© 2015 IJEDR | Volume 3, Issue 2 | ISSN: 2321-9939

Support Vector Machine Technique for Risk Assessment in Patients Suffering From Congestive Heart Failure via Heart Rate Variability Variant

K.Abinaya[1], Dr.M.Vijayakumar[2]
PGScholar[1], Professor and Head[2], Department of CSE,
Nandha college of Technology, Erode.

Abstract - **The congestive heart failure is a major cause of concern among all types of cardiovascular problems and is attributed to imbalance in sympathetic and parasympathetic nervous systems. In an effort to improve the reliability of the detection of congestive heart failure, a method for utilizing phase shifts of cardiac and thoracic acoustics coupled with ECG signals is proposed. In this work, an automatic classifier for risk assessment in patients suffering from congestive heart failure is developed which yields consistency and consensus rate to identify risk. The proposed classifier separates lower risk from higher risk ones, using standard long-term heart rate variability (HRV) measures. Patients are labelled as low risk and high risk based on the depressed HRV. An algorithm for detection of congestive heart failure condition using the variability of instantaneous frequency of intrinsic mode functions obtained using Support vector machine will be determined with optimization of CART (Classification and Regression Tree) System in order to proceed the satisfactory signal quality . The performance of the proposed system will yield better results in terms of sensitivity and specificity rate in identifying the high risk patients.**

I. INTRODUCTION

Heart disease is the leading causes of deaths worldwide, bringing a great burden on individual, national and global economics. The development of a variety of CVDs is associated with autonomic dysfunction. Specifically, declined vagal activity is a powerful indicator of overall mortality, while sympathetic over activity is strongly related to diseases such as hypertension, obesity, and heart failure. Pulse Transit Time (PTT) is a cardiovascular parameter of emerging interest due to its potential to estimate Blood Pressure (BP) continuously and without a cuff. Heart Rate Variability (HRV) is the variation over time of the period between consecutive heartbeats (RR intervals) and is usually extracted from ElectroCardioGraphic signal (ECG) recorded through a non-invasive technique.

II. PROPOSED CONTRIBUTION

Disease Prediction is active medical research based on the multi data types and multi variants. The Classification technique is utilized for the risk assessment in the patient diseases. Risk assessment in patients suffering from congestive heart failure (CHF) is determined through help of supervised and semi supervised technique in literatures. The proposed classifier separates lower risk from higher risk ones, using standard long-term heart rate variability (HRV) measures. Patients are labeled as lower or higher risk according to the New York Heart Association classification (NYHA). A retrospective analysis on two public Holter databases was performed, analyzing the data of few patients suffering from mild CHF (NYHA I and II), labeled as lower risk, and high suffering from severe CHF (NYHA III and IV), labelled as higher risk. Only patients with a fraction of total heartbeats intervals (RR) classified as normal-to-normal (NN) intervals (NN/RR) higher than 80% were selected as eligible in order to have a satisfactory signal quality.

III. METHODS

1. Data processing the ECG data Records
2. Estimating the Heart rate variability for risk assessment
3. Estimation of the Heart failure based on Supervised data through New-York hear Association
4. Developing Classification and Regression tree for Risk assessment using HRV and NYHA
5. Developing a Optimized SVM utilizing HRV and NYHA
6. Performance Evaluation

IV. METHODS DESCRIPTION

1. *Data Processing The Ecg Data Records*

Data processing of the ECG data records has been carried out using ECG recordings were digitized at 128 samples per second, and the beat annotations were obtained by automated analysis, All the computed basic time- and frequency-domain HRV measures were widely used in literature as reported in International Guidelines with manual review and correction. The processing for removing unwanted data in data processing is as following process,

– Data cleaning: fill in missing values, smooth noisy data, identify or remove outliers, and resolve inconsistencies.
– Data integration: using multiple databases, data cubes, or files.

- Data transformation: normalization and aggregation.
- Data reduction: reducing the volume but producing the same or similar analytical results.
- Data discretization: part of data reduction, replacing numerical attributes with nominal ones.

