

Effective F-Score Feature Selection (KFFS) and Fuzzy Neural Network (FNN) to Classify Congestive Heart Failure Patients

S. Iniya Raghavi

Research Scholar, Department of Computer Science,
Hindusthan College of Arts And Science,
Nava India, Coimbatore-641028,
Tamilnadu,
India,

Mrs. P. Lalitha

Assistant Professor, Department of Computer Science,
Hindusthan College of Arts And Science,
Nava India, Coimbatore-641028,
Tamilnadu,
India,

Abstract- In this paper proposed a kernel F-score feature selection (KFFS) to select the features of the heart Rate Variability (HRV) congestive heart failure patients. In the proposed KEFS methods the input HRV data are transformed into linear kernel functions. If the specific F-score value of anyone of the feature in CHF is bigger than this mean value of F-score value of the feature that feature will be selected or else it is removed from the patient records. From this point the unimportant feature are removed. Fuzzy neural network (FNN) was employed to classify the HRV with CHF selected features data into low, middle and high risk to assess the risk of the each patient. It shows that the proposed FNN classification has higher classification results in terms of the sensitivity, precision and a specificity rate in discovering elevated risk patients. Finally, the results achieved by the FNN are understandable and reliable through preceding learning with the intention of depressed HRV for evaluation of the risk for patients who have majorly suffered from CHF.

Key words: Feature Selection, Classification, Heart Rate Variability (HRV), Fuzzy Neural Network (FNN), Radial Basis Function (RBF), kernel F-score feature selection (KFFS), Congestive Heart Failure (CHF) patients.

I. INTRODUCTION

Congestive heart failure (CHF) is a pathophysiological situation occurs due to abnormal condition in the heart based on the blood pressure in the human body. This CHF is measured based on the classification methods. Similarly Heart rate variability (HRV) is occurs based on the differentiation among heartbeats from electrocardiographic signal (ECG) recorded signal for each one of the patient.

Several number of the studies have been proposed in earlier work to measure the relationship between the HRV and CHF [1-2]. The lower level of the CHF is measured between ranges from 0.03 and 0.15 Hz. Guzzetti et al [3] also proposed a non-linear analysis based HRV to analysis the CHF patients simultaneously perform for both time domain and spectral analysis of HRV. Though, HRV parameters were as anticipated concerned through its impulsive variation [4], respiration [5], proposition artifacts [6], and so on. As a result it said that some of the HRV based assessment not used to measure CHF evaluation.

Conservative, the New York Heart Association (NYHA) classification, is one of the majorly used classification method to measure the CHF with HRV rate [7-8].

Recently, small number of the works has been proposed with more attention to analysis the CHF and HRV assessment proves based on the HRV signals from the each one of the patient. Asyali et al [9] developed an efficient time and frequency domain based HRV Bayesian classifier to distinguish CHF disease. The results of this classifier is measured in terms of the classification parameters such as with sensitivity and specificity rate of respectively [9]. Isler et al [10] developed an efficient differentiation method to combine the features of the HRV from the wavelet entropy and measure the HRV parameters through the genetic algorithms; classification is performed by using the k-nearest neighbor classifier.

Several number of the work have been performed to measure the relationship between the HRV and the CHF [11] with NYHA classification scale. [11]. But these classifier not support the HRV based features, it is solved by Yang et al. [12] who included HRV. This method doesn't provide information about the signals.

The main motivation of the research is to investigation the difference of the moment in time epoch among consecutive heart beats, through the altering of the physiological conditions. It measures the intrinsic and extrinsic relationship of the heart rate. It is helpful to appreciate the interplay among the considerate and parasympathetic nervous system, which provide to up and slow down the heart rate, correspondingly.

This work proposed an automatic HRV with CHF based classification system to analysis the risk of the everyone of the patients. Initially the patients records are labeled into low, high and medium risk patients from Holter databases. These mild risk are labeled as NYHA I and II, lower risk and higher risk patients records are labeled as NYHA III and IV. In this work the important features of Congestive Heart Failure (CHF) patient's data are selected based on the kernel F-score feature selection methods. Fuzzy neural network is proposed to classify the CHF patient's records into low, high and middle CHF risks.

