UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Application of Heart Rate Variability (HRV) in Congestive Heart Failure (CHF) Detection and Quantification

by Wenhui Chen

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

under the supervision of Steven Su

University of Technology Sydney
Faculty of Engineering and Information Technology

February 2020

Declaration

I, Wenhui Chen declare that this thesis, is submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the Faculty of Engineering and IT at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise reference or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Production Note: Signature removed prior to publication.

Date: 20/02/2020

Acknowledgements

I would like to express my great appreciation to my supervisor A/Prof. Steven W. Su for his continual support, guidance, help and encouragement during my PhD study. A/Prof. Su has brought me into a higher research level for the topic of congestive heart failure classification, and provided an excellent research platform and brilliant insights into my research works. It is my honor to have a supervisor who mentors and inspires me to create and achieve higher targets. His conscientious and meticulous attitude on research has had a significant influence on my work.

I would like to demonstrate my sincere gratitude to my co-supervisor Dr Steve S.H. Ling, and external supervisor Prof. Hung T. Nguyen (Swinburne University of Technology, Australia), for their solid support in congestive heart failure classification in terms of artificial intelligent algorithm field, which improve my understanding of my research works.

I am grateful to the other co-supervisors: Prof. Qing Jiang (Sun Yet-sen University, China), Dr. Guanzheng Liu (Sun Yet-sen University, China), and Prof. Rong Song (Sun Yet-sen University, China), for their kind help in financial support and mental encouragement. Prof. Yifan Chen (University of Electronic Science and Technology, China), and Prof. Chin M. Chow (University of Sydney, Australia), for their professional and precious comments and suggestions on my research.

Besides, the Australia-China Joint Research Institute for Health Technology and Innovation (2014), which established by Sun Yat-sen University and University of Technology Sydney, give me the treasure opportunity and financial support to study in Australia.

Special thanks to my lovely colleagues in A/Prof. Steven W. Su's research group, in particular, Dr. Tao Zhang, Dr. Lin Ye, Dr. Wentian Zhang, Yao Huang, Taoping Liu, Kairui Guo, and Hairong Yu, for their selfless help and mental support. During my PhD study, it is a precious life memory working together with them.

I am also lucky to meet many lovely friends: Dr. Ye Shi, Dr. Zhichao Sheng, Haimin Zhang, Zhiyuan Shi, Dr. Daniel Roxby, for their warm company.

I also wish to express my appreciation to the staff members in the Centre for Health Technologies, School of biomedical engineering, University of Technology Sydney, especially to Prof. Joanne Tipper and Dr Gyorgy Hutvagner. It's a special and memorizable experience to work with these adorable and inclusive people.

My deepest gratitude goes to my family for their continuous understand and support.

Thanks to all your encouragement!

Abstract

Congestive heart failure (CHF) is one of the most important cardiovascular syndrome and end stage of all kinds of heart diseases. Due to high mortality and morbidity, risk assessment of patient suffering CHF has attracted many attentions. The existing research about CHF assessment mainly focused on disease detection using ECG signals, especially with 24-h/5-min heart rate variability (HRV), both in mechanism analysis and classification. A significant relation between different ECG components and disease condition had been proved. Furthermore, a good classification performance had been achieved in CHF detection using HRV. However, there is not much attention focusing on multilevel assessment of CHF, i.e. disease detection and quantification. Also, sleep apnea and CHF are two of the most common diseases and interrelated, which are hard to differentiate from the syndrome. But there is no research in differentiating the two diseases. Besides, RR intervals are sensitive to physiological activity and rhythm, increasing unstable analysis and results. Thus, this research will devote to ECG analysis and multi-level risk assessment model construction to achieve robust, convenient and accurate CHF detection and quantification, as well as underlying mechanism analysis.

In this research, 116 RR interval data were downloaded from MIT/BIH database, including 72 normal persons and 44 CHF patients. First, we analyzed 24-h RR intervals and proposed a series of novel indices of HRV - dynamic indices - to better describing difference among different risk levels of CHF patients in a day. Then we applied the decision-tree based support vector machine and backward elimination algorithm to construct a 4-level risk assessment model for CHF assessment. Results showed a total accuracy of 96.61% with only two misclassified samples. This demonstrated the stratifying risk assessment model of CHF in our research has the potential to be a

reliable and objective prognostic marker for the routine clinical application (especially daily health nursing) in the future.

Then, we applied 5-min RR intervals into a unsupervised sparse-auto-encoder based deep learning algorithm to explore CHF detection performance under short cycle in big data condition. A total of 30592 5-min RR intervals was obtained from 72 healthy persons and 44 CHF patients. This algorithm first extracts unsupervised features using a sparse auto-encoder neural network from the raw RR intervals. Then a two layers' neural network model was constructed. Various hidden node settings were compared to optimize classification performance. Results showed an accuracy of 72.44% in CHF detection under the constructed 2-layer neural network, and optimal nodes setting is (200, 50). This result indicated that short-term RR intervals have the potential for CHF detection but is sensitive to body condition.

Next, we analyzed different time scale of HRV from 5-min to 24-h to explore optimal time scale for CHF detection and quantification. Statistical analysis between 3-level risk-groups was applied under 10 classical HRV measures to evaluate differentiating power in risk assessment of CHF for the optimal time scale. With the optimal time length, we used classical classifiers with these classical HRV measures in 3-level risk level classification of CHF, to prove the usage in risk assessment. The statistical analysis of HRV measures showed that 2-h RR interval data has the optimal performance in differentiating three risk levels. The classification performance showed that the optimal timescale of 2-hour for CHF assessment, yielded an comparable accuracy of 87.88% and 81.13% for classifying the healthy from patients and lower risk from higher risk patients, respectively. This research demonstrated that the optimal measurement timescale (2h) has potential in providing convenient and reliable CHF assessment in the future application, especially in-home monitoring.

Finally, we analyzed whole night Polysomnography (PSG) data to differentiating congestive heart failure and sleep apnoea patients, which are two of the most tightly interrelated and common diseases in cardiopulmonary system. Twenty whole night PSG data from the Sleep Heart Health Study database were included in this study. The Pan-Tompkins algorithm was applied to the electrocardiograph signal to detect R peaks of the QRS complex. The whole night R peaks data were then manually checked and

segmented into 1895 5-minute epochs to calculate three frequency domain and three nonlinear heart rate variability measures. All these measures were analyzed for their statistical differences between groups (sleep apnea with and without CHF). Finally, a binary support vector machine classifier and extreme search method were performed to construct the model. Results showed that an accuracy of 81.68% was achieved in distinguishing sleep apnea patients with and without CHF. This indicated that HRV measures from PSG had the potential to help distinguish sleep apnea patients with and without CHF reported.

In conclusion, with all these analyses of ECG signals in congestive heart failure classification, we proved the potential of HRV from ECG in robust, accurate and convenient congestive heart failure assessment using intelligent methods, which can be applicated in home-monitoring with wearable ECG measurement equipment.

Publications

The contents of this thesis are based on the following papers that have been published, accepted, or submitted to peer-reviewed journals and conferences.

Journal Papers:

1. Wenhui Chen, Lianrong Zheng, Kunyang Li, Qian Wang, Guanzheng Liu, and Qing Jiang. "A novel and effective method for congestive heart failure detection and quantification using dynamic heart rate variability measurement." PloS one, 11(11), e0165304, 2016.

2. Wenhui Chen, Guanzheng Liu, Steven Su, Qing Jiang, and Hung Nguyen. "A CHF detection method based on deep learning with RR intervals." 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2017.

3. Wenhui Chen, Steven Su, Hung Nguyen, Qing Jiang, and Guanzheng Liu. "Dynamic heart rate variability in autonomic unbalance for congestive heart failure stratification." (Submitted for publication).

4. Wenhui Chen, Qing Jiang, Steven Su, and Hung Nguyen. "Automatic Risk Assessment of Congestive Heart Failure Using ECG at Optimal Time scale." (Submitted for publication).

Conference Papers:

1. Wenhui Chen, Guanzheng Liu, Steven Su, Qing Jiang, and Hung Nguyen. "A CHF detection method based on deep learning with RR intervals." In the 39th

Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC' 2019), pp. 3369-3372, IEEE, 2017.

2. Wenhui Chen, Prof. Yifan Chen, MD B. Uddin, Prof. Hung Nguyen, Chin M. Chow, and Steven W. Su. "An Automatic Method to Differentiate Sleep Apnea Patients with and without Congestive Heart Failure using Polysomnography Records." In The 13th IEEE-EMBS International Summer School and Symposium on Medical Devices and Biosensors (MDBS' 2019), IEEE, 2019.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

This chapter starts with the problem statement, motivation and aims of this thesis and ends with an overview of the dissertation and outlines its contribution to the field.

1.1 Problem Statement

Congestive heart failure (CHF) is a common chronic cardiovascular syndrome along with autonomic nervous system (ANS) abnormality of the heart [1], also called heart failure (HF). CHF is a condition which occurs when the heart is unable to pump enough blood throughout the human body [2]. This disease is characterised by an atrial and ventricular dysfunction [3] and is the end stage of all kinds of cardiac disease [4]. The common symptoms include fatigue, shortness of breath, swelling etc (as shown in Fig.1.1) [5]. All these symptoms give patients pain and decrease their chance of survival. Thus, early detection and well-timed risk assessment of CHF is paramount to help patients to not only lower their pain and suffering but also reduce their financial pressure.

Usually, CHF patients experience no obvious symptoms during its early stages [6]. Furthermore, physicians cannot provide convenient and suitable medical care based on the patient's physical condition and the prognosis [7]. Poor prognosis results in 30 - 40% of diagnosed patients dying in a year [8]. The morbidity of CHF is about 0.9% [9]. According to a report by the Australian Institute of Health and Welfare, 13% of people

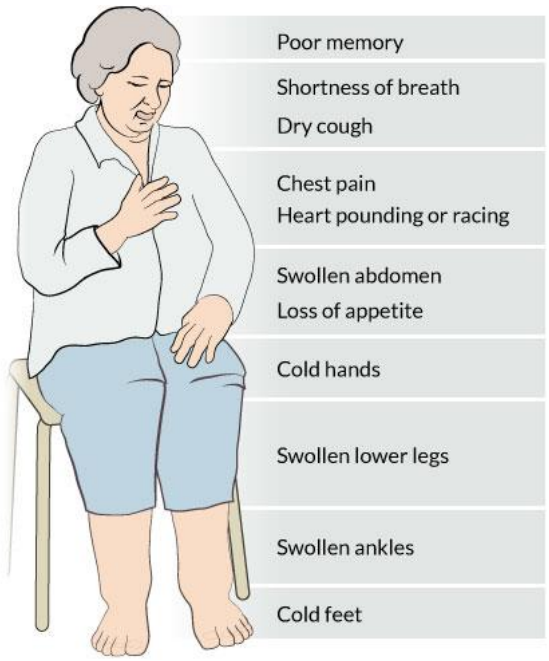

Fig. 1.1 Symptoms of congestive heart failure. (http://pie.uhnresearch.ca/heartfailure/heart-failure/symptoms-heart-failure/)

aged over 65 years old suffer from this disease and this prevalence is increasing due to our aging population (https://www.aihw.gov.au). Furthermore, there are considerable public health costs. it is estimated that it costs the public purse $140 million per year for hospitalisation of heart failure patients and $135 million per year for nursing home care. Even worse, there are nearly 23 million individuals affected by heart failure worldwide [9]. Thus, risk assessment of CHF is essential for saving lives and money.

As one of the most common diseases in the cardiopulmonary system, CHF also has a close connection to sleep apnea. Over the past decades, numerous studies have explored the cardiovascular variation of cardiopulmonary patients, especially in relation to congestive heart failure (CHF) [10–12] and sleep apnea [13–15], which are two of the most common diseases of the cardiopulmonary system [16]. For patients with heart failure, sleep apnea is an under-diagnosed disease condition [17]. For patients with obstructive/central sleep apnea, worsening of their breath disorder results in a failure of the heart to pump blood [18]. While patients with CHF and those with sleep breathing disorders share common symptoms of obstructive and central apnea. Since 1998, many research studies have found a high prevalence (over 50%) of sleep apnea in

the heart failure population [19]. Even worse, both diseases are associated with a high risk of death and financial loss [20, 21].

Sleep apnea includes two types: obstructive sleep apnea (OSA) and central sleep apnea (CSA) [22]. Obstructive sleep apnea (OSA) often causes heart damage [23], while OSA, central apnea and Cheyne-Stokes breathing are consequences of CHF [24]. Many studies have reported clinical and statistical differences between apnea patients and CHF [25–27] and have proved that both diseases have a negative effect on each other, but currently, no autonomic methods are available to distinguish them. Thus, a great deal of work is needed to explore in terms of the classification of CHF and sleep apnea.

The severity of CHF has a well-known measurement, namely, the symptomatic classification scale of the New York Heart Association (NYHA) [28], which has proved to be a very useful tool for the risk assessment of CHF patients [29], as shown in Table 1.1 [30]:

Table 1.1 NYHA functional classification

NYHA Class	**Patients with Cardiac Disease (Description of HF Related Symptoms)**
Class I (Mild)	Patients have no limitation in physical activities even though they have cardiac disease. Ordinary physical activity does not cause symptoms, including undue fatigue, palpitation (rapid or pounding heartbeat), dyspnoea (shortness of breath), or angina pain (chest pain) etc.
Class II (Mild)	Patients with cardiac disease experience slight limitations in terms of physical activity. They are comfortable at rest. Ordinary physical activity results in symptoms, including fatigue, palpitation, dyspnoea, or angina pain.
Class III (Moderate)	Patients with cardiac disease experience marked limitations in terms of physical activity. They are comfortable at rest. Less than ordinary activity causes symptoms including fatigue, palpitation, dyspnoea, or angina pain.
Class IV (Severe)	Patients with cardiac disease are unable to carry on any physical activity without discomfort. Symptoms of heart failure may present even at rest. If any physical activity is undertaken, discomfort increases.

Congestive heart failure (CHF) has attracted has attracted considerable research attention for years. CHF is linked to several complex problems, including end-stage heart suffering [1], that cause chambers of the heart to fail. When symptoms suggest CHF, the gold standard clinical diagnosis is done by a cardiologist in hospital. The diagnosis procedure usually include three steps: physical examination, patient medical history, and various diagnostic tests, some invasive. The final test step is to detect

abnormal function of the left ventricle and/or heart valves [31], as shown in Fig.1.2, which is direct proof of heart failure.

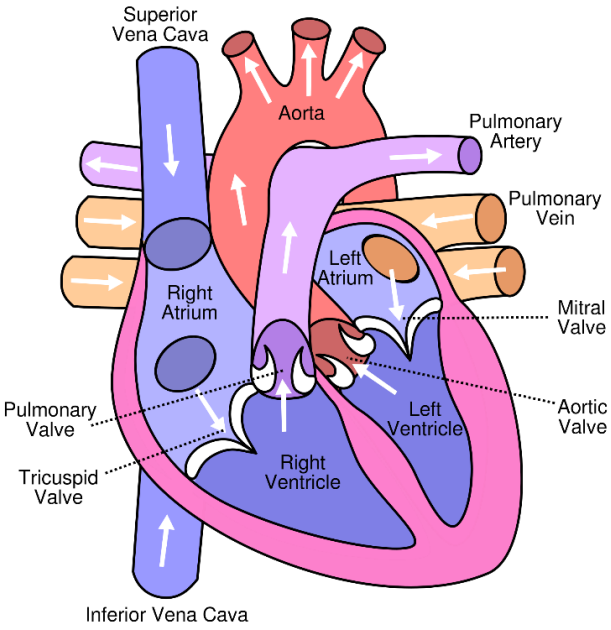

Fig. 1.2 Heart Structure.

The three steps below represent the usual flow during hospital examination:

1. **Physical examination**: The physician looks for an underlying cause and assesses subject's heart function during the physical examination using stethoscope. Here, the stethoscope is applied on the surface of the chest, to detect abnormal heart sounds (murmurs) and fluid accumulation in the lungs. These sounds may indicate a leaky or narrowed (stenotic) valve. The physician also looks for potential symptoms on the subject's body, including enlarged (distended) veins in the neck and swelling (edema) in the legs (particularly the ankles and feet) and/or the abdomen.

2. **Patient medical history**: A patient's medical history includes gathering medical and family information that may relate to heart failure. Usually, information like alcohol and drug use; history of hypertension and diabetes, including treatment; prior chest pains or heart attack; recent viral illness and recent pregnancy will be collected.

3. **Diagnostic tests**: Diagnostic tests are applied to evaluate heart function (e.g., assess ejection fraction), and to detect coronary artery disease, heart attack, and valve dysfunction. Commonly used methods include an electrocardiogram (ECG or EKG), B-type natriuretic peptide (BNP) blood test [32], an echocardiogram [33], chest X-ray [34], cardiac catheterisation, Doppler ultrasound [35], Cardiac magnetic resonance imaging (MRI) [36] and so on.

Normally, for each heart beat, approximately 60% of the blood in the left ventricle is ejected (contracts) [37], which is called left ejection fraction. The left ejection fraction is commonly used to evaluate the severity of heart function depression. According to current research, patients with ejection fractions of approximately 40 - 45% are mildly depressed in terms of heart function [38]; patients with ejection fractions of about 30 - 40% are moderately depressed in terms of heart function [39]; and patients with ejection fractions in the 10 - 25 percent range are severely depressed in terms of heart function [40].

Among the three steps for diagnosis in hospital, the first two steps screen for potential disease while the last step addresses the exact reason for heart function impairment.

According to hospital examination description, several problems exist. First, hospital resources are limited. During hospital examination, no single test can diagnose heart failure. In addition, the diagnostic process described above often results in time delays and high costs. Thus, in-hospital examination is time-consuming and carries a heavy financial burden which has implications in terms of patient health [41]. Second, the diagnosis by a cardiologist requires rich experience and a rigorous attitude. This means possible manual error during diagnosis. Finally, some examinations are invasive and have the potential to hurt patients. Thus, researchers have been focusing on a cheaper, faster, noninvasive and accurate risk assessment method for CHF diagnosis application.

In the third step of hospital examination, an electrocardiogram (ECG or EKG) is the most common signal measured. ECG is a non-invasive test used to measure electrical activity of the heart from the surface. The test shows the patient's heart beating speed and its rhythm (steady or irregular) [42]. Multiple electrical sensors,

called leads, are attached to predetermined positions on the arms, legs, and chest to record electrical activity. These leads collect electrical signals and help assess heart function (https://www.nhlbi.nih.gov/health/health-topics/topics/hf/diagnosis). Due to ECG's advantages, a number of research studies have been based on surface ECG analysis using long-term (24-h)/short-term (5-min) ECG data, and this is described in detail in Section 2.6.

Surface ECG is one of the most simple, accessible and non-invasive signals and is cost-effective to acquire. The traditional ECG signal has many important components, including P wave, QRS complex, QT interval and T wave (as shown in Fig. 1.3) (https://en.wikipedia.org/wiki/QRS_complex).

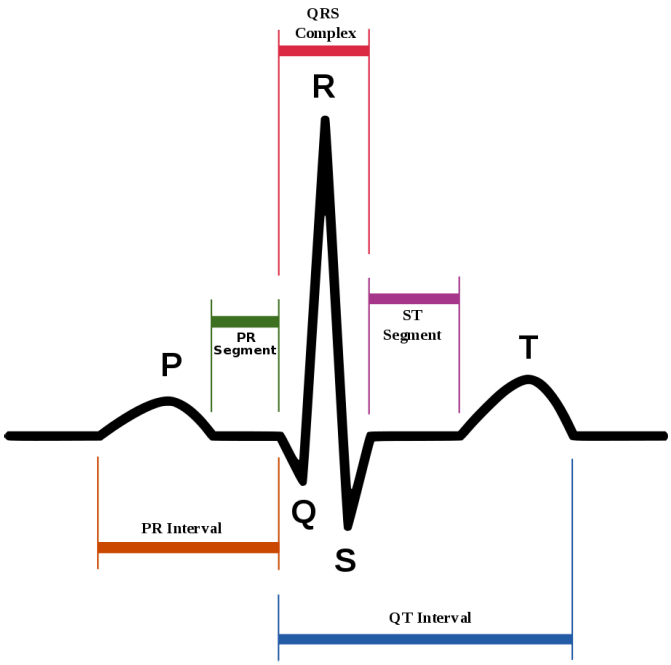

Fig. 1.3 Typical ECG signals with P, Q, R, S, T wave.

Different ECG components show partial information in both heart rate and rhythm (steady or irregular) [1]. For CHF detection, many studies have been done in ECG information analysis of HF patients (see Chapter 2.6) based on different components, including QRS waves [43], QT intervals [44], RR variability [45]. A significant relation between different ECG components and HF has been found.

Among all these components, RR intervals has attracted the most attention. RR interval is the interval of adjacent R peaks, one of the most important and widely used

components of ECG and its variability has great potential in CHF prognosis [30]. This variability is either described by RR variability (i.e. the variation of RR intervals) or heart rate variability (HRV). Researchers have mainly focused on the relationship among HRV measures and patients' severity and autonomic dysfunction. According to the NYHA classification, the severity scale of heart failure depends on the severity of symptoms [46], which are partly modulated by the autonomic nervous system. Over the years, HRV analysis has been confirmed as a reliable and noninvasive tool in the prognosis and risk assessment of CHF, and it is widely used to assess the influence of the ANS on the heart [46].

Based on the statistical analysis of HRV in the short/long term, the classifier based automatic detection method for CHF has achieved accuracy of over 90% [47]. But not much attention has been paid to CHF stratification, which classifies the different severity of CHF patients. Bad precision in lower risk patients (NYHA III-IV) is reported, which requires more exploration in the risk quantification area of CHF [48].

Technically, the majority of researchers have applied three approaches to determining HRV: the time domain analysis, frequency domain analysis, and nonlinear analysis [46]. The time/frequency domain HRV measurements follows International Guidelines [46], and this has been employed by most researchers. Time and frequency domain analyses address variability degrees and the underlying rhythms, respectively [49], while nonlinear analysis processes the dynamic variation features in HRV generated from physiological systems [50], e.g. complexity.

Statistical analysis based on these three approaches of HRV analysis was done between samples from non-CHF people to CHF patients. A significance difference of time/frequency domain HRV measurements has already been studied in statistics difference levels between normal people and CHF patients [51–53]. Decreased time and frequency domain HRV measurement was a marker of sympathoexcitation [54][55]; this measurement is of clinical significance. In addition, HRV measurements of adverse changes in the autonomic function of CHF manifest in altered HRV analysis [56]. Nonlinear HRV indices analysis of HRV were then proposed include appropriate entropy, fuzzy entropy and so on. In recent studies, researchers also proposed to many different nonlinear HRV indices to describe dynamic fluctuations of RR intervals in a certain

cycle [57, 58]. the nonlinear analysis provides an improved decomposition of cycle fluctuation of HRV in a period and achieves significant discrimination power for CHF detection [59]. These provide proof that HRV analysis can be a useful tool for CHF assessment in both statistical level and classification.

Furthermore, as a common syndrome and end stage of different cardiovascular diseases, congestive heart failure (CHF) attracts much attention in terms of using HRV analysis for risk assessment [49]. Two types of HRV analysis have been recommended as a guideline: long-term (24-h) and short term (5-min) [46]. However, there is demand for a more flexible analysis to investigate rapid fluctuations of HRV in a day, rather than single 5-min or 24-h analysis [60]. It has been shown that short term HRV measurement can be used as predictor in congestive heart failure (CHF) detection and autonomic nervous system (ANS) function assessment [61, 47]. However, the main problem is that different 5-min segments do not display the same power in CHF detection. This might indicate that biased data selection was used during the research. Furthermore, the circadian rhythm of HRV was found in CHF patients [62]; and even sleep stages also affect HRV [63].These research findings indicate that HRV alters with daily activity, which also reflects the fact that autonomic nervous function fluctuated under different physiological activity in 24 hours. Thus, timescale analysis in an important issue to explore in the risk assessment of CHF.

As to cardiopulmonary system risk assessment, numerous studies have explored cardiovascular variation in electrocardiograph (ECG) recording of cardiopulmonary patients, especially in sleep apnea [64] and congestive heart failure (CHF) [50]. For patients with obstructive sleep apnea, the accompanying breath disorder condition worsened resulting in the failure of the heart to pump blood [65]. Patients with CHF and those with sleep breathing disorders shared common symptoms of obstructive and central apnea. Worse still, both diseases are associated with a high risk of death and financial loss [65].

As in the diagnosis of apnea, polysomnography (PSG) has been widely used in monitoring sleep breathing and it is the gold standard for the diagnosis of sleep apnea [4]. PSG is a comprehensive sleep study device recording multiple physiological signals, including electroencephalograph (EEG), Electrocardiograph (ECG), respiratory

airflow, electrooculogram (EOG) and electromyogram (EMG). In all these signals, ECG displayed the heart rhythm of the human body during sleep. The variation in the time intervals between heart beats (beat-to-beat interval) is heart rate variability (HRV), which can be derived from the ECG signal. There are many reports using HRV for the detection of sleep apnea and CHF [49, 64, 66, 67].

