Science and Information Conference 2015
July 28-30, 2015 | London, UK

Identifying Best Feature Subset for Cardiac Arrhythmia Classification

Khalid Ahmed Khan Niazi, Shoab Ahmed Khan, Arslan Shaukat, Mahmood Akhtar
College of Electrical & Mechanical Engineering, NUST, Islamabad, Pakistan
khalidniazi@ceme.nust.edu.pk, shoabak@ceme.nust.edu.pk, arslanshaukat@ceme.nust.edu.pk, mahmood.akhtar@ceme.nust.edu.pk

***Abstract*—This paper presents a model for diagnosis of cardiac arrhythmias. The model uses k-Nearest Neighbors (KNN) and support vector machines (SVM) as the classification algorithms and improved F-score and sequential forward search (IFSFS) as the feature selection method. Complete feature selection process presented by this paper comprises two parts i.e. filter part and the wrapper part. Improved F-score is the criterion used in the filter part of the model, and in the wrapper part SFS is used. The setup uses KNN and SVM alternatively in the wrapper part to obtain the best feature subset. 20 fold cross validation was performed on Arrhythmia dataset obtained by UCI (University of California Irvine) machine learning repository. Experiments show that presented model achieves average accuracy of 73.8% in case of KNN, and 68.8% in case of SVM; which makes the model outperform the accuracies achieved by previous methodologies.**

***Keywords*—Support vector Machines (SVM); K-nearest Neighbors (KNN); improved F-score and sequential forward search (IFSFS); Arrhythmias**

I. INTRODUCTION

Since the very beginning of mankind heart diseases have posed a greater threat to life than any other disease; which in turn has paved ways for many research organizations to conduct research which helps in diagnosis of the disease in early stages so as to minimize annual deaths caused by heart diseases. As a result many instruments have been made to perform analysis of heart behavior. One method to identify heart's performance is to calculate inter beat intervals from an ECG signal. This provides an insight of the influence of nervous system over the cardio vascular system [1].

Machine learning has revolutionized the domain of biomedical Sciences and has paved ways for facilitation of diagnosis of different kind of diseases. This paper presents a model for the diagnosis of Cardiac Arrhythmias from a 12 lead standard Electrocardiogram (ECG) Machine. The proposed model uses both feature selection and classification methods to achieve accurate results. An ECG signal can be better comprehended by making further divisions of the complete wave signal. The three main divisions are; P wave is representative of sequential activation of the left and right atria, QRS duration or QRS complex is representative of ventricular muscle depolarization and the T wave represents repolarization of the ventricular [2] as in Fig. 1.

Dataset used in this research is obtained by UCI machine learning repository and was formulated by use of a 12 lead standard ECG machine and first used by [3] for the purpose of classification of arrhythmias. [3] Proposed an algorithm VF15 for classification of arrhythmia which is a supervised technique based on majority voting criteria.

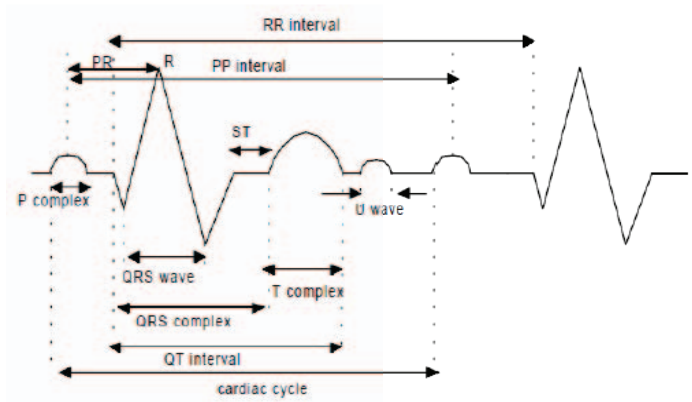

Fig. 1. ECG Sub Waves Explained [2]

[4] Has used fuzzy logic and Markov models to address the problem of cardiac arrhythmias; the proposed model first uses fuzzy logic and then uses Markov model to finally provide the classification results. [5] Used Bayesian (ANN) Artificial Neural Network for arrhythmia classification. [6] Uses Wavelet theory for classification of arrhythmia. [7] Combines radial basis function neural networks with wavelet analysis. [2] Has used KNN, Naïve Bayes and Decision Tree for comparison of Arrhythmia classification accuracy. [1] Has reduced features using GDA and then applied SVM for classification of cardiac arrhythmias.

