Machine learning based congestive heart failure detection using feature importance ranking of multimodal features

6 authors, including:

Lal Hussain
University of Azad Jammu and Kashmir
54 PUBLICATIONS   596 CITATIONS

SEE PROFILE

Wajid Aziz
University of Azad Jammu and Kashmir
83 PUBLICATIONS   956 CITATIONS

SEE PROFILE

Ishtiaq Rasool Khan
University of Jeddah
93 PUBLICATIONS   721 CITATIONS

SEE PROFILE

Haider Alkinani
32 PUBLICATIONS   153 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Extraction of Dynamical Information Encoded in Biological Signals Using Complexity Analysis Tools and Techniques View project

Project   Human Machine Interaction View project

Research article

Machine learning based congestive heart failure detection using feature importance ranking of multimodal features

Lal Hussain [2,3], Wajid Aziz [1*], Ishtiaq Rasool Khan[1], Monagi H. Alkinani [1] and Jalal S. Alowibdi [1]

[1] Department of Computer & AI, University of Jeddah, Jeddah, 23890, Saudi Arabia

[2] Department of Computer Science & IT, University of Azad Jammu and Kashmir, King Abdullah Campus, 13100, Muzaffarabad, Pakistan

[3] Department of Computer Science & IT, University of Azad Jammu and Kashmir, Neelum Campus, 13230, Muzaffarabad, Pakistan

* Correspondence: Email: wloun@uj.edu.sa.

Abstract: In this study, we ranked the Multimodal Features extracted from Congestive Heart Failure (CHF) and Normal Sinus Rhythm (NSR) subjects. We categorized the ranked features into 1 to 5 categories based on Empirical Receiver Operating Characteristics (EROC) values. Instead of using all multimodal features, we use high ranking features for detection of CHF and normal subjects. We employed powerful machine learning techniques such as Decision Tree (DT), Naïve Bayes (NB), SVM Gaussian, SVM RBF and SVM Polynomial. The performance was measured in terms of Sensitivity, Specificity, Positive Predictive Value (PPV), Negative Predictive Value (NPV), Accuracy, False Positive Rate (FPR), and area under the Receiver Operating characteristic Curve (AUC). The highest detection performance in terms of accuracy and AUC was obtained with all multimodal features using SVM Gaussian with Sensitivity (93.06%), Specificity (81.82%), Accuracy (88.79%) and AUC (0.95). Using the top five ranked features, the highest performance was obtained with SVM Gaussian yields accuracy (84.48%), AUC (0.86); top nine ranked features using Decision Tree and Naïve Bayes got accuracy (84.48%), AUC (0.88); last thirteen ranked features using SVM polynomial obtained accuracy (80.17%), AUC (0.84). The findings indicate that proposed approach with feature ranking can be very useful for automatic detection of congestive heart failure patients and can be very helpful for further decision making by the clinicians and physicians in order to decrease the mortality rate.

Keywords: multimodal features; empirical receiver operating characteristics; machine learning; congestive heart failure; decision tree; normal sinus rhythm; support vector machine

1.   Introduction

Heart Rate Variability (HRV) is a convenient non-invasive tool for the measurement of autonomous cardiac function by sympathetic and parasympathetic branches of the nervous system i.e. to introduce electrocardiography (ECG) time series analysis [1] and complex systems and the technology of variability analysis [2]. Conventional techniques used for the quantification of various HRV signals by employing linear strategies have represented that decrease in the variability have direct association with increase in the heart failure mortality. However, in some situations the HRV data cannot be evaluated by using linear methods [3].

In the recent studies, the researchers developed and employed different techniques for the detection of congestive heart failure (CHF) subjects using Inter-beat Interval (IBI) time series extracted from ECG signals including symbolic time series was used by [4] to study the dynamics of interbeat heart interval [5], threshold dependent symbolic entropy to classify healthy and pathological subjects, wavelet based soft decision technique to detect congestive heart failure [6] and combined classical HRV indices with wavelet entropy to detect congestive heart failure [3]. Isler and Kuntalp [3] considered wavelet entropy features and HRV features along with KNN classifier to distinguish regular subjects and CHF subjects. Hossen and Al-Ghunaimi [6], used wavelet-based soft determination methodology for estimating the spectral density of average power of IBI time series for screening of CHF subjects. Thuraisingham [7] proposed a novel technique using features from KNN classification method and second order difference plot of IBI time series to distinguish CHF and normal subjects. Yu and Lee [8] proposed a mutual information based featured to detect congestive heart failure. Pecchia et al. [9] proposed short term power features along with very simple threshold based classifier for the detection of CHF. Aziz et al. [5] used symbolic time series analysis for distinguishing healthy subjects from CHF patients. Altan et al. [10] extracted features from IBI time series using Hilbert-Huang transform and multilayer perceptron neural network used to classify normal subjects, CHF and coronary artery disease subjects. Awan et al. [4] introduced multiscale simplified improved Shannon entropy for extracting features from IBI time and used different classifiers for discriminating NSR and CHF subjects. Choudhary et al. [11] proposed grouped horizontal visibility graph entropy for discriminating normal, CHF and atrial fibrillation subjects. Recently, Isler et al. [12] applied Multi-stage classification of congestive heart failure based on short-term heart rate variability. Moreover, Narin et al. [13] predicted paroxysmal atrial fibrillation based on short-term heart rate variability. The researchers [14] tested the irregularity or very short electrocardiogram (ECG) signals as a method for predicting a successful defibrillation in patients with ventricular fibrillation.

The machine learning algorithms rely on the type and relevancy of feature extraction approach. The classification efficiency can be enhanced by extracting the most relevant features which is a hot topic in machine learning and signal & image processing problems. In the past, researchers have obtained numerous characteristics from different physiological signals and systems. Wang et al. [15] proposed multi-domain feature extracting approach to identify the accurate epileptic seizure detection. Hussain [31] proposed multi-modal (multi-domain, and nonlinear) feature extracting approach for epileptic seizure detection, arrhythmia detection [16] and Rathore et al. proposed hybrid feature to detect colon cancer [17]. After extracting the features, all features are not contributing equally, their importance can be determined by ranking the features based on different ranking algorithms. The management for feature selection and follow-up for the relevant features information processing, the Feature Importance Ranking (FIR) plays an important role in the area feature selection methods based

on mutual information criteria of max dependency [18], feature ranking algorithm to detect cardiac arrythmia [19], feature ranking to reconstruct dynamical network [20], feature selection to assess thyroid cancer pronosis [22], and Multi-objective-based radiomic feature selection for lesion malignancy classification [21]. The main objective of the FIR is to arrange the features in accordance with their relative significant value. Depending to which the labels on training samples are used, all methods are divided into supervised and unsupervised approaches, whereas the supervised process labels are used [21]. From a technical point of view, some approaches, such as the Wilcoxon rank-sum test and t-test, use statistical analysis and class separability parameters to measure inter-feature relationships, and some other approaches investigate reciprocal knowledge [18], sparse regression, spectral analysis, and include some classification efficiency into accounts and the selection of classifiers for machine learning [21]. Leguia et al. [20] used the Random Forest and Relief-F to ranked the feature importance of each node to predict the value of each other node. Karnan et al. [19] proposed feature ranking score (FRS) algorithms on different statistical parameters to select the optimal parameters for classification of signals from public domain MIT-BIH arrhythmia data. These optimal features are provided to the least square support vector. Mourad et al. [22] combined the features selection algorithms and machine learning algorithms (Kruskal-Wallis' analysis, Relief-F and Fisher's discriminant ratio) to analyse the specific attributes of de-identified thyroid cancer patients in the SEER sample.

In this study, we employed FIR for extracting most contributing factors based on the feature ranking categorized (1 to 5) for detection of healthy and CHF subjects for clinical decision making. The category value 1 depicts that feature is most important and the value 5 reveals that feature least important. Moreover, the greatest ROC value indicate that feature is more important and as the ROC value decreases, the importance of the feature deceases accordingly. We first extracted multimodal features from CHF and NSR subjects and then ranked them based on EROC and random classifier slop [23], which ranks features dependent on the criterion for class separability of the region between EROC and Classifier at periodic intervals slope.

2. Material and methods

2.1. Dataset

The RR time series interval data were taken from the Physionet databases [24]. Data from the Normal Sinus Rhythm (NSR) subject, Congestive Heart Failure (CHF) subjects and Atrial Fibrillation (AF) subjects in the cardiac interval (RR interval) time series were analysed [24]. The heart activity data from NSR subjects had been taken from 24-Hour recordings of 72 subjects by using Holter monitor system. The dataset consist of 35 Males and 37 Females (54 from the NSR subjects of the RR-interval and 18 from the Normal sinus rhythm RR internal Database used in the study [25]. The age range of 20–78 years for the measured population was 54.6 ± 16.2 (mean $\pm$ SD). At 128 Hz, ECG data was sampled. The CHF group consisted of 44 participants aged from 22–78 years, 29 Males and 15 Females aged 55.5 ± 11.4 and the data from the RR CHF interval and 15 years from the Congestive Heart Failure RR interval Database used in the study [26] were collected for 29 CHF subjects [24]. According to the practical classification system of the New York Heart Association (NYHA), CHF subjects can be divided into four categories. This method categorizes patients by the signs of the

patient's regular behaviour and quality of life. In this study we used 20,000 samples of each subject to distinguish CHF from NSR patients.

