-Related fMRI Signal Patterns", PLOS ONE | DOI:10.1371/journal.pone.0119089 March 23, 2015.

[19] Yook Chin Chia, Hooi Min Lim, Siew Mooi Ching," Use of Chronic Kidney Disease to Enhance Prediction of Cardiovascular Risk in Those at Medium Risk", PLOS ONE | DOI:10.1371/journal.pone.0141344 October 23, 2015.