This paper comprises six sections. Section II briefly explains the feature selection process in general, while explaining the feature selection techniques used in this research work. Section III describes the classification techniques used in this research work. Section IV explains the arrhythmia dataset in detail and then further presents the proposed methodology. Section V discusses, in detail, the results obtained by this work and compares these results with priory existent techniques to serve the purpose. Section VI finally concludes the research by providing conclusive analysis of the document and points out the future avenues in the domain.

II. FEATURE SELECTION TECHNIQUES

Correct selection of features is very important in building of any classification system [8]. Feature selection helps enhance classification accuracy and at the same time it reduces computation cost of the model hence saving the processing resources. Generally, Feature selection Algorithms are of two categories i.e. Filter methods and wrapper methods [9]. Filter methods identify best features without assistance from any classification algorithm; this is done by any of presumed criteria. On the other hand the wrapper method seeks assistance of a classification algorithm to identify the best feature subset amongst the superset of features. The feature selection model presented in this paper takes advantage of both filter, as well as wrapper techniques of feature selection. In the filter part improved F-Score of each feature is calculated for identifying relative importance of each feature [10]. In the wrapper part SFS is used to finally obtain the best feature subset [11]. Following subsection will present the Improved F-Score and Sequential Forward Search.

A. *Improved F-Score*

Improved F-Score resolves the same problem as does the F-Score but it is enhanced to cater for more than two sets of real numbers and provides discrimination between the sets [10][12]. Consider a vector x_k where $k=1, 2, 3,$ ----, m and total number of subsets are represented by l where$(l \geq 2)$, if the size of thej_{th} subset is n_j. Where $k=1, 2, 3,$ ----,l. then the improved F-Score of i_{th} feature is defined as:

$$F_i = \frac{\sum_{j=1}^{l}(\bar{x}_i^{(j)} - \bar{x}_i)^2}{\sum_{j=1}^{l}\frac{1}{n_j-1}\sum_{k=1}^{n_j}(x_{k,i}^{(j)} - \bar{x}_i^{j})^2} \quad (1)$$

Where $\bar{x}_i$, $\bar{x}_i^{(j)}$ are the averages of i_{th} feature in whole dataset and in j_{th} subset respectively; whereas $x_{k,i}^{(j)}$ is the value of i_{th} feature at the k_{th} instance in the j_{th} subset. In this case numerator presents discrimination between l subsets whereas the denominator presents discrimination with in that j_{th} subset. Finally those with greater Improved F-Score values are expected to provide maximum discrimination.

B. *Sequential Forward Search*

SFS works to find out the feature models starting from a model comprising one best feature with highest Improved F-Score and then with two best features and so on, until the model contains all features. The process of SFS is explained below [13]:

1) Select one best feature.

2) Select the best feature set which includes the feature of Step 1.

3) Repeat the process until all features are included in the feature model.

III. Classification Algorithms

A. *K-Nearest Neighbours (KNN)*

The intuition behind KNN is that the nearest neighbors in any plane are supposed to be of same kind i.e. they belong to a same class. Class assignment in KNN is made by labeling the test sample with the majority class among the K neighbors. K is always greater or equal to 1. It is a supervised learning approach and hence has a training set which is used for identification of the input test sample. The training data is labeled with true class labels. It is a lazy learning algorithm hence it trains only when it has some test sample to be classified [16]. For classification of a test sample, KNN performs following steps:

1) It first calculates Euclidian distance between the test sample and all training samples.

2) Then sorts them to obtain k nearest samples.

3) Next the algorithm pinpoints, majority class in the nearest k neighbors.

4) Finally the test sample is assigned to the majority class in step 3.

Selection of k is an important problem in case of KNN, as it has to have different values with different kind of data. If the data contains very few samples of some classes compared to a large number from a single class then for 'k' having greater value the results will be poor. Similarly if the data has nearly same number of instances for each class then the classification results will be better, with relatively greater number of K neighbors [17].