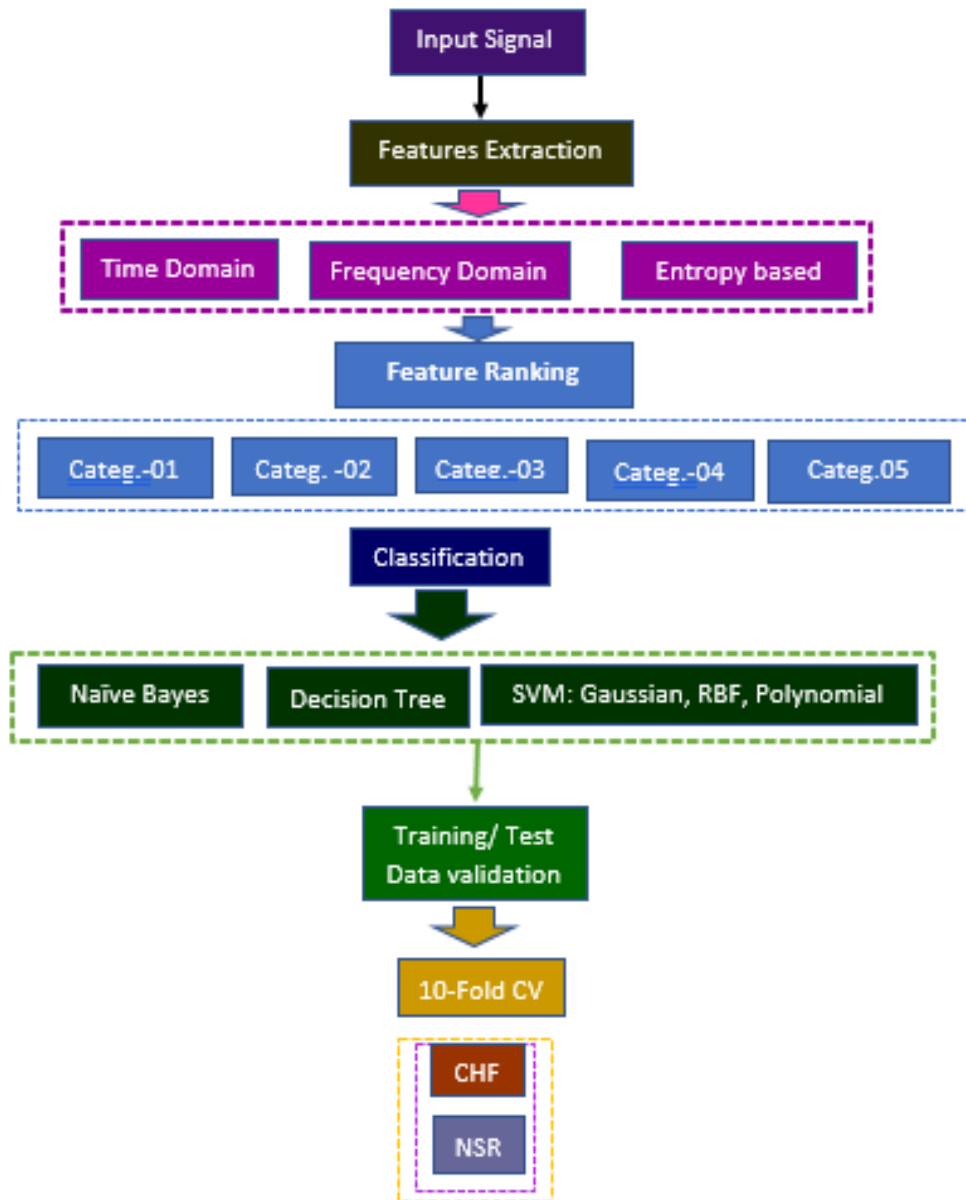### 2.2. *Feature extraction*

Figure 1. Schematic Diagram to extract Multimodal features (i.e. Time domain, frequency domain and entropy-based features) to detect CHF and then applying the feature ranking to determine the feature importance. The classification performance was computed based on multimodal features and ranked features with categorify one to five.

After extracting features, another important criterion is also to get the most appropriate features with high ranks. This can be done by ranking the features based on various criteria. Firstly, from CHF and NSR subjects, we extracted multimodal features. We then ranked the features to differentiate the CHF from NSR subjects based on receiver operating curve (ROC) value. We then applied the machine

learning classification techniques based on the different inputs of the ranked features to evaluate the detection performance.

The Figure 1 shows the schematic diagram for CHF detection. In the first step, we extracted the general multimodal features from NSR subject and the CHF subjects. In the second step, we ranked the extracted features based on ROC values. In the third step, we employed different machine learning algorithms such as Decision Tree, SVM along with its kernels tricks, and Naïve Bayes approach on five different categories of ranked features i.e. Category 01 with all extracted features, Category 02 with top 05 ranked features with higher ROC values obtained, Category 03 top nine ranked features, Category 04 last thirteen ranked features and Category 05 the last two ranked features with very low ROC values. Finally, for testing and training of data validation, we employed standard 10-fold cross validation.

2.2.1. Time and frequency domain features

The time and frequency domain approaches are commonly used to collect the time series and heart spectral dynamics control of these signals vibrations to quantify the heterogeneity of physiological signals (i.e. EEG or ECG) caused by various pathologies. The techniques of the time domain are used for the tracking of the short-term, medium-term and long-term fluctuations of physiological signals and processes, while preserving the effects of various spectrums. The definition for patients suffering from various heterogeneity dysfunctions is detailed in (Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology and Electrophysiology 1996; Seely and Macklem 2004) including heart rate variability in insomnia patients [27], Ultra-shortened time-domain HRV parameters [28], evaluating the homeostasis assessment model insulin resistance and the Cardiac autonomic System in bariatric surgery patients [29] and short-term measurement of heart rate variability during spontaneous breathing in people with chronic obstructive pulmonary disease [30]. etc. We have used same time domain, frequency domain, nonlinear entropy based and wavelet based features in previous studies to detect epileptic seizure [31], congestive heart failure [32] and arrhytymia detection [16].

2.2.2. Entropy and wavelet-based features

Biological signals are production of beating heart and several muscle interacting components that display complex patterns variation and rhythms on monitoring devices. To analyse the fundamental mechanisms of these processes, these rhythmical shifts and patterns provide very valuable secret details. Extracting useful knowledge using conventional methods of data mining is impractical. The complexity of the physiological processes that are degraded by ageing and disease consists of systemic components and coupling between them. The researchers in the past applied various complexity based methods such as epileptic seizure detection using multi-modal features [31], seizure detection using symbolic entropy [33], lung cancer detection based on refined fuzzy entropy [34], arrythmia detection using refined fuzzy entropy [16], electroencephalographic (EEG) signals with motor movement using multi scale sample entropy [35], EEG alcoholic and control subjects using multiscale entropy with KD tree algorithmic approach [36], regression analysis to detect seizure [37]. The healthy subjects are more complex than pathological subjects. In the healthy subjects, all the structural elements and integrated functions within the structural elements are properly functional and linked for inter-communication, thus increasing their complexity computation value and entropy values. But, due to the weakening of

the coupling between the structural elements, the computed complexity value and entropy value of the diseased subjects is decreased.

2.2.2.1. Approximate entropy

Pincus in 1991 proposed approximate entropy (ApEn) [38] to compute the regularity presence in the bio-signal time series recording data. The measurement of the entropy indicates that the probability of related or similar patterns does not repeat in observation. Mathematically,

$$ApEn(m, r, N) = C^m(r) - C^{m+1}(r) \tag{2.1}$$

The $C^m(r)$ and $C^{m+1}(r)$ are being computed as detailed in [36]. Two parameters are set to measure the average entropy, i.e. m, which is the length of the window, and r, the criterion of similarity. We selected m = 3 and r = 0.15 times the standard deviation of data in this analysis as given in [38].

2.2.2.2. Fast sample entropy with KD tree algorithmic approach

Sample entropy (SampEn) as proposed by [39], which is a revamped form the approximate entropy. In contrast to the average entropy, sample entropy is more stable then approximate entropy since it is independent to estimate the randomness of data duration and trouble-free execution. Recently, researchers used a sample entropy version based on the KD tree algorithmic approach, which is more stable in terms of time and space complexity as detailed in [36].

In 1975, Bentley design a binary tree algorithm known as the K-Dimensional (KD) space partition tree. A rectangle "Bv" is connected with each of its 'v' nodes. The 'v' would be the leaf node if "Bv" does not have any point in its interior. In other examples, by creating a vertical and horizontal line such that each rectangle comprises at no more than half of the lines, "Bv" can be separated into two rectangles. Details of the KD tree algorithm are computed by Hussain et al [36]. Using the following steps, the complexity of time and space is minimized.

Step 1. Original time series of RR-Interval was transformed to spaced points set from $\{X\} = \{x1, x2, x3, \ldots xN\}$ be a time series of length N. The time and memory cost are O (N).

Step 2. Using N-m points of the series for which total cost is O (N log N) and memory is O(N)), the K-dimensional tree is constructed. Here construct time is O (N log N) for k-d tree.

Step 3. Range query: The time cost is $N\, O(N^{1\frac{1}{d}})$ for N queries for d-dimensional k-d search and the memory cost are O(N). Where $O(N^{1\frac{1}{d}})$ is search time for k-d tree.