In general, statistical analysis of ECG components for CHF detection have proved adverse changes between CHF patients and normal people. No work applied ECG components into CHF classification other than HRV. Furthermore, HRV based CHF detection has achieved over 90% accuracy with linear/non-linear measures of short-term (5-min)/long-term (24-h) RR intervals data [31]. However, due to its sensitivity along with clinical condition changes, the robustness of the HRV-based approaches is still an issue to be addressed. The dynamic fluctuation of HRV in 24 hours is dynamically changed along with conditions, in accordance with RR intervals fluctuation. Thus, it is necessary to improve the robustness of CHF detection with RR intervals data, in both the algorithm and indices. While searching for research on CHF quantifications, it seems not much attention has been paid to classification. There are many works about the significance analysis of different risk level CHF patients which have sought to prove the difference among different levels but no clear monotonic change has been found along with severity. One paper described the classification performance in two-level CHF quantification but the results were poor. Finally, most of these studies were based on 5-min/24-h time length. There was no detailed discussion about different timescales when assessing the risk of CHF patients. As to cardiopulmonary system assessment, there are many papers which have achieved good performance in terms of CHF detection and sleep apnea classification but no work which has incorporated the classification of CHF and sleep apnea.

1.2 Motivation and Aims

The focus point of this dissertation is to build a more robust and convenient model for congestive heart failure (CHF) assessment based on ECG signal and intelligent classification methods, which includes innovative dynamic indices analysis, 4-level

risk stratification using 24-h data, deep learning based CHF detection using 5-min data, and optimal timescale analysis for 3-level CHF classification and classification of CHF and sleep apnea patients for comprehensive cardiopulmonary assessment system construction. The final cardiopulmonary system is targeted to be implemented on a wearable device, which could collect high quality surface ECG signals or R peaks. The device featuring our algorithm could help to give a fast and reliable risk assessment to CHF patients which could reduce hospitalisation and improve their quality of life.

The dissertation has several aims:

1) To propose new heart rate variability (HRV) measures to describe cardiac autonomic nervous system function fluctuation in 24-h daily life and explore its potential in congestive heart failure assessment;

2)To develop a multilevel risk assessment model of CHF using long-term (24-h) RR intervals data to explore the classification power of 24-h ECG data in CHF detection and stratification;

3) To develop a multilevel neural network with a deep learning algorithm and short-term RR intervals to explore the disease detection ability of 5-min RR intervals;

4) To search for optimal time length for multilevel risk assessment of CHF to achieve robust and fast CHF classification;

5) To construct a comprehensive CHF and sleep apnea classification model for the purposes of achieving robust, reliable and comprehensive CHF assessment based on whole night ECG data.

1.3 Dissertation Contribution

This dissertation presents the research and development of a novel automated congestive heart failure risk assessment method using ECG signals and intelligent classification algorithms with application in noninvasive and convenient monitoring. The primary contribution of this dissertation has five aspects, as listed below:

1. This dissertation creatively proposes a new series of HRV measures - dynamic indices - to evaluate heart function fluctuation under different daily life and the severity evaluation of CHF. Furthermore, mechanism analysis proves three dynamic indices (mean of ratio of low to high frequency power (MLH), standard deviation of mean (SDM) and fuzzy entropy of ratio of low to high frequency power (FELH)) display a monotonic and significance difference (p<0.001) reduction in severity, and indicate good discrimination power in quantifying three risk groups of CHF (detailed described in Chapter 3). The reduction in these indices might be related to potential cardiovascular autonomic unbalance risk, and could be helpful for prognostic information in future clinical application.

2. This dissertation firstly builds a four-level risk assessment model for CHF detection and quantification based on the decision tree based support vector machine (DT-SVM) algorithm. The model succeeds with a total accuracy of 96.61 %, in risk assessment among individuals of N (no risk), P1 (mild risk), P2 (moderate risk), and P3 (severe risk). This long term data based model provides a powerful predictor between predicted and actual ratings, and it could serve as a clinically meaningful outcome in relation to providing an early assessment and a prognostic marker for CHF patients.

3. In the performance analysis of severity evaluation of CHF based on the support vector machine (SVM), it is clear that classic HRV is more powerful in the disease detection of CHF than dynamic indices with an accuracy of 98.31 %. This is consistent with the results of prior research regarding disease detection; the dynamic indices of HRV are obviously superior to the classic ones in CHF stratification, increasing the accuracy from 73.91 % to 91.30 %. This is consistent with the mechanism analysis results and helps to improve the fur-level risk assessment model performance of CHF. It also proves that our proposed dynamic indices of HRV is better in CHF quantification while classic HRV is better in CHF detection.

4. The DT-SVM based multistage risk assessment model proposed in this work significantly improves discrimination power from 76.27 % to 96.61 % when compared to classic SVM. The performance of our classifier improved based on

the combination of the backward elimination algorithm and significance difference to get an optimal tree node structure and feature set at each node.

5. This dissertation firstly applies a sparse anto encoder based deep learning (SAEDL) algorithm into CHF detection and achieves an accuracy of 71.44% using 5-min RR data. This unsupervised detection method proves that 5-min RR intervals have potential in CHF detection but cannot fully reflect dynamic change in 24-h; the normal performance is also consistent with the clinical fact that 5-min RR data is sensitive to clinical condition changes.

6. This dissertation firstly finds the optimal time scale is 2-h when using HRV for 3-level risk assessment, based on significance analyses among data at different timescales from 5-min to 24-h. We apply HRV measurements at the optimal time scale (2-h) to construct a 3-level risk assessment model of CHF. The model achieves good precision in healthy persons (NR), lower risk patients (LR) and higher risk patients (HR) at 91.89 %, 65 % and 90.91 %. Compared with previous research using both 5-min and 24-h data , the 2-h data shows comparable performance both in CHF detection and quantification. Indeed, the CHF auto assessment model based on the optimal time scale (2h) data constructed in this study provides a credible and convenient risk assessment, which can meaningfully contribute to family monitoring.

7. In this dissertation, an autonomic model is firstly constructed to distinguish between sleep apnea patients with and without congestive heart failure and this model achieves an accuracy of 81.68 %. The result indicates that HRV from PSG records has the potential to distinguish between apnea and CHF patients. This can serve as a potential risk factor in risk assessment for comprehensive cardiopulmonary system in the future.

1.4 Dissertation Outline

The outline of this dissertation is as follows:

Chapter 1

This chapter give an introduction to the problem statement, motivation and aims of this thesis and give a introduction of the contribution and outlines of this dissertation.

Chapter 2

This chapter firstly gives a brief introduction to the classic heart rate variability measurement, the statistical analysis method used in this dissertation, feature selection methods, intelligent classification methods and widely used model performance evaluation measures. Then introduces a literature review of related research topics to give a clear understanding to the current situation,

Chapter 3

This chapter describes in detail the mechanism analysis of dynamic heart rate variability in autonomic unbalance for congestive heart failure stratification.

Stratification of patients with congestive heart failure (CHF) has been based on the traditional heart rate variability (HRV) measurement, mainly in disease detection. The aim of this study is to investigate new quantifying indices of HRV for samples from no risk, low risk, and high risk to evaluate the underlying autonomic modulation mechanism. 116 electrocardiographic data from MIT/BIH database were analysed, including 72 normal samples and 44 samples with CHF. These samples were labeled into 3 risk groups: no risk (NR, normal samples; n = 72), low risk (LR, patients in NYHA I-II; n = 12) and high risk (HR, patients in NYHA III-IV; n = 32). Classical HRV measures and dynamic indices were obtained from nominal 24-hour recordings in three perspectives: time domain, frequency domain and complexity to quantify risk levels. Significance difference analysis was performed among the groups with these indices.

The analysis results showed that 30 out of 40 indices showed significance difference ($p<0.05$) power among groups. Three significant quantifying indices were found: mean of ratio of low frequency power to high frequency (MLH), standard deviation of mean (SDM), and fuzzy entropy of LH (FELH). These indices showed clear difference ($p<0.001$) and monotonic reduction in proportion to different risk levels of NR, LR and HR. These three dynamic indices indicated a powerful ability to differentiate the risk level of samples. The monotonic reduction of these indices might be related to

potential cardiovascular autonomic unbalance risk, and could be helpful for prognostic information.

Chapter 4

This chapter describes in detail our novel and effective method for congestive heart failure detection and quantification using dynamic heart rate variability measurement.

Risk assessment of congestive heart failure (CHF) is essential for detection, especially helping patients make informed decisions about medications, devices, transplantation, and end-of-life care. The majority of studies have focused on disease detection between CHF patients and normal subjects using short-/long-term heart rate variability (HRV) measures but not much on quantification. We downloaded 116 nominal 24-hour RR interval records from the MIT/BIH database, including 72 normal people and 44 CHF patients. These records were analysed under a 4-level risk assessment model: no risk (normal people, N), mild risk (patients with New York Heart Association (NYHA) class I-II, P1), moderate risk (patients with NYHA III, P2), and severe risk (patients with NYHA III-IV, P3). A novel multistage classification approach is proposed for risk assessment and rating CHF using the non-equilibrium decision-tree-based support vector machine classifier. We propose dynamic indices of HRV to capture the dynamics of 5-minute short term HRV measurements for quantifying autonomic activity changes of CHF. We extracted 54 classical measures and 126 dynamic indices and selected from these using backward elimination to detect and quantify CHF patients. Experimental results show that the multistage risk assessment model can realise CHF detection and quantification analysis with a total accuracy of 96.61 %. The multistage model provides a powerful predictor between predicted and actual ratings, and it could serve as a clinically meaningful outcome providing an early assessment and a prognostic marker for CHF patients.

Chapter 5

This chapter describes in detail the proposed CHF detection method based on deep learning with 5-min RR intervals.

There are extensive studies investigating congestive heart failure (CHF) detection based on heart rate variability. Although a high level of accuracy has been achieved, its robustness under different conditions is not guaranteed. To improve the robustness, we

applied a sparse auto-encoder-based deep learning algorithm in CHF detection with RR intervals. A total data size of 30,592 (5-min RR interval) was obtained from 72 healthy persons and 44 CHF patients. The deep learning algorithm first extracts unsupervised features using a sparse auto-encoder from raw RR intervals, then constructs a deep neural network model with various hidden nodes combinations. Results showed that the model achieved 72.41 % accuracy. This demonstrates that RR intervals have potential in CHF detection but cannot fully reflect dynamic change in 24-h.

Chapter 6

This chapter describes in detail the automatic risk assessment of congestive heart failure using ECG at optimal timescale.

Risk assessment of congestive heart failure (CHF) is essential because of its high morbidity and high cost. Most of the studies related to CHF detection have focused on 5-minute/24-hour RR interval analysis and different automatic classification algorithms. The problem is that the 5-minute RR interval data is sensitive to physiological changes, while the 24-hour data is difficult to obtain. In this chapter, our goal is to explore the best time scale for the electrical data of the CHF assessment center and to achieve robust and accurate predictions.

This chapter analysed 72 healthy samples from the MIT-BIH database, 12 low-risk patients in NYHA I-II, and 32 high-risk patients in NYHA III. It firstly calculated 10 classical heart rate variability indicators based on ECG data at different time scales (from 1h to 24h). A statistical analysis with a significant level of 0.05 was then performed to compare the discriminating power of ECG data at different time scales. Finally, a CHF autonomous evaluation model was constructed based on the support vector machine and limit search feature selection algorithm.

The results showed that 2 hours was the best time scale for CHF assessment, and the classification precision for healthy and high-risk patients was 91.89 % and 90.91 %, respectively. The study shows that the CHF auto assessment based on the optimal time scale (2h) data constructed in this chapter provides a credible and convenient risk assessment, which may contribute to family monitoring.

Chapter 7

This chapter It firstly calculated an automatic method to differentiate sleep apnea patients with and without congestive heart failure using Polysomnography records.

Sleep breathing disorders and congestive heart failure (CHF) are common diseases of the cardiopulmonary system, both of which are associated with a high risk of death. Obstructive sleep apnea (OSA) often causes heart damage, while OSA, central apnea and Cheyne-Stokes breathing are consequences of CHF. Many studies have reported clinical and statistical differences between apnea patients with and without CHF, but currently, no autonomic methods are available to distinguish them. Polysomnography (PSG) is the gold standard for the diagnosis of sleep apnea with heart rate data. Twenty whole night PSG data from the Sleep Heart Health Study database were included in this study. The Pan-Tompkins algorithm was applied to the electrocardiograph signal to detect R peaks of the QRS complex. The whole night R peaks data were then manually checked and segmented into 1895 5-minute epochs to calculate three frequency domain and three nonlinear heart rate variability measures. All these measures were analysed for their statistical differences between groups (sleep apnea with and without CHF). Finally, a binary support vector machine classifier and extreme search method were performed to construct the model. Results showed that an accuracy of 81.68% was achieved in distinguishing sleep apnea patients with and without CHF. This indicated that HRV measures from PSG had the potential to help distinguish sleep apnea patients with and without CHF reported.

Chapter 8

This chapter provides a conclusion to this project, and suggests areas for future work in this field.

Chapter 2

Background and Literature Review

In this chapter, we first provide a brief introduction and understanding of classical heart rate variability (HRV) measures for risk assessment of congestive heart failure (CHF) using ECG. The statistical analysis methods are briefly introduced. Then we introduce the feature selection methods used in this dissertation to reduce feature space and improve classification performance. Applied intelligent algorithms are explained for CHF detection and stratification, as the main foundation of robust model construction. Next, we examine traditional measures to evaluate the performance of the constructed the model. Finally, a literature review of related research topics is provided to give a clear understanding of the current situation,

2.1 Classical HRV Measures for Risk Assessment of Congestive Heart Failure

In the analysis for CHF detection and quantification, multiple HRV measures were calculated in this project. Usually, the most widely used and classical measurement for HRV analysis is grouped into three different methods: time-domain and frequency-domain and non-linear methods, as described below.

1) Time-domain methods

The time domain methods of HRV analysis means to calculate the statistic measures from the beat-to-beat intervals (also called RR intervals) of ECG signal. Some most widely used measures include:

- Mean, the mean value of RR intervals.
- SDNN, the standard deviation of RR intervals. This is often calculated over a window of 5 minutes. SDNN reflects all the cyclic components responsible for variability in the period of recording. As such, it represents total variability.
- RMSSD, the root mean square of successive RR intervals differences.
- CV, the coefficient variation of RR intervals over a period.
- SDANN, the standard deviation of the average RR intervals calculated over short windows of 5 minutes.
- SDSD, the standard deviation of the standard deviation of RR intervals over short windows of 5 minutes.
- NN50, the number of pairs of successive RRs that differ by more than 50 ms.
- pNN50, the proportion of RR50 divided by the total number of RRs.

2) Frequency-domain methods

The frequency domain methods assign multiple bands of frequency and then count the number of RR intervals that match each band. The bands are typically divided into three parts: high frequency (HF) from 0.15 to 0.4 Hz, low frequency (LF) from 0.04 to 0.15 Hz, and the very low frequency (VLF) from 0.0033 to 0.04 Hz. Thus, there are multiple frequency domain measures achieved:

- VLF: the very low frequency power from 0.0033 to 0.04 Hz.
- LF: the low frequency power from 0.04 to 0.15 Hz.
- HF: the high frequency power from 0.15 to 0.4 Hz.
- LH: the ratio of low to high frequency power.
- TP: the total frequency power.

The frequency domain analysis methods are called power spectral density (PSD) analysis. PSD used parametric or nonparametric calculation methods to provide power distribution information across frequencies. Both parametric or nonparametric methods provide comparable results in frequency domain analysis.

Both methods have their own advantages. The advantages of the nonparametric methods are the simplicity of the algorithm used (fast Fourier transform (FFT) in most of the cases) and the high processing speed. But the nonparametric methods has limitation in uneven signal. While the advantages of the parametric methods are smoother spectral components with better distinguished components independent of preselected frequency bands, easy postprocessing of the spectrum (automatic calculation of different frequency power components with an easy identification of the central frequency of each component), and accurate estimation on a small sample number on the stationarity signal. The basic disadvantage of parametric methods is owing to the chosen calculation model. For the chosen model, the user needs to verify suitability and complexity (i.e. the model order).

In addition to classical FFT-based PSD methods used for the calculation of frequency domain parameters of HRV, a more appropriate PSD estimation method is the Lomb-Scargle (LS) periodogram [68]. Analysis has shown that the LS periodogram displayed more accurate PSD estimation than FFT methods for typical RR data [68]. This advantage relates to the fact that LS based PSD method is able to be used without the need to resample and detrend the RR data. This is more suitable for unevenly sampled data, like RR interval data, as opposed to FFT-based methods.

In recent years, a new frequency domain HRV measurement has been proposed, called wavelet entropy. This is an alter understanding aspect in bands power to the above frequency domain measure. Wavelet entropy measures are calculated with a three-step procedure defined in the literature [68]. First, the wavelet packet algorithm is applied using a chosen mother wavelet. In this project, the Daubechies 4 (DB4) function was used as the mother wavelet with a scale of 7. Based on the mother wavelet, the algorithm calculates the wavelet coefficient at different bands. Once the wavelet coefficients are obtained, the energy for each coefficient is calculated as described in the literature[68]. Then calculation of the

normalised values of wavelet energies is done, which represent the relative wavelet energy (or the probability distribution), the wavelet entropies are obtained using the definition of entropy given by Shannon [69].

3) Non-linear methods

Recent researchers proposed multiple different nonlinear HRV measures based on the complexity of the mechanisms regulating heart rate. As a signal from the cardiac system, it is reasonable to assume that non-linear HRV analysis gives the potential to yield valuable cardiac system fluctuation information. Although chaotic behaviour has been discussed for ECG records under different daily activities, more rigorous testing has shown that HRV measurent can be described as a higher dimensional chaotic process [70]. Moreover, application of chaotic global to HRV has been shown to predict diabetes status [71]. One of the most commonly used non-linear measures for HRV analysis is the Poincare plot. In this plot, each data point represents a pair of successive beats: the x-axis is the current RR interval, while the y-axis is the previous RR interval. Besides this, HRV can also be quantified by fitting mathematically defined geometric shapes to the data [72]. Other methods used include the approximate entropy [57], correlation dimension, detrended fluctuation analysis [73], nonlinear predictability [70], memory length (based on inverse statistical analysis) [74], multiscale entropy analysis [58], pointwise correlation dimension [75], sample entropy [57], sample asymmetry [76] and so on. More and more non-linear HRV measurements are explored by researchers to represent long-range correlations geometrically [77].

All the HRV measurement calculations were done using the software MATLAB 7.11.0 (version R2010b, The MathWorks, Inc., Natick, MA, USA).

2.2 Statistic Analysis Methods

In this dissertation, all measurements are presented as mean±SD. Differences between groups are assessed using one-way analysis of variance (one-way ANOVA). One-way ANOVA determines whether there are any statistically significant differences between

the means of two or more independent groups. As mentioned before, this project focuses on multiple risk groups of congestive heart failure (CHF). This method evaluates the classification performance of each heart rate variability (HRV) measure among the risk groups. Statistical tests are conducted with the significant level set at 0.05. A p value <0.05 is considered statistically significant. All the significance analysis has been done using the SPSS software (version 19, SPSS Inc., Chicago, IL, USA).

2.3 Feature Selection Methods for Congestive Heart Failure Assessment

In this dissertation, multiple measures were calculated. Calculation were performed on software MATLAB 7.11.0 (version R2010b, The MathWorks, Inc., Natick, MA, USA). Owing to their high correlation, directly using all features for classification might not give the best performance [78]; thus, using an appropriate selection method for feature subset discrimination improves classifier performance. We applied backward elimination (BE) and exhaustive search [79] in the feature selection.

2.3.1 Backward Elimination

The backward elimination (BE) algorithm begins with all features and iteratively removes them one by one until the remaining features reach the highest precision. The BE algorithm was combined with a support vector machine classifier to evaluate the performance of all possible feature subsets, as detailed below.

Let's say there are N features from M participants which are labelled into 0,1. First, all features were inputted into a support vector machine to evaluate the classification power of current feature subset, set as ACC^N. Then we deleted one feature, which leads to $N-1$ possible feature subsets. The corresponding classification performance of these feature subsets was set as $\{ACC^0, ACC^1, ..., ACC^{N-1}\}$. Here, we need to compare ACC^N and $\{ACC^0, ACC^1, ..., ACC^{N-1}\}$ and find the biggest number. If ACC^N is the biggest one, we stop; if not, we set the biggest one as ACC^k, which

corresponds to the kth feature subset. Then we update the kth feature subset as the new feature set, and repeat all the steps. Finally, the optimal subset is achieved.

2.3.2 Exhaustive Search

Exhaustive search method is a simple way to evaluate all possible feature subsets on current feature space, usually applied in small feature space. Let's say there are N features in the current feature space, then we can calculate the number of possible feature subsets as $C_N^1 + C_N^2 + C_N^3 + ... + C_N^N$. Then we apply a classifier, like the support vector machine, to evaluate the classification power of all these subsets $\{ACC^0, ACC^1, ..., ACC^{C_N^1+C_N^2+C_N^3+...+C_N^N}\}$. The biggest number among these classification powers indicates the optimal feature subset.

2.4 Classification Algorithm for Congestive Heart Failure Assessment

2.4.1 Support Vector Machine

Support vector machine (SVM) is a widely used classifier in machine learning. The advantage of SVM is in binary classification assigning new samples to one category or the other [80]. The SVM algorithm will separate input features by a hyperplane to reach the largest margin between different groups, also called maximum-margin hyperplane [81].

The hyperplane was calculated based on the function:

$$f(x) = W^T x + b, \begin{cases} f(x) > 0, then & x \in class \quad 1 \\ f(x) < 0, then & x \in class \quad 2 \end{cases} \tag{2.1}$$

Here, W and b are the weight vector and bias that maximise the margin, respectively. x is the feature vectors of one sample. The W and b are then decided by the Lagrangian function.

2.4.2 Decision Tree based Support Vector Machine

DT-SVM is an effective way of combining an SVM and a decision tree for solving multi-class problems [82]. It is a modified method of the classical SVM for dealing with its difficulty in multi-class problems, but DT-SVM brings another danger: cumulative error. This error is caused by sample misjudgment at the upper node of the decision tree and lasts throughout the rest of the classifier without elimination. The main idea of this algorithm is the conversion of multiclass classification problems into multilevel binary classification problems. Each level includes two nodes to be classified, and each node includes one or several classes. At every node, a decision is made to assign the samples. This step is repeated until all the samples reach a leaf node, to which only one class of samples is assigned. In this way, a hierarchy is formed [83].

2.4.3 Sparse Auto Encoder based Deep Learning

Denote input data as $RR = rr(i), i = 1, \cdots, N, rr(i) \in R^M$. Here N is the data sample number of one class, and M is the data samples length. First, we input this unlabelled data set into a sparse auto encoder (SAE) for automatic feature learning to constrain features. At hidden layer l, the feature is expressed as $F(rr(i), W, b), i = 1, \cdots, N$, where W denotes the weights between two neighbouring layers and b is the bias. This means that the required feature is controlled by the input and connection parameters (W, b).

The unlabelled segment data RR are first used to train unsupervised features automatically with SAE through the following steps:

1. Initial settings including the sparsity penalty term, decay parameter, desired average activation of the hidden units, and initialise connections W and b randomly close to 0.

2. Use batch gradient descent to train the neural network in the forward propagation algorithm to compute the sparse cost function of each iteration of the layer as:

$$Cost\,(w, b) = \frac{1}{n}\sum_{i}^{N}(\frac{1}{2}||F_{w,b}\,(rr\,(i))\,||^2) + \lambda] + \beta\sum_{j=1}^{s2} KL(\rho||\rho_j) \qquad (2.2)$$

Here $\rho_j = \frac{1}{n}\sum_{i=1}^{n} a_j\left(rr\left(i\right)\right)$ is the average activation of hidden unit j; $KL(\rho|\left|\rho_j\right) = \rho log\frac{\rho}{\rho_j} + (1-\rho)log\frac{1-\rho}{1-\rho_j}$ is a standard measuring function of the difference of the two distributions; λ is the weight decay parameter; $s2$ is the number of hidden neurons in the hidden layer.

3. Update the parameters W and b at each iteration of the gradient descent:

$$W_{ij}\left(l\right) = W_{ij}\left(l\right) - \alpha\frac{\partial}{\partial W_{ij}\left(l\right)}Cost\left(W,b\right) \tag{2.3}$$

$$b_i\left(l\right) = b_i\left(l\right) - \alpha\frac{\partial}{\partial b_i\left(l\right)}Cost\left(W,b\right) \tag{2.4}$$

Here α is the learning rate/step size, optimised by the linear search optimiser; l is hidden layer (here is 1).

4. Repeat step 2-4 for the second hidden layer for each iteration until the iteration is done (set to 300); or we have reached the minimum cost value.

Thus, the learnt feature set F is extracted and fed into deep learning (DL) for classification.

1) Select the values of the parameters (W,b) from feature learning to initialise the DL neural network.

2) Set up the training parameters, and conduct the forward propagation algorithm to construct the classification model based on a Softmax classifier.

3) Compute the mean square error for the cost function of the DL using Eq.2.2.

4) Conduct the back-propagation algorithm to update the connections and fine-tune the entire network for each iteration until the iteration is done (set to 300); or we have reached the minimum cost value.

In such a way, a deep neural network is conducted for unsupervised classification.

2.5 Performance Evaluation Measures

In this dissertation, during the model construction procedure, we input the training set into a classifier to obtain a model for risk assessment of CHF and we used a validation set to evaluate its performance. In the binary classification, all subjects were usually labelled into two classes, 0 (negative) or 1 (positive). To evaluate the performance of a constructed model, some widely used parameters were defined. Let n be the number of subjects labelled as negative, and p be the number of subjects labelled as positive. Then tp is the number of true positives, tn is the number of true negatives, fp is the number of false positives, and fn is the number of false negatives. Then, the $Accuracy$, $Sensitivity$, $Specificity$, and Auc (area under the curve) can be calculated:

$$Accuracy = (tp + tn)/(tp + tn + fp + fn) \tag{2.5}$$

$$Sensitivity = tp/(tp + fn) \tag{2.6}$$

$$Specificity = tp/(tn + fp) \tag{2.7}$$

$$Auc = \frac{(Sensitivity + Specificity)}{2} \tag{2.8}$$

2.6 Literature Review

In this dissertation, to build a noninvasive, convenient, robust and comprehensive model for congestive heart failure (CHF), former related work is separately introduced and discussed below.