B. *Support Vector Machine (SVM)*

SVM was proposed by [14] in 1995 for classification of data. While training, SVM only considers the closest instances amongst the classes to draw a boundary condition, the intuition behind this idea is that if the closest of the samples are separated correctly then it would be obvious that all others will be classified correctly. SVM uses VC (Vapnik–Chervonenkis dimension) Dimension and minimizes the structural risk [18]. SVM tries to enhance generalization by finding optimized hyper plane for separation amongst the classes.

Consider a training sample$\{(x_i, y_i)|x_i \in R^N \in \{-1,1\}, i = 1, \dots. n\}$, every hyper-plane needs to satisfy the following conditions for separation of both classes:

$$(w.x_i) + b \geq +1 = \varepsilon_i,\ \text{If } y_i = +1 \quad (2)$$

$$(w.x_i) + b \leq -1 = \varepsilon_i,\ \text{If } y_i = -1 \quad (3)$$

Inequalities in equation 2 and 3 are equivalent to 4 below:

$$y_i[(w.x_i) + b] \geq 1 - \varepsilon_i,\ i = 1, \dots., n \quad (4)$$

Where $\varepsilon_i \geq 0$ and ε_i is a slack variable. For finding optimal hyper-plane (5) is minimized subject to (4):

$$||w||^2 + C\sum_{i=1}^{n}\varepsilon_i \quad (5)$$

Here C is a positive constant parameter and helps control the tradeoff between classification accuracy and complexity. By use of Lagrange multipliers optimized (6) is obtained.

$$\text{Maximize} \quad \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j(x_i, x_j)$$

$$\text{Subject to} \quad \sum_{i=1}^{n}\alpha_i y_i = 0, \alpha_i \geq 0, \forall i, \quad (6)$$

Decision function as per α_i Lagrange multiples is shown below in (7)

$$f(x) = sgn\left(\sum_{i=1}^{n}\alpha_i y_i(x_i, x) + b\right) \quad (7)$$

For a nonlinear case, SVM maps the input vector into a higher dimensional feature space using a kernel function. So the decision function becomes as follows:

$$f(x) = sgn\left(\sum_{i=1}^{n} \alpha_i y_i K(x_i, x) + b\right) \tag{8}$$

IV. RESEARCH DESIGN

A. *Arrhythmia Dataset*

Diagnosis of cardiac arrhythmias by use of ECG signals is an important problem in domain of biomedical sciences. Dataset used in this study contains record of 452 patients and the division of these patients is among 16 classes representing any of the heart rhythms. The type of arrhythmias presented in dataset are Atrial Fibrillation (Flutter), Left ventricular hypertrophy, 1st degree AtrioVentricular block, 2nd degree AV block,3rd degree AV block, Right bundle branch block, Left bundle branch block, Supraventricular Premature Contraction, Ventricular Premature Contraction (PVC), Sinus bradycardia, Sinus tachycardia, Old Inferior Myocardial Infarction, Old Anterior Myocardial Infarction, Ischemic changes (Coronary Artery Disease), Other Unidentified Rhythms. The features contained in dataset are 279. Feature f1 is Age, f2 is Sex, f3 is Height, f4 is Weight, f5 is average duration of QRS in msec, f6 is duration in msec between onset of P and Q waves, f7 is duration in msec between onset of Q and offset of T waves, f8 is average duration in msec between 2 consecutive T waves, f9 is average duration in msec between 2 consecutive P waves. Feature f10 is the vector angle in degrees on front plane of QRS, likewise f11 is angle for T, f12 is angle for P, f13 is angle for QRST and f14 is angle for j. The 11 features mentioned below are measured from the DI channel; f15 is heart pulse per minute, f16 is the average width of Q wave in msec. f17 is R waves average width in msec, f18 is S waves average width in msec, f19 is R' waves average width in msec, f20 is S' waves average width in msec, f21 is number of intrinsic deviation, f22 represents presence of dysphasic R waves (Boolean), f23 is representative of presence of notched R waves (Boolean), f24 represents presence or absence of notched P wave (Boolean), f25 represents presence or absence of dysphasic P wave (Boolean), f26 represents presence or absence of notched T wave (Boolean), f27 represents presence or absence of dysphasic T wave (Boolean). These features measured for DI are also measured for (f28-f39) DII, (f40-f51) DIII, (f52-f63) AVR, (f64-f75) AVL, (f76-f87) AVF, (f88-f99) V1, (f100-f111) V2, (f112-f123) V3, (f124-f135) V4, (f136-f147) V5 and (f148-f159) V6. Next 10 features are calculated from DI Channel, f160 represents J point depression in (millivolts), f161 represents Amplitude of Q wave in (millivolts), f162 represents R wave's Amplitude in (millivolts), f163 represents S wave's Amplitude in (millivolts), f164 represents R' wave's Amplitude in (millivolts), f165 represents S' wave's Amplitude in (millivolts), f166 represents P wave's Amplitude in (millivolts), f167 represents T wave's Amplitude in (millivolts), f168 represents QRSA (sum of area of all segment divided by 10) and f169 represents QRSTA i.e. equivalent to (QRSA+0.5 * width of T wave * 0.1 * Height of T wave). Similarly features for other channels are; (f170-f179) DII, (f180-f189) DIII, (f190-f199) AVR, (f200-f209) AVL, (f210-f219) AVF, (f220-f229) V1, (f230- f239) V2, (f240-f249) V3, (f250-f259) V4, (f260-f269) V5 and (f270-f279) V6 [3].