2.2.2.3. Wavelet entropy

Wavelet-based entropic measurements were also computed by researchers in the past to identify the nonlinearity presence in the results. The most widespread wavelet entropy techniques [40] include Shannon, Threshold, Log Energy, Sure and Norm etc. Shannon entropy [40] was used by calculating wavelet coefficients which is created from the wavelet packets (WPT) to calculate the signal intensity, where maximum values indicate a high uncertainty in the CHF or NSR subjects and hence greater complexity. In addition, wavelet entropy was used [41] to capture the underlying dynamical mechanism connected with the bio-signal. The entropy 'E' must be a cost function of additive information such that E (0) = 0.

$$E(S) = \sum_i E(S_i) \tag{2.2}$$

Where S is a signal and (S_i) are signal coefficients on an orthonormal basis. The function E (S) is defined as wavelet entropy as expressed inequation (5).

2.2.2.4. Shannon entropy

The Claude Shannon first suggested the entropy of Shannon in 1948 [42] and is most commonly used in information science. In addition, it is the measurement of the vulnerability associated with a randomness of the data space. Shannon entropy precisely estimate the predicted value of the results found in a packet. We can describe the Shannon entropy of a random variable S as follows:

$$E(S) = -\sum_{i=1}^{n} S_i^2 log_2(S_i^2) \tag{2.3}$$

Where Si represents coefficients of signal S in an orthonormal basis. If the entropy value is greater than one, the component has a potential to reveal more information about the signal and it needs to be decomposed further in order to obtain simple frequency component of the signal [43]. By using the entropy, it gave a useful criterion for comparing and selection the best basis.

2.2.2.5. Wavelet entropy

This entropy measure was proposed by [44] can mathematically defined such as:

$$E(S) = \frac{\sum_i |S_i|^p}{N} \tag{2.4}$$

Where p is the power, the terminal node signal must be $1 \ll P < 2$ and (Si) I is the terminal waveform signal.

2.2.2.6. Threshold entropy

E(Si) = 1 if $|Si| > p$ and 0 elsewhere so $E(s) = \#$ {i such that $|Si| > p$} is the number of time instants when the signal is greater than a threshold p.

The threshold entropy value was determined using a value of 0.2.

2.2.2.7. Sure entropy

The threshold of the parameter P and the values of $P \geq 0$ are used.

$$E(s) = n - \#\{i \ such \ that \ |si| \leq p\} + \sum_i min_{(si^2, p^2)} \tag{2.5}$$

Where, the discrete wavelet entropy E is a real number, s is the terminal node signal and (si) i the waveform of terminal node signals. In Sure entropy, p is a positive threshold value and must be $p \geqslant 2$ [45].

The entropy of Sure was measured at threshold 3.

2.2.2.8. Norm entropy

The P is used in Normal Entropy as the power and value of $P \geq 1$. The intensity in l^p norm entropy is:

$$E(si) = |s_i|^p \tag{2.6}$$

$$so\ E(s) = \sum_i |s_i|^p\ _i = ||S||_p^p$$

The entropy of the norm was estimated at 1.1 with power.

The wavelet norm entropy represents the ordering of nonstationarity of time series fluctuation.

2.2.2.9. Log energy

$$H_{logEn}(B) = -\sum_{i=0}^{N-1}(log_2(\ Pi(B)))^2 \qquad\qquad (2.7)$$

Where $\boldsymbol{Pi(B)}$ denotes the function of probability distribution and is a logarithmic amount of the distribution square of these probabilities.

2.3. Feature ranking algorithms

Feature ranking algorithms are mostly used for ranking features independently without using any supervised or unsupervised learning algorithm. A specific method is used for feature ranking in which each feature is assigned a scoring value, then selection of features will be made purely on the basis of these scoring values [46]. The finally selected distinct and stable features can be ranked according to these scores and redundant features can be eliminated for further classification. To perform this step, feature selection algorithms such as wrapper method and filter method can be used. As filter method is an unsupervised technique that analyse the inherent distribution properties of the features, on the other hand wrapper method correlates the features properties with the class labels [47]. In the past, multiple experiments studies have shown that every well-known function discovery algorithm that exposes the rating to errors has been used to pick features [48]. Feature ranking can be affected by the selection of algorithm for feature selection for classification purpose.

2.3.1. Filter methods

Radiomic feature ranking is a type of feature ranking method that is used to select features based on their high scoring values. The algorithm described in [49] selects the features that are showing minimum correlation with each other. However, Laplacian score [50] calculates a scoring value for each individual feature which shows the locality preserving power of a feature. In greedy feature selection algorithm [51], for all the chosen features a nearest neighbour graph is used and reconstruction error is repeatedly calculated for assigning ranks to the selected features subset. Mitra et al. [52] proposed a minimum information index for feature ranking. Multi-cluster feature selection (MCFS) [53] algorithm is proposed to measure the correlations between various features and then select and rank the features accordingly. Zeng and Cheung [54] proposed another clustering algorithm that takes into account the correlation between each feature by employing Local Learning Based Clustering (LLC) method. Zhao et al. [55] proposed a normalized Laplacian matrix method obtained from the similarity graph of pair-wise features.

2.3.2. Wrapper methods

Feature selection phase can be repeated by using wrapper method [56]. Relief-F algorithm [57] sort the features into one group that have similar values for the closest neighbours with the same binary class and higher characteristics based on the associated values that shows various values for the closest

neighbours in various groups. Fisher Score [47] is another algorithm which assigns a scoring value to each feature by calculating the intra-class variance and inter-class separation ratio. Feature based Neighbourhood Component Analysis (FNCA) [56] learns weights of features to minimize the objective function which is used to calculate the cumulative lack of regression over training data leaves one out.. The Infinite Latent Feature Selection (ILFS) [54] technique is also an impressive algorithm that is used in rank assigning to the features by estimating the relevancy using conditional probability of all subsets of the features. Features Selection via Eigenvector Centrality [58] is another technique used for ranking the features by connecting them to a graph of clustering, then discovering the correlation between separate pairs of features. Concave Optimization [59] is another novel method that is used in feature selection and ranking. Two function classes are defined in this approach by using a separate plane that is generated by using a series of features that can discern between a pair of classes.

2.3.3. Final feature ranking

The above-mentioned algorithms can be used for ranking the radiomic and any other features individually by usage of filter methods and wrapping techniques. The scoring values that is assigned to each feature by ranking group methods to get the final ranking scoring values of all the functions, they were summed. To get more precise scoring values, the key objective is averaging the features scoring value and to give equal weight to all rating algorithms. Then top 25 selected features having average scoring values can be calculated from the filter methods and wrapper methods [47].

Let us consider an example that illustrates the feature ranking methodology [20]; we have an equation

$$y = f(x_1, x_2, \dots) = x_1^2 + x_2 + 2 \qquad (2.8)$$

Assume that f is an unknown function, but y depends on several variables of x_i is known. Simply we can say that, x_i represents the features and y represents the target variable. This task can be solved by employing machine learning algorithm $M, i.e.$

$$f \approx \hat{f} = M(D) \qquad (2.9)$$

Where, D represents the data set and f shows the prediction model for any observation (x_1, x_2, x_3) can be used for predicting the value of $y, \hat{y} = \hat{f}(x_1, x_2, x_3)$. Now, consider a data set D comprised of L attributes tuples $(x_1, x_2, x_3; y)$ and we want to reconstruct f. Remember that, in data we have x_3 feature, which does not influence y, because only available data is collected via features selection and ranking that "may or may not" influence the target variable.

To check the feature importance, there are more than 30 algorithms developed. In this study, we computed ROC for feature importance ranking (FIR) as detailed in [60]. This method ranks the features based on the class separability criteria of the area between the empirical receiver operating characteristic curve (EROC) and the random classifier slop [23]. In this study, we extracted 22 multimodal features from CHF and NSR subjects. We then ranked these features based on above criteria, and Figure 2 below sorted the multimodal features based on their importance obtained. We then categorized these features based on the ROC values to further classify the CHF and NSR subjects to see the overall detection performance based on ranked features instead of using all the features. The top 5 features even show the higher detection than 82%. This will further help the clinicians to make the decision for future diagnosis and treatment of the patients. The highest ROC value indicates the highest ranked and highly important feature and as the ROC value decreased the feature importance decreased accordingly. In this study, the feature importance is depicted in descending order based on