II. BACKGROUND STUDY

In the twenty years ago, Sayers as well as others mainly focused attention existence of physiological rhythms. It is imbedded in beat-to-beat heart rate signal. Frequency-domain analyses throw in accepts automatic background RR interval fluctuations in the heart rate record [13]. The clinical substance of HRV is a strong furthermore self-determining predictor of humanity follow an acute myocardial infarction. AMONG the accessibility of new, digital, high frequency, 24-hours multi-channel electrocardiographic recorders, HRV potential on the way to provide further valuable insight interested in physiological along with pathological conditions also to enhance risk stratification.

To measure the results of the HRV the series of the regular normal NN intervals have been converted into the geometric form of the pattern using the sample density distribution, if finds the difference between the Normal NN intervals and Lorenz plot of NN intervals, etc., and a straightforward method is used which board of judges the unpredictability founded based on the geometric properties. For the most part geometric methods necessitate the NN interval succession to be transformed to a disconnected scale which is not too very well and which authorize the structure of round histograms.

The most important key advantage of the geometric methods is that it fabrication in their comparative inattentiveness to the investigative value of the progression of NN intervals [14]. But the major problem of this works is that it requires at least 20 min to create a geometric pattern to measure the HRV and CHR V results .In realtime applications it becomes not a easy task so these methods are not easily adaptable to smallest range of the HRV and CHF.

A. Spectral components:

In generally there are three major spectral components are specified in this work to measures the Heart Rate Variability (HRV) measured from the normal short-term recordings between two to five min [15] intervals very low frequency (VLF), low frequency (LF), and high frequency (HF) components. The values of the LF and HF are not fixed based on any threshold values, but it may be varied based on the inflection of the heart rate time [15]. The measurement of the VLF component is used less in earlier works and presented VLF component are mostly attribute dependent able to the HRV.

B. Rhythm pattern analysis:

In earlier 'peak-valley' procedures are proposed to measure the HRV rate, it is based on the procedure of the summit and the lowest point of oscillations [16]. In generally the procedure of the summit is easy to extend the longer variations to measure the HRV measurements. The variations of the HRV are distinguished based on the slowing, frequency, the wavelength and amplitude. The most majority of the short- to mid-term recordings, the outcomes are measured based on the frequency and time domain components of Heart Rate Variability [17].

C. Non-linear methods

These methods are generally used to measure the results from genesis of HRV. It is determined based on the multifaceted relations of haemodynamic, electrophysiological and human being variables through essential worried system. It has been hypothesize with the intention for the analysis of HRV with non linear method with the intention to obtain important information and evaluation of the possibility in unexpected death. It includes the following parameters to include the 1/f scaling Poincarè division and attractor course [18].

D. Stability and reproducibility of HRV measurement

Numeral of methods have been proposed in earlier work or earlier research to measure the Heart Rate Variability (HRV) for each one of the patients after the transient perturbations stimulate through temporary coronary occlusion, etc. Further commanding stimulus, such as most do exercises might consequence in a great deal more prolonged time interval ahead of the control values. There are some of the few HRV measures also attained from continuously monitoring 24-h time interval in both usual subjects [19] populations.

Recently, small number of the works has been proposed with more attention to analysis the CHF and HRV assessment proves based on the HRV signals from the each one of the patient. Asyali et al [9] developed an efficient time and frequency domain based HRV Bayesian classifier to distinguish CHF disease .The results of this classifier is measured in terms of the classification parameters such as with sensitivity and specificity rate of respectively [9]. Isler et al [10] developed an efficient differentiation method to combine the features of the HRV from the wavelet entropy and measure the HRV parameters through the genetic algorithms; classification is performed by using the k-nearest neighbor classifier

III. PROPOSED KERNEL FEATURE SELECTION AND FUZZY NEURAL NETWORK METHODOLOGY

In this paper, we have presents a new kernel F-score feature selection to select important feature of the CHF including heart disease to improve classification accuracy .In the proposed KEFS methods the input HRV data are transformed into Linear kernel functions. And then, compute F-score values to each CHF patient record features. For that F –Score value [19] for each CHF feature the mean value should be calculated. If the specific F-score value of anyone of the feature in CHF is bigger than this mean value of F-score value of the feature that feature will be selected or else it is removed from the patient records. This type of the situation happens if the dataset becomes small and the unbalanced or imbalance dataset, in order to deal with this problem the tree creation is performed to all the data in the dataset it measure k out of all N features.