2.6.1 Risk Assessment of CHF with Physiological Measures

Monitoring and risk assessment of congestive heart failure (CHF) aimed at accurate and robust detection of CHF. As a progressive disease, the adverse trend of heart function is one of the significant features among CHF patients. Cardiac autonomic nervous dysfunction of CHF patients appears with sympathetic nervous tension and/or parasympathetic nervous inhibition. To assess the heart functional deterioration degree,

there are many physiological measurement analyses, especially cardiopulmonary signal analysis related to CHF assessment.

In 2006, a model called the Seattle Heart Failure Model was derived with the use of a multivariate Cox model to estimate 1-, 2-, 3-year survival in heart failure patients with the use of characteristics relating to clinical status, therapy, and laboratory parameters (see Fig.2.1) [84]. The overall receiver operating characteristic area under the curve was 0.729 (95% CI, 0.714 to 0.744). This indicated that clinical characteristics were the primary factors in heart failure prognosis. Also, more researchers explored accurate CHF detection with pulmonary and cardiac signals.

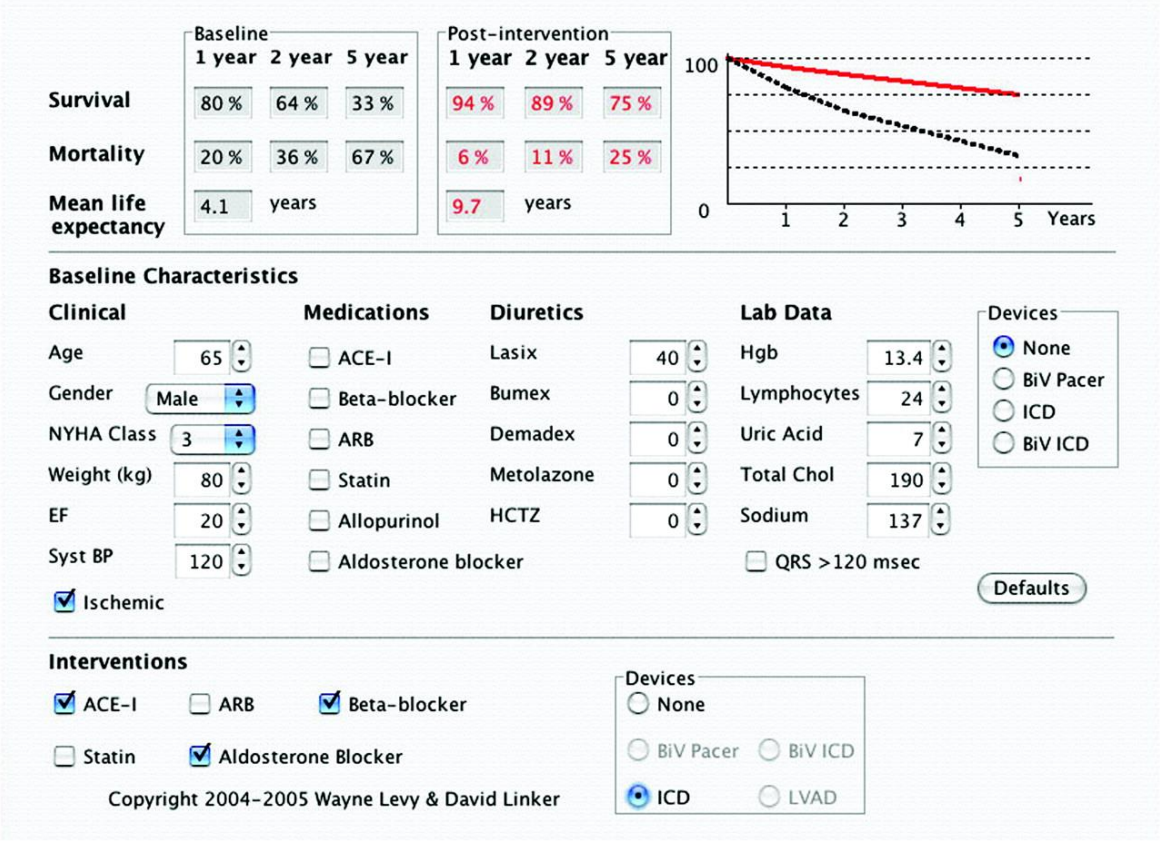

Fig. 2.1 The Seattle Heart Failure Model has been implemented as an interactive program that employs the Seattle Heart Failure Score to estimate mean, 1-, 2-, and 5-year survival and the benefit of adding medications and/or devices for an individual patient. This model is available at www.SeattleHeartFailureModel.org. ACE-I indicates ACE inhibitor; ARB, angiotensin receptor blocker; HCTZ, hydrochlorothiazide; Hgb, hemoglobin; and BiV, biventricular.

While in oxygen analysis of respiratory signals, Tabet et al. compared the prognostic value of the mitral inflow pattern and peak oxygen uptake in one hundred patients

with heart failure who underwent exercise testing (ejection fraction <45%) [85]. The mean follow-up of 17 months comparison indicated that prognostic value of peak oxygen uptake was more than that of Doppler variables and ejection fraction (P = 0.0004)(see Fig.2.2). This indicates that the Peak oxygen uptake remains a more powerful prognostic variable in patients with systolic heart failure and is independent of the left ventricular ejection fraction.

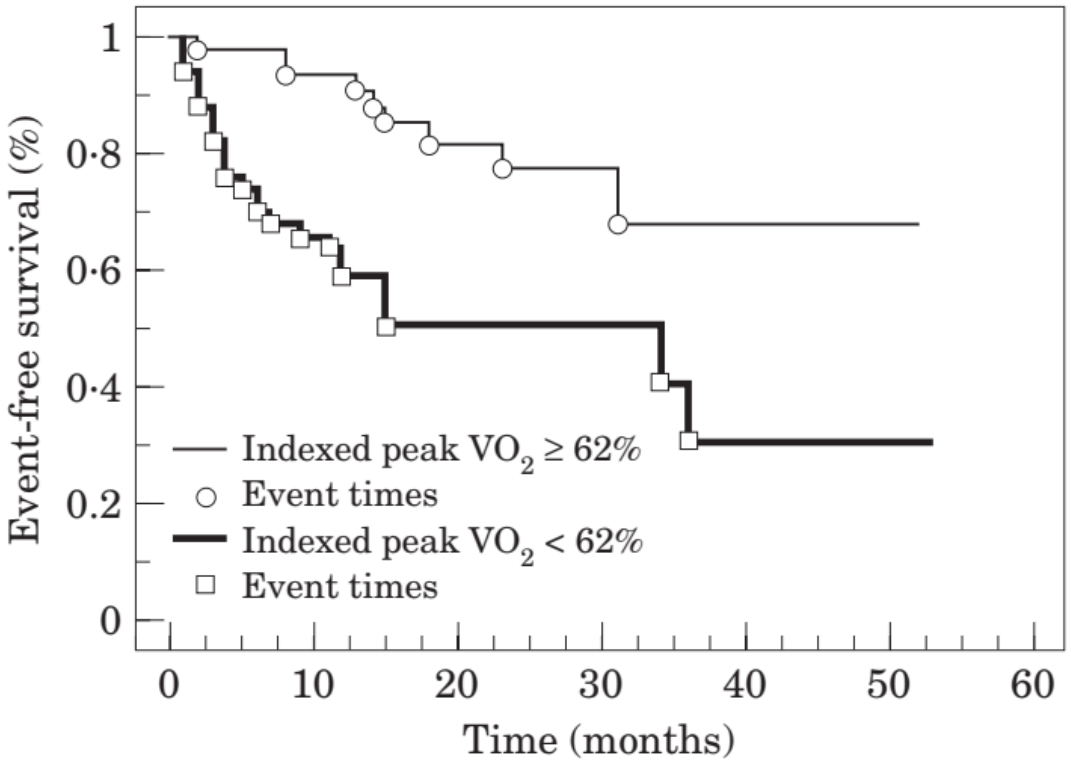

Fig. 2.2 Event-free survival curves according to baseline percentage of predicted peak oxygen uptake, greater or less than 62% of the predicted values (median value)(p=0.0004).

In addition, a lung image based algorithm application was done in CHF detection. Rezaeieh et al. used a foam-embedded wideband antenna array to build a portable system for congestive heart failure (CHF) detection [86]. The system includes a portable vector network analyser (VNA), a switching system, and a laptop with control, signal processing, and image formation algorithms. The system was successfully tested on an artificial phantom and a phantom with a pair of lamb lungs to validate reliability in the early detection of CHF. This indicated that this lung-image based system has

potential as a portable preclinical device for future human tests in noninvasive early stage CHF detection.

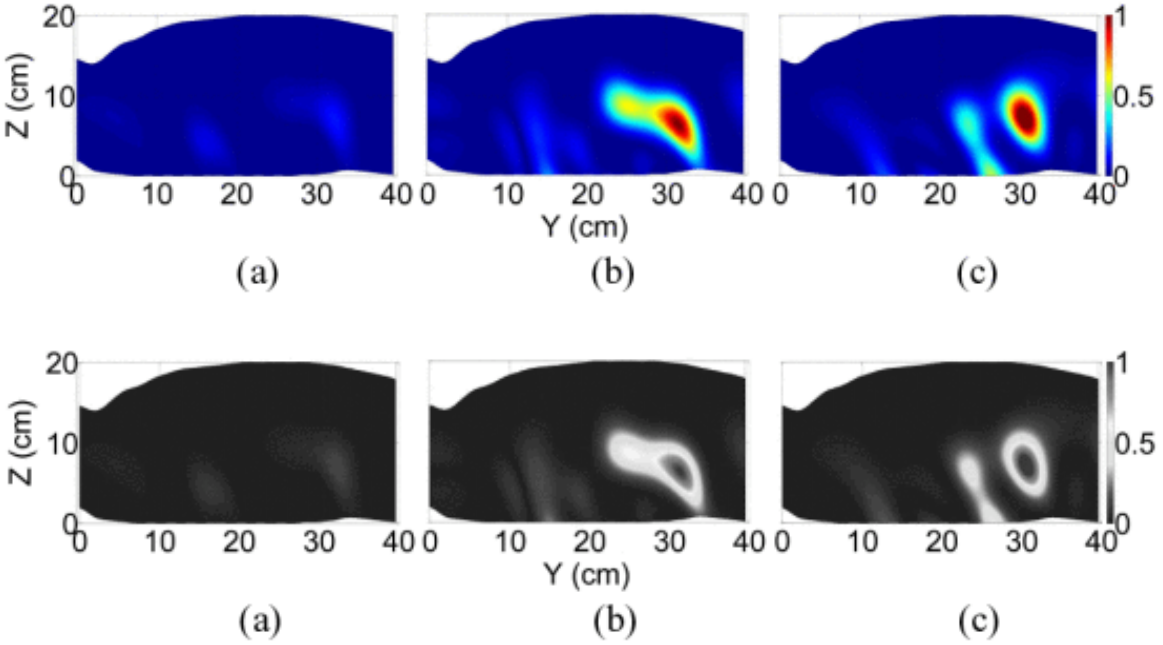

Fig. 2.3 Images for (a) healthy case, (b) 2 mL water injected in right lung, and (c) water injected n both lungs with difference of 2 mL.

Finally, cardiac signals is another focus in CHF detection. Zhao et al. utilised fuzzy measure entropy and sample entropy of long-term RR intervals to distinguish CHF patients from normal persons [87]. Performance of fuzzy measure entropy was better under different combinations of parameter setting.

These researches proved the power of cardiac and pulmonary physiological measurement in CHF detection. Among these measurements, the majority of reports are about the application of ECG signals in CHF detection, i.e. different ECG components. The following sections will introduce different components of ECG signals in the risk assessment of CHF.

2.6.2 Risk Assessment of CHF with HRV

Heart rate variability (HRV) is a measurement of heart rate variation, while heart rate is one of the most important components of ECG. Major studies focus on using short/long term HRV measures combined with a suitable classification algorithm and feature selection methods, to achieve a better performance in congestive heart failure (CHF) detection and stratification. Here, CHF detection indicates the detection of CHF patients amongst healthy people, while CHF stratification means classification among different risk level CHF patients according to the NYHA classification. All works are described below.

In 1991, American scholar Binkley analysed frequency domain HRV measures for the relation between sympathetic/parasympathetic nervous system rhythm variations and proved there was a significant difference [88]between the healthy people and congestive heart failure (CHF) patients. In 1992, Takase et al. proved a significance difference between CHF patients and normal people via time domain HRV measurement [89]. Many key works were done in mechanism analysis between healthy people and CHF patients using HRV measures, to explore the relationship between HRV and cardiac autonomic function damage. These reports all proved that cardiac autonomic nervous system was adversely related to HRV variation, and indicated the potential power of HRV in CHF detection [51, 77]. All these reports support the application prospect of HRV in the risk assessment of CHF.

As far back as 1996, the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology published standards on statistical analysis of short-/long-term HRV measurements [46]. Many studies about classification between CHF patients and normal person commenced. Different classifier algorithms, and different HRV measures using 5-min/24-h ECG records were applied.

In 2003, Asyali et al. analysed 52 normal person and 22 CHF patients' 24-h RR intervals data with nine time-frequency domain features [90]. Best classification power of standard deviation was found via linear discriminate analysis. Finally, he applied Bayesian classifiers for CHF detection achieving an accuracy of 93.24%. In 2007, Isler et al. utilised classical HRV parameters and wavelet entropy with k-nearest-neighbour (KNN) classifiers for CHF prognosis [68]. Short-term HRV measurement of 29 CHF

patients and 54 healthy samples was extracted and an accuracy of 96.39% was achieved after using the genetic algorithm feature section method. These studies proved that using classical HRV measures and the classical classifier gave a good performance in terms of CHF detection.

Based on former research studies, more innovative HRV measures and classification algorithms were proposed. In 2011, Pecchia et al. applied two additional non-standard measures-δAVNN (average of RR intervals) and δLF/HF (average of LF/HF)-in CHF detection combined with classical HRV measures [61]. A new proposed classification and regression tree (CART) was applied to these features with an extreme search feature selection method and this resulted in an accuracy of 96.4%. In 2012, Yu et al. applied a support vector machine (SVM) classifier and genetic algorithm (GA) into CHF recognition based on bi-spectral HRV analysis [91]. Classification included 29 CHF patients and 54 normal persons' long-term RR intervals data and achieved a detection accuracy of 98.79%. In 2014, Liu et al. extracted 47 5-min RR intervals data sample for time/frequency domain features and three entropies (appropriate entropy, sample entropy, and fuzzy entropy) [47]. Then three combined features were proposed based on these features and this achieved 100% accuracy via the support vector machine. In 2017, Janjarasjitt analysed the spectral-based features of the long-term RR interval data of 18 normal persons and 15 CHF patients using the wavelet-based approach [92]. Linear least-square regression was applied to disease detection with 100% accuracy. In recent years, more complex HRV analysis was proposed and also achieved good performance in CHF detection [93, 94]. Among the results, many of these reports could distinguish CHF patients from normal people with accuracies of more than 90%. All these support that 1)non-classical measures of HRV provide different ways to describe the HRV fluctuation of CHF patients; and 2) new classification methods have great potential to provide better discrimination performance for CHF detection.

Our research for risk assessment of CHF is based on the MIT/BIH database, which is a well-known open access ECG database. Therefore, we mainly focused on works based on the same database for the purposes of comparison. Table 2.1 is a summary of the research mentioned above. All these studies focused on 5-min (short-term) and 24-h (long-term) ECG data using different classifiers and feature selection for performance improvement. Furthermore, these studies which mainly focused on the

overall level condition of autonomic function by classical indices of HRV measurements for disease detection and high performance (accuracy over 90%) were easy to access. This is consistent with the fact that the HRV can discern the autonomic dysfunction of CHF patients from normal function [56]. However, relatively little attention has been paid to assessing the autonomic activity change among CHF patients according to a literature search, i.e. CHF stratification.

Table 2.1 Related work on congestive heart failure (CHF) detection

Timescale	Reference	Samples ∗ Times	Feature Selection	Classifier	Accuracy
Long term	Asyali et al.	74 ∗ 24h	N/A	Bayesian	93.24%
	Yu et al.	83 ∗ 68min	GA	SVM	96.38 %
	Janjarasjitt et al.	33 ∗ 24h	ESM	CART	85.40 %
Short term	Isler et al.	83 ∗ 5min	GA	KNN	96.39 %
	Pecchia et al.	83 ∗ 5min	ESM	CART	96.4%
	Liu et al.	47 ∗ 5min	SA	SVM	100%

N: normal samples; P: CHF patients, in which 1 is of NYHA I-II, 2 is of NYHA III, 3 is of NYHA III-IV; SI: static indices; DI: dynamic indices; GA: genetic algorithm; ESM: exhaustive search method; BE: backward elimination; SVM: support vector machine; KNN: k-nearest neighbour.

In congestive heart failure (CHF) stratification based on NYHA classification, there are only a few papers found (shown in Table 2.2). In 2013, Melillo et al. first tried to assess the severity of CHF disease by using long-term HRV measurements of 44 24-h CHF patients RR intervals [95]. The classification and regression tree (CART) classifier was used to separate lower-risk (NYHA I-II) patients from higher-risk (NYHA III-IV) patients with a relatively low accuracy (i.e., 85.4%). Sensitivity was over 90% while specificity was only under 70%. Though this result might be influenced by a small amount of un-balanced data, the HRV measurement used and classification method might need to be improved. Later, Shahbazi et al. applied generalised discriminant analysis for CHF risk assessment based on long-term HRV measurement to discriminate between lower risk and higher risk patients [96]. 14 linear and 11 nonlinear HRV measures from 10 patients suffering from mild CHF (NYHA classes I and II) and 29 patients suffering from severe CHF (NYHA classes III and IV). and a k-nearest-neighbour (KNN) classifier based feature selection method were used to access the optimal feature subset with the smallest numbers. The least number feature subset achieved a sensitivity and a specificity of 100%, respectively. The leave-one-out

method was repeated for all 39 subjects for cross validation. This report showed a good performance in CHF stratification, though it was limited by the unbalanced and small dataset. Both reports used long-term data, which require more exploration in timescale for the purposes of CHF stratification.

Table 2.2 Related work on congestive heart failure (CHF) stratification

Timescale	Reference	Groups	Samples ⋆ Timescale	Classifier	Feature Selection	Accuracy/%
Long term	Melillo et al.	NYHA I-II vs. NYHA III-IV	44 ⋆ 24-h	CART	N\A	85.4
	Shahbazi et al.	NYHA I-II vs. NYHA III-IV	39 ⋆ 24-h	GDA	KNN	100

NYHA: New York Heart Association classification; CART: classification and regression tree; GDA: generalized discriminant analysis; KNN: k-nearest-neighbour classifier; N/A means not mentioned.

In general, risk assessment of CHF with short/long term HRV measurement has made progress but still needs to be improved. First, CHF detection was widely analysed with high performance. Potential problems might exist (including validation, and manual error during data pre-processing), which can lead to unstable results. Also, CHF quantification had only attracted limited research with unsatisfying performance results. Finally, the present work mainly focused on 5-min/24-h classical HRV measurement. The difficulty is that high quality 24-h RR interval data is hard to access, while 5-min data is easy to access but sensitive to body condition. Thus, more analysis of new measures of HRV needs to be done to better describe the cardiac autonomic nervous system and RR intervals fluctuation among different risk levels under different clinical conditions.

2.6.3 Risk Assessment of CHF with Other ECG Components

As described in Section 1, there are many components in ECG signals to explore the change of beat-to-beat records. There are also many studies using other types of ECG components except for HRV in risk assessment of congestive heart failure (CHF) (as shown in Table 2.3).

In the year of 2004, Dixen et al. explored 74 patients' ECG records for prolonged signal-averaged P wave duration (SAPWD). The correlation between the SAPWD and

Table 2.3 Related work using the other ECG components except heart rate variability (HRV) in risk assessment of congestive heart failure (CHF)

ECG components	Reference	Highlight
P wave	Dixen et al.	a significant correlation between prolonged signal-averaged P wave duration and atrial size parameters
	Proietti et al.	prolonged P wave duration (more than 110 ms) characterizing an abnormal delay arterial activation of interatrial block
PR intervals	Magnani et al.	a significant relationship (46% increased) of PR interval > 200 ms to 46 % increased risk of incident heart failure (95% confidence interval, 1.11-1.93)
QRS complex	Shenkman et al.	a linear relationship between increased QRS duration and decreased ejection fraction ($p<0.01$) during the first year of diagnosis
	Dhingra et al.	the association between the prolonged duration and disease with 23% increased CHF risk
	Madias	voltage of P wave can help as additional information of QRS complexes in fluid overload heart disease
QT intervals	Tereshchenko et al.	abnormally augmented QTV1 separated 97.5% of the health from HF patients
ECG spectrum	Isler et al.	statistical significance ($p<0.05$) of Poincare measures derived from ECG in the diagnosis of CHF patients
	Kamath	sequential spectrum analysis of ECG signals with statistical significance and 100% accuracy

the left atrial diameter (LAD), left atrial volume (LAV), right atrial volume (RAV), and total atrial volume (TAV) was analysed by linear regression analyses. The linear analysis showed a significant correlation between prolonged signal-averaged P wave duration and atrial size parameters (LAD, LAV, RAV and TAV) [97]. This indicated that the prolonged signal-averaged P wave had potential as a factor to show the atrial size change of heart failure patients. Proietti et al. reported a CHF patient showed prolonged P wave duration (more than 110 ms) characterising an abnormal delay arterial activation of interatrial block, which occurs in CHF while poorly investigated at hospital admission [98]. These reports proved that prolonged P wave duration from ECG records is an important factor to indicate the atrial change of CHF patients.

Another important component of ECG records is PR interval. Magnani et al. used multivariable Cox proportional hazards models to 2722 ECG records of participants from the Health, Aging, and Body Composition Study. This analysis identified a significant relationship (46% increased) of PR interval > 200 ms to 46 % increased risk of incident heart failure (95% confidence interval, 1.11-1.93) [99]. This helps to show that PR intervals can be used as an clinical factor in future CHF assessment.

QRS complex is one of the most important components of the ECG record with rich information on heart function fluctuation. Shenkman et al. explored QRS prolongation based on a large heart failure population, finding a linear relationship between increased QRS duration and decreased ejection fraction ($p<0.01$) during the first year of diagnosis [100]. Dhingra et al. analysed prolonged QRS duration in CHF patients and found an association between prolonged duration and disease with 23% increased CHF risk [43]. Madias reported 3 patients and revealed the possibility that voltage of P wave can help as additional information of QRS complexes in fluid overload heart disease [101]. The association between the QRS complex and CHF risk factors revealed the potential for its application in classification between CHF and the healthy person.

Tereshchenko et al. combined beat-to-beat QT variability with heart rate variance in heart failure and claimed that abnormally augmented QTV1 separated 97.5% of the healthy patients from the HF patients [44].

Isler et al. applied Poincare measures derived from ECG in the diagnosis of CHF patients and found statistical significance ($p<0.05$) [102]. Kamath detected CHF using sequential spectrum analysis of ECG signals with statistical significance and 100% accuracy [103].

All these reports give positive evidence of the power of different ECG components in risk assessment of CHF from the mechanism level to classification application. A larger, good quality ECG dataset will help to extend the exploration of noninvasive and comprehensive CHF assessment based on ECG signals.

2.6.4 Risk Assessment of CHF and Sleep Apnea

Risk assessment of congestive heart failure (CHF) using ECG has been introduced in detail in the above subsection. The classification performance of risk assessment of CHF using ECG is over 90% in disease detection. While in sleep apnea detection using heart rate variability, Ravelo et al. (2015) used the permutation entropy of HRV and quadratic discriminant analysis and reached an area under curve (AUC) of 91.7% [64]. Li et al. (2018) used the classic HRV and deep neural network and hidden Markov model to detect sleep apnea, achieving an accuracy of 85% based on 1-min ECG data

[67]. Taken together, these studies show that there is the potential to apply HRV for detection of sleep apnea and CHF. However, currently, there are no autonomic methods for distinguishing between sleep apnea patients with and without CHF.

Chapter 3

Dynamic HRV in Autonomic Unbalance for CHF Stratification

This Chapter described a newly proposed dynamic heart rate variability (HRV) measures in Congestive Heart Faliure (CHF) assessment to better describing dynamic fluctuation of cardiac function in daily life.

3.1 Introduction

Over the past decades, numerous studies have examined variations in electrocardiograph (ECG) recording of heartbeat intervals, namely, heart rate variability (HRV) [104]. HRV has proven to be an important noninvasive indicator of the autonomic nervous system [1] and has been successfully applied in analyzing several cardiovascular diseases, including hypertension [105], diabetes mellitus [106], sudden cardiac death [107], coronary artery disease [108], and heart failure [109].

As a common syndrome and end stage of different cardiovascular diseases, congestive heart failure (CHF) attracts much attention in using HRV analysis for risk assessment [49]. Two types of HRV analysis have been recommended as Guideline: long-term (24-h) and short term (5-min) [46]. However, there is demand for a more flexible analysis to investigate rapid fluctuations of HRV in a day, rather than single 5-min or 24-h analysis [60]. It has been proven that short term HRV measurement can be used

as predictor in congestive heart failure (CHF) detection and autonomic nervous system (ANS) function assessment [48, 110]. However, the main problem is that different 5-min segments cannot show the same power in CHF detection. This might indicated biased data selection existed during research. Furthermore, the circadian rhythm of HRV was found in CHF patients [62]; and even sleep stages also effect HRV [63].These research findings indicate that HRV alters with daily activity, which also reflects autonomic nervous function fluctuated under different physiological activity in 24 hours.