the ROC values obtained. Figure 2 shows the importance of ranked Multimodal features based on the class separability criteria of the area between EROC and random classifier slope.
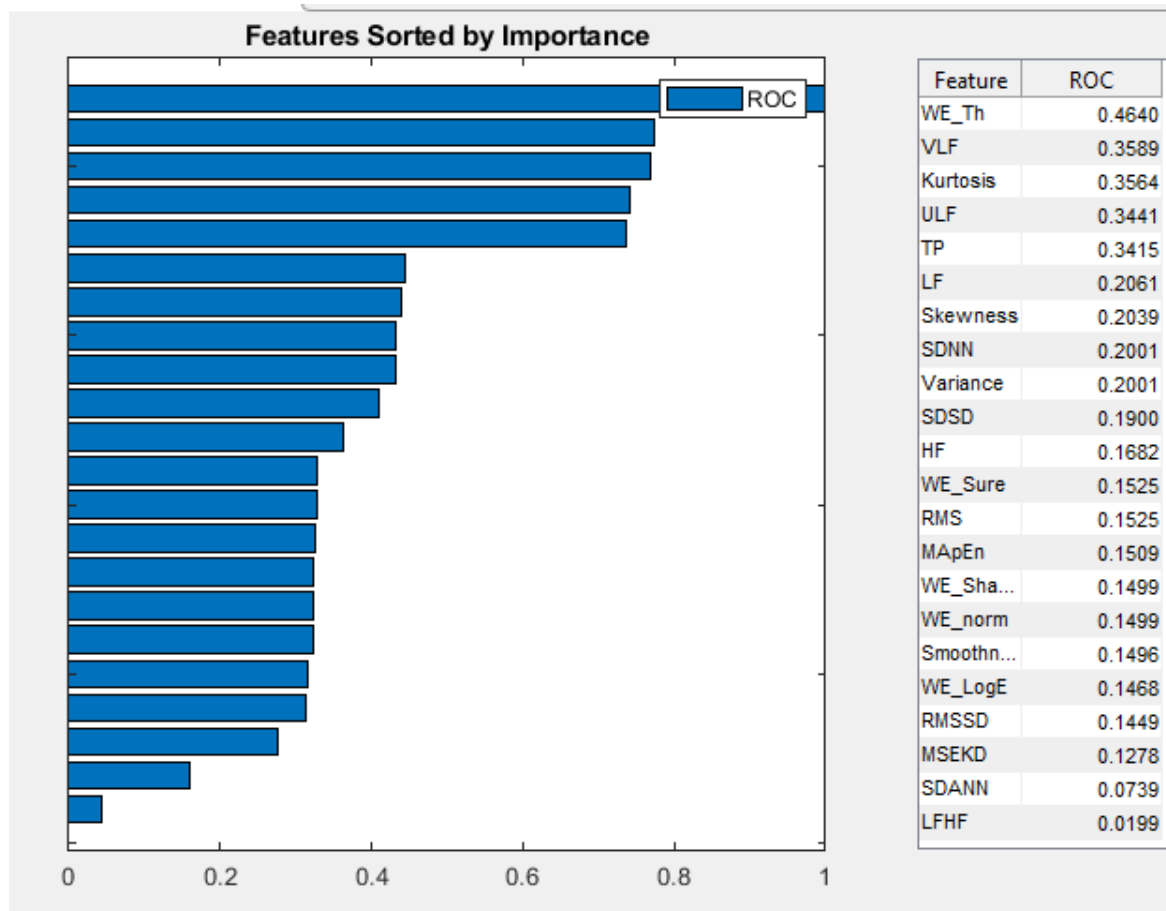

Feature	ROC
WE_Th	0.4640
VLF	0.3589
Kurtosis	0.3564
ULF	0.3441
TP	0.3415
LF	0.2061
Skewness	0.2039
SDNN	0.2001
Variance	0.2001
SDSD	0.1900
HF	0.1682
WE_Sure	0.1525
RMS	0.1525
MApEn	0.1509
WE_Sha...	0.1499
WE_norm	0.1499
Smoothn...	0.1496
WE_LogE	0.1468
RMSSD	0.1449
MSEKD	0.1278
SDANN	0.0739
LFHF	0.0199

Figure 2. Multimodal features ranking based on Receiver operating Curve (ROC) to distinguish the CHF from normal NSR. The multimodal features are sorted based on the ranking values obtained using the ROC values obtained. The greatest ROC value denotes the highest ranked followed by lower ranked values. The ranked features are represented in descending order.

2.4. *Classification Methods*

2.4.1. Support vector machine (SVM)

SVM is one of the most versatile approaches used for classification purposes of supervised learning techniques. SVM has recently been used excellently for concerns of graphical pattern recognition [61], Artificial Intelligence (Machine Learning) [62] and Computer aided medical diagnosis Health problems [63]. In addition, SVM is used in numerous applications in several field, such as identification and detection, text recognition, retrial of content-based images, bioinformatics, voice recognition, etc. In infinite or high dimensional space, SVM creates a hyperplane or series of hyper-plan that could be used to classify a successful separation while using this hyper-plane that has the greatest distance to the closest training instances in each class (also called as the functional margin), typically the greater the margin implies the classifier's relatively generalization classification error.

SVM attempts to determine the hyper-plane that provides the training example with the greatest minimum width. This concept is also known as margin in SVM theory. The optimum margin is obtained for the maximised hyperplane. SVM has another significant function that offers the higher efficiency of generalization. Basically, SVM is a two-class classifier that relies on nonlinear training instances or a maximum dimension to transform data into a hyperplane. Let us define $x.w + b = 0$ a hyperplane, where w is its normal. The linearly separable instances is labelled as:

$$\{x_i, y_i\}, x_i \in R^N d, y_i \in \{-1,1\}, \qquad \text{i=1, 2, ... , N}$$

Here y_i is the class label of two (Positive, Negative) class of SVM. The optimal limit with full margin is achieved by decreasing the objective function. i.e. $E = w^2$ subject to:

$$x_i.w + b \geq 1 \; for \; y_i = +1$$

$$x_i.w + b \leq 1 \; for \; y_i = -1 \qquad (2.10)$$

This can be incorporated into a series of disparities as follows:

$$(x_i.b + b)y_i \geq 1 \text{ for all i}$$

2.4.2. Decision tree (DT)

Decision Tree determined the series regularity and similarities of the dataset which can be verified by the classifier and grouped into different classes. Liu et al. [64] used DT to assign data based on the option of an attribute that maximizes and improves the division of data. The characteristics are divided into multiple divisions before the termination conditions are fulfilled. The DT algorithm is mathematically developed using the following equations:

$$\overline{X} = \{X_1, X_2, X_{3,\ldots\ldots}X_m\}^T$$

$$X_i = \{x_1, x_2, x_3, \ldots\ldots, x_{ij}, \ldots\ldots x_{in}\}$$

$$S = \{S_1, S_2, S_3, \ldots\ldots, S_i, \ldots\ldots S_m\} \qquad (2.11)$$

Where **m** corresponds to the number of observations available, **n** represents the several independent variable, **S** use the m-dimension vector of the variable projected from $\overline{X}$. X_i is the ith component of autonomous n-dimension variables, $x_1, x_2, x_3, \ldots\ldots, x_{ij}, \ldots\ldots x_{in}$ of the X_i vector pattern and T is the transpose notation.

The aim of DTs is to predict the $\overline{X}$ observations. It is possible to construct multiple DTs from $\overline{X}$ to various precision levels; although, the desirable DT is difficult since search space has a broad parameter dimension. Reasonable algorithms should be built for DT to represent the negotiate-off between precision and complexity. In this situation, the partitioning of the dataset $\overline{X}$ using DT algorithms uses a collection of local ideal decisions on the function instances. Which according to corresponding optimization method, Optimal DT, T_{k0} is built.

$$\hat{R}(T_{k0}) = min\{\hat{R}(T_{k0})\}, \qquad k = 1,2,3, \ldots\ldots K$$

$$\hat{R}(T) = \sum_{t \in T}^{k}\{r(t)p(t)\} \qquad (2.12)$$

Where $\hat{R}(T)$ denotes the uncertainty level during most of the T_k tree misclassification, T_{k0} represent the desirable DT that decrease the classification error in the binary tree misclassification, T represent the binary tree $\in, \{T_1, T_2, T_3, \ldots \ldots T_k, t_1\}$. The tree index is represented by **k**, tree node by **t**, root node by t_1, error resubstituting by r(t) misclassifying node **t,** likelihood that p(t) denotes some decrease in node **t**. The sub-trees of the right and left partitions are denoted by T^L and T^R. The tree T is created by portioning the feature plan. For larger datasets, there is classification problems as these data sets are these circumstances, the decision tree is a suitable strategy and contains errors. The objects are taken as input and the output in form of yes/ no decision is provided by the algorithm. The decision tree algorithms use Boolean function [65] and sample selection [66]. The decision tree algorithms are used in many applications such as bioinformatics economics, medical diagnoses problems and other scientific situations etc. [67].

2.4.3. Naïve Bayes

Naive Bayes is among the simplest probabilistic classifiers. In many real-world implementations, it also performs remarkably well, considering the firm presumption that, provided the class, all functions are conditionally independent. Pearl's (1988) proposed Bayesian Networks (BNs) are high-level description of distributed probabilities over a set of parameters $X = \{X_1, X_2, \ldots \ldots X_n\}$ used by a learning method. The learning method of the NBs is split into two steps: learning constructs and learning parameters. A directed acyclic graph from the set X is being built by the former. Every node refers to the parameter in the graph, and each Arc represents a probabilistic interaction between two parameters, whereas the Arc path implies the causality direction. The probabilistic node is called the parent of the other node when two nodes are connected by an arc, and another is called the child. To denote both the vector (feature) and its respective node, we use X_i, and $P_a(X_i)$ to denote the parent set of the X-i node. The discovery of probability distributions, class probabilities and conditional probabilities associated with each component is called parameter learning, provided a framework [68].

2.4.4. K-nearest neighbor (KNN)

The KNN classification was built from the need for discriminant analysis where it is unclear or difficult to establish accurate parametric estimates of probability densities. In the world of machine learning, KNN is the most commonly used algorithm for pattern recognition and many other fields used for classification problems. This algorithm is also known as an example-based algorithm (lazy learning). A model or classifier is not created automatically, but all samples of training data are preserved and kept until it is appropriate to identify new observations. This lazy learning algorithm feature makes it easier than eager learning to create a classifier until it is required to classify new observations.

In the world of machine learning, KNN is the most commonly used algorithm for pattern recognition and many other fields used for classification problems [69]. This algorithm is also recognizing as an example-based algorithm. A predictive algorithm is not created automatically, however all samples of training instances are preserved and kept until it is appropriate to identify new observations. This KNN algorithm feature makes it easier and simpler to construct a classifier than eager learning before new insights need to be listed. Where complex data must be modified and revised more easily, this algorithm is even more significant. KNN with various distance metrics was used [70]. The KNN algorithm operates using the Euclidean distance theorem in conjunction with the following steps.