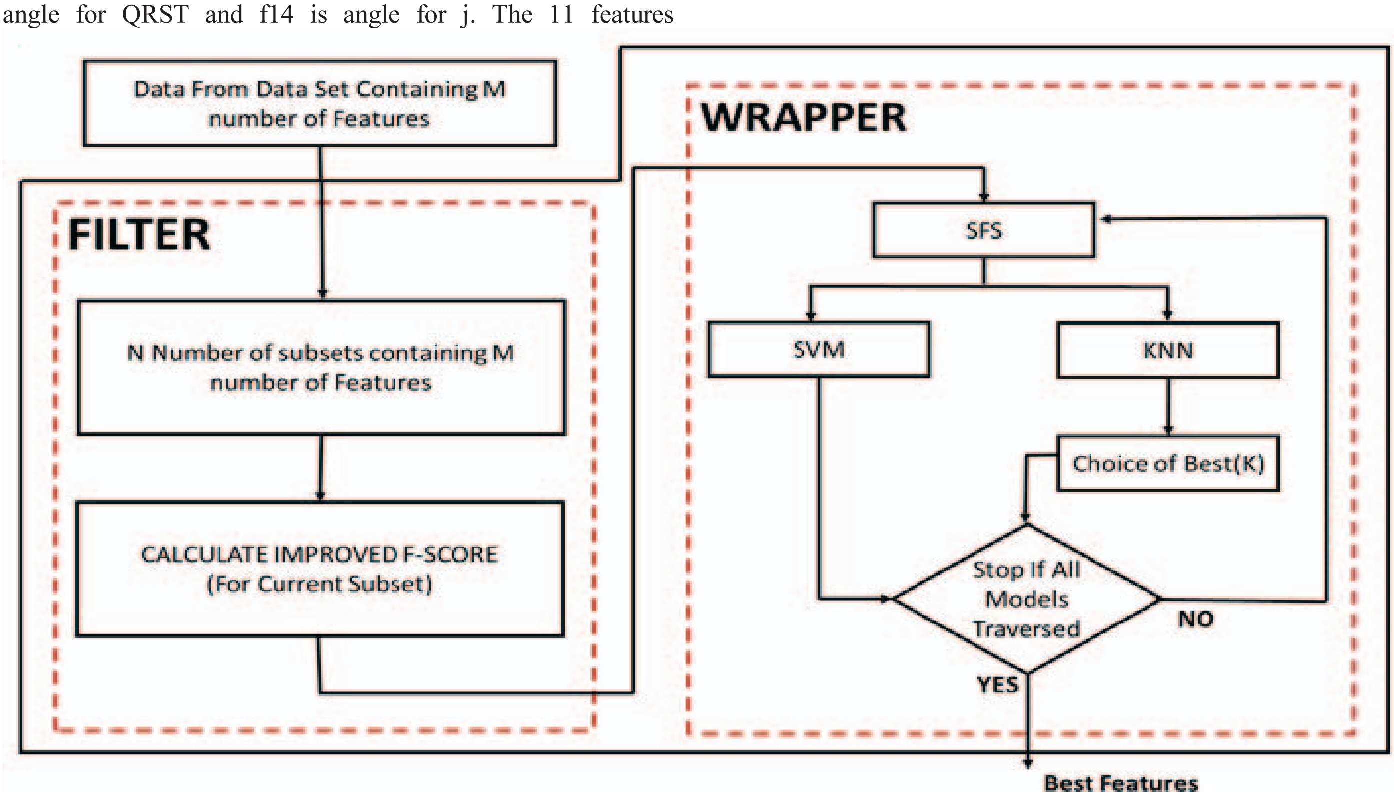

Fig. 2. Proposed Model for Feature Selection

B. Proposed Methodology

Proposed model is shown in Fig 2. In the filter part, model uses improved F-score to identify the feature importance. These Scores are then sorted in descending order; next all possible feature subset are identified by using SFS. This process is repeated in the wrapper part until all features of training set appear in the subset.

Next, the performance of model for all subsets is calculated using SVM and KNN Alternatively, considering classification accuracy as the performance metric. This process is repeated for every feature subset for all 20 folds until the model identifies best feature subset for both SVM and KNN. In case of KNN most suitable value for 'k' is identified as well. Finally the subset which has maximum classification accuracy and contains the least number of features is selected.

V. EXPERIMENT RESULTS AND ANALYSIS

20 fold cross validation was performed, for testing the performance of proposed model, for classification of cardiac arrhythmias. First the improved F-Score of every feature is calculated to identify its importance; those having greater improved F-Score are considered to be more discriminant than others.

Further the SFS Algorithm generates 279 feature models on the basis of feature importance suggested by the improved F-Score. Each model possesses one feature more than the previous model until all features are present in the model. Finally two Classification algorithms (i.e. SVM and KNN) are used alternatively to obtain best feature model for arrhythmia classification. Best feature model in case of both the Algorithms is shown in table I.