Technically, a majority of researchers have applied three approaches to determining short/long term HRV: time domain analysis, frequency domain analysis, and nonlinear analysis [46]. Time and frequency domain analyses address variability degrees and underlying rhythms, respectively [49], while nonlinear analysis processes the dynamic variation features in HRV generated from physiological systems [50], e.g. complexity. Statistical analysis based on these three approaches of short/long term HRV analysis showed significant difference between samples from non-CHF people and CHF patients. Decreased time and frequency domain HRV measurement was a marker of sympathoexcitation [54, 55]. This sympathoexcitation was often measured clinically. Besides, nonlinear analysis provides an improved decomposition of cycle flucturation of HRV in 24-h and achieves significant discrimination power between normal samples and patient samples [59]. However, it is still difficult to find robust quantifying measure for autonomic unbalance estimation and risk stratification of CHF patients.

In this study, we hypothesise that autonomic modulation balance changes at different short cycle in 24 hours. 5 minute period is a steady cardiac cycle to explore HRV variation in daily life [46]. 10 widely used HRV measures are applied for autonomic function evaluation of 5-min segments and 24-h data. Then, dynamic indices are calculated to describe ANS fluctuation from two perspectives: statistical level and complexity. Statistical level dynamic indices consist of mean and standard deviation of each 5-min HRV measures reflecting variation degree, while complexity dynamic indices consist of fuzzy entropy of each 5-min HRV measures, estimating cardiac autonomic system regulation. These dynamic indices are intended to show more flexible fluctuation in RR intervals and cardiac autonomic tone change. Finally, 40 indices are statistically analyzed among NR, LR and HR in search of risk quantification indices. The underlying

mechanism between risk quantification indices and cardiac autonomic modulation will be discussed for potential clinical usage.

3.2 Methods

3.2.1 Samples Data

The data used in this work was from the widely-used MIT/BIH database in PhysioNet [111]. The study associated with this database was approved by the Institutional Review Boards of Beth Israel Deaconess Medical Center (Boston, MA) and the Massachusetts Institute of Technology (Cambridge, MA).

We obtained nominal 24-h RR interval record of 116 subjects from four RR interval databases: (1) 72 normal samples (aged 20 to 76); (2) and 44 CHF patient samples (aged 22 to 79). The data of normal people came from two databases: the MIT/BIH Normal Sinus Rhythm Database and the Normal Sinus Rhythm RR Interval Database [111]. The data of the CHF patients came from the Congestive Heart Failure RR Interval Database and BIDMC Congestive Heart Failure Database [112]. These records were all manually reviewed and corrected by experts.

We regrouped the 116 subjects into three groups according to clinical severity:

- NR: normal people labeled as no risk (n=72)
- LR: patients in NYHA I-II labeled as mild risk (n=12)
- HR: patients in NYHA III-IV labeled as moderate risk (n=32)

The subjects' gender information was partially abridged, so there was no description about gender. All these ECG data and subject information can be downloaded from http://www.physionet.org/cgi-bin/atm/ATM [111].

3.2.2 Data Analysis

After removing the leading and trailing value of each recording, we excluded RR intervals longer than 3 seconds [1] These steps were performed in case of unstable measurement conditions and artificial error. Then, we divided the 24-h data into multiple 5-min segments in sequence. Dynamic indices and classical indices of 24-h data and 5-min segments were then calculated to quantitatively evaluate cardiovascular autonomic function.

Here, 10 classical HRV measures were calculated with 5-min segments and 24-h data, including: (1) time domain features: mean value of RR intervals (MEAN); standard deviation of RR intervals (SDNN); root mean square of the successive RR intervals (RMSSD) [110]; coefficient variation of RR intervals (CVRR) [113]; (2) frequency features: low frequency (0.04-0.15Hz) power of RR intervals (LF); high frequency (0.15-0.4Hz) power of RR intervals (HF); ratio of LF to HF (LH) [47]; (3) entropy features: appropriate entropy of RR intervals (AE); sample entropy of RR intervals (SE) [47]; fuzzy entropy of RR intervals (FE) [114]. Fast Fourier transform - based power spectral density measurement was calculated for different spectral components [46]. All these measurements were described in Ref. [45].

In this study, 24-h data's 10 classical HRV measures were obtained to analyse the overall condition of different groups in 24 hours. 10 classical HRV measures of 5-min segments were also analyzed to obtain relevant dynamic indices. Dynamic indices have two perspective: statistical level and complexity to track dynamic change trend of subject in 24 hours. Statistical level dynamic indices include two kind: mean and standard deviation (SD) of each classical measures of 5-min segments. This two kind of statistical level indices measure the central tendency of RR intervals and evaluate ANS function change trend.

Complexity dynamic indices are fuzzy entropy (FE) of each classical measures of 5-min segments FE, which was applied to measure system complexity [115] and the degree to which a sample belongs to a given group [116]. Furthermore, FE can track qualitative changes in vector and estimates the general behaviour of a vector in noisy background. Hence, complexity dynamic indices were calculated to help track changes of 5-min segments in 24-h.

All of the calculations were done using the software MATLAB 7.11.0 (version R2010b, The MathWorks, Inc., Natick, MA, USA).

3.2.3 Statistical Analysis

All measurements are presented as mean±SD (see Table 3.1 and 3.2). Differences between groups are assessed using one-way analysis of variance (one-way ANOVA). Statistical tests are conducted with the significant level set at 0.05. A p value <0.05 is considered statistically significant. All the significance analysis are done using the SPSS software (version 19, SPSS Inc., Chicago, IL, USA).

3.3 Results

The difference among groups of these indices was shown in Table 3.1, 3.2, 3.3 and 3.4, as well as the indices values (mean±SD). Significant difference ($p<0.05$) showed in most indices among three groups (30 out of 40). 15 out of 40 indices showed a stepwise decrease or increase change along with severity. Furthermore, a clear monotonic decrease appeared in mean of ratio of low frequency power to high frequency (MLH), standard deviation of mean (SDM) and fuzzy entropy of ratio of low frequency power to high frequency (FELH) in accordance with severity, as well as remarkable significance value among groups. Thus, a good differentiation power for all groups with the three indices from each other was observed according to results.

3.3.1 Analysis among Groups

Decreased trend of HRV existed in all indices of 24-h, described in Table 3.1. Six out of ten 24-h indices showed significance ($p<0.05$) among groups, and stepwise decrease in proportion to severity showed in SD, RMSSD, CVRR, LF, HF, LH, SE and FE. There was no 24-h indices which showed a significant difference between LR vs. HR. Time-domain indices of 24-h RR data showed remarkable difference (p value lower or close to 0.001) among groups, while frequency-domain indices reflects no difference($p>0.05$). entropy indices had different results as p value of appropriate

entropy is above 0.05 and p value of sample entropy and fuzzy entropy are lower than 0.05.

Table 3.1 Analysis result of classical heart rate variability (HRV) features of 24-h data according to severity

Feature	NR(n=72)	LR(n=12)	HR(n=32)	p value		
				NR-LR	LR-HR	NR-HR
MEAN	0.8023±0.0821	0.6786±0.0860	0.6900±0.1056	**<0.001**	0.710	**<0.001**
SDNN	0.4695±0.5985	0.1143±0.0544	0.0798±0.0297	**0.020**	0.830	**<0.001**
RMSSD	0.5857±0.8851	0.0958±0.0557	0.0749±0.0513	**0.030**	0.930	**<0.001**
CVRR	0.5804±0.7584	0.1653±0.0692	0.1160±0.0430	**0.030**	0.810	**<0.001**
LF	0.1002±0.2655	0.0015±0.0017	0.0006±0.0007	0.140	0.990	**0.030**
HF	0.1329±0.4389	0.0012±0.0014	0.0005±0.0007	0.230	1.000	0.080
LH	1.3813±0.7653	1.2929±0.3617	1.2211±0.6469	0.690	0.760	0.290
AE	-0.1605±0.0829	-0.1373±0.0275	-0.1401±0.0421	0.290	0.910	0.170
SE	0.0273±0.0178	0.0125±0.0063	0.0087±0.0048	**<0.001**	0.440	**<0.001**
FE	0.2524±0.0443	0.2368±0.0685	0.2221±0.0628	0.340	0.410	**0.010**

Values are mean±SD. NR: no risk normal samples; LR: CHF patients with NYHA I-II; HR: CHF patients with NYHA III-IV.

24 out of 30 dynamic indices showed significant difference among three risk level groups; 10 of them were extremely significant (p<0.001), as shown in Table 3.2, 3.3 and 3.4. There was 11 dynamic indices which showed a significant difference between LR vs. HR. Furthermore, three quantifying indices were found, including mean of LH, SD of mean and FE of LH. Statistical level indices represented a unstable change trend among groups, without regular direction. Mean and SD of the ten indices showed similar fluctuations. Most indices showed good power in discriminating NR, LR and HR (16 out of 20, shown in Table 3.2 and 3.3). In particular, 8 statistical indices showed extremely significant difference among groups. Two indices showed monotonic decrease along with severity increasing with extremely significant difference, i.e. Mean of LH and SD of Mean. Analysis result in N vs. LR was clearly significant, as p value of mean of LH was 0.009 and SD of mean was lower than 0.001. In the analysis of LR vs. HR, the p value of Mean of LH was 0.031 and SD of Mean was 0.009 (see Fig.3.1 and Fig.3.2).

8 out of 10 complexity indices showed good differentiate power in significance analysis (p<0.05), in which half was lower than 0.001 (Table 3.4). Only one complexity index - fuzzy entropy of LH showed clear monotonic difference with p<0.001. The

Table 3.2 Analysis result of dynamic heart rate variability (HRV) features - mean and standard deviation (SD) of classical features of 5-min segments in 24-h according to severity

Mean feature	NR(n=72)	LR(n=12)	HR(n=32)	*p* value		
				NR-LR	LR-HR	NR-HR
MEAN	0.8102±0.0812	0.6901±0.0941	0.6944±0.1065	**<0.001**	0.890	**<0.001**
SDNN	0.0579±0.0187	0.0680±0.0385	0.0489±0.027	0.180	**0.020**	0.080
RMSSD	0.0399±0.0250	0.0811±0.0530	0.0642±0.0469	**<0.001**	0.160	**<0.001**
CVRR	0.0716±0.0195	0.0985±0.0538	0.0702±0.0408	**0.010**	**0.010**	**0.840**
LF	0.0011±0.0008	0.0015±0.0017	0.0006±0.0007	0.150	**<0.001**	**0.010**
HF	0.0006±0.0006	0.0012±0.0015	0.0005±0.0007	**0.010**	**0.010**	0.920
LH	3.2308±1.0343	2.3442±1.3761	1.5496±1.0439	**0.010**	**0.030**	**<0.001**
AE	-0.1851±0.0220	-0.2357±0.0465	-0.2503±0.0742	**<0.001**	0.340	**<0.001**
SE	0.4017±0.0676	0.2832±0.0940	0.2947±0.1153	**<0.001**	0.690	**<0.001**
FE	0.0283±0.0223	0.0444±0.0413	0.0341±0.0334	0.070	0.280	0.330

SD: standard deviation; values are mean±SD. NR: no risk normal samples; LR: CHF patients with NYHA I-II; HR: CHF

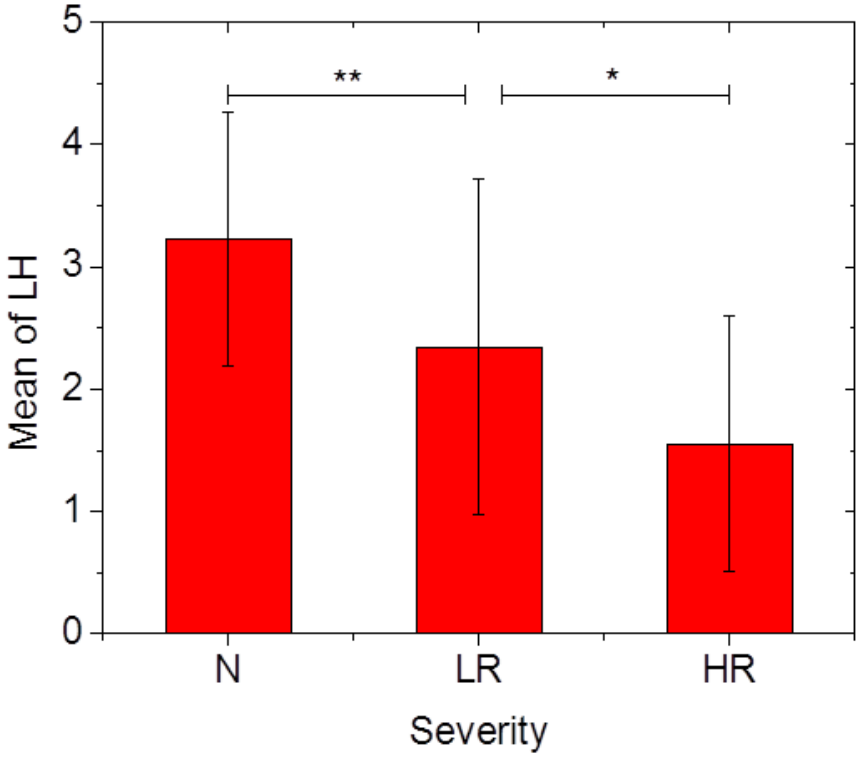

Fig. 3.1 Difference in mean of low to high frequency (LH) power among different groups according to severity. N: normal samples; LR: CHF patients with NYHA I-II; HR: CHF patients with NYHA I-II. *, ** and *** represent $p<0.05$, $p<0.01$ and $p<0.001$, respectively.

Table 3.3 Analysis result of dynamic heart rate variability (HRV) features - mean and standard deviation (SD) of classical features of 5-min segments in 24-h according to severity

SD feature	NR(n=72)	LR(n=12)	HR(n=32)	*p* value		
				NR-LR	LR-HR	NR-HR
MEAN	0.1257±0.0337	0.0818±0.0501	0.0518±0.0224	**<0.001**	0.010	**<0.001**
SDNN	0.0267±0.0110	0.0345±0.018	0.0263±0.0183 0.080	0.090	0.890	
RMSSD	0.0243±0.0206	0.0477±0.023	0.0364±0.0273	**<0.001**	0.150	0.010
CVRR	0.0316±0.0117	0.0505±0.0325	0.0372±0.0279	**<0.001**	0.050	0.190
LF	0.0011±0.0008	0.002±0.0022	0.0011±0.0013	**0.020**	**0.020**	0.920
HF	0.0007±0.0007	0.0017±0.0019	0.001±0.0014	**0.010**	0.100	0.130
LH	1.5494±0.6280	1.5043±1.1245	0.8225±0.692	0.840	**0.010**	**<0.001**
AE	0.0717±0.0254	0.1360±0.0474	0.1451±0.0457	**<0.001**	0.440	**<0.001**
SE	0.1936±0.0384	0.1658±0.0515	0.1985±0.0761	0.090	0.070	0.670
FE	0.0201±0.0195	0.0310±0.0214	0.0246±0.0238	0.100	0.360	0.320

SD: standard deviation; values are mean±SD. NR: no risk normal samples; LR: CHF patients with NYHA I-II; HR: CHF

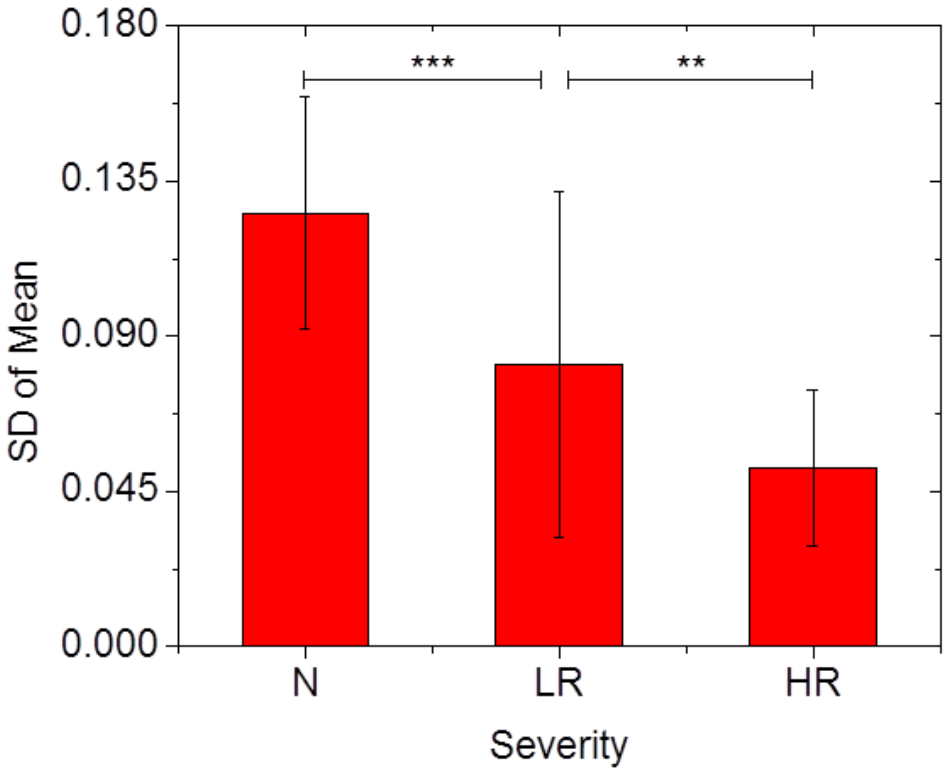

Fig. 3.2 Difference in stabdard deviation (SD) of mean among different groups according to severity. N: normal samples; LR: CHF patients with NYHA I-II; HR: CHF patients with NYHA I-II. *, ** and *** represent $p<0.05$, $p<0.01$ and $p<0.001$, respectively.

fuzzy entropy of LH reduced along with clinical scales, and significant difference existed in N vs. LR and LR vs, HR (p value are 0.007 and 0.028, respectively; see Fig.3.3).

Table 3.4 Analysis result of dynamic heart rate variability (HRV) features - fuzzy entropy of classical features of 5-min segments in 24-h according to severity

FE feature	NR(n=72)	LR(n=12)	HR(n=32)	p value		
				NR-LR	LR-HR	NR-HR
MEAN	0.0261±0.0087	0.0223±0.0136	0.0173±0.0091	0.200	0.120	**<0.001**
SDNN	0.0296±0.0092	0.0323±0.0178	0.0258±0.0177	0.500	0.140	0.180
RMSSD	0.0203±0.0155	0.0408±0.0185	0.0332±0.0234	**<0.001**	0.230	**<0.001**
CVRR	0.0398±0.0107	0.0474±0.0244	0.0368±0.0236	0.150	0.060	0.410
LF	0.0010±0.0009	0.0022±0.003	0.0014±0.0018	**0.020**	0.130	0.280
HF	0.0006±0.0007	0.0020±0.0027	0.0014±0.0019	**<0.001**	0.190	**0.020**
LH	0.6414±0.1704	0.4779±0.3034	0.3327±0.189	**0.010**	**0.030**	**<0.001**
AE	0.0584±0.0267	0.1007±0.0499	0.1051±0.0462	**<0.001**	0.710	**<0.001**
SE	0.1625±0.0438	0.1295±0.0524	0.1311±0.0568	**0.030**	0.920	**<0.001**
FE	0.0110±0.0072	0.0180±0.0117	0.0127±0.0077	**0.010**	**0.050**	0.310

SD: standard deviation; values are mean±SD. NR: no risk normal samples; LR: CHF patients with NYHA I-II; HR: CHF

3.3.2 Subgroup Analysis

Though the significant monotonic change was found among groups for these indices, group analysis among these indices showed a different situation. In the discrimination power of N from LR, more than half indices (23 out of 40) showed good ability (p<0.05 at least), while only nine out of 40 indices can differentiate LR from HR. Furthermore, no significant index was found in differentiating LR from HR in 24-h indices.

3.4 Conclusion

In this study, we aim to explore risk stratification of samples from CHF and normal people. A total of 116 samples were analysed, including 72 no risk (NR), 12 low risk (LR) and 32 high risk (HR) samples. The results of this study showed that dynamic indices found in this study can be considered as a marker in risk stratification of CHF patients with the following characteristics:

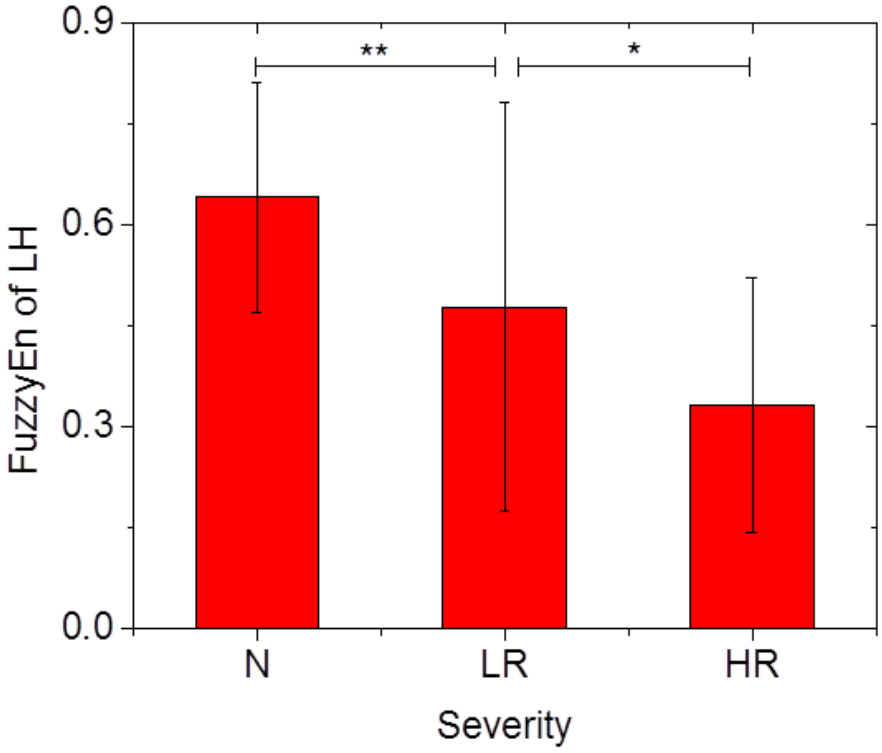

Fig. 3.3 Difference in fuzzy entropy (FuzzyEn) of low to high frequency (LH) power among different groups according to severity. N: normal samples; LR: CHF patients with NYHA I-II; HR: CHF patients with NYHA I-II. *, ** and *** represent $p<0.05$, $p<0.01$ and $p<0.001$, respectively.

(1) Three dynamic indices (MLH, SDM and FELH) had a monotonic and significant different decrease along with severity, which indicated good discrimination power in quantifying three risk groups;

(2) Dynamic indices showed good power in differentiating LR and HR, while classical 24h indices cannot;

(3) Three quarters indices showed clear difference among three risk levels.

Time and frequency domain analyses of HRV have been calculated to evaluate cardiovascular autonomic tone of CHF patients in literature [104]. Significant differences between healthy to patients have been proven [66]. However, the time domain analysis is unable to include dynamic structral information to fully delineate time-space varying property [59]. Furthermore, the limitation of frequency domain analysis was found for estimation of cardiac sympathetic activity in heart failure [117]. Thus, analysis of time/frequency domain can not properly describe dynamic variation and intrinsic fluctuation in HRV. On the other hand, we should pay attention to the risk stratification according to NYHA classification scales [56]. Significant correlation between standard HRV measures and NYHA was analysed on human, as well as autonomic modulation analysis [118]. According to our research, most papers have shown decreased HRV is associated with heart failure [56], stating an analysis on HRV with robust and markable result in 4-level risk stratification of CHF patients has not been found.

It is widely accepted that HRV analysis is a powerful tool in CHF prognostic and cardiovascular ANS function evaluation. Decreased HRV is a result of altered cardiac autonomic regulation [119]. This study hypothesises that relevant indices could differentiate three risk level groups with decreased/increased trend and aimed at investigating the autonomic mechanism among groups.

Forty indices were analyzed in this study for HRV measurement deviation in 24 hours. In daily life, a person' s physiological situation changed along with physical activity and environment, which might influence its cardiovascular function. To evaluate these change, these deviation were measured in statistical level and complexity aspects (Subsection 3.2.2). According to the result, three indices showed the ability in assessing risk level for patients among these indices. These may connect with cardiovascular

autonomic function impairment along with disease. CHF is a continuous disease towards deterioration, and a decreased ANS function is observed [56]. Functional differences among groups in our paper were clear for three indices which found significant difference existed among each two-groups.

24-h indices:The decreased HRV trend concurs with previous research. All time-domain measurements and most entropy feature showed good performance in disease detection (i.e. NR vs. LR, NR vs. HR). But no 24-h indices showed power in differentiating LR and HR. This concurred with previous literature review that robust quantifying features were difficult to find.

Statistical level indices: Fluctuations of 10 classical indices of 5-min segments in 24 hour were expressed in mean and SD value (Table 3.2 and Table 3.3). LH reflected interactions of both types of autonomic modulation [120], and dynamic indices of LH marks sympathovagal unbalance levels. Thus, decreased LH dynamic measures (i.e. mean of LH) represented a decreased sympathovagal balance. Standard deviation of mean was a common used measurement, i.e. SDANN, as SDANN is an estimate of the changes in heart rate due to cycles longer than 5-min. decreased monotonic change at different risk level related to adverse heart rate change and risk increase.