Step I: To train and validate the model, provide the extracted feature set to KNN.

Step II: Measure distance using Euclidean distance formula.

$$d(x_i, y_i) = \sum_{i=1}^{n} \sqrt{(x_i - y_i)^2} \qquad (2.13)$$

Step III: Sort the values calculated using Euclidean distance using $d_i \leq d_i + 1$ where i=1,2,3,...k

Step IV: Depending on the quality of the results, apply the means or vote.

Step V: The K value (i.e. the number of nearest neighbors) depends on the sum and type of the KNN data supplied. The value of k is retained as large for large data, while the value of k is still kept tiny for small data.

2.5. *Training/testing data formulation*

For data training and testing formulation of the parameter, the Jack-knife k-fold cross validation methodology was used. 10-fold CVs is used in this research to test the efficiency of classifiers for various methods of extracting features. The most widely used and well-known methodology to test the output of classifiers is the 10-fold CV. The data is divided into 10 folds using 10-fold CV, 9 folds are involved in preparation, and sections of samples of remaining folds are expected based on the 9-fold testing. The research samples in the research fold are purely inaccessible to the qualified models. The entire process is replicated 10 times and is estimated appropriately by each class study. For other CVs, a corresponding approach is used. Finally, the projected labels for unseen samples are used to determine the precision of the designation. For any combination of system parameters, this procedure is repeated, and classification output for the sample has been recorded.

2.6. *Receiver operating characteristic curve (ROC)*

The ROC is graphed against the true positive rate (TPR), i.e. sensitivity and false positive rate (FPR), i.e. the CHF and NSR subjects' specificity values. The mean values of features for NSR subjects are graded as 1 and 0 for CHF subjects. The ROC function is then transferred to this vector, which plots each sample value against the values of specificity and sensitivity. ROC is one of the popular methods of calculating success in order to diagnose and interpret the efficacy of a classifier [71]. The TPR is graphed against the y-axis, and the x-axis is graphed against the FPR. The portion of a square unit is represented by the area under the curve (AUC). Its value varies from 0 to 1. The distinction is shown by AUC > 0.5. The superior diagnostic tool is shown by the greater AUC. TPR represents right positive cases calculated by dividing the total positive cases, while FPR represents negative cases expected as positive, calculated by dividing the total number of negative cases.

3. **Results and discussions**

In this study we extracted 22 multimodal features i.e. frequency domain features [very low frequency (VLF), ultra-low frequency (ULF), low frequency (LF), high frequency (HF), ratio of LF and HF (LFHF), total power (TP)], , time domain features [standard deviation of the consecutive intervals in each segment (SDNN), Standard deviation of the averages of EEG intervals (SDANN), standard deviation of differences between adjacent time series intervals in each segment (SDSD),

square root of the mean squared differences of N successive EEG time series intervals (RMSSD)]; statistical features (Skewness, smoothness, kurtosis, variance, root mean square), Nonlinear entropy features (approximate entropy, sample entropy based on KD tree) and wavelet entropy features (Threshold, Shannon, sure, norm, log energy) from Congestive Heart Failure (CHF) and Normal Sinus Rhythm (NSR) signals. We ranked the features and categorized the ranking based on ROC values obtained Category 01: all ranked features, Category 02: top five ranking features (Wavelet threshold, VLF, Kurtosis, ULF, TP), Category 03: top 09 ranking features (Wavelet threshold, VLF, Kurtosis, ULF, TP, LF, Skewness, SDNN, Variance), Category 04: last 13 ranked features (SDSD, HF, Wavelet sure, RMS, Approx. Entropy, Wavelet Shannon, Wavelet norm, smoothness, wavelet log energy, RMSSD, MSEKD, SDANN, LFHF) and Category 05: last two ranked features (SDANN, LFHF). The feature ranks are categorized on the value of ROC obtained using ranking method. We then employed machine learning techniques such as Decision Tree, Naïve Bayes, SVM Gaussian, RBF and Polynomial. Based on the all multimodal features, the highest performance was obtained using SVM Gaussian with sensitivity (93.06%), specificity (81.82%), PPV (89.33%), NPV (87.80%), accuracy (88.79%), FPR (0.1818) and AUC (0.9441) followed by Naïve Bayes with accuracy (88.79%), AUC (0.9296); SVM RBF with accuracy (87.93%) and AUC (0.9347); SVM polynomial with accuracy (87.93%), AUC (0.9343); decision tree with accuracy (78.45%), AUC (0.9296) as reflected in Table 1.

Based on the category 02 ranking features, the overall highest detection performance was obtained using Naïve followed by SVM Gaussian, decision tree and SVM polynomial. Using the category 03 features, we obtained highest performance using Naïve Bayes followed by SVM Gaussian, decision tree and SVM polynomial. Using the category 04 features, the highest performance was obtained using SVM polynomial followed by decision tree, SVM RBF Gaussian and Naïve Bayes. Based on the category 05 features, the highest detection performance was obtained using Naïve Bayes followed by SVM polynomial, SVM RBF, SVM Gaussian and decision tree.

Figure 3 (a–e) shows the area under the receiver operating curve to distinguish the CHF subjects from NSR subjects by extracting multimodal features with feature ranking using robust machine learning classifiers. We categorized the AUC performance based on category 01 ranked a) all 22 multimodal features, category 02 features b) Top five Ranked features, category 03 features c) Top nine ranked features, category 04 features d) last ranked 13 features and category 05 features e) last ranked two features.

Using the category 01 ranked features, the highest separation was obtained using SVM Gaussian with AUC (0.9441) followed by SVM RBF with AUC (0.9347), SVM polynomial with AUC (0.9343); Naïve Bayes and decision tree with AUC (0.9296). By using category 02 ranked features, the highest separation was obtained using decision tree & Naïve Bayes with AUC (0.8722) followed by SVM Gaussian with AUC (0.8633), SVM RBF with AUC (0.8204), SVM polynomial with AUC (0.7869). Based on the category 03 ranked features, the highest AUC was obtained using NB & decision tree. Likewise, based on the category 04 ranked features, the highest separation was obtained using SVM. Moreover, based on the category 05 ranked features, the highest separation was obtained using Naïve Bayes and decision tree. The AUC values of Naïve Bayes and Decision tree are same so are merged with one color.

Table 1. Detection of congestive heart failure detection performance with multimodal features-based features ranking using robust machine learning techniques.