Fig. 3. Shows Classification accuracy obtained by SVM for each Feature Model which suggests that best accuracy is obtained by feature model 148 containing 148 features.

Fig. 4 shows Classification accuracy obtained by KNN where K is varied from 1 to 5 for each Feature Model which suggests that best accuracy is obtained for k=3 by feature model 60 containing 60 features.

Table II provides comparison of provided model with previous techniques. It is evident that proposed model offers better classification accuracy compared to any of the previous models.

TABLE I. BEST FEATURE MODELS

Classification Algo	**Features Selected**
KNN	272,262,124,4,273,150,236,161,263,16,138,40,190,93,149,76,226,224,251,221,88,185,47,223,160,100,253,125,136,141,45,1,99,213,246,172,173,233,91,31,46,154,212,112,192,6,49,71,111,135,127,44,225,277,191,39,66,92,231,57
SVM	272,262,124,4,273,150,236,161,263,16,138,40,190,93,149,76,226,224,251,221,88,185,47,223,160,100,253,125,136,141,45,1,99,213,246,172,173,233,91,31,46,154,212,112,192,6,49,71,111,135,127,44,225,277,191,39,66,92,231,57,156,279,103,131,240,252,241,3,174,227,170,28,211,181,50,98,23,58,115,216,269,184,117,232,66,107,96,63,110,106,121,27,109,256,229,129,183,113,139,137,244,276,255,143,147,74,122,120,94,151,97,43,274,80,21,104,19,128,239,271,186,54,250,235,215,230,202,254,203,267,59,25,119,26,130,51,62,60,37,134,61,38,82,193,175,116,167,270

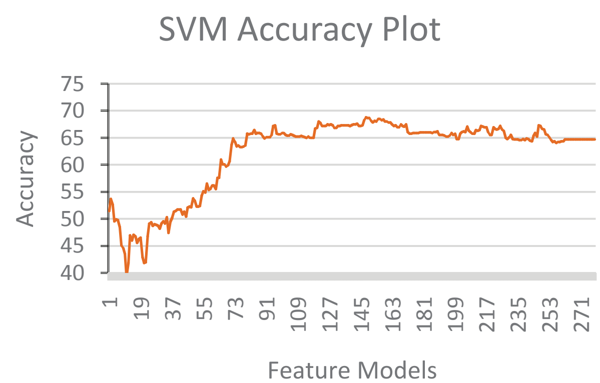

Fig. 3. SVM Accuracy Plot

TABLE II. CLASSIFICATION ACCURACY

Methodology	**Classification Accuracy**	**Author**
VF15	62%	[3]
KNN	66.9645%	[2]
Decision Tree	59.7696	[2]
Naïve Bayes	50%	[3]
KNN with Improved F-Score	73.80%	Our Model
SVM with Improved F-Score	68.80%	Our Model