Complexity indices: The fuzzy entropy of 10 standard and meaningful short-term HRV measures, which reflect cardiovascular system autonomic balance in different aspects [56]. Fuzzy entropy can be used to assess the temporal regularity of short-term HRV series with its anti-interference performance. Thus, complexity indices could be used to explain cardiovascular system dynamic regularity in cardiovascular autonomic regulation. Here, monotonic change of complexity indices manifested that ANS regulation change clearly existed among groups, which could be a potential indicator for CHF quantification in clinical studies.

In order to evaluate risk stratification power of the three indices, discriminate analysis were done using SPSS software. The results showed that 92 out of total 116 samples were correctly identified, including 60 NR, 6 LR and 26 HR samples. However, 12 NR samples were misclassified into LR; 4 LR samples were misclassified as HR; and 4 HR were misclassified as LR.Therefore, the total accuracy was 79.31%, and

the precisions of NR, LR and HR were 83.33%, 50% and 81.25%, respectively. These results enhanced the three dynamic indices' risk stratification potential.

In conclusion, three dynamic indices with significant between-group difference and monotonic decrease were found, which suggested that the dynamic indices studied here could be useful for risk stratification and quantitatively evaluating ANS function. They might potentially be applied in clinic as a useful complement to clinical scales.

Limits: Our data from MIT/BIH database lack of some information about samples, for examples, age, gender and body weight, etc. this might limit our findings in a smaller scale without discussion about them. Also, the small number of 116 samples could be a limitation. A bigger data collection could be needed. In addition, the difference among patients with CHF was not as strong as we anticipated. In the future, a deeper research should be done to reach better understanding. Finally, previous studies claimed that entropy could be limited by sensitivity to both N and r [121]. Its influence was not discussed in this study.

Chapter 4

Multilevel CHF Detection and Quantification using Dynamic HRV Measurement

This Chapter described a novel method for 4-level risk assessment of CHF detection and quantification using dynamic HRV measurement and decision tree based classification method, to achieve a more robust and accuracy home monitoring .

4.1 Introduction

congestive heart failure (CHF) is a common chronic cardiovascular syndrome along with autonomic nervous system (ANS) abnormality of the heart [1]. Patients experience no obvious symptoms during its early stages. Once diagnosed, physicians still cannot provide convenient suitable medical care based on prognosis according to the patient's physical condition. Furthermore, poor prognosis results in 30-40 % of diagnosed patients dying in a year [8]. Thus, risk assessment of CHF is essential for saving lives and money. The severity of CHF has a well-known measurement, namely, the symptomatic classification scale of the New York Heart Association (NYHA) [28], which has proved to be a very useful factor for risk assessment of CHF patients [29].

According to the NYHA classification, the severity scale of heart failure depends on the severity of symptoms [31], which are partly modulated by the autonomic nervous system. Heart rate variability (HRV) analysis has been confirmed as a reliable and noninvasive tool in the prognosis and risk assessment of CHF, and it is widely used to assess the influence of the ANS on the heart [66]. HRV measurements (time/frequency domain and non-linear) of 5-minute/24-hour (5-min/24-h) data have already been studied in statistic difference levels between normal people and CHF patients [51–53]. Measurements of adverse changes in the autonomic function of CHF manifest in altered HRV analysis [56]. In this study, we redefined short-/long-term (i.e., 5-min/24-h) HRV measurements as static indices (SI) to assess the autonomic function of the recording.

As far back as 1996, the Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology published standards on statistical analysis of short-/long-term HRV measurements [66]. In 2003, Asyali et al. applied Bayesian classifiers to classical time/frequency HRV parameters of long-term measurements for CHF discrimination with an accuracy of 93.24 % [90]. In 2007, Isler et al. utilized wavelet entropy and classical HRV parameters with k-nearest-neighbor (KNN) classifiers for CHF diagnosis and achieved an accuracy of 96.39 % [68]. In 2011, Pecchia et al. applied two additional non-standard measures-$\triangle$AVNN (average of RR intervals) and ?LF/HF (average of LF/HF)-in CHF detection with an accuracy of 96.4 % [48]. In 2012, Yu et al. applied a support vector machine (SVM) classifier and genetic algorithm (GA) into CHF recognition based on bi-spectral HRV analysis and achieved an accuracy of 98.79 % [91]. These studies mainly focused on the overall level condition of autonomic function by static indices of HRV measurements for disease detection; however, relatively little attention have been paid to assessing the autonomic activity change among CHF patients. Among the results, many of these reports could distinguish CHF patients from normal people with accuracies of more than 95 %. This is consistent with the fact that the redefined SI can discern the autonomic dysfunction of CHF patients from normal function [56].

By 2013, Melillo et al. first tried to assess the severity of CHF disease by using long-term HRV measurements. The classification and regression tree (CART) classifier was used to separate lower-risk patients from higher-risk patients with a relatively low accuracy (i.e., 85.4 %) [95]. Two reasons may explain this result. First, the performance

of the classifier needed to be improved. Second, the static HRV measurement might not fully quantify trend changes in the autonomic activity of CHF patients during different daily activity [56]. Thus, we proposed a new measurement of HRV-dynamic indices (DI)-for the stratifying estimate. DI reflects the dynamic of 5-min segments' HRV measurement in 24 hours, described in HRV measurement Part. The functional class of CHF patients tends to deteriorate unevenly over time and this indices can demonstrated this fluctuation with low individual difference [31].

In our present study, we developed a multistage CHF risk assessment model. The work presented in this study involves the following contributions:

1 We creatively establish a four-level risk assessment model for CHF detection and quantification, including no risk (normal people, N), mild risk (patients with NYHA I-II, P1), moderate risk (patients with NYHA III, P2), and severe risk (patients with NYHA III-IV, P3).

2) We extract dynamic HRV measurements to improve the precision of the model, especially in disease quantification. These dynamic HRV measures better reflect the autonomic function change during different daily activity for individuals with CHF.

3) We apply the decision-tree-based support vector machine (DT-SVM) classifier to take advantage of SI and DI in CHF detection and quantification, respectively. We improve the performance of the classifier by integrating backward elimination (BE) with significance difference.

4.2 Multilevel Risk Assessment with 24-h RR Data

Fig.4.1 presents a flowchart of the entire work.

4.2.1 Data

We obtained the data used in this work from the widely-used MIT/BIH database in PhysioNet [111]. All subjects provided informed written consent. The study was

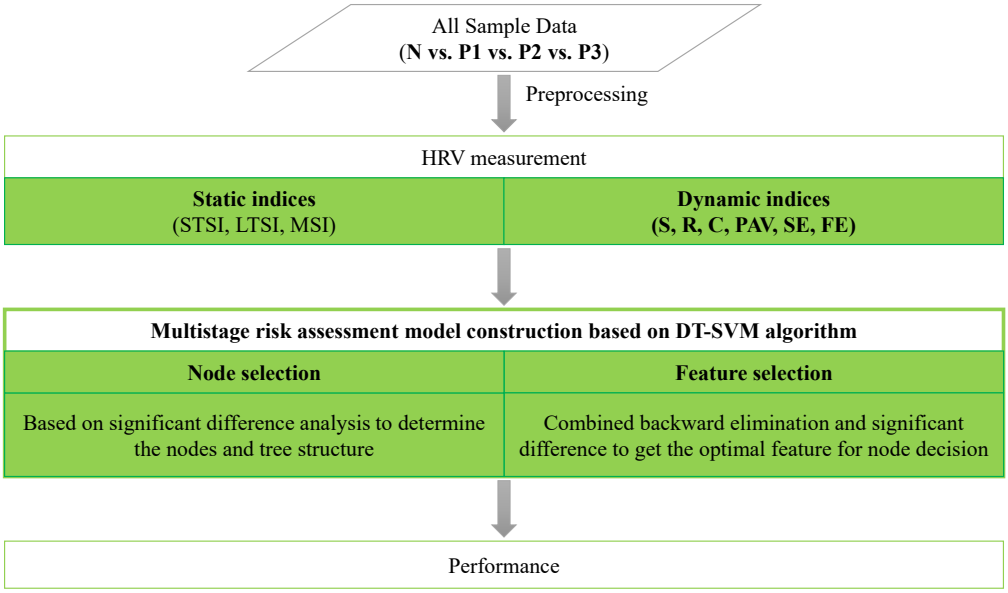

Fig. 4.1 Flowchart of entire work. N: normal people; P: CHF patients, in which 1 is of New York Heart Association (NYHA) I-II, 2 is of NYHA III, 3 is of NYHA III-IV; S1: basic measures of 24-h RR interval data, which reflect long-term data variation); S2: basic measures of the second 5-min segment, which representing a stable measurement condition of short-term data; S3: mid-value of basic measures of 5-min segments, which showing an intermediate state of short-term data; D1: mean value of basic measures of 5-min segments, for robustness improvement; D2: standard deviation of each basic measure of 5-min segments; D3: root mean square of each basic measure of 5-min segments; D4: coefficient variation of each basic measure of 5-min segments; D5: percentage of abnormal value (value intervening M±S) of each basic measure of 5-min segments; D6: sample entropy of each basic measure of 5-min segments; D7: fuzzy entropy of each basic measure of 5-min segments; DT-SVM: decision tree based support vector machine.

approved by the Institutional Review Boards of Beth Israel Deaconess Medical Center (Boston, MA) and the Massachusetts Institute of Technology (Cambridge, MA). We chose four RR interval databases, obtaining 116 nominal 24-h RR interval records: 72 normal person samples (N, aged 20 to 76) and 44 CHF patient samples (P, aged 22 to 79). The data of normal people came from two databases: the MIT/BIH Normal Sinus Rhythm Database and the Normal Sinus Rhythm RR Interval Database [66]. The data of the CHF patients came from the Congestive Heart Failure RR Interval Database and BIDMC Congestive Heart Failure Database [112]. These records were all manually reviewed and corrected by experts.

All samples were grouped into four stages according to severity: s

- 72 normal people labeled as no risk (N, aged 54.62±16.03 years);
- 12 patients in NYHA I-II labeled as mild risk (P1, aged 52.5±14.25 years);

- 17 patients in NYHA III labeled as moderate risk (P2, aged 57.24±9.28 years);
- 15 patients in NYHA III-IV labeled as severe risk (P3, aged56±11.50 years with one sample unknown).

Patients in group P3 were receiving medical therapy; therefore, we considered it as a special type with a higher risk of mortality different from groups P1 and P2. The subjects' gender information was partially abridged, so there was no description of gender. All these data can be downloaded online from http://www.physionet.org/cgi-bin/atm/ATM [66] for free.

Before feature extraction, we preprocessed all these data:

1) deleting the first and the last RR interval;
2) excluding RR intervals longer than 3 seconds [66];
3) dividing the 24-h data into multiple 5-min segments saved in a sequence.

The first two steps were performed in case of unstable measurement conditions and artificial error. The third step was for feature extraction.

4.2.2 HRV Measurement

In this study, dynamic and static indices of HRV measurement were analyzed from 116 pre-processed RR interval data, both 24-h and 5-min segment RR intervals. This analysis processing included two steps: classical HRV measurement calculation and our HRV measurement calculation.

1) Classical HRV measurement calculation: After pre-processing, we had two types of data: nominal 24-h RR interval records and 5-min segment RR interval data. With these two types of data, we calculated 18 classical HRV measurements, which included:

Time Domain (T1∼T5): average of RR intervals (T1); standard deviation of RR intervals (T2); root mean square of successive RR interval difference (T3); percentage

of successive RR interval difference larger than 50ms (T4) [66]; coefficient variation (ratio of T2 to T1) of RR intervals (T5) [113];

Frequency Domain (F1∼F4): power of RR intervals in 0.04-0.15 Hz (F1); power of RR intervals in 0.15-0.4 Hz (F2); ratio of F1 to F2 (F3); total power (F4) [66];

Nonlinear (E1∼E9): low frequency wavelet entropy (E1); high frequency wavelet entropy (E2); normalized low frequency wavelet entropy (E3); normalized high frequency wavelet entropy (E4); ratio of E1 to E2 (E5); total power wavelet entropy (E6) [68]; approximate entropy (E7); sample entropy (E8) [47]; fuzzy entropy (E9) [114].

The time/frequency domain HRV measurements in our work followed International Guidelines [66], and frequency domain HRV measurement was calculated based on Fast Fourier Transform. The nonlinear HRV measurements are fully introduced in Chapter 4 Section 4.2 Subsection 4.2.2 [68, 47, 114]. The 18 Classical HRV measurement were calculated for both 24-h and 5-min segment RR intervals as ***basic measures (T1∼E9)***.

2) Our HRV measurement calculation: Based on the classical measurements, two types of HRV measurement SI and DI were calculated, as we defined hereinafter.

The SI was calculated from HRV measurements of data in a period (5-min/24-h). This series of indices demonstrated the global or average level of cardiovascular autonomic activity, composed of four series with a total of 54 (=18 basic measures*3 series) indices:

S1: basic measures of 24-h RR interval data, which reflect long-term data variation (S1T1∼S1E9);

S2: basic measures of the second 5-min segment, which representing a stable measurement condition of short-term data (S2T1∼S2E9);

S3: mid-value of basic measures of 5-min segments, which shows an intermediate state of short-term data (S3T1∼S3E9).

In contrast, the DI was calculated from 5-min segment' basic measures in the nominal 24-h data to evaluate the dynamic changes of symptoms and autonomic function during different activities. Here, we analyzed each basic measure of 5-min

segments from 6 aspects; thus, DI has six series with a total of 126 (=18 basic measures $*$ 7 series) indices:

D1: mean value of basic measures of 5-min segments, for robustness improvement (S4T1∼S4E9).

D2: standard deviation of each basic measure of 5-min segments (D1T1∼D1E9);

D3: root mean square of each basic measure of 5-min segments (D2T1∼D2E9);

D4: coefficient variation of each basic measure of 5-min segments (D3T1∼D3E9);

D5: percentage of abnormal value (value intervening M±S) of each basic measure of 5-min segments (D4T1∼D4E9);

D6: sample entropy of each basic measure of 5-min segments (D5T1∼D5E9);

D7: fuzzy entropy of each basic measure of 5-min segments (D6T1∼D6E9).

Thus, 180 HRV measures, comprising 54 SI and 126 DI, were extracted from 72 normal person samples and 44 CHF patient samples.

4.2.3 DT-SVM Algorithm based Multistage Risk Assessment Model Construction

In our work, we constructed a multistage risk assessment model for CHF detection and quantification (shown in Fig.4.3) based on the DT-SVM algorithm.

DT-SVM is an effective way of combining an SVM and a decision tree for solving multi-class problems [82]. It is a modified method of the classical SVM for dealing with its difficulty in multi-class problems, but DT-SVM brings another danger: cumulative error. This error is caused by sample misjudgment at the upper node of the decision tree and lasts throughout the rest of the classifier without elimination. The main idea of this algorithm is the conversion of multiclass classification problems into multilevel binary classification problems. Each level includes two nodes to be classified, and each node includes one or several classes. At every node, a decision is made to assign the samples. This step is repeated until all the samples reach a leaf node, to which only one class of samples is assigned. In this way, a hierarchy is formed [83].

In our research, DT-SVM was applied to establish the tree-structure of the risk assessment classifier. Performance of the classifier was determined by the tree structure and node decision [83], which depended on the nodes and input features selection.

1) Node Selection: Usually, there are two kinds of tree structures: balanced or unbalanced tree architectures [83]. Furthermore, the most separable classes should be separated at the upper nodes [83]. In this study, one-way analysis of variance (one-way ANOVA) was used to calculate the significant difference of indices among two-groups [113] as separability measurement [83] on SPSS software (version 19, SPSS Inc., Chicago, IL, USA). The rule of node selection at each level is that a larger number of features in a smaller p value scale indicate a higher separability for binary separation. The corresponding pair of groups was used as the suitable nodes for the level.

In our work, the samples were divided into four stages, corresponding to a four-leaf-node tree. The nodes for a level were selected as follows: First, we denote the feature set as the data set $X(i) : i = 1, 2, ..., N$ and N as the feature number. We then define the significant differences of $X(i)$ as $P(i)$, which are calculated by one-way ANOVA between all possible two-groups, where

$$P(i) = p(i, k), k = 1, 2, ..., M \tag{4.1}$$

where $p(i, k)$ represents the p value of the ith feature for the kth pair. In this study, M is 7 for level 1 (the seven pairs are shown in Table 4.3).

Then, we define the number of features in the particular significance value range as *count*:

$$count(k) = \sum_{i=1}^{n} d(i, k) \tag{4.2}$$

where d is the sign function discrimination matrix:

$$d(i, k) = \begin{cases} 1, & if \quad p(i, k) \leq \varepsilon \\ 0, & if \quad p(i, k) > \varepsilon \end{cases} \tag{4.3}$$

and ε is the particular significance value range. In this study, ε is initially 0.001.

Finally, we define *num* as the maximum value of *count*:

$$num = max(count) \tag{4.4}$$

The pair corresponding to *num* yields the selected nodes for the level. If the maximum value is associated with more than one pair, we repeat this procedure, sequentially changing the value of ε to 0.01, 0.05, and 0.1, until only one pair is determined for the level.

We iterated these steps for the remaining levels until each group consisted of only one class; thus, the tree structure was determined (shown in Fig.4.2).

2) Feature selection: Owing to their high correlation, directly using all features for classification might not give the best performance [78]; thus, using an appropriate selection method for feature subset discrimination improves classifier performance. We applied BE [79] into feature selection at each node (shown in Fig.4.2).

First, we performed feature pre-screening to remove invalid characters and improve algorithm performance. The significance level was computed among three pairs (shown in Fig.4.2). Considering the physiological rule, features with a high significance level (i.e., p>0.1) were rejected; additionally, SI was used for disease detection and DI for quantification.

Then, we applied backward feature selection to the pre-screened features (shown in Fig.4.2). The BE algorithm begins with all features and iteratively removes them one-by-one until the remaining features reach the highest precision. The feature selection was based on the following iteration below.

We denote $\{X(i,j) : i = 1, 2, ..., M; j = 1, 2, ..., N\}$ as the feature matrix, in which M and N are the numbers of samples and features, respectively. The samples are labeled by $y_i \in \{-1, 1\}$. We define r as the rate of the training set and testing set and then select the training and testing set from $X(i,j)$ randomly in proportion. We define the line numbers of training and testing set in the feature matrix as tr and te, respectively.

We define $Y^m = X(i,k), i = 1, 2, ..., M; m \in \{0, 1, 2, ..., N\}$ as submatrices of the feature matrix, where m represents the mth feature deleted from the feature matrix.

Thus, $N+1$ submatrices are formed, where k is the volume number of remaining features. Here, Y^0 is the feature matrix without deletion.

Thus, the training and testing sets are defined as $Training^m$ and $Testing^m$:

$$Training^m = Y^m(tr, k), \quad \forall \quad tr \in \{1, 2, ..., M\} \tag{4.5}$$

$$Testing^m = Y^m(te, k), \quad \forall \quad te \in \{1, 2, ..., M\} \tag{4.6}$$

We input the training set into the SVM classifier and validate with the testing set. Here, we regard the SVM as a black box to score different feature combinations according to their predictive power [122]. We define the accuracy of the testing subset as ACC^m, in which:

$$ACC^m = \frac{number(correctly \quad classified \quad samples)}{number(all \quad samples)} \tag{4.7}$$

We define S as the difference between ACC^0 and the maximum of the other ACC^m:

$$S = ACC^0 \quad - \quad max(ACC^1, ACC^2, ..., ACC^N) \tag{4.8}$$

If $S > 0$, the algorithm ends, and the final accuracy is ACC^0; however, if $S = 0$, we need to refresh the feature matrix to the associated submatrix of $max(ACC^1, ACC^2, ..., ACC^N)$ and continue the iteration.

The submatrix of the highest accuracy is the optimal feature subset. We repeat this process for the remaining levels. With the selected optimal subset, we have built the hierarchical model for CHF detection and quantification.

Finally, we have achieved the risk assessment model. The decision hyperplane functions of the three nodes were calculated. We define the hyperplane function for node decision as $f(x)$:

$$f(x) = \omega^T x + b, \quad \begin{cases} if \quad f(x) > 0, \quad then \quad x \in class1 \\ if \quad f(x) < 0, \quad then \quad x \in class2 \end{cases} \tag{4.9}$$

Where ω and b are the weight vector and the bias that maximize the margin [123], respectively; x is the feature vector. The weight vector ω and bias b are decided by the Lagrangian function:

$$J(\omega, b, a) = \frac{1}{2}\omega^T\omega - \sum_{i=1}^{M} a_i[y_i(\omega x_i + b) - 1] \tag{4.10}$$

Where a_i is the Lagrange multiplier and x_i is the training set. The weight vector and bias are determined from the partial derivative of the function J as

$$\omega = \sum_{i=1}^{M} a_i y_i x_i \tag{4.11}$$

$$b = -\frac{max_{i:y_i=-1}\omega^T x_i + min_{i:y_i=1}\omega^T x_i}{2} \tag{4.12}$$

We named the resulting hyperplane functions according to their roles: disease screening function (DSF; for separating normal heart function from abnormal heart function), risk assessment function I (RAF I; for discriminating the higher risk from the lower risk), and risk assessment function II (RAF II; for distinguishing of moderate risk and mild risk). During this part, the SVM operated under a linear kernel with the same parameters.

4.2.4 Validation and Performance

In our work, we randomly divided the data into training and testing sets at a ratio of approximately 1:1 (i.e., 57 samples for the training set and 59 samples for the testing set). To measure the performance of the tree-structured multistage classifier for risk assessment, we used confusion matrixes (CM) [124]. From these matrixes, we computed the widely used parameters [66] for binary classification to make a comparison with others. Calculation was performed on software MATLAB 7.11.0 (version R2010b, The MathWorks, Inc., Natick, MA, USA)

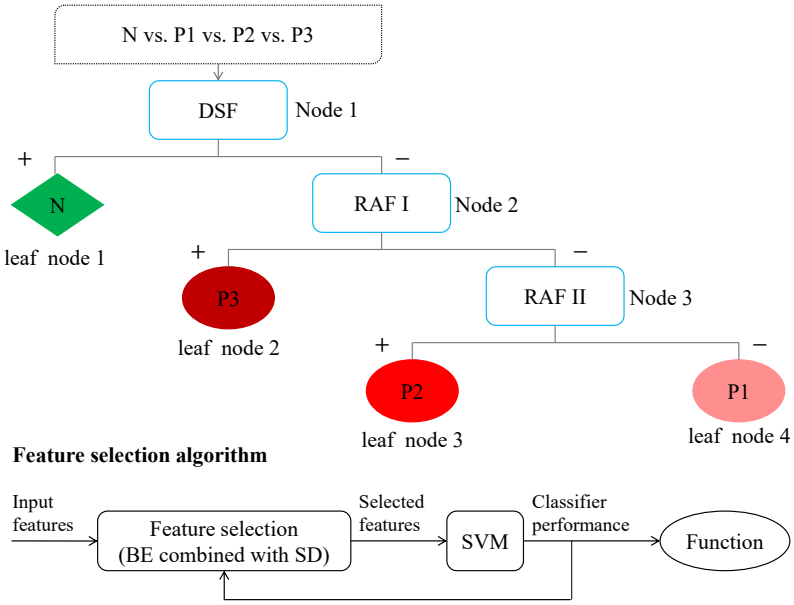

Fig. 4.2 Multistage classification algorithm based on DT-SVM for risk assessment. Upper diagram: tree-structured classifier. Lower diagram: wrappers for feature selection. N: normal samples; P: CHF patients, in which 1 is of New York Heart Associaiton (NYHA) I-II, 2 is of NYHA III, 3 is of NYHA III-IV; DSF: disease screening function; RAF: risk assessment function, in which I is for discriminating the higher risk from the lower risk, II is for distinction of moderate risk and mild risk; BE: backward elimination; SD: significance difference.

4.3 Results

4.3.1 Feature Performance Analysis with C-SVM

With the static and dynamic HRV indices introduced in Method Section, we tested their performance in 4-level risk assessment with linear kernel classical support vector machine (C-SVM) in Table 4.1. The input features were SI, DI and SI+DI, all with p value under 0.1. Feature selection method of the two classifiers was the same, i.e. the backward elimination method. The accuracy rate of C-SVM classifier was close under different input feature combination, all under 80 %. Results were only 76.27 % in 4-level risk assessment of CHF with SI/SI+DI inputted; while only DI inputted, the result was even 10% lower.

While taking apart 4-level risk assessment into disease detection (N vs. P) and disease quantification (P1 vs. P2&P3), our features presented different from above. Table 4.2 compares our results with SI and DI in disease detection and quantification based on C-SVM classifier. Performance with different feature combinations was

Table 4.1 Classification performance of classical support vector machine (C-SVM) in 4-level risk assessment of congestive heart failure (CHF)

Method	Input Feature*	Accuracy (%)
Classical SVM	SI	76.27
	DI	67.80
	SI+DI	76.27

SI: static indices; DI: dynamic indices; * represents that significance value of features were under 0.1.

introduced in the table among different pairs under the linear kernel SVM. The p values of the input features in Table 4.2 were lower than 0.1, and all the features were selected with the BE method. For the discrimination of normal people and CHF patients (Table 4.2: N vs. P), the CHF detection accuracy of SI was 98.31 %, which was higher than DI and DI + SI by over 11 % and 8 %, respectively.

In contrast, prominent diversity existed when distinguishing between higher risk (P2&P3) and lower risk (P1) CHF. The disease quantification accuracy was 91.30 % with DI or DI + SI inputs, but the accuracy of SI dropped by nearly 20 % from the omission of DI.