| Method	Sensitivity	Specificity	PPV	NPV	Accuracy	FPR	AUC
Category 01: All Ranked features							
Naïve Bayes	0.8889	0.8864	0.9275	0.8298	0.8879	0.1136	0.9296
Decision Tree	0.8194	0.7273	0.831	0.7111	0.7845	0.2727	0.9296
SVM Gaussian	0.9306	0.8182	0.8933	0.878	0.8879	0.1818	0.9441
SVM RBF	0.9028	0.8409	0.9028	0.8409	0.8793	0.1591	0.9347
SVM poly.	0.9722	0.7273	0.8537	0.9412	0.8793	0.2727	0.9343
Category 02: Top five Ranked Features							
Naïve Bayes	0.8611	0.7727	0.8611	0.7727	0.8276	0.2273	0.8722
Decision Tree	0.8889	0.5	0.7442	0.7333	0.7414	0.5	0.8722
SVM Gaussian	0.9722	0.5909	0.7955	0.9286	0.8276	0.4091	0.8633
SVM RBF	0.9306	0.5909	0.7882	0.8387	0.8017	0.4091	0.8204
SVM poly.	0.8611	0.5455	0.7561	0.7059	0.7414	0.4545	0.7869
Category 03: Top Nine Ranked Features							
Naïve Bayes	0.8889	0.7727	0.8649	0.8095	0.8448	0.2273	0.8767
Decision Tree	0.8472	0.6364	0.7922	0.7179	0.7672	0.3636	0.8767
SVM Gaussian	0.8611	0.6818	0.8158	0.75	0.7931	0.3182	0.8523
SVM RBF	0.8472	0.6591	0.8026	0.725	0.7759	0.3409	0.8479
SVM poly.	0.875	0.5455	0.759	0.7273	0.75	0.4545	0.6979
Category 04: Last thirteen Ranked Features							
Naïve Bayes	0.8056	0.7045	0.8169	0.6889	0.7672	0.2955	0.828
Decision Tree	0.8333	0.7955	0.8696	0.7447	0.819	0.2045	0.828
SVM Gaussian	0.875	0.5909	0.7778	0.7429	0.7672	0.4091	0.8422
SVM RBF	0.875	0.6136	0.7875	0.75	0.7759	0.3864	0.8469
SVM poly.	0.9167	0.6136	0.7952	0.8182	0.8017	0.3864	0.8374
Category 05: Last Two Ranked Features							
Naïve Bayes	0.9722	0.4318	0.7368	0.9048	0.7672	0.5682	0.761
Decision Tree	0.7361	0.5455	0.726	0.5581	0.6638	0.4545	0.761
SVM Gaussian	0.8889	0.4545	0.7273	0.7143	0.7241	0.5455	0.7301
SVM RBF	0.8889	0.4545	0.7273	0.7143	0.7241	0.5455	0.72
SVM poly.	0.8889	0.4545	0.7273	0.7143	0.7241	0.5455	0.7336
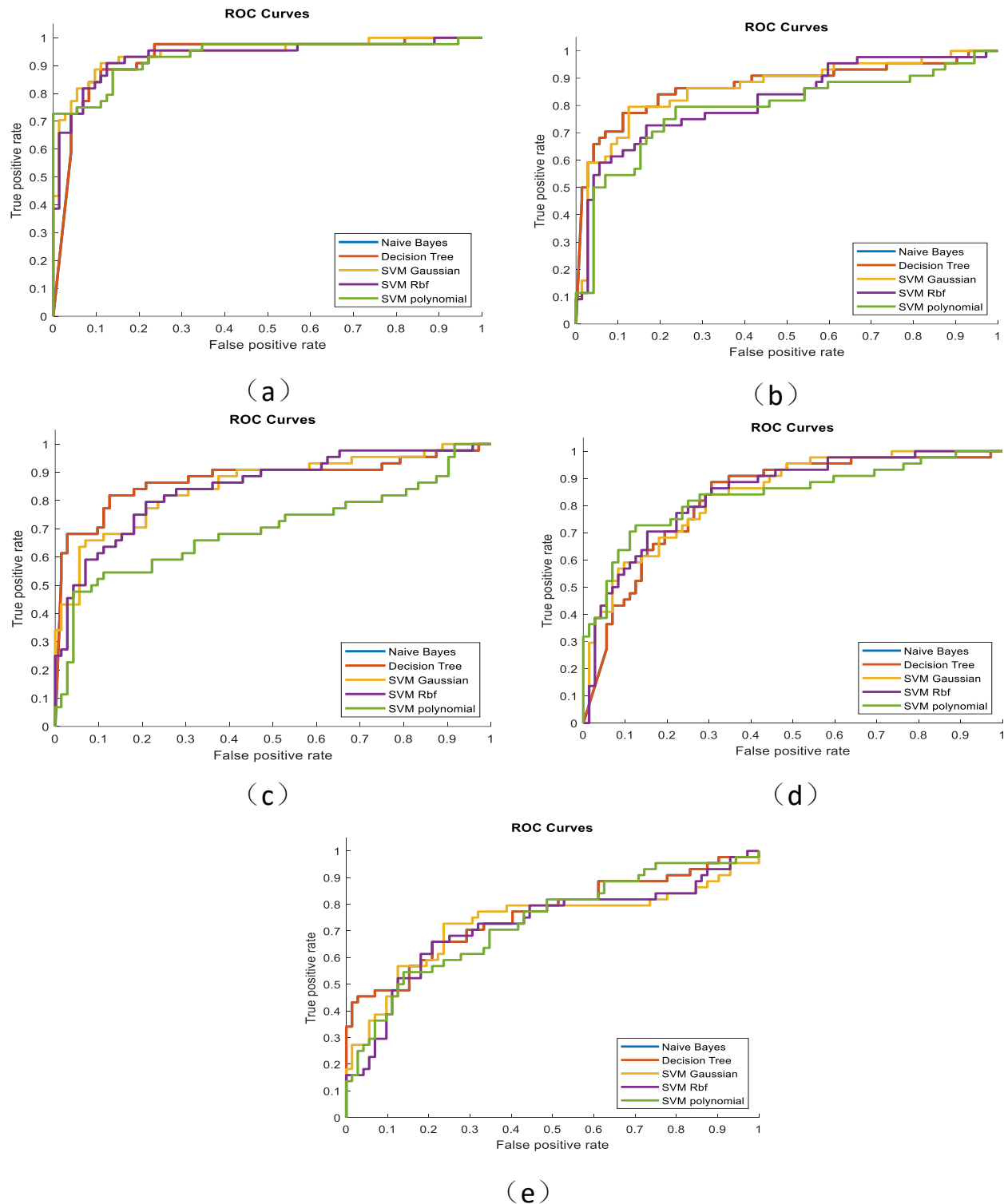
Figure 3. Area under the receiver operating characteristics (ROC) curve to determine the separation between CHF and NSR subjects using a) All features, b) Top five Ranked features, c) Top nine ranked features, d) last ranked 13 features, e) last ranked two features. The highest AUC values reflect the highest separation using robust machine learning classifiers such as Naïve Bayes, decision tree, SVM with Gaussian, RFB and polynomial kernels.

Table 2. Comparison of results with previous studies.

Author	Method	Performance
Li et al. [72]	Convolutional Neural network	Accuracy = 81.9%; Accuracy = 81.92%
Isler Kuntalp [3]	K-nearest Neighbor	Sensitivity = 82.74%; Specificity = 96.27%
Narin et al. [73]	Support Vector Machine	Sensitivity = 79.33%; Specificity = 94.47%
Isler Kuntalp., [74]	K-nearest Neighbor	Sensitivity = 82.72%; Specificity = 100.0%
Pecchia et al. [9]	Classification and regression tree	Sensitivity = 89.75%; Specificity = 100.0%
Elfadil Ibrahim. [75]	Spectral Neural network	Accuracy = 83.65%
Yang et al. [76]	Support vector machine Naïve Bayes	Accuracy = 74.42%
Son et al. [77]	Decision tree	Sensitivity = 97.53%
	SVM Gaussian	Sensitivity (93.06%); Specificity (81.82%) Accuracy (88.79%); AUC (0.9441)
This work	Naïve Bayes	Accuracy (82.76%); AUC (0.8722)
	SVM Gaussian	Accuracy (76.72%); AUC (0.828)
	Naïve Bayes	Accuracy (76.72%); AUC (0.761)

The Table 2 reflect the summary of results obtained for different feature extracting strategies and classification algorithms. We aimed to check the features importance in detecting the congestive heart failure by ranking the features. This ranking will help the clinicians that which features are more import for them to make further decision. From the results depicted in Table 2, it is important to note that using all 22 multimodal features together, the highest accuracy (88.79%), AUC (0.9441) was obtained, while using only top five ranked features there was a very low decrease in performance with accuracy (82.76%), AUC (0.8722). This indicates that these five ranked features are more important than the other all extracted features in decision making. Similarly, the low ranked features further yielded the low detection performance.

The heart rate dynamics are highly complex and nonlinear. The patients admitted in the emergency department with complain of shortness of breathing, increase of lower extremity edema, dyspnea on exertion, lower extremity edema, and or worsening fatigue should have heart failure, which require differential diagnosis. The temporal and spectral dynamics can be analyzed the time domain and frequency domain methods. The dynamics of complex systems also degraded due to aging and disease. To capture these dynamics, we extracted the nonlinear entropy and wavelet-based entropy measures. The researchers in the past extracted multidomain and modal features to detect epileptic seizure [31,33], arrhythmia detection [16], seizure detection using time frequency representation methods [78] and cancer detection such as lung cancer dynamics using refined fuzzy entropy methods [34], lung cancer detection based on multimodal features [79] and colon cancer based on hybrid feature extracting strategy [17]. Recently Singh et al. [80] analysed that coronary heart disease with diabetes mellitus patients get significant results in clinical symptoms with improvement in the quality of life. To detect the heart rate variability, they employed SVM with RBF and decision tree [81]. The results obtained their studies revealed very good detection performance [82].

This study is aimed to compute the congestive heart failure detection performance by ranking the multimodal features. The feature ranking may help the clinicians that which features are most suitable for further decision making. The feature ranking method also ranked the feature importance. There are

different feature ranking methods, we ranked the importance of Multimodal features based on the class separability criteria of the area between EROC and random classifier slope. The ranked features are than categorized based on ROC values achieved i.e. high ROC values, medium, low and very small ROC values. These categories can also help us to determine the detection performance by using all extracted features and categorized ROC values. We have observed that among the 22 multimodal features, the top ranked features gained a reasonable high detection performance. The top five selected features based on ranking methods were from wavelet, frequency domain and few statistical such as wavelet threshold, very low frequency, kurtosis, ultra-low frequency, total power. This indicates that these features are very helpful in detection the congestive heart failure. Moreover, with the lowest ROC ranked features, the detection performance was decreased. With the very low ranked features such as SDANN and LFHF, the performance was decreased further. These different categories also helped to determine the detection performance in a better way.

4.   Conclusion

Heart rate variability analysis is a non-invasive tool used for assessing the cardiac autonomic control of nervous system. The congestive heart failure is the major problem worldwide. The researchers are developing efficient tools to improve the detection performance. In the past, researchers used different features extraction approaches. However, feature ranking also plays a vital role to judge the importance of features based on various factors. The important features can be very helpful for clinicians and radiologists to make the early decision. In this study, we extracted the multimodal features from both CHF and NSR subjects. We then ranked the features based on ROC values. The performance was measured based on categorized ranked features from one to five different ranking categories in order to see the performance results with top, medium, and low ranked features. Based on all features used, the highest performance with accuracy (88.79%), AUC (0.9441) was obtained using SVM Gaussian. Based on the top five ranked features (i.e. wavelet entropy threshold, VLF, kurtosis, ULF, TP) with ROC value > 3 yielded highest detection performance with accuracy (82.76%), AUC (0.822) using, whereas top 9 features with ROC value between 2–3 yielded an accuracy (84.48%), AUC (0.8767) using Naïve Bayes. The ranked features based on their importance will greatly be helpful for clinicians for further decision making and can greatly impact in reducing the mortality rate. The results with top ranked features contributed a lot, while the performance results with low ranked features dramatically decreased.