For selected features then perform classification task using FNN methods .The proposed classification methods consists of the three major steps such as creation of the

basic structure using KNN and addition of more number of rules to each one of the selected CHF patients data records with until lesser RMSE error values is obtained. In the phase 2 number rules in the FNN is optimized using the genetic algorithm(GA) which removes the number of irrelevant neurons in the FNN to attain less RMSE error value .Mean shift algorithm is proposed to select optimum nodes in the KNN ,the architecture of the KNN which is shown in Figure.1. The details of the above mentioned three steps are detail studied as follows.

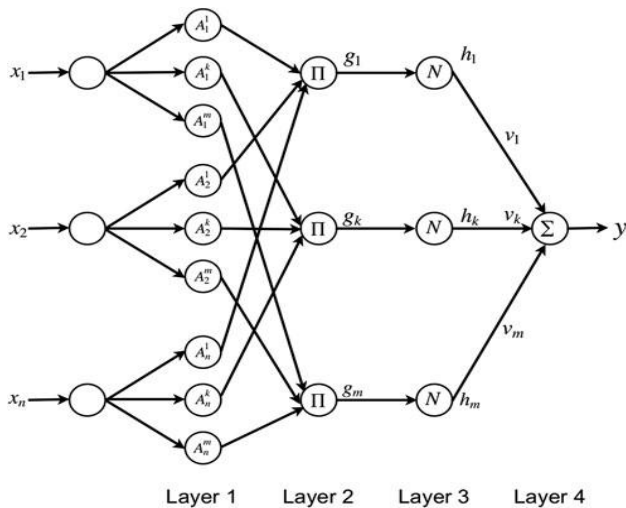


Figure 1: FNN architecture

In *K nearest neighbor (KNN)* learning or classification method the total number of rules which is created for classification, the calculation of the center values and measurement vectors designed for each rule must be known. The classification algorithm has two major phases, Rule generation and rule reduction.

In the rule generation phase, KNN follows the procedure of the error measurement algorithm to measure the performance of the FNN methods with less RMSE error. To reduce number of rules in the rule phase we perform the rule reduction based on the GA which identifies the best rule for CHF patient feature data points and the finds the best consequent parameters in every step of the process.

In rule generation the number of rules is created to each one of the CHF patient records by using IF-THEN fuzzy rules in the K nearest neighbor (KNN) algorithm. Nonlinear function is capable to approximation of local optimums through regularly distributed samples. It tries to find nearest local optimum results through K nearest neighbors. For a selected feature training points of CHF patient records is represented as (x_1, x_2, \dots, x_n) , nearest feature selected data point is denoted as A_x with K training input points with smallest Euclidean distance space to each feature dimensional space. IF then fuzzy rules for CHF feature points rules as follows,

$$M = \{x = (x_1, \dots, x_n) \in P|y(x) < y(A_x) \text{ or } y(x) > y(A_x)\} \quad (1)$$

To measure the performance of the proposed FNN classification system Gaussian function with a mean and width is choosed as fuzzy membership function ,width value of the fuzzy membership function is prespecified as $\sigma_0 = (\sigma_{01}, \dots, \sigma_{0n})$ in FNN learning algorithm for classification of HRV data. The width values of the fuzzy sets in equation is calculated from K nearest neighbors. Consequent parameters of the FNN learning should be identified using Least-Square algorithm and it is specified as $C = (c_0^1, c_1^1, \dots, c_n^1)$, that is:

$$C = (H^T H)^{-1} H^T Y \quad (2)$$

Newly generated fuzzy rules are measured based on the root mean square error(RMSE). Highest error based HRV training dataset is selected for that data new rules are created until less error values are found in the classification phase. If the results not achieved in this stage those fuzzy rules are reduced by using genetic algorithm on them. Initial population are created for each fuzzy rules and genetic operators like crossover and mutation are applied to that fuzzy membership values ,so the number of irrelevant fuzzy rules are reduced , a fitness scaling is calculated to each fuzzy membership function . it fitness value of the fuzzy membership rule is one means the specific neuron is selected or else it is not selected. Then define our fitness function as follows:

$$\text{fitness} = \begin{cases} \text{RMSE} * M & \text{if RMSE} > 2 * \tau \\ L & \text{Otherwise} \end{cases} \quad (3)$$