Table 4.2 Performance of different feature combinations for disease detection and quantification of congestive heart failure (CHF)

Groups	Accuracy			Destination	Method
	SI*	DI*	SI* + DI*		
N vs. P	**98.31**	86.44	89.83	**Disease detection**	C-SVM
P1 vs. P2&P3	73.91	**91.30**	**91.30**	**Disease quantification**	

N: normal samples; P: CHF patients, in which 1 is of New York Heart Association (NYHA) I-II, 2 is of NYHA III, 3 is of NYHA III-IV; SI: static indices; DI: dynamic indices; * represents that significance level of features were under 0.1; C-SVM: classical SVM.

4.4 Multistage Risk Assessment Model Construction based on DT-SVM

On account of feature performance analysis result, 4-level risk assessment model was constructed under DT-SVM algorithm and significance analysis, which was described in Section 4.2. For each binary SVM of DT-SVM, linear kernel and other default

parameter was the same. Thus, DT-SVM classifier construction included two parts: node and input feature selection and model construction, as below.

4.4.1 Node and Feature Selection

Table 4.3 shows the results of node selection based on the significance difference analysis. For the first node, there were seven possible pairs, and it was apparent that the majority of features were extremely significant between the normal person samples and the patient samples (N vs. rest, 57 features with p<0.001) among all these possible pairs. The number of features of this pair was larger than that of all other pairs at each significance level range (79 with p<0.01, 98 with p<0.05, and 105 with p<0.1). Thus, the first node decision was between normal people and CHF patients (N vs. rest).

Table 4.3 Result of node selection for level 1 among all samples

Node	number of p<0.001	number of p<0.01	number of p<0.05	number of p<0.1	number of p>0.1
N vs. rest	**57**	**79**	**98**	**105**	**75**
P1 vs. rest	0	17	39	51	129
P2 vs. rest	15	41	73	86	94
P3 vs. rest	28	52	68	83	97
N&P1 vs. rest	48	65	88	102	78
N&P2 vs. rest	29	51	71	86	94
N&P3 vs. rest	33	61	85	97	83

N: normal samples; P: CHF patients, in which 1 is of New York Heart Association (NYHA) I-II, 2 is of NYHA III, 3 is of NYHA III-IV.

Based on the result of node 1, Table 4.4 shows the numbers of features at different significance level ranges between the three possible pairs for node 2. The higher risk (P3) patients had one feature with $p < 0.001$ corresponding with the lower risk (P1 & P2) patients and no other pairs had this extremely significant feature. In other ranges of p values, the separability of P3 also performed well. Thus, the second node decision was made between higher risk and lower risk (P3 vs. rest) patients. Therefore, the third node decision was between P1 and P2.

Overall, an unbalanced tree was formed under our rules (shown in Fig.4.2) for multistage risk prediction of CHF. The first node was a binary tree between normal people and CHF patients (N vs. P); the second node was a binary tree between higher

Table 4.4 Result of node selection for level 2 among CHF patients

Node	number of p<0.001	number of p<0.01	number of p<0.05	number of p<0.1	number of p>0.1
P1 vs. rest	0	3	19	39	141
P2 vs. rest	0	0	7	22	157
P3 vs. rest	**1**	**9**	**35**	**49**	131

P: CHF patients, in which 1 is of NYHA I-II, 2 is of NYHA III, 3 is of NYHA III-IV.

risk and lower risk (P3 vs. P2 & P1); the third was a binary tree between moderate risk and mild risk (P2 vs. P1).

The selected features for decisions of each node are shown in Table 4.5. Input features for each node are in sequence from SI with a p value under 0.1, DI with a p value under 0.1, and SI + DI with a p value under 0.1. With the aforementioned feature selection method (Section 4.2), the optimal features were selected from the input. The effectiveness of features also showed that a portion of the input features could fully represent them all with relatively low percentages of 7.21 %, 4.08 %, and 33.33 %. During the feature selection of the first node, we found a subset with 23 features that provided maximum discrimination power (i.e., the accuracy was 100 %) in classification between N and P. Considering the computational cost, the number of input features of node 1 ($S1T1, S1T5, S1F4, S2T1, S3T1, S3T3, S2F3, and S3E8$) was chosen to be similar to the other nodes (shown in Table 4.5). These features were fully described in Section 4.2.

Table 4.5 Selected optimal feature subsets for each level with backward elimination algorithm

Node	Input feature numbers	Optimal feature subsets *	Effectiveness (%)
Node 1	111	S1T1,S1T5,S1F4,S2T1,S3T1,S3T3,S2F3,S3E8	7.21
Node 2	49	D5F4,D5E1	4.08
Node 3	12	D2T4,D3F3,D4T1,D5E5	33.33

*: Meaning of features were defined in HRV Measurement; Effectiveness is ratio of number of selected features to number of input features at each node.

4.4.2 DT-SVM based 4-level Risk Assessment Model

As the tree structure and optimal feature subsets of each node had already been obtained, it was easy to construct the final classification model. Fig.4.3 is the unbalanced

multistage classification model of CHF with hyperplane functions, with input and outputs at each binary choice. These functions were described in Section 4.2.

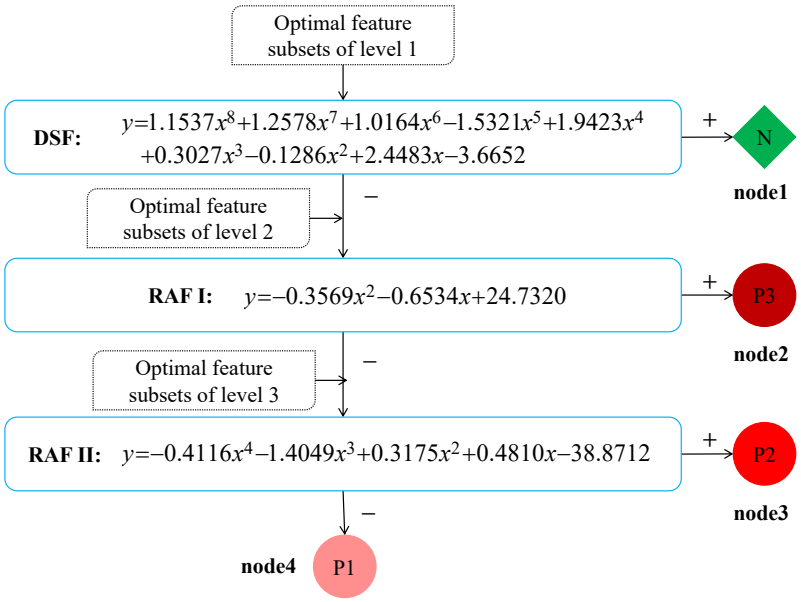

Fig. 4.3 Multistage risk assessment model of CHF. DSF: disease screening function to detect normal people from CHF patients; RAF: risk assessment function, in which I is for discriminating the higher risk from the lower risk, II is for distinction of moderate risk and mild risk; N: normal samples; P: CHF patients, in which 1 is of New York Heart Association (NYHA) I-II, 2 is of NYHA III, 3 is of NYHA III-IV.

All these functions were computed by SVM with the same linear kernel. The three calculated hyperplane functions are shown in Fig.4.3:

$$DSF: \quad y = \omega 1 * X - 3.6652 \tag{4.13}$$

$$RAF \quad I: \quad y = \omega 2 * X + 24.7320 \tag{4.14}$$

$$RAF \quad II: \quad y = \omega 3 * X - 38.8712 \tag{4.15}$$

where the parameters were:

$\omega 1 = [1.1537, 1.2578, 1.0164, -1.5321, 1.9423, 0.3027, -0.1286, 2.4483]$

$\omega 2 = [-0.3569, -0.6534]$

$\omega 3 = [-0.4116, -1.4049, 0.3175, 0.4810]$

The parameter X is the input feature subsets, as showed in Table 4.5; the parameter y is the output value of the function, for which $y > 0$ indicates that the output type is positive (+) and vice versa.

4.5 Validation

The performance of the DT-SVM-based multistage classifier for 4-level risk assessment of CHF is shown in Fig.4.4 and Table 4.6. Fig.4.4 shows the CM of the three binary trees and the multistage classifier. According to the CM, classification error only occurred at the classification of moderate risk (P2), in which one was misclassified as N at node 1 and one misclassified as P3 at node 2.

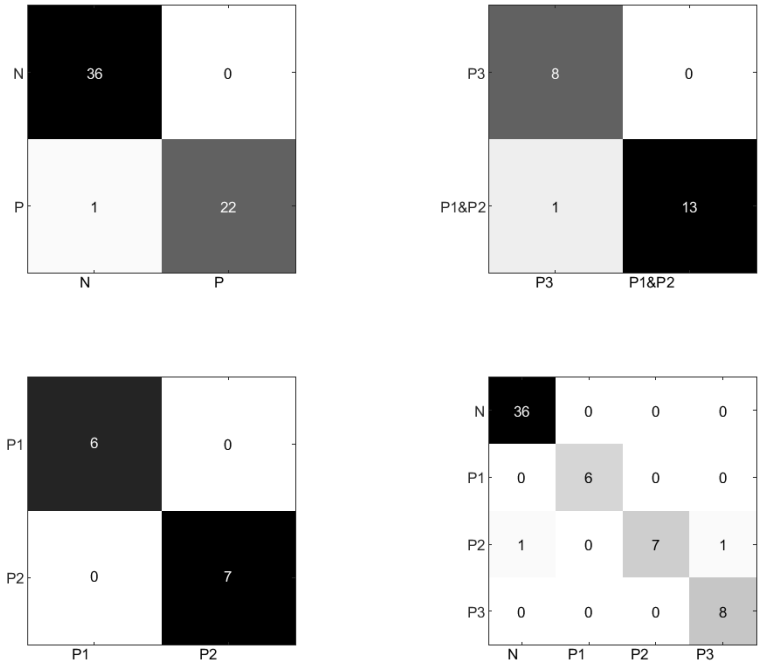

Fig. 4.4 Confusion matrices. N: normal samples; P: CHF patients, in which 1 is of New York Heart Association (NYHA) I-II, 2 is of NYHA III, 3 is of NYHA III-IV.

Table 4.6 contains some common measures computed from the CM-accuracy (ACC), sensitivity (SEN), specificity (SPE), precision (PRE), and area under the curve (AUC). With the DT-SVM algorithm and backward feature selection method, we achieved

a high total accuracy of 96.61 %. The accuracies of nodes 1, 2, and 3 were 98.31 %, 95.45 %, and 100 % respectively. Moreover, the precision (PRE) of each node was 100 %, which demonstrated a lack of false positives. Meanwhile, the values of AUC are over 90 % for every node.

Table 4.6 Classification performance

Node	TP	TN	FP	FN	ACC (%)	SEN (%)	SPE (%)	PRE (%)	AUC (%)	Total ACC (%)
Node 1	36	22	0	1	98.31	97.3	100	100	98.65	
Node 2	8	13	0	1	95.45	88.89	100	100	94.45	**96.61**
Node 3	6	7	0	0	100	100	100	100	100	

TP: true positive, TN: true negative, FP: false positive, FN: false negative;
$ACC = ((TP + TN))/((TP + TN + FP + FN))$,$SEN = TP/((TP + FN))$,$SPE = TN/((TN + FP))$,
$PRE = TP/((TP + FP))$,$AUC = 1/2(SEN + SPE)$

4.6 Discussion

In this study, we applied dynamic indices (DI) and static indices (SI) of HRV measurements to construct a multistage risk assessment model of CHF. This model integrated a DT-SVM classifier, the backward feature selection method, and significance difference analysis. The final calculated hyperplane functions (shown in Fig.4.3) achieved a total accuracy of 96.61 %.

4.6.1 Comparison with Others

As the changes of HRV are influenced under interact of nervous and humoral regulation, autonomic nervous system is part of the reason [125]. Mortara et al. found that static HRV was significantly lower in CHF patients with abnormal heart autonomic nerve function [126]. The adverse change of autonomic function in CHF patients was confirmed [56]. Autonomic nerve dysfunction may increase the incidence of sudden cardiac death in patients with CHF by altering ventricular electrophysiology (correlation with HRV) [127]. We speculated that CHF related to autonomic nerve abnormality. HRV measurements have been confirmed as a noninvasive tool for assessing autonomic nerve function [128]. Thus, the CHF risk assessment model of HRV proposed in this study can serve as a noninvasive and reliable predictor of the incident risk of CHF.

Table 4.7 highlights the comparison with related studies. Yu et al. applied a SVM classifier and GA into CHF detection based on static (bi-spectral HRV) analysis and achieved an accuracy of 96.38 % [91]. Isler et al. utilized static HRV measurements with a KNN classifier a KNN classifier for CHF detection, resulting in an accuracy of 96.39 % [102]. These studies demonstrated that static HRV measurements could distinguish CHF patients from normal people with accuracies of more than 95%. However, Melillo et al. first attempted to distinguish low risk CHF patients from higher risk ones with a relative low accuracy (i.e., 85.4%) [95]. Question that arose then was what caused the low prediction accuracy? Here we analyzed in two aspects: features (static indices vs. dynamic indices) and algorithm (C-SVM vs. DT-SVM) in our paper.

Table 4.7 Highlight

Reference	Classes	Samples∗Time	Feature	Feature Selection	Classifier	Accuracy	Highlight
Yu et al.	**N vs. P**	83∗68min	SI	GA	SVM	96.38 %	CHF detection based on bi-spectral HRV analysis and genetic algorithm
Isler et al.	**(N vs. P**	83∗5min	SI	GA	KNN	96.39 %	CHF detection by combining classical HRV with wavelet entropy measures
Melillo et al.	**P1 vs. (P2&P3)**	44∗24h	SI	ESM	CART	85.40 %	2-level CHF quantification in patients with CHF via long-term HRV and CART algorithm
Our work	**N vs. P1 vs. P2 vs. P3**	**116∗24h**	**SI,DI**	BE	**DT-SVM**	**96.61** %	**4-level CHF detection and quantification using dynamic HRV measures and DT-SVM algorithm**

N: normal samples; P: CHF patients, in which 1 is of NYHA I-II, 2 is of NYHA III, 3 is of NYHA III-IV; SI: static indices; DI: dynamic indices; GA: genetic algorithm; ESM: exhaustive search method; BE: backward elimination; SVM: support vector machine; KNN: k-nearest neighbor; DT-SVM: decision tree based support vector machine.

4.6.2 HRV Measurement Analysis

According to NYHA classification [7], the severity of heart failure is related to the severity of symptoms. For example, NYHA I: no symptoms and no limitation in ordinary physical activity; NYHA III: marked limitation in activity due to symptoms, even during less-than-ordinary activity; NYHA IV: experiences symptoms even while at rest. After diagnosis, the heart conditions of certain classes of patients fluctuate according to the treatment, physical condition, etc. [66].

It has been recognized that autonomic imbalance happens in heart failure, which leads to further worsening of the condition. Thus, autonomic dysfunction in CHF patients was confirmed [1]. To this end, HRV analysis of SI has already served as a powerful tool in autonomic nerve function assessment. However, we presumed that static HRV measurements could not fully reflect the fluctuation of autonomic nerve function over time among different classes of patients. As a result, Melillo et al. only reached an accuracy of 85.4% for 2-level disease quantification when using static HRV analysis [95]. Thus, in this work, we proposed dynamic HRV measurements to rate CHF risk. Comparing performance of SI and DI in disease detection and quantification based on C-SVM classifier with previous research, we can conclude that:

1. For the discrimination of normal people and CHF patients (Table 4.1: N vs. P), the CHF detection accuracy of SI was higher than DI and DI + SI. Based on this result, we can conclude that DI does not help in this respect. Furthermore, static HRV measurements have been proven by former studies to be reliable for high-accuracy detection of CHF [47, 48, 90, 68, 129].

2. In contrast, nearly 20 % diversity existed when distinguishing between higher risk (P2&P3) and lower risk (P1) CHF from the omission of DI. This demonstrates that DI are more important in discriminating higher risk patients from lower risk ones. Furthermore, inclusion of SI does not improve performance in disease quantification.

Therefore, this analysis demonstrates that: 1) Dynamic HRV measurements have an obvious advantage over SI in disease quantification; 2) DI offer no help in disease detection; 3) Static HRV measurements have excellent performance in disease detection.

4.6.3 Classifier Analysis

From the SI and DI analysis above, the multistage classifier (e.g. DT-SVM) is suitable to build the CHF risk assessment model, for using certain feature in certain purpose. We compared the performance of multistage DT-SVM classifier of our paper and classical SVM one in 4-level risk assessment in Table 4.1. The DT-SVM classifier revealed an excellent power in 4-level risk assessment of CHF than C-SVM classifier.

The result of C-SVM in 4-level risk assessment was not satisfied for this low precision. According to the performance of DI and SI in disease detection and quantification, a relative high precision on risk assessment was possible using combination of static and dynamic HRV measurement. This was cause by the reason that the classical SVM cannot take the advantage of SI and DI described at Subsection 4.6.2 into risk assessment progress, as DT-SVM did.

Review the whole work, our multistage CHF risk assessment model has the following advantages:

1. The multistage model could fully combine respective advantages from static HRV measures and dynamic HRV measures. According to the conclusion in Section 4.7 about HRV measurement, SI was more suitable for disease detection and DI for disease quantification. Thus the stratified structure could make full use of this trait. In our work specific features were inputted for specific nodes: SI with $p < 0.1$ for disease detection and DI with $p < 0.1$ for disease quantification.

2. The parameter setting of multistage risk assessment model also conformed to the NYHA classification. From the perspective of physiological law, adverse change in ANS activity is a hallmark characteristic of CHF [56]. And CHF patients showed weakness in vagal mechanisms to counteract sympathetic activation [56]. This dysfunction worsens along with disease exacerbation [1]. Thus, the result that the first leaf node was N and the second was P3 was conformed to their difference in ANS function.

3. The DT-SVM classifier we applied was modified by backward feature selection algorithm combined with significance difference. Significance difference analysis helped to decide the suitable tree structure and nodes. Backward feature selection method improved the efficiency of whole classifier, with features that p value as under 0.1.

4.6.4 Clinical Significance

The multistage CHF risk assessment model achieved an accuracy of 96.61% between predicted and actual ratings. Compared with NYHA classification according to the

limitations/symptoms during physical activity [31], our multistage risk model with HRV analysis is a noninvasive and objective CHF rating method. This helps reduce diagnosis mistakes caused by various physician-related factors (e.g., limited experience, work stress, fatigue) [130], ignorance of circadian clinical features caused by unprompted modulation of autonomic activity, therapy, etc.. Moreover, combining with 24-h ECG recording, our model is more universal and stable than most short-/long-term HRV measurements (e.g. [47, 129]).

The ANS can modulate the sino-aerial nodal depolarizations to adjust the needs of the body. HRV analysis with SI has been used in prior research to assess ANS function of CHF patients [66]. Relative to static HRV measurements, the dynamic HRV measurements presented in this study could better reflect the fluctuation of autonomic nerve dysfunction during physical activity. Additionally, the modified DT-SVM algorithm can fully combine both advantages and improve performance. Autonomic nerve functions trend worse in CHF patients, and HRV dynamic analysis demonstrates this small time-scale variation. Finally, no previous studies have been tested in a comprehensive multistage model. Researchers have stated that the NYHA method remains arguably the most important prognostic marker in routine clinical use in heart failure today [131]. Thus, in view of these advantages, our multistage CHF risk assessment model could serve as a clinically meaningful outcome by providing an objective and timely assessment and rating of autonomic function change trends for CHF patients in the future, especially for those with in-home monitors.

Beyond all these benefits, our study still had some limitations. First, a larger and richer data collection is needed for further study. Moreover, dynamic HRV measurements should be analyzed and explained in future work; algorithm parameter setting needs further analysis. In the future, we plan to access an early-warning risk model with more elements (e.g., therapy, clinical state) included for survival prediction.

4.7 Conclusion

To construct a multistage risk assessment model of CHF, we applied dynamic HRV measurements to CHF detection and quantification. DT-SVM algorithm and feature

selection method based on BE and significance difference was included. The data used for this study consisted of 126 DI and 54 SI of HRV measurements obtained from 116 samples of 24-h ECG records from the MIT/BIH database. According to the study, we reached several conclusions:

1). In this study, we build a four-level risk assessment model for CHF detection and quantification based on the DT-SVM algorithm. The model succeeded with a total accuracy of 96.61 %, in risk assessment among individuals of N (no risk), P1 (mild risk), P2 (moderate risk), and P3 (severe risk).

2). We creatively proposed dynamic indices for the severity evaluation of CHF. In CHF quantification (Table 4.1), the DI are obviously superior to the SI, increasing the accuracy from 73.91% to 91.30%.

3). The DT-SVM-based multistage risk assessment model proposed in this work significantly improved discrimination power from 76.27 % to 96.61 % when compared to C-SVM. The performance of our classifier improved based on the combination of the BE algorithm and significance difference.

4). According to the analysis of SI, it was clear that SI of HRV were more powerful in disease detection than DI with an accuracy of 98.31 %. This is consistent with the results of prior research regarding disease detection.

In light of these advantages, the stratifying CHF risk assessment model will be a reliable and objective prognostic marker for routine clinical application (especially daily health nursing) in the future.

Chapter 5

Unsupervised CHF Detection method using SAE-based DL and 5-min RR intervals

This Chapter firstly applied an Unsupervised sparse auto-encoder-based deep learning (SAE-based DL) algorithm into short term RR intervals for CHF detection in fast assessment.

5.1 Introduction

As a common chronic cardiovascular syndrome, congestive heart failure (CHF) has attracted many researchers' attention. CHF is linked to several complex problems, including end stage heart suffering [1], that cause chambers of the heart to fail. When symptoms suggest CHF, the diagnosis is usually confirmed by physical examination, patient history and various tests to detect abnormal function of the left ventricle and/or heart valves [31]. This diagnosis process often results in time delay and high costs.

Electrocardiogram (ECG) is a commonly used physiological signal in monitoring cardiovascular condition, as it' s non-invasive, portable and easy to use. ECG shows both heart rate and rhythm (steady or irregular) [1].

For CHF detection, many researchers focused on ECG analysis, including QRS waves [43], QT intervals [44], RR variability [45], and so on. RR interval is one of the important components of ECG and has attracted great attention as its variability has great potential in CHF prognosis [45]. This variability is either described by RR variability (i.e. variation of RR intervals) or heart rate variability (HRV). HRV based CHF detection has achieved over 95% accuracy with linear/non-linear measures of short-term (5-min)/long-term (24-h) RR intervals data [45]. However, due to its sensitivity along with clinical condition changes [31], the robustness of the HRV-based approaches is still an issue to be addressed. In our previous work [5], we proposed dynamic HRV to describe dynamic fluctuation of HRV in 24 hours. The analysis revealed that HRV is dynamically changed along with conditions, in accordance with RR intervals fluctuation. Thus, it is necessary to improve the robustness of CHF detection with RR intervals data.

Although there is a guideline [31], manual error and characteristic preprocessing of data are still used. Furthermore, all these former CHF detection methods with HRV are supervised learning. This may require a large amount of labelled data. In this study we propose to construct an automatic CHF detection without manual extraction and supervision. This process is based on a sparse auto-encoder (SAE)-based deep learning (DL) algorithm.

Deep learning has been applied using unsupervised features [132] and achieved remarkable results in many fields like image recognition [133]. In this work, SAE was applied to learn features automatically from raw unsupervised RR intervals data. Then a deep learning neural network was trained to construct a model to discriminate CHF. In order to improve the neural network performance, various hidden nodes were searched in proportion to input data length. The features in our algorithm are constructed to reflect as much fluctuations in RR intervals.

This work contributes to CHF detection model construction based on SAE-based DL algorithm. It analyses potential mechanism behind between dynamically changed RR intervals. We intend to achieve an automatic method for CHF detection.

The organization of this paper is as follows: in Section 5.2, procedure of model construction of SAE based DL is introduced. The analysis results as well as potential

physiological mechanism are discussed in Section 5.3. Finally, Section 5.4 provides the conclusion.

5.2 Methods

5.2.1 Data

In our work, we applied 116 subjects' RR interval data from PhysioNet [111, 112]. All subjects provided informed written consent. The study was approved by the institutional Review Boards of Beth Israel Deaconess Medical Center (Boston, MA) and the Massachusetts Institute of Technology (Cambridge, MA). We downloaded the 24-h RR interval data of 72 healthy persons and 44 CHF patients. The healthy person data came from two databases: the MIT/BIH Normal Sinus Rhythm Database and the Normal Sinus Rhythm RR Interval Database [111]. The data of the CHF patients came from the Congestive Heart Failure RR Interval Database and BIDMC Congestive Heart Failure Database [112]. The data were manually reviewed and corrected by experts.

5.2.2 Deep Learning based CHF Detection Algorithm

In this work, we applied sparse auto-encoder-based deep learning (SAE-based DL) algorithm for CHF detection (shown in Fig.5.1).

Sparse auto encoder is a neural network that can learn automatically sparse features from data by minimizing reconstruction error [134]. As one of many famous unsupervised feature learning methods, SAE combines with DL to realize its effectiveness. The aim of this process is to reconstruct input data at the output layer by a sparse penalty term β [135].

Detailed steps of our algorithm are below:

1. Unsupervised feature automatically learning with SAE

Denote input data as $\{RR = rr(i), i = 1, \cdots, N, rr(i) \in R^M\}$. Here N is data sample number of one class, and M is data samples length. First, we input this

unlabeled data set into SAE for automatic feature learning to constrain features. At hidden layer l, the feature was expressed as $\{F(rr(i), W, b), i = 1, \cdots, N\}$, where W denotes the weights between two neighboring layers and b is the bias. This means that the required feature was controlled by input and connection parameters (W, b).