2. *Estimating The Heart Rate Variability For Risk Assessment*

Heart rate variability (HRV) is the variation over time of the period between consecutive heartbeats (RR intervals) and is usually extracted from electrocardiographic signal (ECG) recorded through a non-invasive technique. HRV is commonly used to assess the influence of the autonomic nervous system (ANS) on the heart. HRV has been widely studied in patients suffering from CHF (Congestive heart failure). The frequency-domain HRV measures rely on the estimation of power spectral density (PSD) computed in this work by Lomb–Scargle (LS) periodogram

$$\text{Heart rate Variability} = \sum_{k=0}^{n} \binom{n}{k} x^{RR} a^{n-k}$$

Fast implementation and further detail of LS period gram can be found elsewhere. After PSD estimation, six standard frequency domain HRV measures were calculated: total spectral power of all NN intervals up to 0.4 Hz (TOTPWR); between 0 and 0.003 Hz (ULF); between 0.003 and 0.04 Hz (VLF); between 0.04 and 0.15 Hz (LF); between 0.15 and 0.4 Hz (HF); and ratio of low-to high-frequency power (LF/HF).

3. *Estimation Of The Heart Failure Based On Supervised Data Through New-York Heart Association*

Difficulties include the acquisition, collection and organization of the data that will be used for training the system. This becomes a major problem especially when the system requires large data sets over long periods of time, which in most cases are not available due to the lack of an efficient recording system. This is usual in most medical tasks as a different degree of significance may be required for the system's performance on each class. The cardiac events that occur from the beginning of one heartbeat to the beginning of the next are called the *cardiac cycle which* consists of a period of relaxation called *diastole,* during which the heart fills with blood, followed by a period of contraction called *systole.* In R-R interval features extraction, we use empirical mode decomposition (EMD) to decompose each subject's R-R interval signal into several intrinsic mode function (IMF), and use singular value decomposition (SVD) to extract the ranked singular values for each subject's IMF. The ranked singular values obtained are input to the support vector machine (SVM) for classification of physiological state (health or CHF).

4. *Developing Classification And Regression Tree For Risk Assessment Using Hrv And Nyha*

CART implements a stepwise feature selection for pulse value prediction in the ECG data, it may happen that one feature is excluded because other variables masked its effect. This could be particularly critical in small and unbalanced dataset. In order to deal with masking and to be sure that the tree included the best subset of features, the misclassification error estimated by tenfold cross-validation approach, as described in the following paragraph. Even if the cross-validated estimate could provide limited information in small and imbalanced dataset, the feature selection improves the classification since it enables to consider combination.

The CART algorithm consists of two stages: tree growing and tree pruning. In the former stage, the tree grows by selecting among all the possible splits, which generate the "purer" child nodes where the purest node is the one containing elements of only one class. The outcome of this step is further referred to as the large tree (LT). Among different functions that have been proposed for the measure of the impurity of each node
"*t*" , we adopted the Gini index criterion which for binary classification can be computed. LT is pruned according to a minimal cost-complexity function, which relies on the tree size and the misclassification error. The misclassification error is estimated by the inner tenfold cross-validation of the CART. The dataset is randomly divided into ten subsets. One of the subsets is used as independent testing dataset while the other nine subsets are used as training dataset. The tree-growing and pruning procedure is repeated several times, each time with one of the ten different subsets used as testing set. The misclassification error is calculated as the percentage of misclassified cases averaged over all the ten subsets.This procedure is repeated pruning the tree, and for each subtree, the cost complexity function is computed as a linear combination of the number of nodes and of the cross-validated estimated of the misclassification error. The outcome of the stage is referred further to as the best subtree (BST), which is the subtree achieving the lowest value of the cost-complexity function. Further details about minimal cost-complexity pruning can be found in Breiman. The most common measures for binary classification were estimated by tenfold cross-validation, using the formulas reported in particular *F*1 is one of the most suitable metric for imbalanced class problem, as it is the harmonic mean of precision and sensitivity.