In the above step KNN is replaced with Mean-Shift [20] which finds the best local neurons for each width based on the estimation of the density function for each heart rate variability training data ,width value of the HRV rate is meaured based on the predetermined threshold value ,error values are using RMSE and fuzzy rules are reduced using GA. The best fuzzy ruleset are created by the changing the usual fuzzy membership function to triangular membership function is found as follows:

$$i = \arg \max \left\{ \prod_{j=1}^{D_i} t(x, a_j^i, b_j^i, c_j^i) \right\} \quad (4)$$

where $t(x; a, b, c)$ is denoted as triangular-shaped membership function, and D_i is denoted as the dimensional value of the fuzzy sets for HRV data . a_j^i, b_j^i, c_j^i are three scalars values such as left, right and center positions of j^{th} fuzzy membership values respectively and it is vectorely represented as $i, C_i = (c_1^1, c_1^2, \dots, c_1^n)^T, A_i = (a_1^1, a_1^2, \dots, a_1^n)^T$ and $B_i = (b_1^1, b_1^2, \dots, b_1^n)^T$. A training HRV with highest error rate is $x^* = (x_1^*, x_2^*)$ is defined to each fuzzy rule i with the following center vector is generated:

$$C^{M+1} = [c_1^{M+1}, \dots, c_r^{M+1}] \quad (5)$$

$$c_j^{M+1} = \begin{cases} c_j^i & j \neq r \\ x_r & j = r \end{cases} \quad (6)$$

IV. EXPERIMENTATION RESULTS

In order to measure the accuracy of the experimentation results between the KFSS-FNN and existing CART Classification methods for the HRV data, the important features of the HRV are extracted and examined by using the software PhysioNet's HRV Toolkit [22]. Select this software as the open source free available and a thoroughly authenticate using the package available from the software. The proposed KEFS and the classification method FNN is implemented with the help of the software MATLAB version R2009b. In particular, FNN was implemented by utilizing the functions and the classes of the HRV for CHF analysis of the each one of the patient's records.

Proposed system analysis the 44 nominal 24-h recordings from both category of the 12 from lower risk patients (LRP) who have already affected by mild CHF for class I and Class II and 34 from higher risk patients (HRPs) who have already affected by CHF specifically for Classes III and IV. The original data records for experimentation results are taken from Congestive Heart Failure RR Interval Database [22].

The dataset of the CHF consists of the normal RR regular intervals extracted from twenty four hours ECG record of the each one of the patients (8 men, 2 women, and 19 unknown-gender subjects) who has aged between the 34-79 years. The ECG record of each one of the patients is digitized at 128 samples for seconds. At later it consists of the long-term ECG records signals along with 11 men and 4 women, whose ages is between the 22-71 years, the ECG signals are digitalized at 250 samples per second. The extracted ECG signals were annotated automatically using the automatic methods and analyzed using the classification and feature selection methods.

In this section measure the performance accuracy of proposed KEFS-FNN classification result of CHF class with HRV related sample dataset through a normal CART. The uncommon class (mild CHF) was oversampled through generate original artificial uncommon class. To measure performance accuracy of classification result through the various parameters such as Accuracy, Precision, Sensitivity and specificity was selected. For various records the accuracy of the KEFS-FNN and CART methods results were analyzed and were tested using the parameters mentioned in Table 1.

Table 1: Performance comparison measurements

Measure	Abbreviations	Formula
Accuracy	ACC	$\frac{TP + TN}{TP + FP + FN + TN}$
Precision	PRE	$\frac{TP}{TP + FP}$
Sensitivity	SEN	$\frac{TP}{TP + FN}$
Specificity	SPE	$\frac{TN}{FP + TN}$

In the table 1, TP-denotes the true positive which is number of classes which is correctly classified as yes, TN-denotes the true negative which is number of classes which is incorrectly classified as no, FP-denotes the False Positive which is number of classes which is falsely classified as yes, FN-denotes the False Negative which is number of classes which is falsely classified as no.