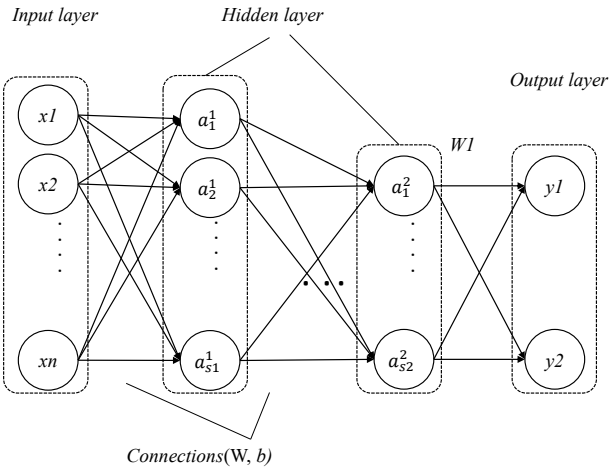

Fig. 5.1 Typical neural network of sparse auto encoder (SAE) based deep learning (DL) structure. Each circle of hidden layers is a hidden node; input layer is learnt features with SAE network.

The unlabelled segment data RR are first used to train unsupervised features automatically with SAE through the following steps:

1. Set up the sparsity penalty term as 3, decay parameter as 10^{-4} and desired average activation of the hidden units as 0.1, and initialize connections W and b randomly close to 0.

2. Use batch gradient descent to train the neural network in the forward propagation algorithm to compute the sparse cost function of each iteration of the layer as:

$$Cost\,(w, b) = \frac{1}{n}\sum_{i}^{N}(\frac{1}{2}||F_{w,b}\,(rr\,(i))\,||^2) + \lambda] + \beta\sum_{j=1}^{s2} KL(\rho||\rho_j) \qquad (5.1)$$

Here $\rho_j = \frac{1}{n}\sum_{i=1}^{n} a_j\,(rr\,(i))$ is the average activation of hidden unit j; $KL(\rho|\,|\rho_j) = \rho log\frac{\rho}{\rho_j} + (1-\rho)log\frac{1-\rho}{1-\rho_j}$ is a standard measuring function of the difference of

the two distributions; λ is the weight decay parameter; $s2$ is the hidden neuron numbers in the hidden layer.

3. Update the parameters W and b at each iteration of gradient descent updating:

$$W_{ij}(l) = W_{ij}(l) - \alpha \frac{\partial}{\partial W_{ij}(l)} Cost(W, b) \tag{5.2}$$

$$b_i(l) = b_i(l) - \alpha \frac{\partial}{\partial b_i(l)} Cost(W, b) \tag{5.3}$$

Here α is the learning rate/step size, optimized by linear search optimizer; l is hidden layer (here is 1).

4. Repeat step 2-4 for the second hidden layer for each iteration until iteration is done (set to 300); or we reached the minimum cost value.

2. Deep neural network construction with DL

The feature set F was extracted and fed into DL for classification.

1) Select the values of the parameters (W, b) to initialize the DL neural network.

2) Set up the training parameters, and conduct the forward propagation algorithm to construct CHF detection model based on Softmax classifier.

3) Compute the mean square error for the cost function of the DL using Eq.5.1.

4) Conduct the back-propagation algorithm to update the connections and fine-tune the entire network for each iteration until iteration is done (set to 300); or we reached the minimum cost value.

Thus, a deep neural network was conducted for CHF classification. Finally, the test data set is used to verify the effectiveness of the presented SAE-based DL classifier. Classifier performance is assessed by

$$Accuracy = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}} \tag{5.4}$$

where N_{TP} is defined as the number of true positive, N_{FN} is the number of false negative, N_{FP} is the number of false positive, and N_{TN} is the number of true negative.

All above steps were done with different hidden node combination to find a suitable network construction. Here, we tested different hidden nodes combination in proportion with input length at range of [10, 200]. Train set and test set were randomly selected with ratio of 1:1. Performance of train set and test set were all verified to avoid over-training.

5.3 Results and Discussion

To construct SAE-based deep learning neural network in CHF detection, we downloaded 116 24-h RR interval records. In our work, a neural network was built with hidden node optimization.

Before model construction, all the 24-h data were preprocessed:

1) deleting the first and the last interval;

2) excluding RR intervals longer than 3 seconds [66];

3) dividing 24-h data into multiple 5-min segments as a circle reflecting cardiovascular condition [31] and saved in sequence.

The first two steps were in case of unstable measurement conditions and artificial error; the third step was preparing for CHF detection. After preprocessing, 30592 segments were accessed. After preprocessing, all segments were interpolated into equal length (to biggest segment length 779) according to algorithm requirement.

Then, 5-min RR intervals segments were applied for CHF detection which were reconstructed by SAE algorithm. This reconstructed feature learning happens automatically.

Since the architecture of SAE-based DL, including the size and parameter setting, influences its performance, it is critical to carefully select hidden node numbers and layers. In this work, the hidden layer was chosen as two while hidden nodes of each layer were tested proportionally as shown in Table 5.1. Classification results shown in Table 5.1 were around 72 % among various hidden node combinations. The train and test accuracy was calculated as in Table 5.1, which indicates that 72.44 % testing

accuracy is achievable even without an explicit supervised feature detection stage. The result also shows the variation of testing conditions in a whole day influenced CHF detection. This is consistent with the existing research. In addition, the accuracy of the train and that of the test are very close, indicating that there is no over-training.

Table 5.1 Performance of train and test set with different hidden nodes combinations

Hidden nodes of layer 1		200	200	200	200	200	200	150	150	150	150
Hidden nodes of layer 2		200	150	100	50	30	10	150	100	50	30
Train	**Accuracy (%)**	72.36	72.21	72.48	72.02	72.45	71.27	72.18	71.98	72.58	72.05
	Sensitivity (%)	48.78	49.32	49.41	49.38	42.13	39.83	48.6	48.38	46.64	40.95
	Specificity (%)	85.72	85.18	85.55	84.84	89.62	89.07	85.53	85.35	87.27	89.66
Test	**Accuracy (%)**	72.07	72.49	72.21	72.44	72.45	71.27	72.13	71.95	72.37	71.39
	Sensitivity (%)	48.83	49.61	48.43	50.39	42.42	39.01	48.18	47.58	46.33	39.97
	Specificity (%)	85.24	85.44	85.68	84.93	89.45	89.54	85.69	85.75	87.12	89.97
Hidden nodes of layer 1		150	100	100	100	100	50	50	50	30	30
Hidden nodes of layer 2		10	100	50	30	10	50	30	10	30	10
Train	**Accuracy (%)**	71.18	72.16	71.8	72.79	71.22	72.22	71.7	71.41	70.65	71.37
	Sensitivity (%)	41.19	48.7	41.57	46.55	41.19	48.6	48.58	41.44	47.58	41.12
	Specificity (%)	88.16	85.44	88.92	87.66	89.38	85.6	84.8	88.37	83.72	88.51
Test	**Accuracy (%)**	71.03	72.41	72.43	72.23	71.28	72.86	72.4	71.23	71.47	71.26
	Sensitivity (%)	41.41	49.09	42.2	45.64	40.01	49.09	49.28	41.67	48.9	40.75
	Specificity (%)	87.81	85.62	89.57	87.29	89.52	86.33	85.49	87.97	84.25	88.53

Considering trade-off between classifier performance and algorithm effectiveness, a two-layer deep neural network with hidden nodes setting (200,50) was selected as optimal for this study. With this combination structure, the accuracy of train and test were 72.02 % and 72.44 %, respectively. Performance lost along with descent of nodes. It should be noted that considering the features are automatically extracted in an implicit way, the achieved accuracy is still acceptable.

As many works have shown good results in CHF detection with HRV (described in Section 5.1), RR intervals do maintain useful information for differentiating CHF patients. We also noticed that difference existed while using long-term (24-h) HRV and short-term (5-min) HRV about its reliability [136]. To the best of our knowledge, there has not been any research which achieves stable CHF detection with only RR variability. This might demonstrate that RR variability is not quite stable as it is sensitive to test conditions, e.g., exercise condition, sleep condition, psychological status and so on [137]. Existing research development is mainly focused on feature extraction methods and

statistical analysis [31]. Although these works did reveal potential linkages between disease risk and RR interval based features, it involves a high level of manual operation, which leads to potential human error and unstable results. Thus, there is an urgent need for a fully automated CHF detection method without manual feature extraction.

In this study, we applied SAE-based deep learning method into CHF detection. Essentially, features were extracted by the algorithm itself, also called auto-learning. All segments were trained to extract 50 features judged by Equation 5.1. These 50 learnt features were inputted into a softmax classifier, which is based on a two-layer neural network.

It has been reported that RR variability has potential in CHF detection, even prognosis. In our former work, we introduced dynamic measures of RR variability into CHF detection [45] and concluded that different conditions influence RR intervals classification performance, which might be one of the reasons that the detection accuracy in this study is difficult to improve further. There is no special technology recommended for optimal parameter selection. Thus, parameter setting selection will be the next task in the future to further improve the proposed approach.

It is certain that RR intervals contain information about cardiovascular condition. Actually, when analyzing classification results of segments of each person, it showed that less than 30 % segments of 90.28 % (65/72) healthy persons has been classified as negative, and higher than 30 % segments of 77.27 % (34/44) CHF patients has classified as negative. This supports that RR intervals can be used as CHF detection indicator, but this is sensitive to test condition variations.

5.4 Conclusion

In this study, a two-layer deep neural network model was constructed for CHF detection based on SAE-based DL algorithm. The results indicated that RR intervals have potential for CHF detection but is sensitive to body condition. Though the classification accuracy for the overall segments only reached 72.41 % , further analysis demonstrated the efficiency of the proposed method and interpreted the rationality of the achieved classification results in CHF detection. This also demonstrated the dynamic change

with body condition may not fully reflect by only analyzing short/long term RR variability. More analysis on RR intervals is needed. Thus, a dynamic assessment method is required for reliable CHF detection and prognosis. For future work, we will explore the dynamic analysis of the underlying mechanisms.

Chapter 6

Automatic Risk Assessment of CHF Using ECG at Optimal Time Scale

This Chapter described about optimal time scale analysis for 3-level risk assessment of Congestive heart failure (CHF) based on heart rate variability, to achieve a more convenient and robust assessment.

6.1 Introduction

Congestive heart failure (CHF) is a typical syndrome and end stage of various heart diseases [49]. Clinically, CHF is caused by heart disease and is characterized by left ventricular dysfunction and neurohormonal regulation disorder [138]. Clinical symptoms include respiratory distress and decreased exercise capacity or tolerance [139]. Usually, there is no obvious symptom in the early stage of congestive heart failure, but once diagnosed, 30% of CHF patients die after a year [8]. Therefore, early prognosis and timely assessment of congestive heart failure are essential. One of the most commonly used clinical risk assessment methods is New York Heart Association classification, which classified CHF patients into I-IV risk levels according to physiological activity limits [28, 29].

Overall, most studies have focused on ECG analysis in CHF assessments, especially on heart rate variability (HRV) analysis. Classical HRV analysis assesses changes

in cardiac autonomic function by calculating time domain, frequency domain, and nonlinear features. The European Society of Cardiology published the International Guideline to guide the calculation of HRV features such as time and frequency domain features in 1996 [66]. Time and frequency domain analyses reflect the degree of variability and underlying rhythms of cardiac autonomic nervous system, respectively [49]. while nonlinear analysis processes the dynamic variation in HRV from physiological systems within a certain period [50], which showed in dynamic of HRV. The time domain features of HRV measurements have been proved to have a good discrimination in CHF detection (accuracy over 90%), especially the standard deviation [45, 51]. While analyzing the frequency domain features of HRV measurements, a decreased trend showed, which indicated the deterioration of cardiac autonomic nervous system function [54, 140, 141]. Frequency domain HRV measures-based CHF screening studied has achieved an accuracy of over 90% [45]. In recent years, researchers focused more on the nonlinear feature analysis of HRV, include the geometric and informatics knowledge describing dynamic cycle fluctuation, and proposed different calculation methods [45, 142]. Among these nonlinear feature analyses, entropy measures have attracted many attentions. Various entropies have been proposed in previous studies to describe the complexity of the cardiopulmonary system, including appropriate entropy, sample entropy, fuzzy entropy, etc. [45]. These entropies show their ability in CHF detection (accuracy over 90%), even with the limitation in parameter setting [45, 114]. Those studies highlight the potential of HRV analysis in CHF risk assessments.

Most researchers perform HRV analysis based on short-term (5-minute) or long-term (24-hour) ECG data for CHF risk assessment. Long-term HRV analysis presents abundant physiological information of individuals in a day after various daily activities. Long-term HRV measurements based CHF assessments show good performance with an accuracy of over 90% [45, 53, 91]. However, it is difficult to obtain long-term ECG data with good quality in the clinical context. The advantage of short term HRV analysis is in easy-access and many researchers have demonstrated its ability in CHF detection [48, 90, 68]. The main problem is that short-term ECG signals are sensitive to physiological changes, including emotion [143], fatigue [144], and exercise condition [145]. This will affect the robustness of the short-term HRV analysis in CHF assessment. In addition, one of our previous publications using short-term RR intervals for CHF

detection also illustrates this deficiency [146]. Therefore, we will explore the optimal time scale of ECG for CHF risk assessment based on HRV analysis in this study.

In this study, the CHF risk assessment involves three risk groups: healthy people, CHF patients with NYHA I-II, and CHF patients with NYHA III-IV, labeled as no risk (NR), lower risk (LR), and higher risk (HR), respectively. HRV analysis at multiple time scales (from 1h to 24h) were compared to explore their performance in CHF risk assessment. First, ten HRV measurements were calculated based on ECG data at different time scales, as described in Section 6.2.2. Then the significance analysis was performed to compare the discrimination power of ECG data at these time scales. Criteria for comparison were shown in Section 6.2.3. Based on the Criteria, the optimal time scale for HRV analysis in CHF risk assessment will be obtained. Next, the support vector machine will be applied for CHF assessment using HRV measurements of the optimal time scale ECG data to achieve a convenient and robust CHF risk assessment model. To improve classifier performance, exhaustive search methods will be applied to get optimal feature subset. Finally, cross validation and classical metrics of confusion matrix will be used to assess classification performance.

6.2 Methods

6.2.1 Data

We obtained the data used in this work from the widely used MIT/BIH database in PhysioNet [111]. All subjects provided informed written consent. The study was approved by the Institutional Review Boards of Beth Israel Deaconess Medical Centre (Boston, MA) and the Massachusetts Institute of Technology (Cambridge, MA). We chose four RR interval databases, obtaining 116 nominal 24-h RR interval records: 72 healthy samples (aged 20 to 76) and 44 CHF patient samples (aged 22 to 79). The data of healthy people came from two databases: the MIT/BIH Normal Sinus Rhythm Database and the Normal Sinus Rhythm RR Interval Database [66]. The data of the CHF patients came from the Congestive Heart Failure RR Interval Database and

BIDMC Congestive Heart Failure Database [112]. These records were all manually reviewed and corrected by experts.

All samples were grouped into four stages according to severity:

- 72 healthy people labeled as no risk (NR, aged 54.62±16.03 years);
- 12 patients in NYHA I-II labeled as lower risk (LR, aged 52.5±14.25 years);
- 32 patients in NYHA III-IV labeled as higher risk (HR, aged 56±11.50 years with one sample unknown).

The subjects' gender information was partially abridged, so there was no description of gender. All these data of the 116 subjects can be downloaded online for free from http://www.physionet.org/cgi-bin/atm/ATM [111] .

6.2.2 HRV Measurement

In this study, ten classical HRV measurement were extracted from 116 RR interval data.

1) Data Preprocessing

Before performing feature extraction, we pre-processed all these data by:

1) deleting the first and the last RR intervals;
2) excluding RR intervals longer than 3 seconds [111];
3) dividing the 24-h data into multiple segments at different time scales sequentially.

The first two steps were performed in case of unstable measurement conditions and artificial error. In the raw RR interval data, the first and last RR intervals were found large and abnormal, so they were deleted in case of measurement error. According to The International Guideline [66], RR intervals longer than 2.5 seconds leads to a serious heart attack. Thus, RR intervals longer than 3 seconds were excluded as the patients need medical assistance in this condition and cannot be recorded. The third

step was associated with feature extraction. The time scales include 1h, 2h, 3h, 4h, 6h, 8h, and 24h.

2) Feature Calculation

In this study, ten classical HRV measurements were analyzed from 116 pre-processed RR interval data among all segmented RR intervals. After pre-processing, we had seven types of data according to time scales.

With these types of data, we calculated 10 classical HRV measurements, which included:

- **Time Domain**: the average of RR intervals (M); standard deviation of RR intervals (S); root mean square of successive RR interval difference (R); coefficient variation (ratio of S to M) of RR intervals (CV) [113];
- **Frequency Domain**: the power of RR intervals in 0.04-0.15 Hz (LF); the power of RR intervals in 0.15-0.4 Hz (HF); the ratio of LF to HF (LF);
- **Nonlinear**: approximate entropy (AE); sample entropy (SE) [45]; fuzzy entropy (FE) [114].

The time/frequency domain HRV measurements in our work followed International Guidelines [66], and frequency domain HRV measurement was calculated based on Fast Fourier Transform. The nonlinear HRV measurements are introduced in Ref [45]. All of the calculations were done using the software MATLAB 7.11.0 (version R2010b, The MathWorks, Inc., Natick, MA, USA).

6.2.3 Optimal Time Scale Analysis

After feature extraction, statistical analyses involving three risk levels and each-two risk level were obtained. Differences between groups are assessed using one-way analysis of variance (one-way ANOVA). Statistical tests are conducted with a significant level set at 0.05. All the significance analysis is calculated using the SPSS software (version 19, SPSS Inc., Chicago, IL, USA). The overall statistical analysis was performed on each time scales of data.

Through the statistical analysis, three p values between every two groups (NR-LR, NR-HR, LR-HR) were obtained. To compare the classification performance of data at different time scales, we selected the biggest p-value of these three group-pairs as a comparative index, as it shows the worst differentiating performance. To guarantee the comparability of data at different time scales, sample numbers of any two-comparisons were balanced using multiple random selections (8 times in this work). Here, we compared data at different time scales from small to large successively. First, 1h and 2h of HRV were compared and the better one was compared with 3h of HRV, and the same process continued for data at other time scales. In this study, eight random sample selections were performed and various powers of the ten features at different time scales were compared based on their p values. To compare differentiating powers among data at these time scales, several criteria were applied, as follow:

- **Rule 1: stability**. For different time scales of data, whether the significant level of features at the eight random selections stays stable (p-value always under 0.05); if so goes to Rule 2, if not, the differentiating power is worse;
- **Rule 2: separability**. For features with higher stability, a lower significant level represents higher separability. Here, a lower significance level was defined as a higher mean p-value of eight random selections. If data at a time scale have more features with higher separability, the differentiating power is better, if the same number of features, goes to Rule 3, if a smaller number of features, the differentiating power is better;
- **Rule 3: availability**. For data at different time scales, a shorter time scale represents better availability, as it means easier access in the application. Thus, shorter time scale indicated a higher priority.

Using the above criteria, the optimal time scale of HRV analysis for CHF assessment can be obtained.

6.2.4 3-level Risk Assessment Model Construction

After optimal time scale analysis, the optimal time scale was found. To verify the assessment ability, the classical HRV features at the optimal time scale were then applied into a 3-level model construction for CHF risk assessment using a support vector machine (SVM). The flow chart of this process can be seen in Fig.6.1.

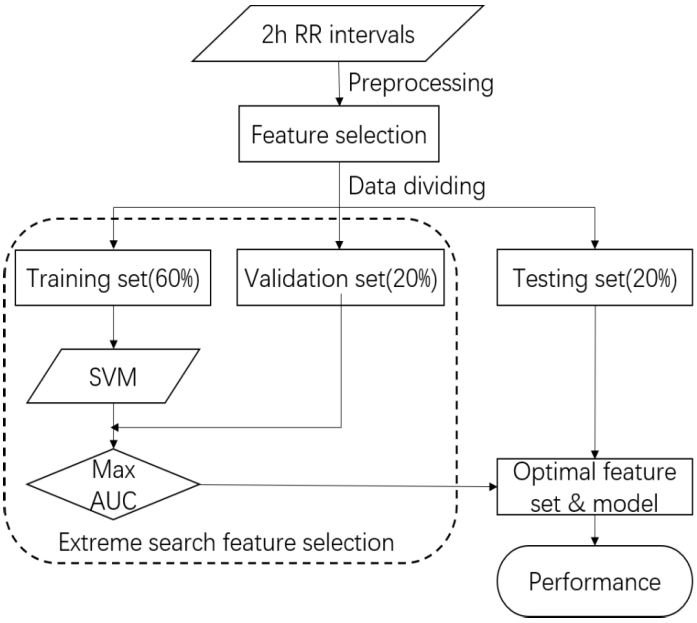

Fig. 6.1 Flow chart of 3-level congestive heart failure (CHF) assessment using optimal timescale (2h) and support vector machine.

1) Data division

The sample sizes of each risk levels were unbalanced in this research. These samples consist of 18 healthy samples (no risk: NR) from MIT-BIH Normal Sinus Rhythm Database, 12 lower risk CHF patients (LR) in NYHA I-II from Congestive Heart Failure RR Interval Database, and 17 higher risk CHF patients (HR) in NYHA III from Congestive Heart Failure RR Interval Database.

With these RR interval records, multiple segments at optimal time scale (2h, as described in Section 6.3) were obtained. These 2h segments were divided into three sets randomly in a ratio of 3:1:1 for training, validation and test procedures respectively, as shown below.

Table 6.1 Sample information about 2h segments

	#NR	#LR	#HR	Total number
Training set	109	74	103	286
Validation set	36	24	34	94
Testing set	37	26	36	99
Total number	182	124	173	479

means the number of; NR: healthy (no risk) samples; LR: lower risk samples; HR: higher risk samples.

2) 3-level CHF assessment model construction

In this study, we applied support vector machine for 3-level CHF assessment using optimal time scale data, i.e. 2-h RR intervals.

Support vector machine (SVM) is one of the most commonly used classifiers in machine learning. The advantage of SVM is in binary classification assigning new samples to one category or the other [80]. One of the most important parameters influenced the performance of the SVM classifier is kernel function. The introduction of kernel function extends linear classification to non-linear classification, via mapping inputs into high-dimensional feature space [147]. In the mapped feature space, samples will be separated by a hyperplane, which tends to reach the largest margin between two classes, also called maximum-margin hyperplane [81]. Commonly used kernel functions in the SVM algorithm include linear, Gaussian and polynomial kernel. In this study, a Gaussian kernel function was used.

As mentioned in the Subsection 6.2.1, there are three classes of samples: healthy samples (no risk: NR), lower risk samples (LR) and higher risk samples (HR). Thus, we organized the 3-level classification into three different binary classifications: 1) NR vs. (LR vs HR); 2) LR vs. (NR vs. HR); 3) HR vs. (NR vs. LR). Here, the 3-level CHF assessment was organized into a two-layer binary classification. For each subject, ten classical HRV measurements were calculated as described in Section 6.2. All these features were applied into risk assessment model construction procedure.

During the model construction procedure at each layer, the training set was inputted into an SVM classifier to obtain a model for risk assessment and the validation set was used to evaluate its performance. In the binary classification of each layer, all subjects were labeled into two classes, 0 (negative) or 1 (positive). To evaluate the

performance of a constructed model, some widely used parameters were defined. Let n the number of subjects labeled as negative, and p the number of subjects labeled as positive. Then tp is the number of true positives, tn is the number of true negatives, fp is the number of false positives, and fn is the number of false negatives. Then, the $Accuracy$, $Sensitivity$, $Specificity$, and Auc (area under the curve) can be calculated:

$$Accuracy = (tp + tn)/(tp + tn + fp + fn) \tag{6.1}$$

$$Sensitivity = tp/(tp + fn) \tag{6.2}$$

$$Specificity = tp/(tn + fp) \tag{6.3}$$

$$Auc = \frac{(Sensitivity + Specificity)}{2} \tag{6.4}$$

Hence, the 3-level risk assessment of CHF could be constructed with optimal classification order and features. The optimal classification order was analyzed via inputting different possible classification combinations at each layer and comparing the performance of corresponding models. For layer 1, the possible classification combinations include NR vs. LR&HR, LR vs. NR&HR and HR vs. NR&LR. The binary classification of layer 2 was selected among corresponding classification combinations of LR vs. HR, NR vs. HR and NR vs. LR.

The optimal feature subset was obtained using an exhaustive search feature selection method. For each subject, there are 10 features that were inputted for model construction Thus, there are a total of 1023 possible feature combinations to be analyzed. Each feature combination was evaluated by the SVM classifier using Auc, as it is an arithmetic mean value and can balance sensitivity and specificity. The feature subset with the highest Auc is the optimal feature subset.

After this feature selection process, the optimal feature subset under different classification combinations at each layer was found. Combining the classification performance of layer 1 and 2, three 3-level risk assessment models were constructed. The test set will then be applied to estimate the overall performance of these models. The final model for 3-level CHF assessment was one with the best performance among the three models.

6.2.5 Validation and Performance

In our work, we randomly divided the segment data into training, validation and test sets at a ratio of 3:1:1. The training, validation and test set were mutual independent. The training set was applied to construct multiple initial models under different parameter settings, include SVM classifier parameters and feature combinations. Then, the validation set was applied in model construction and provided an unbiased evaluation of model fitting during parameter tuning to avoid over-fitting. Finally, the test set was applied to evaluate the final model using optimal parameters. To measure the performance of the optimal time scale of data with the SVM classifier for CHF assessment, we computed some widely used parameters [124] for binary classification to make a comparison with others, as described above. All calculations were performed using the software MATLAB 9.3.0.71359 (version R2017b, The MathWorks, Inc., Natick, MA, USA).