5. *Developing A Optimized Svm Utilizing Hrv And Nyha*

Support vector machine constructs a Hyperplane or set of hyper planes in a high dimensional space or infinite-dimensional space for the ECG training data, which can be used for classification, regression, or other tasks like estimating the high heart rate. Intuitively, a good separation is achieved by the hyper plane that has the largest distance to the nearest training data point of any class (so-called functional margin) for discriminating the high Risk(HR) and Low Risk(LR), since in general the larger the margin the lower the generalization error of the classifier.

Whereas the original problem may be stated in a finite dimensional space, and often happens that the sets to discriminate are not linear separability in that dataset. For this reason, it was proposed that the original finite-dimensional space be mapped into a much higher-dimensional space, making the separation easier in that space. To keep with computational load reasonable, the mappings used by SVM schemes are designed to ensure that dot product may be computed easily in terms of the variables in the original space, by defining in terms of a positive definite kernel selected to suit the problem. The hyper planes in the higher-dimensional space are defined as the set of points whose dot product with a vector in that space is constant to enable the disease prediction in the optimized way

$$F(x)= 1+\frac{nx}{1!}+\frac{n(n-1)x^2}{2!}+\cdots$$

Where x is the heart rate value in ECG

SVM classifier predicts the unknown labels in a efficient manner using the training knowledge.

6. Performance Evaluation

Performance of the CART is compared with optimized SVM for classification of the High Risk patients with low risk patients. Optimized SVM yields better results in terms of precision and recall; it also proves that Optimized SVM determines the Heart disease through the ECG monitored data.

Measure	Units	LRPs			HRPs			Wilcoxon test
		25th	50th	75th	25th	50th	75th	p-value
AVNN	ms	631	674	785	620	697	778	0.803
SDNN	ms	184	399	525	67	90	180	**0.003**
SDANN	ms	63	87	206	47	60	76	**0.026**
SDNNIDX	ms	79	177	295	25	50	87	**0.005**
RMSDD	ms	242	515	675	18	110	234	**0.003**
pNN10		26.78	39.66	55.33	27.50	33.48	43.76	0.385
pNN50		0.52	3.02	7.25	0.40	1.70	2.94	0.257
TOTPWR	ms^2	20145	132516	184566	3660	6718	27751	**0.005**
ULF	ms^2	3991	8914	28656	1852	3575	6430	**0.013**
VLF	ms^2	1874	12553	16577	202	622	2604	**0.005**
LF	ms^2	4328	28813	45652	133	595	5234	**0.003**
HF	ms^2	9013	60783	102470	101	1691	11678	**0.003**
LF/HF		0.44	0.45	0.47	0.44	0.46	0.64	0.222

25th: 1st quartile / 25th percentile; 50th: median / 2nd quartile / 50th percentile; 75th: 3rd quartile / 75th percentile.

Table 1 comparison of HRV between two groups

V. CONCLUSION

The long-term HRV measures enable higher risk patients to be distinguished from lower risk ones using support vector machine technique with decision trees. Finally, the results will be consistent with the consensus that depressed HRV values are associated with higher cardiovascular risk.

VI. REFERENCES

[1] ElioConte (2014)'A New Method for Analysis of Heart Rate Variability, Asymmetry and BRS'Nova Science Publishers, vol.8, no 1.

[2] Faezehmarzbanrad, Yoshitakakimura, Kiyoefunamoto, Rikasugibayashi, Miyukiendo, Takuyaito, Marimuthupalaniswamiand Ahsan H.Khandoker (2014)'Automated Estimation of Fetal Cardiac Timing Events from Doppler Ultrasound Signal using Hybrid Models' IEEE Journal of Biomedical and Health Informatics, vol. 18, no. 4.

[3] François Tavard, Antoine Simon, Christophe Leclercq, Erwan Donal, Alfredo I. Hernández, and Mireille Garreau(2014),'Multimodal Registration and Data Fusion for CardiacResynchronization Therapy Optimization', IEEE Transactions on Medical Imaging, vol. 33, no. 6.

[4] George d. o'clock, yong wan lee, Jongwong leeand warren j. Warwick (2012),'High-Frequency and Low-Frequency Chest Compression: Effects on Lung Water Secretion, Mucus Transport, Heart Rate and Blood Pressure using a Trapezoidal Source Pressure Waveform'IEEE Transactions on Biomedical Engineering, vol. 59, no. 1

[5] Leandro pecchia, Melillo and Marcello Bracale(2011)'Remote Health Monitoring of Heart Failure with Data Mining via CART Method on HRV Measures'.