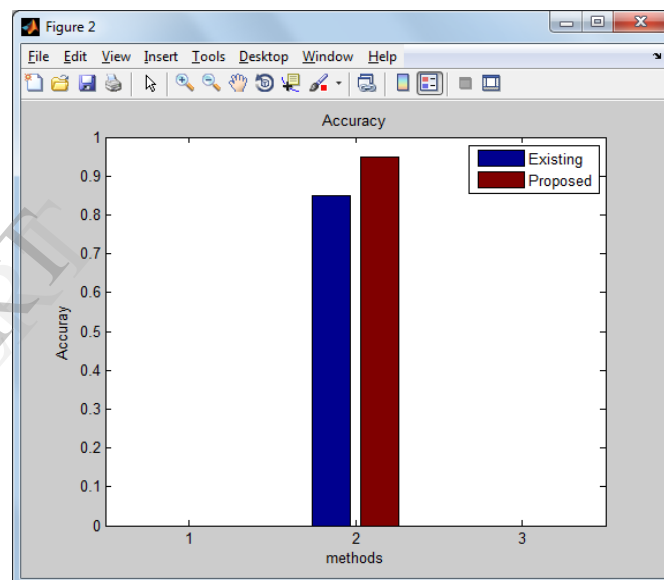


Figure 2: Accuracy comparison

Figure 2 measure the performance comparisons accuracy results of the proposed KEFS-FNN classification and the existing CART classification regression tree method. From the experimentation it shows that the performance comparison results of the proposed KEFS-FNN classification method have higher classification accuracy because of the best feature are selected from the KEFS feature selection methods, than the existing cart classification method, it highly analysis the CHF failure rate in different classes.

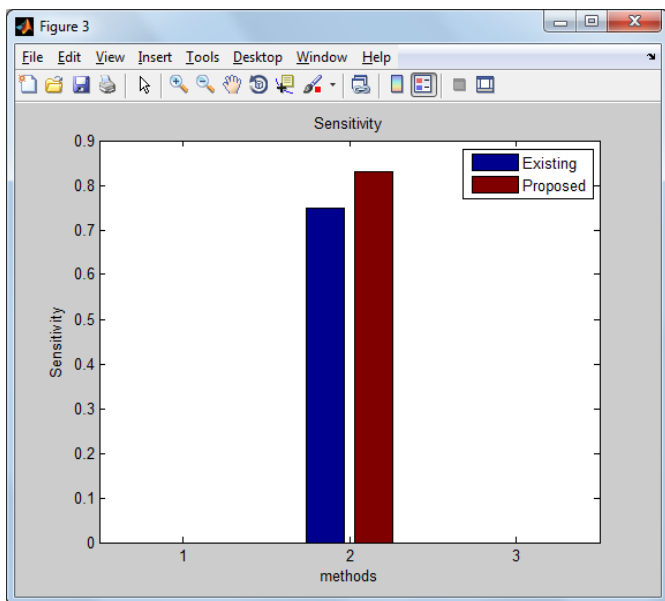


Figure 3: Sensitivity comparison

Figure 3 measure the performance sensitivity comparison results of the proposed KEFS-FNN classification method and existing CART classification method .The true positive rate of the proposed system is high since it correctly identification of the number of the classes for CHF results becomes high with efficient feature selection methods when compare to existing methods .