Limitation and future recommendations

Currently, we have used the dataset with small sample size and lack of clinical information. In future, we will acquire big data and clinical profile of the patients. Moreover, we will explore more relationship to determine the feature importance base on different ranking methods and associations among the features. We will also extract and ranked these features for New York Heart Association (NYHA) functional classes and compute associations and ranks accordingly. We will also explore the association between different multimodal extracted features by computing the strengths and coupling relation which will further assist the clinicians to find the association and strength between and among the extracted features. The ranked features will further help the clinicians for further diagnosis and treatment of the patient.

Acknowledgements

Conflicts of interest

The authors declare no conflict of interest in this paper.

References

1. H. F. Jelinek, D. J. Cornforth, A. H. Khandoker, ECG Time Series Variability Analysis, *CRC Press*, 2017.

2. A. J. Seely, P. T. Macklem, Complex systems and the technology of variability analysis, *Crit. Care.*, **8** (2004), R367.

3. Y. İşler, M. Kuntalp, Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure, *Comput. Biol. Med.,* **37** (2007), 1502–1510.

4. I. Awan, W. Aziz, I. H. Shah, N. Habib, J. S. Alowibdi, S. Saeed, et al., Studying the dynamics of interbeat interval time series of healthy and congestive heart failure subjects using scale based symbolic entropy analysis, *PLoS One.*, **13** (2018), e0196823.

5. W. Aziz, M. Rafique, I. Ahmad, M. Arif, N. Habib, M. Nadeem, Classification of heart rate signals of healthy and pathological subjects using threshold based symbolic entropy, *Acta Biol. Hung.*, **65** (2014), 252–264.

6. A. Hossen, B. Al-Ghunaimi, A wavelet-based soft decision technique for screening of patients with congestive heart failure, *Biomed. Signal Process. Control.*, **2** (2007), 135–143.

7. R. A. Thuraisingham, A classification system to detect congestive heart failure uing second-order difference plot of RR intervals, *Cardiol. Res. Pract.*, **2009** (2009), 1–7.

8. S. N. Yu, M. Y. Lee, Conditional mutual information-based feature selection for congestive heart failure recognition using heart rate variability, *Comput. Methods Programs Biomed.*, **108** (2012), 299–309.

9. L. Pecchia, P. Melillo, M. Sansone, M. Bracale, Discrimination power of short-term heart rate variability easures for CHF assessment, *IEEE Trans. Inf. Technol. Biomed.*, **15** (2011), 40–46.

10. G. Altan, Y. Kutlu, N. Allahverdi, A new approach to early diagnosis of congestive heart failure disease by using Hilbert–Huang transform, *Comput. Methods Programs Biomed.*, **137** (2016), 23–34.

11. G. I. Choudhary, W. Aziz, I. R. Khan, S. Rahardja, P. Franti, Analysing the dynamics of interbeat interval time series using grouped horizontal visibility graph, *IEEE Access*, **7** (2019), 9926–9934.

12. Y. Isler, A. Narin, M. Ozer, M. Perc, Multi-stage classification of congestive heart failure based on short-term heart rate variability, *Chaos Solitons Fractals*, **118** (2019), 145–151.

13. A. Narin, Y. Isler, M. Ozer, M. Perc, Early prediction of paroxysmal atrial fibrillation based on short-term heart rate variability, *Phys. A Stat. Mech. Its Appl.*, **509** (2018), 56–65.

14. T. Jagrič, M. Marhl, D. Štajer, Š. T. Kocjančič, T. Jagrič, M. Podbregar, et al., Irregularity test for very short electrocardiogram (ECG) signals as a method for predicting a successful defibrillation in patients with ventricular fibrillation, *Transl. Res.*, **149** (2007), 145–151.

15. L. Wang, W. Xue, Y. Li, M. Luo, J. Huang, W. Cui, et al., Automatic Epileptic Seizure Detection in EEG Signals Using Multi-Domain Feature Extraction and Nonlinear Analysis, *Entropy*, **19**

(2017), 222.

16. L. Hussain, W. Aziz, S. Saeed, I. A. Awan, A. A. Abbasi, N. Maroof, Arrhythmia detection by extracting hybrid features based on refined Fuzzy entropy (FuzEn) approach and employing machine learning techniques, *Waves Random Complex Media,* **30** (2020), 656–686.

17. S. Rathore, M. Hussain, A. Khan, Automated colon cancer detection using hybrid of novel geometric features and some traditional features, *Comput. Biol. Med.,* **65** (2015), 279–296.

18. H. Peng, F. Long, C. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy, *IEEE Trans. Pattern Anal. Mach. Intell.,* **27** (2005), 1226–1238.

19. H. Karnan, N. Sivakumaran, R. Manivel, An efficient cardiac arrhythmia onset detection technique using a novel feature rank score algorithm, *J. Med. Syst.,* **43** (2019), 167.

20. M. G. Leguia, Z. Levnajić, L. Todorovski, B. Ženko, Reconstructing dynamical networks via feature ranking, *Chaos Interd. J. Nonlinear Sci.,* **29** (2019), 093107.

21. Z. Zhou, S. Li, G. Qin, M. Folkert, S. Jiang, J. Wang, Multi-objective-based radiomic feature selection for lesion malignancy classification, *IEEE J. Biomed. Heal. Inform.,* **24** (2020), 194–204.

22. M. Mourad, S. Moubayed, A. Dezube, Y. Mourad, K. Park, A. Torreblanca-Zanca, et al., Machine learning and feature selection applied to SEER data to reliably assess thyroid cancer prognosis, *Sci. Rep.,* **10** (2020), 5176.

23. A. P. Bradley, The use of the area under the ROC curve in the evaluation of machine learning algorithms, *Patt. Recogn.,* **30** (1997), 1145–1159.

24. A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, et al., PhysioBank, PhysioToolkit, and PhysioNet, *Circulation,* **101** (2000).

25. J. T. Bigger, J. L. Fleiss, R. C. Steinman, L. M. Rolnitzky, W. J. Schneider, P. K. Stein, RR variability in healthy, middle-aged persons compared with patients with chronic coronary heart disease or recent acute myocardial infarction, *Circulation,* **91** (1995), 1936–1943.

26. J. E. Mietus, The pNNx files: Re-examining a widely used heart rate variability measure, *Heart,* **88** (2002), 378–380.

27. K. L. Dodds, C. B. Miller, S. D. Kyle, N. S. Marshall, C. J. Gordon, Heart rate variability in insomnia patients: A critical review of the literature, *Sleep Med. Rev.,* **33** (2017), 88–100.

28. M. R. Esco, H. N. Williford, A. A. Flatt, T. J. Freeborn, F. Y. Nakamura, Ultra-shortened time-domain HRV parameters at rest and following exercise in athletes: An alternative to frequency computation of sympathovagal balance, *Eur. J. Appl. Physiol.,* **118** (2018), 175–184.

29. S. A. Geronikolou, K. Albanopoulos, G. Chrousos, D. Cokkinos, Evaluating the homeostasis assessment model insulin resistance and the cardiac autonomic system in bariatric surgery patients: A meta-analysis, *in: P. Vlamos (Ed.), Springer International Publishing,* Cham, 2017, 249–259.

30. C. A. Sima, J. A. Inskip, A. W. Sheel, S. F. van Eeden, W. D. Reid, P. G. Camp, The reliability of short-term measurement of heart rate variability during spontaneous breathing in people with chronic obstructive pulmonary disease, *Rev. Port. Pneumol. (English Ed.),* **23** (2017), 338–342.

31. L. Hussain, Detecting epileptic seizure with different feature extracting strategies using robust machine learning classification techniques by applying advance parameter optimization approach, *Cogn. Neurodyn.,* **12** (2018), 271–294.

32. L. Hussain, I. A. Awan, W. Aziz, S. Saeed, A. Ali, F. Zeeshan, et al., Detecting congestive heart failure by extracting multimodal features and employing machine learning techniques, *Biomed. Res. Int.,* **2020** (2020), 1–19.

33. L. Hussain, W. Aziz, J. S. Alowibdi, N. Habib, M. Rafique, S. Saeed, et al., Symbolic time series analysis of electroencephalographic (EEG) epileptic seizure and brain dynamics with eye-open and eye-closed subjects during resting states, *J. Physiol. Anthropol.*, **36** (2017).

34. L. Hussain, W. Aziz, A. A. Alshdadi, M. S. A. Nadeem, I. R. Khan, Q. U. A. Chaudhry, Analyzing the dynamics of lung cancer imaging data using refined fuzzy entropy methods by extracting different features, *IEEE Access,* **7** (2019), 64704–64721.

35. L. Hussain, W. Aziz, S. Saeed, S. A. Shah, M. S. A. Nadeem, A. Awan, et al., Complexity analysis of EEG motor movement with eye open and close subjects using multiscale permutation entropy (MPE) technique, *Biomed. Res.*, **28** (2017), 104–7111.

36. L. Hussain, W. Aziz, S. Saeed, S. A. Shah, M. S. A. Nadeem, I. A. Awan, et al., Quantifying the dynamics of electroencephalographic (EEG) signals to distinguish alcoholic and non-alcoholic subjects using an MSE based K-d tree algorithm, *Biomed. Eng. Biomed. Tech.*, **63** (2018), 481–490.

37. L. Hussain, S. Saeed, A. Idris, I. A. Awan, S. A. Shah, A. Majid, et al., Regression analysis for detecting epileptic seizure with different feature extracting strategies, *Biomed. Eng. Biomed. Tech.*, **64** (2019), 619–642.

38. M. Pincus, Approximate entropy as a measure of system complexity, *Proc. Natl. Acad. Sci.*, **88** (1991), 2297–2301.