[6] Li-napu, zezhao, and yuan-ting zhang, (2012), 'Investigation On Cardiovascular Risk Prediction using Genetic Information 'IEEE Transactions on Information Technology in Biomedicine, vol. 16, no. 5.

[7] Linda Pattini, Roberto Sassi, and Sergio Cerutti (2014) 'Dissecting HeartFailure through the Multiscale Approach of Systems Medicine', IEEE Transactions on Biomedical Engineering, vol. 61, no. 5

[8] Murukesan L, Murugappan M.,Iqbal, M. andDr.Saravanan K. (2014)'Machine Learning Approach for Sudden Cardiac Arrest PredictionBased onOptimal Heart Rate Variability Features' Journal of Medical Imaging and Health Informatics ,vol. 4,1–12.

[9] Musialik-Lydka A., Sredniawa, and Pasyk S., Kardiol. Pol, (2003), 'Heart Rate Variabilityin Heart Failure',vol. 58, no. 1, pp. 10–16.

[10] Pablo Laguna Juan Bolea (2013)'Prediction of sudden cardiac Death in chronic Heart Failure Patient by Analysis of Restitution Dispersion'Computing in Cardiology vol.40 no 1.4.

[11] Paolo Melillo, Nicola De Luca, Marcello Bracale, and Leandro Pecchia.(2013)'Classification Tree for Risk Assessment in Patients Suffering From Congestive Heart Failure via Long-Term Heart Rate Variability' IEEE Journal of Biomedical and Health Informatics, vol. 17, no. 3.

[12] Pecchia L., Melillo P., Sansone M., and Bracale M., (2011)'Discrimination Power of Short-Term Heart Rate Variability Measures for CHF Assessment', IEEE Trans Inf Technol Biomed. 40-6.

[13] Qing Liu, Bryan P. Yan, Cheuk-Man Yu, Yuan-Ting Zhang, and Carmen C. Y. Poon(2014),' Attenuation of Systolic Blood Pressure and PulseTransit Time Hysteresis During Exercise and Recovery in cardiovascular Patients'IEEE Transactions on Biomedical Engineering, vol. 61, no. 2.

[14] Reps, Jonathan M. Garibaldi M., UweAickelin, Daniele Soria, Jenna (2014),'Novel Semisupervised Algorithm forRare PrescriptionSide Effect Discovery'IEEE Journal of Biomedical and Health Informatics.vol. 18, no. 2

[15] Shantanu sarkar and Jodi Koehler(2013)'A Dynamic risk score to identify increased risk for Heart Failure Decomposition',IEEE Transactions on BioMedical Engineering vol.60,no.1

[16] Sudhakar K and .Manimekalai M.(2014)'Study of Heart Disease Prediction using Data Mining'International Journal on Advanced Research in Computer Science and Software Engineering,vol.4,no.1.

[17] Susana Merino-Caviedes, Lucilio Cordero-Grande, Ana Revilla-Orodea, Teresa Sevilla-Ruiz, M. Teresa Pérez, Marcos Mart´ın-Fernández, and Carlos3-D Framework to Determine Myocardial Thickness and Transmurality in Late Enhancement Images' IEEE Transactions on Medical Imaging, vol. 33, no. 1. Alberola-López(2014), 'Multi-Stencil Streamline Fast Marching.

[18] Sumit Bhatia, Praveen Prakash, and G.N. Pillai(2008) 'SVM Based Decision Support System for heart disease classification with integer-coded Genetic Algorithm to Select Critical Features', Proceedings of the World Congress on Engineering and Computer Science 2008 WCECS 2008, October 22 - 24, 2008, San Francisco, USA

[19] Valentina Bono, Evangelos B. Mazomenos, Taihai Chen, James A. Rosengarten, Amit Acharyya, Koushik Maharatna, , John M. Morgan, and Nick Curzen(2014),' Development of an Automated Updated Selvester QRS.