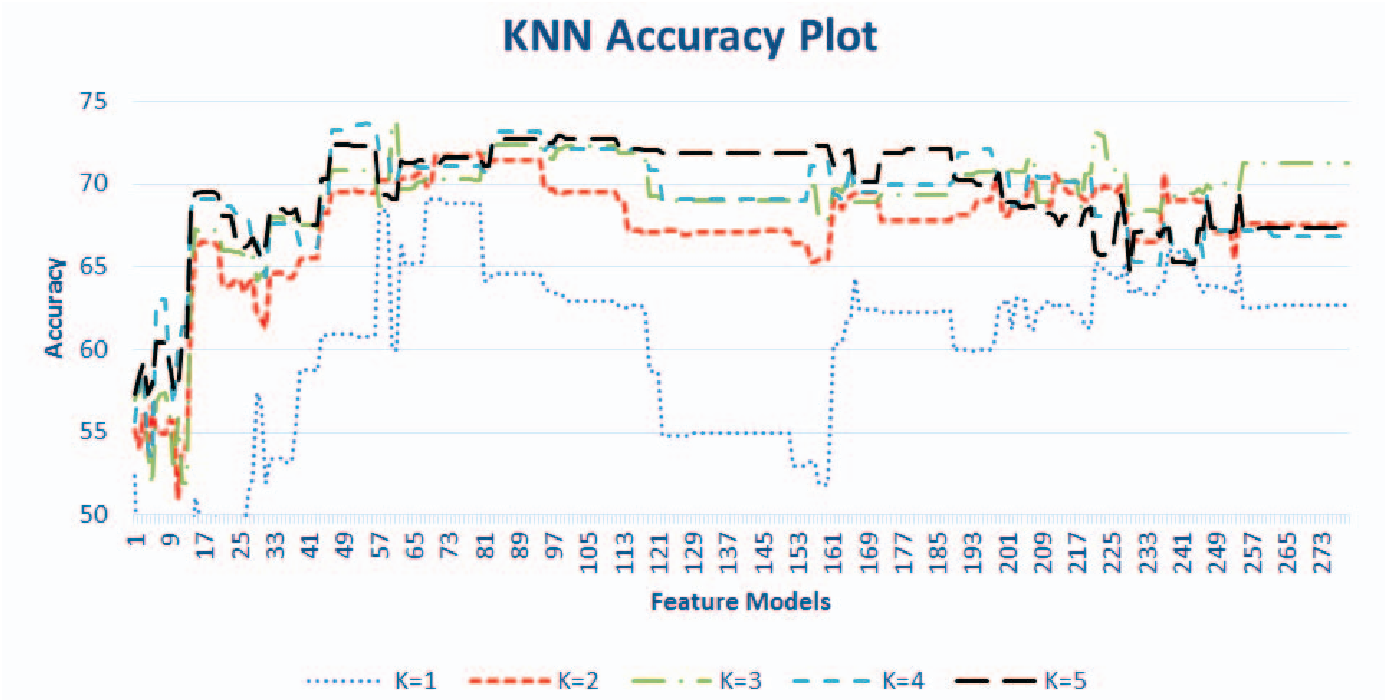

Fig. 4. KNN Accuracy Plot

VI. CONCLUSION AND FUTURE WORK

Proposed model for classification of arrhythmias generates promising results compared to previous studies that have been conducted in the past. It relies on the best features selected by the help of feature selection process. The feature selection process comprised of two parts; the wrapper component uses SFS to obtain the feature subsets while the filter component utilizes improved F-score. Then, further, SVM and KNN are used alternatively to perform classification. It turned out that KNN generates better results than SVM (i.e. 73.80%) for K=3 and that too with a lesser number of features (i.e. 60 features). SVM also performed well when compared to previous methodologies and generated best results of 68.80% for feature model containing 148 features.