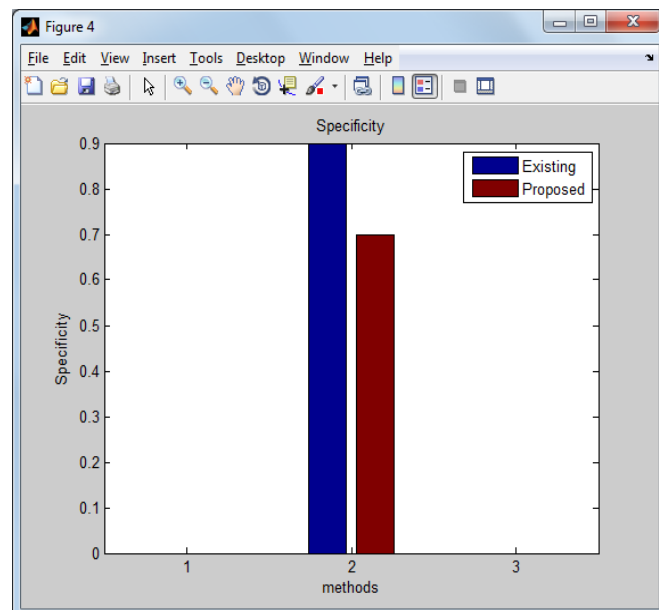


Figure 5: Specificity comparison

Figure 5 measure the performance specificity comparison results of the proposed KEFS-FNN classification method and existing CART classification method is high when compare to KEFS-FNN classification methods .

V. CONCLUSION AND FUTURE WORK

In this work, proposed a new feature selection method for congestive heart failure patients called kernel F-score feature selection (KFFS), then Fuzzy neural network (FNN) was employed to classify the CHF patient records. In relation to the method, the thorough investigation used for KEFS feature selection to improve the FNN classification performance and it can be compared to existing classification methods. Totally there 25 higher risk and 12 lower risk patients records were selected to examine the results of the existing and proposed KEFS-FNN classification method. The experimentation study with larger dataset have been handled to measure result and higher HRV patients results were found in proposed classification system with high classification ,precision ,sensitivity and less specificity results and compare with existing CART classification method. Our present KEFS-FNN classification method has the following limitations while applying this to Holter databases:

- 1) A little and unequal dataset;
- 2) The differentiation in the sample occurrence of ECG recordings;
- 3) The various feature extraction and different NN intervals were manually examined and some of the incorrect RR intervals may occur in the dataset).

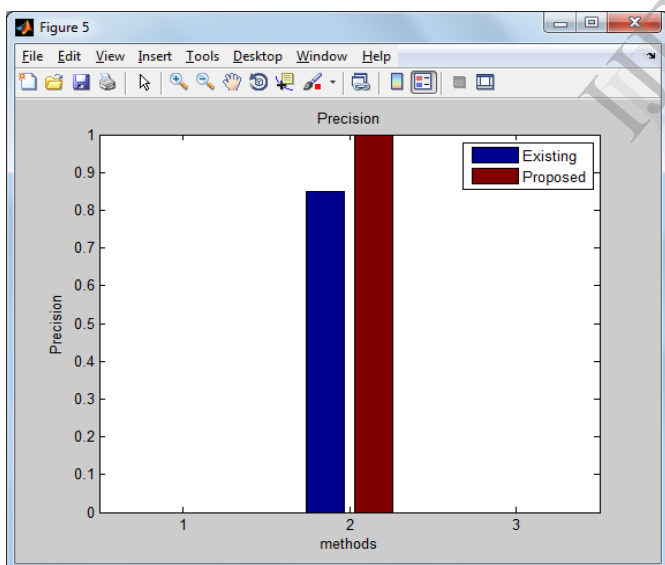


Figure 4: Precision comparison

Figure 4 measure the performance precision comparison results of the proposed KEFS-FNN system and the existing CART classification system ,it shows the number of correctly classified results of the proposed KEFS-FNN for CHF patients is high when compare to existing CART classification methods .