39. J. S. Richman, J. R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, *Am. J. Physiol. Circ. Physiol.*, **278** (2000), H2039–H2049.

40. D. Wang, D. Miao, C. Xie, Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection, *Expert Syst. Appl.*, **38** (2011), 14314–14320.

41. O. A. Rosso, S. Blanco, J. Yordanova, V. Kolev, A. Figliola, M. Schürmann, et al., Wavelet entropy: A new tool for analysis of short duration brain electrical signals, *J. Neurosci. Methods*, **105** (2001), 65–75.

42. Y. Wu, Y. Zhou, G. Saveriades, S. Agaian, J.P. Noonan, P. Natarajan, Local Shannon entropy measure with statistical tests for image randomness, *Inf. Sci. (Ny).,* **222** (2013), 323–342.

43. S. Ekici, S. Yildirim, M. Poyraz, Energy and entropy-based feature extraction for locating fault on transmission lines by using neural network and wavelet packet decomposition, *Expert Syst. Appl.*, **34** (2008), 2937–2944.

44. E. Avci, D. Hanbay, A. Varol, An expert discrete wavelet adaptive network based fuzzy inference system for digital modulation recognition, *Expert Syst. Appl.*, **33** (2007), 582–589.

45. I. Turkoglu, A. Arslan, E. Ilkay, An intelligent system for diagnosis of the heart valve diseases with wavelet packet neural networks, *Comput. Biol. Med.*, **33** (2003), 319–331.

46. H. Wang, T. M. Khoshgoftaar, K. Gao, A comparative study of filter-based feature ranking techniques, *in: 2010 IEEE Int. Conf. Inf. Reuse Integr.,* IEEE, 2010, 43–48.

47. H. Shakir, Y. Deng, H. Rasheed, T. M. R. Khan, Radiomics based likelihood functions for cancer diagnosis, *Sci. Rep.*, **9** (2019), 9501.

48. W. Wu, C. Parmar, P. Grossmann, J. Quackenbush, P. Lambin, J. Bussink, et al, Exploratory study to identify radiomics classifiers for lung cancer histology, *Front. Oncol.,* **6** (2016), 187–194.

49. Y. Saeys, I. Inza, P. Larranaga, A review of feature selection techniques in bioinformatics, *Bioinformatics,* **23** (2007), 2507–2517.

50. L. Zhu, L. Miao, D. Zhang, Iterative laplacian score for feature selection, *in: Proc. 18th Int. Conf. Neural Inf. Process. Syst.*, 2012, 80–87.

51. A. K. Farahat, A. Ghodsi, M. S. Kamel, Efficient greedy feature selection for unsupervised

learning, *Knowl. Inf. Syst.*, **35** (2013), 285–310.

52. P. Mitra, C. A. Murthy, S. K. Pal, Unsupervised feature selection using feature similarity, *IEEE Trans. Pattern Anal. Mach. Intell.*, **24** (2002), 301–312.

53. D. Cai, C. Zhang, X. He, Unsupervised feature selection for multi-cluster data, *in: Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. - KDD '10, ACM Press, New York, New York, USA, New York, USA*, 2010, 333.

54. H. Zeng, Y. Cheung, Feature selection and Kernel learning for local learning-based clustering, *IEEE Trans. Pattern Anal. Mach. Intell.*, **33** (2011),1532–1547.

55. Z. Zhao, H. Liu, Spectral feature selection for supervised and unsupervised learning, *in: Proc. 24th Int. Conf. Mach. Learn. - ICML '07, ACM Press, New York, New York, USA*, 2007, 1151–1157.

56. W. Yang, K. Wang, W. Zuo, Neighborhood component feature selection for high-dimensional data, *J. Comput.*, **7** (2012), 161–168.

57. I. Kononenko, E. Šimec, M. Robnik-Šikonja, Overcoming the myopia of inductive learning algorithms with RELIEFF, *Appl. Intell.*, **7** (1997), 39–55.

58. G. Rofo, S. Melzi, Ranking to learn: Feature ranking and selection via eigenvector centrality, in: A. Appice, M. Ceci, C. Loglisci, E. Masciari and Z. Ras, (eds.) *New Frontiers in Mining Complex Patterns: 5th International Workshop*, NFMCP 2016.

59. P. Bradley, O. Mangasarian, Feature selection via concave minimization and support vector machines, *in: Proc. Int. Conf. Mach. Learn.*, 1998,: pp. 82–90.

60. S. Yu, Z. Zhang, X. Liang, J. Wu, E. Zhang, W. Qin, et al., A Matlab toolbox for feature importance ranking, *in: 2019 Int. Conf. Med. Imaging Phys. Eng., IEEE*, 2019,: pp. 1–6.

61. V. N. Vapnik, An overview of statistical learning theory, *IEEE Trans. Neural Networks*, **10** (1999), 988–999.

62. P. Toccaceli, A. Gammerman, Combination of conformal predictors for classification, *Proc. Sixth Work. Conform. Probabilistic Predict. Appl.*, **60** (2017), 39–61.

63. A. Subasi, Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders, *Comput. Biol. Med.*, **43** (2013), 576–586.

64. W. Liu, S. Chawla, D. A. Cieslak, N. V. Chawla, A robust decision tree algorithm for imbalanced data sets, *in: Proc. 2010 SIAM Int. Conf. Data Min., Society for Industrial and Applied Mathematics, Philadelphia, PA*, 2010, 766–777.

65. M. J. Aitkenhead, A co-evolving decision tree classification method, *Expert Syst. Appl.*, **34** (2008), 18–25.

66. R.Wang, S. Kwong, X.Wang, Q. Jiang, Segment based decision tree induction with continuous valued attributes, *IEEE Trans. Cybern.*, **45** (2015), 1262–1275.

67. J. J. Rissanen, Fisher information and stochastic complexity, *IEEE Trans. Inf. Theory.*, **42** (1996), 40–47.

68. A. Zaidi, B. Ould Bouamama, M. Tagina, Bayesian reliability models of Weibull systems: State of the art, *Int. J. Appl. Math. Comput. Sci.*, **22** (2012), 585–600.

69. P. Zhang, B.J. Gao, X. Zhu, L. Guo, Enabling fast lazy learning for data streams, *in: 2011 IEEE 11th Int. Conf. Data Min., IEEE*, 2011, 932–941.

70. F. Schwenker, E. Trentin, Pattern classification and clustering: A review of partially supervised learning approaches, *Pattern Recognit. Lett.*, **37** (2014), 4–14.

71. K. Hajian-Tilaki, Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation, *Casp. J. Intern. Med.*, **4** (2013), 627–635.

72. Y. Li, Y. Zhang, L. Zhao, Y. Zhang, C. Liu, L. Zhang, et al., Combining convolutional neural network and distance distribution matrix for identification of congestive heart failure, *IEEE Access*, **6** (2018), 39734–39744.

73. A. Narin, Y. Isler, M. Ozer, Investigating the performance improvement of HRV Indices in CHF using feature selection methods based on backward elimination and statistical significance, *Comput. Biol. Med.*, **45** (2014), 72–79.

74. Y. Işler, M. Kuntalp, Heart rate normalization in the analysis of heart rate variability in congestive heart failure, *Proc. Inst. Mech. Eng. Part H J. Eng. Med.*, **224** (2010), 453–463.

75. N. Elfadil, I. Ibrahim, Self organizing neural network approach for identification of patients with Congestive Heart Failure, *in: 2011 Int. Conf. Multimed. Comput. Syst.,* IEEE, 2011, 1–6.

76. G. Yang, Y. Ren, Q. Pan, G. Ning, S. Gong, G. Cai, et al., A heart failure diagnosis model based on support vector machine, *in: 2010 3rd Int. Conf. Biomed. Eng. Informatics,* IEEE, 2010, 1105–1108.

77. C. S. Son, Y. N. Kim, H. S. Kim, H. S. Park, M. S. Kim, Decision-making model for early diagnosis of congestive heart failure using rough set and decision tree approaches, *J. Biomed. Inform.*, **45** (2012), 999–1008

78. L. Hussain, W. Aziz, S. Saeed, A. Idris, I.A. Awan, S.A. Shah, et al., Spatial wavelet-based coherence and coupling in EEG signals with eye open and closed during resting state, *IEEE Access*, **6** (2018), 37003–37022.

79. L. Hussain, S. Rathore, A. A. Abbasi, S. Saeed, Automated lung cancer detection based on multimodal features extracting strategy using machine learning techniques, in: H. Bosmans, G. H. Chen, T. Gilat Schmidt (Eds.), *Med. Imaging 2019 Phys. Med. Imaging,* SPIE, 2019, 134.

80. V. Singh, G. Kumari, B. Chhajer, A. K. Jhingan, S. Dahiya, Effectiveness of enhanced external counter pulsation on clinical profile and health-related quality of life in patients with coronary heart disease: A systematic review, *Acta Angiol.*, **24** (2018), 105–122.

81. Y. Isler, A. Narin, M. Ozer, Comparison of the effects of cross-validation methods on determining performances of classifiers used in diagnosing congestive heart failure, *Meas. Sci. Rev.*, **15** (2015), 196–201.

82. R. Han, X. Liu, M. Zheng, R. Zhao, X. Liu, X. Yin, et al., Effect of remote ischemic preconditioning on left atrial remodeling and prothrombotic response after radiofrequency catheter ablation for atrial fibrillation, *Pacing Clin. Electrophysiol.*, **41** (2018), 246–254