In future pre-processing of the data acquired by ECG wave signal needs to be performed, so as to reduce biasedness of feature selection approaches in predicting relative feature importance; This may be achieved by normalizing the feature value ranges. Further, these same techniques can be applied on the EEG signals for the diagnosis of brain related disorders to accurately identify the type of disorder.

ACKNOWLEDGMENT

Authors are thankful to the author who provided the benchmark dataset and UCI machine learning repository for hosting the dataset; also at the same time cooperation from NUST's faculty and staff was one main reason which translated this research effort into a success.

REFERENCES

[1] Babak Mohammadzadeh Asl, Seyed Kamaledin Setarehdan, Maryam Mohebbi, *Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal*. Artificial Intelligence in Medicine (2008) 44, 51—64

[2] Saleha Samad, Shoab A. Khan et al, *Classification of Arrhythmia.* International Journal of Electrical Energy, Vol. 2, No. 1, March 2014.

[3] H. A. Guvenir, S. Acar, G. Demiroz, and A. Cekin, "*A supervised machine learning algorithm for arrhythmia analysis*," in Proc. IEEE conf. on Computers in Cardiology, 1997, pp. 433- 436.

[4] M. G. Tsipouras, Y. Goletsis, and D. I. Fotiadis, A method for arrhythmic episode classification in ECGs using fuzzy logic and markov models," in Proc. IEEE conf. on Computers in Cardiology, 2004, pp. 361-364.

[5] D. Gao, M. Madden, D. Chambers, and G. Lyons, "*Bayesian ANN classifier for ECG arrhythmia diagnostic system: A comparison study*," in Proc. IEEE International Joint Conference on Neural Networks, 2005, pp. 2383-2388.

[6] Khadra L, Al-Fahoum AS, Al-Nashash H, *Detection of life threatening cardiac arrhythmias using wavelet transformation*. Med Biol Eng Comput 1997;35(6):626—32.

[7] Al-Fahoum AS, Howitt I. Combined wavelet transformation and radial basis neural networks for classifying life-threatening cardiac arrhythmias. Med Biol Eng Comput 1999; 37(1):566—73.

[8] Hua, J. P., Tembe, W. D., & Dougherty. *Performance of feature-selection methods in the classification of high-dimension data*. E. R. (2009) Pattern Recognition, 42, 409–424.

[9] Talavera. *An evaluation of filter and wrapper methods for feature selection in categorical clustering*. L. (2005). In Proceedings of 6th international symposium on intelligent data analysis, Madrid, Spain (pp. 440–451).

[10] Juanying Xie a,b,., Chunxia Wangc. Using support vector machines with a novel hybrid feature selection method for diagnosis of erythemato-squamous diseases Expert Systems with Applications 38 (2011) 5809–5815.

[11] Nakariyakul, S., & Casasent. *Adaptive branch and bound algorithm for selecting optimal features*. D. P. (2007). Pattern Recognition Letters, 28, 1415–1427.

[12] Juanying Xi, Jinhu Lei, et al. Two-stage hybrid feature selection algorithms for diagnosing erythemato-squamous diseases. Health Information Science & Systems 2013, 1:10.

[13] Gunal, S., Gerek, O. N., Ece, D. G., & Edizkan. *The search for optimal feature set in power quality event classification*. R. (2009) Expert Systems with Applications, 36, 10266–10273.

[14] Vapnik, V. N. *The nature of statistical learning theory*. (1995) New York: Springer.

[15] N. S. Altman. *An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression*. The American Statistician, Vol. 46, No. 3. (Aug., 1992), pp. 175-185.

[16] D. Bremner, E. Demaine, J. Erickson, J. Iacono, S. Langerman, P.Morin, and G. Toussaint, "*Output-sensitive algorithms for computation of nearest-neighbor decision boundaries,"* International Journal of Discrete and Computational Geometry, vol. 33, pp. 593-604, April 2005.

[17] G. Toussaint, "*Geometric proximity graphs for improving nearest neighbor methods in instance-based learning and data mining*," International Journal of Computational Geometry and Applications, vol. 15, pp. 101–150, April 2005.

[18] Burges, C. J. C. (1998). *A tutorial on support vector machines for pattern recognition*. Data Mining and Knowledge Discovery, 2, 121–167