REFERENCES

- [1] Moore RKG, Groves D, Kearney MT, Eckberg DL, Callahan TS, et al (2004) HRV spectral power and mortality in chronic heart failure (CHF): 5 year results of the UK heart study. *Heart* 90: A6.
- [2] Guzzetti S, Magatelli E, Borroni E, Mezzetti S (2001) Heart rate variability in chronic heart failure. *Autonomic Neuroscience: Basic and Clinical* 90: 102–105.
- [3] Guzzetti S, Mezzetti S, Magatelli R, Porta A, De Angelis G, et al (2000) Linear and non-linear 24 h heart rate variability in chronic heart failure. *Autonomic neuroscience basic & clinical* 86(1–2): 114–119.
- [4] Hu J, Gao JB, Tung WW, Cao YH (2010) Multiscale Analysis of Heart Rate Variability: A Comparison of Different Complexity Measures. *Annals of biomedical engineering* 38(3): 854–864.
- [5] Liu GZ, Huang BY, Wang L (2011) A Wearable Respiratory Biofeedback System Based on Generalized Body Sensor Network. *Telemedicine and e-health* 17(5): 348–357.
- [6] Liu GZ, Guo YW, Zhu QS, Huang BY, Wang L (2011) Estimation of Respiration Rate from Three-Dimensional Acceleration Data Based on Body Sensor Network. *Telemedicine and e-health* 17(9): 705–711.
- [7] Dorgin M (1994) Nomenclature and Criteria for classification of Diseases of the Heart and Great Vessels. New York: Little Brown and Company. 12. Jessup M, Abraham WT, Casey DE, Theodore GG, Mariell J, et al (2009)
- [8] Focused update: ACCF/AHA guidelines for the classification and management of heart failure in adults a report of the American college of cardiology foundation/American heart association task force on practice guidelines. *Circulation* 119(14): 1977–2016.
- [9] Asyali MH (2003) Discrimination power of long-term heart rate variability measures. *Proc 25th Annu Int Conf IEEE Eng Med Biol Soc*, pp. 200–203.
- [10] Isler Y, Kuntalp M (2007) Combining classical HRV indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure. *Comput Biol Med* 37(10): 1502–1510.
- [11] C. Paggetti, C. Nugent, et al., Eds. Berlin/Heidelberg, Germany: Springer, 2012, pp. 278–281
- [12] Y. Guiqiu, R. Yinzi, P. Qing, N. Gangmin, G. Shijin, C. Guolong, Z. Zhaocai, L. Li, and Y. Jing, “A heart failure diagnosis model based on support vector machine,” in *Proc. 3rd Int. Conf. Biomed. Eng. Inf. (BMEI 2010)*, pp. 1105–1108.
- [13] Pagani M, Lombardi F, Guzzetti S et al. Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog. *Circ Res* 1986; 59: 178–93.
- [14] Malik M, Xia R, Odemuyiwa O, Staunton A, Poloniecki J, Camm AJ. Influence of the recognition artefact in the automatic analysis of long-term electrocardiograms on time domain measurement of heart rate variability. *Med Biol Eng Comput* 1993; 31: 539–44.
- [15] Malliani A, Pagani M, Lombardi F, Cerutti S. Cardiovascular neural regulation explored in the frequency domain. *Circulation* 1991; 84: 1482–92.
- [16] Courmel Ph, Hermida JS, Wennerblöm B, Leenhardt A, Maison-Blanche P, Cauchemez B. Heart rate variability in myocardial hypertrophy and heart failure, and the effects of beta-blocking therapy. A non-spectral analysis of heart rate oscillations. *Eur Heart J* 1991; 12: 412–22.
- [17] Grossman P, Van Beek J, Wientjes C. A comparison of three quantification methods for estimation of respiratory sinus arrhythmia. *Psychophysiology* 1990; 27: 702–14.
- [18] Bigger JT, Fleiss JL, Rolnitzsky LM, Steinman RC. Stability over time of heart period variability in patients with previous myocardial infarction and ventricular arrhythmias. *Am J Cardiol* 1992; 69: 718–23.
- [19] Chen, Yi-Wei., Lin, Chih-Jen., (2003). Combining SVMs with various feature selection strategies, *NIPS 2003 feature selection challenge*, 1-10
- [20] Fukunaga K, Hostetler LD (2002) The estimation of the gradient of a density function, with applications in pattern recognition. *IEEE Trans Pattern Anal Mach Intell* 24:603-619
- [21] Wang JS, Lee CSG (2001) Efficient neuro-fuzzy control systems for autonomous underwater vehicle control. In: *IEEE international conference on robotics and automation*, pp 2986-2991.
- [22] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, “PhysioBank, PhysioToolkit, and PhysioNet : Components of a new research resource for complex physiologic signals,” *Circulation*, vol. 101, no. 23, pp. e215-e220, Jun. 13, 2000