6.3 Results

6.3.1 The Optimal Time Scale for CHF Risk Assessment

With the significant analysis of HRV measurement at different time scales (1h, 2h, 3h, 4h, 6h, 8h, and 24h), the discrimination power of data at different time scales was compared. All these features were described in Subsection 6.2.2. Among the three risk groups: no risk (NR), lower risk (LR) and higher risk (HR), p values between every two groups (NR-LR, NR-HR, LR-HR) were achieved and the biggest one was used as the comparative indices. The significant statistical level is at 0.05. In this study, eight random selections of segment data were done to equalize sample size.

First, we compared risk assessment ability of 1h and 2h data using statistical analysis among three risk groups. The 1h data were analyzed at each random selection and the biggest p-value among three risk groups was compared with 2h data. Fig.6.2 is the bar graph of the biggest p-value of HRV measurements among 3-level risk groups at 1h and 2h. There were five features showed the biggest p-value under 0.05 for both 1h and 2h data, i.e. R, CV, LF, SE, FE, as shown in Fig.6.2. Among these five features,

four of them (R, CV, SE, FE) showed stable significant difference with all biggest p-value < 0.05 at each random selection, except for LF. This means that only four features (R, CV, SE, FE) fit Rule 1 and 2. Table 6.2 showed the mean of the biggest p-value of the four HRV measurement at 1h and 2h (=average of 8 random selection). According to Table 6.2, R, CV, and SE at 2h had a lower p-value than at 1h; while FE of both 2h and 1h showed a highly significant characteristic with $p<0.001$. According to the Criteria, these results demonstrated that HRV measurements at 2h have better classification capability in CHF assessment than those at 1h.

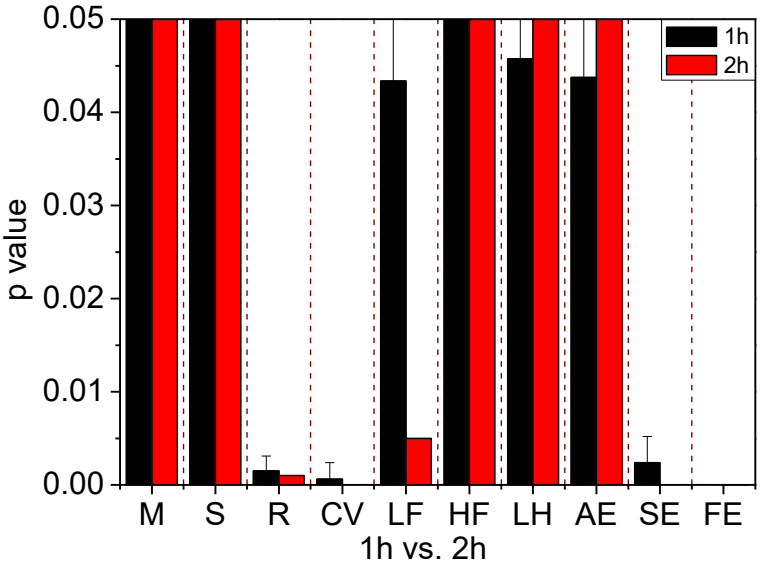

Fig. 6.2 Bar graph of the biggest p-value of features between 1h and 2h. The x-axis is the ten HRV measures described in Section 6.2; the y-axis is the biggest p-value of the features among groups.

Table 6.2 Mean of biggest p-value of features at 1h and 2h.

	R	CV	SE	FE
1h	0.002	0.001	0.002	<0.001
2h	0.001	<0.001	<0.001	<0.001

Meaning of all Features described in Method Part

A similar comparison process was applied to data at other time scales. As segment data at 2h were better in CHF assessment, we compared the statistical analysis results of HRV measurements at 2h and 3h. here, the 2h data were randomly selected to

equalize sample size and the statistical results were presented with the biggest p-value. Fig.6.3 is the bar graph of the biggest p-value of HRV measurements among 3-level risk groups at 2h and 3h. There are five features with the biggest p values under 0.05 for both 2h and 3h data, i.e. R, CV, LF, SE, FE, as shown in Fig.6.3. Four of them showed stable significance as all the biggest p-value is under 0.05 at each random selection, except for LF. This means that only four features (R, CV, SE, FE) fit Rule 1 and 2. Table 6.3 showed the mean of the biggest p-value of the four HRV measurements at 2h (=average of 8 random selection) and 3h. According to Table 6.3, CV and SE at 2h had lower p-value than at 3h; while p-value of R at 2h was higher than 3h. For FE, both 2h and 3h data showed a significant difference with $p<0.001$. Thus, these results demonstrated that HRV measurements at 2h have better CHF assessment capability than at 3h.

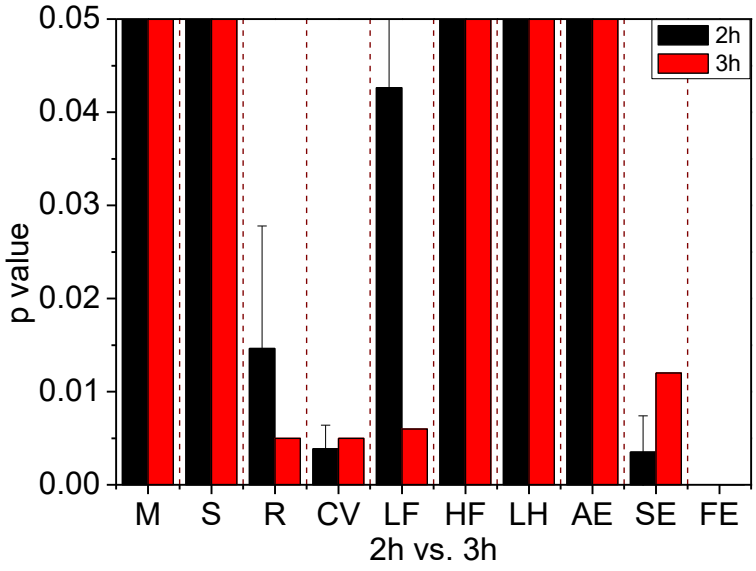

Fig. 6.3 Bar graph of the biggest p-value of features between 2h and 3h. The x-axis is the ten HRV measures described in Section 6.2; the y-axis is the biggest p-value of the features among groups.

Table 6.3 Mean of biggest p-value of features at 2h and 3h

	R	CV	SE	FE
3h	0.005	0.005	0.012	<0.001
2h	0.015	0.004	0.004	<0.001

Meaning of all Features described in Method Part

Then a comparison of HRV measurement happened between 2h and 4h. Here, The 2h data were randomly selected to equalize sample size and the statistical results were presented with the mean of the biggest p-value. Fig.6.4 is the bar graph of the biggest p-value of HRV measurements among 3-level risk groups at 2h and 4h. There were four features with the mean of the biggest p values under 0.05 for both 2h and 4h data, i.e. R, CV, SE, FE, as shown in Fig.6.4. Two of them (SE, FE) showed stable significance as all biggest p-value is under 0.05 at each random selection, see Fig.6.4. This means that only the two features (SE, FE) fit Rule 1 and 2. Table 6.4 showed the biggest p-value of all HRV measurement at 2h (=average of 8 random selection) and 4h. SE of 2h data has lower p-value than at 4h and significant values of FE were extremely small (<0.001) at both 2h and 4h. Besides, 2h is shorter than 4h. Thus, HRV measurement at 2h has better ability in CHF assessment than at 4h.

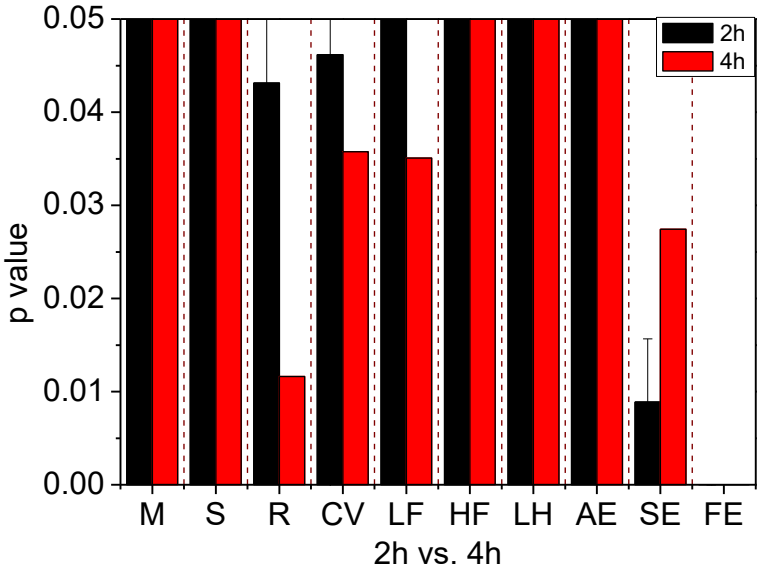

Fig. 6.4 Bar graph of the biggest p-value of features between 2h and 4h. The x-axis is the ten HRV measures described in Section 6.2; the y-axis is the biggest p-value of the features among groups.

Table 6.4 Mean of biggest p-value of features at 2h and 4h

	SE	FE
4h	0.027	<0.001
2h	0.009	<0.001

Meaning of all Features described in Method Part

Next, comparison of HRV measurement happened between 2h and 6h. Here, the 2h data were randomly selected to equalize sample size and the statistical results were presented with the mean of the biggest p-value. Fig.6.5 is the bar graph of the biggest p-value of HRV measurements among 3-level risk groups at 2h and 6h. Only FE showed a stable significance with biggest p values under 0.05 for both 2h and 6h data, as shown in Fig.6.5. FE showed remarkable significant (<0.01) at both 2h and 6h. Furthermore, 2h is much shorter than 6h. Thus, HRV measurement at 2h has better ability in CHF assessment than at 6h.

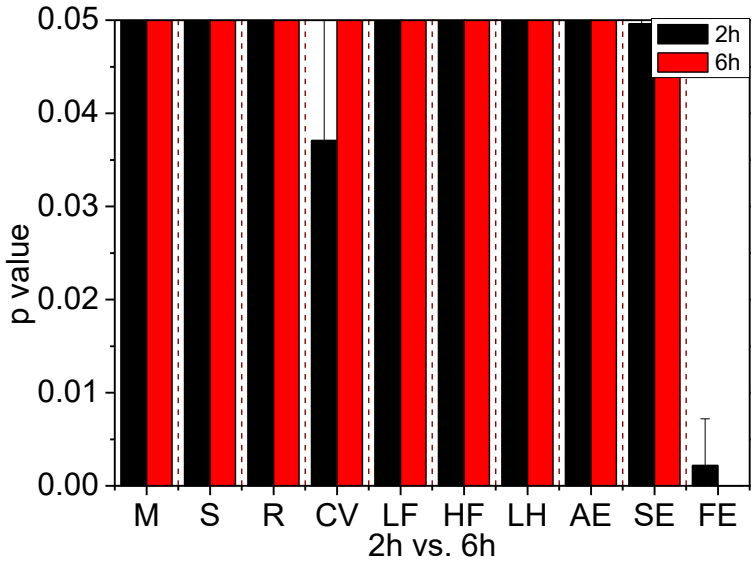

Fig. 6.5 Bar graph of the biggest p-value of features between 2h and 6h. The x-axis is the ten HRV measures described in Section 6.2; the y-axis is the biggest p-value of the features among groups.

Fig.6.6 is the bar graph of the biggest p-value of HRV measurements among 3-level risk groups at 2h and 8h. Here, 2h data were randomly selected to balance sample size and the statistical results were presented with the mean of the biggest p-value. Statistical analysis of HRV measurement of 2h and 8h showed that no single feature yielded a stable significant difference among the three risk groups. As 2h is much shorter than 8h, HRV measurement at 2h is more practical in CHF assessment than 8h.

The final comparison is between 2h and 24h. Fig.6.7 is the bar graph of the biggest p-value of HRV measurements among 3-level risk groups at 2h and 24h. Here, The 2h

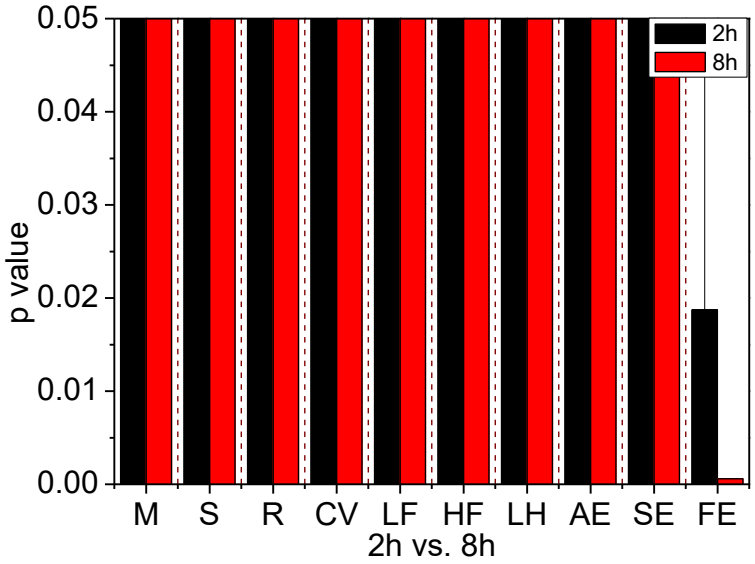

Fig. 6.6 Bar graph of the biggest p-value of features between 2h and 8ph. The x-axis is the ten HRV measures described in Section 6.2; the y-axis is the biggest p-value of the features among groups.

data were randomly selected to equalize sample size and the statistical results were presented with the mean of the biggest p-value. There was no feature had the biggest p-value under 0.05 for both 2h and 24h data, as shown in Fig.6.7. Considering that 2h is much shorter than 24h, these results demonstrated that HRV measurement at 2h has better ability in CHF assessment than at 24h.

In conclusion, we have found HRV measurements at 2h has optimal CHF assessment capability in comparison with at other time scales.

6.3.2 Classification Performance of 2h Data in 3-level CHF Assessment

With optimal time scale analysis for CHF assessment, 2h data showed the best performance in risk assessment among No risk (NR), lower risk (LR), and higher risk (HR) patients. Then HRV measurements at 2h were applied into the 3-level CHF assessment using support vector machine (SVM) classifier and exhaustive search feature selection method.

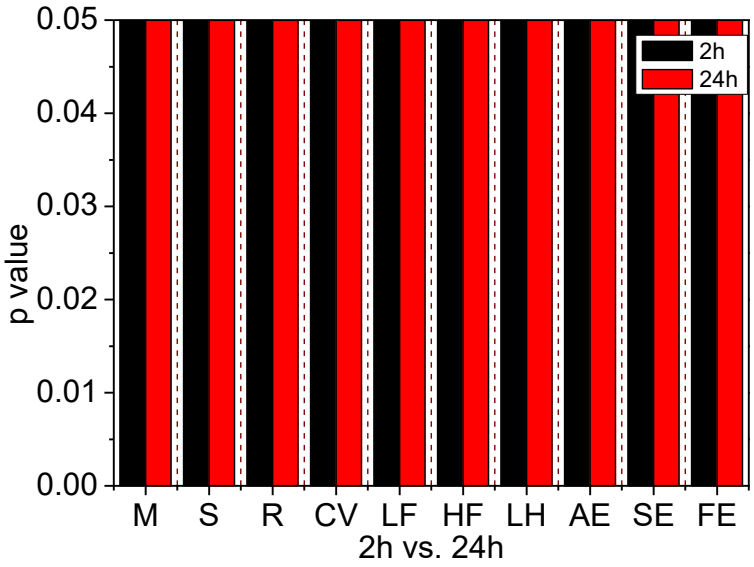

Fig. 6.7 Bar graph of the biggest p-value of features between 2h and 24h. The x-axis is the ten HRV measures described in Section 6.2; the y-axis is the biggest p-value of the features among groups.

In this study, the 3-level risk assessment was done as two binary classifications based on SVM. As described in the Section 6.2, there are three possible binary combinations, which indicated a 2-layer classification. All possible 2-level binary classification combinations included NR vs. (LR vs. HR), LR vs. (NR vs. HR) and HR vs. (NR vs. LR). The input of all classifications consists of the optimal feature subset from the exhaustive search feature selection. Results of the validation set and test set were shown using the widely-used parameters described in the Section 6.2 under each combination.

1) Classification performance at layer 1

Table 6.5 shows the performance of validation and test set using SVM classifier of all 3 possible binary classification combinations in layer 1: NR vs. LR&HR, LR vs. NR&HR and HR vs. NR&LR. From results in Table 6.5, it is clear that the optimal performance accessed in classification between NR vs. LR&HR with an accuracy of 87.88%, sensitivity of 85.48% and specificity of 91.89%. The classification performance between LR and NR&HR was the worst, as the lowest accuracy (<80%) and large difference existing between sensitivity and specificity (over 50%). The classification performance between HR and NR&LR is worse than the classification performance

between NR and LR&HR with a difference of accuracy < 8%. In addition, the performance of validation and testing set under three binary combinations were close with a difference under 10%.

Table 6.5 Performance of validation and testing set under possible binary classification combinations at layer 1

Layer 1		Acc/%	Sen/%	Spe/%	n	p	tp	tn	fp	fn
NR vs. rest	**Validation**	91.49	87.93	97.22	36	58	51	35	1	7
	Testing	**87.88**	**85.48**	**91.89**	**37**	**62**	**53**	**34**	**3**	**9**
LR vs. rest	**Validation**	84.04	97.14	45.83	24	70	68	11	13	2
	Testing	77.78	91.78	38.46	26	73	67	10	16	6
HR vs. rest	**Validation**	81.91	81.67	82.35	34	60	49	28	6	11
	Testing	80.81	84.13	75	36	63	53	27	9	10

Performance in the table was with optimal feature subset after exhaustive search feature selection NR: healthy samples (no risk); LR: lower risk samples; HR: higher risk samples; Acc: accuracy; Sen: sensitivity; Spe: specificity; Pre: precision; n: negative; p: positive; tp: true positive; tn: true negative; fp: false positive; fn: false negative.

2) Classification performance at layer 2

According to the three binary classifications in layer 1, there are three corresponding classification combinations: LR vs. HR, NR vs. HR, NR vs. LR at layer 2. The performance of validation and test set under the three combinations was shown in Table 6.6. As previously described, the input of all classifications was the optimal feature subset.

Table 6.6 Performance of validation and testing set with optimal feature subset at layer 2

Layer 2		Acc/%	Sen/%	Spe/%	n	p	tp	tn	fp	fn
LR vs. HR	**Validation**	75.86	94.12	50	24	34	32	12	12	2
	Testing	81.13	90.91	65	20	33	30	13	7	9
NR vs. HR	**Validation**	92.86	85.29	100	36	34	29	36	0	2
	Testing	**92.54**	**87.5**	**97.14**	**35**	**32**	**28**	**34**	**1**	**6**
NR vs. LR	**Validation**	98.33	100	97.22	36	24	24	35	1	11
	Testing	86.79	75	91.89	37	16	12	34	3	10

NR: healthy sample (no risk); LR: lower risk samples; HR: higher risk samples; Acc: accuracy; Sen: sensitivity; Spe: specificity; Pre: precision; n: negative; p: positive; tp: true positive; tn: true negative; fp: false positive; fn: false negative.

From the results in Table 6.6, the classification between NR and HR performed best among the three classification combinations. The accuracy, sensitivity, and specificity were 92.54%, 87.5% and 97.14%, respectively. All of these three classifications reached an accuracy of over 80%. The classification performance between NR vs. LR was better

than between LR and HR with a slight difference (<7% difference in accuracy). In addition, the performance between validation and test set under the three combinations was also close.

3) Classification performance of the final model for 3-level risk assessment

With the analysis of the 2-layer binary classification above, three different risk assessment models of CHF have achieved: NR vs. (LR vs. HR), LR vs. (NR vs. HR) and HR vs. (NR vs. LR). Combining the performance of the three models at layer 1 and 2, the overall performance of each model was shown in Table 6.7. Table 6.7 is the overall performance of validation and test set using the SVM classifier and exhaustive search in three models. This table was obtained via generating results in Table 6.5 and 6.6 to find the optimal model for a 3-level risk assessment of CHF. The results of the three models were shown using the accuracy and precision of each risk level.

Table 6.7 The overall performance of validation and testing set with optimal feature subset

	NR vs. (LR vs. HR)		**LR vs. (NR vs. HR)**		**HR vs. (NR vs. LR)**	
	Validation set	**Testing set**	**Validation set**	**Testing set**	**Validation set**	**Testing set**
Acc%	84.04	**77.78**	80.85	72.73	92.55	73.74
Pre_NR%	97.22	**91.89**	100	97.14	97.22	91.89
Pre_LR%	50	**65**	45.83	38.46	100	75
Pre_HR%	94.12	**90.91**	85.29	87.5	82.35	75

NR: healthy samples (no risk); LR: lower risk samples; HR: higher risk samples; Acc: accuracy; Pre: precision.

According to Table 6.7, the overall classification performance of the model in an order of NR vs. (LR vs. HR) was the best compared to the other two models. Precisions of NR and HR were good in this order, which was 91.89% and 90.91%, respectively. Furthermore, the classification ability among risk levels differed in the 3 models. The precisions of the healthy sample (NR) in all three classification models were good (all >90%) and the highest value was found in the model with the classification order of LR vs. (NR vs. HR) with 97.14%. The best accuracy in higher-risk patients (HR) was associated with a classification order of NR vs. (LR vs. HR) (90.91%). The accuracy in lower risk patients (LR) under all classification was not good (under 80%). The best recognition ability of LR was obtained while the classification order was HR vs. (NR vs. LR) with an accuracy of only 75 %.

6.4 Discussion

In this study, the optimal time scale for a 3-level risk assessment of congestive heart failure (CHF) was proved to be 2h and a 2-layer model was then constructed using 2h ECG data. The optimal time scale analysis was performed using statistical analysis while the model construction was based on support vector machine (SVM) and exhaustive search feature selection method. 116 ECG records from MIT/BIH database were included. The final model achieved precisions in the healthy sample (no risk: NR), lower risk patients (LR) and higher risk patients (HR) as 91.89%, 65%, and 90.91% respectively (see Table 6.7).

As a common syndrome, CHF has been analyzed by many researchers, from risk factors to prognosis. Risk assessment of CHF had attracted a lot of attention and researchers analyzed various physiological signals associated with CHF, in particular, ECG. Among various measurements of ECG, heart rate variability (HRV) presented excellent CHF assessment and indicated a deteriorating change in the cardiac autonomic nervous system (ANS) [56]. Mortara et al. found that HRV was significantly lower in CHF patients with abnormal heart autonomic nerve function [126]. As described in the Section 6.1, the relation between HRV measures and the ANS system was confirmed. For CHF patients, the balance between sympathetic and parasympathetic nervous systems of the cardiovascular system was impaired, especially in patients in end-stage, which is related to HRV decrease [119]. Thus, we hypothesize that classical HRV measures can be applied to 3-level risk assessment as well as the autonomic imbalance. In this study, 10 classical HRV measures were calculated for 3-level risk assessments.

Table 6.8 highlights the comparison with related studies using HRV in CHF assessments using the same database. There are two main areas of CHF assessment: CHF detection (classification between healthy people and CHF patients) and CHF quantification (classification between different risk level CHF patients).

In CHF detection, 5-min/24-h HRV analysis confirmed its risk assessment capability. Isler et al. utilized 5-min HRV measurements with a KNN classifier for CHF detection, resulting in an accuracy of 96.39 % [68]. In 2012, Yu et al. applied a support vector machine (SVM) classifier and genetic algorithm (GA) into CHF recognition based on

Table 6.8 Highlight

Reference	Class	Sample	Time scale	Feature Selection	Classifier
Isler et al.	**NR vs. P**	83	5min	GA	KNN
Yu et al.	**NR vs. P**	83	24h	GA	SVM
Melillo et al.	**LR vs. HR**	44	24h	ESM	CART
This work	**NR vs. LR vs. HR**	47	2h	ESM	SVM

NR: healthy people (no risk); P: CHF patient; LR: lower risk CHF patients; HR: higher risk CHF patients; GA: genetic algorithm; ESM: exhaustive search method; KNN: k-nearest neighbor; SVM: support vector machine; CART: classification and regression tree.

bi-spectral HRV analysis and achieved an accuracy of 98.79% [91]. The performance in CHF detection was excellent (over 90% of accuracy) and validated the potential of HRV measures in detection. As to CHF quantification, little research work has been done. Melillo et al. first attempted to use 24-h HRV measures in distinguishing low-risk CHF patients from higher risk ones with a sensitivity and specificity of 93.3% and 63.6% [95]. The result showed that HRV can be used in CHF quantification but not good in detection. Thus, we concluded that previous work indicated that HRV at 5-min/24-h could be used in a 3-level risk assessment of CHF. The main problem is that if we used 5-min data into risk assessment, it would be hard to decide which 5-min segment is appropriate to assess a subject's condition as 5-min RR intervals are sensitive to physiological change, such as emotion, exercise, and so on [143–145]. Besides, it has been proven that the circadian variation existed in CHF patients [62]. One of our previous work applied all 5-min segments of 116 24-h RR intervals into CHF detection but resulted in the only accuracy of 72.44%. This proved that the information obtained from the 5-min segment was not stable and insufficient to describe changes. These characteristics limited the application of 5-min HRV. While using 24-h data, the HRV fluctuation will be covered in a day, as discussed in our previous work [45]. Besides, it is difficult to obtain 24-h RR intervals data of CHF patients with good quality and sufficient size. Thus, we intended to find a suitable time scale of HRV analysis for 3-level risk assessment. 10 classical HRV measures were then calculated and analyzed using statistical analysis and 3-level classification.

The statistical analysis was done at seven kind of time scales of data, including 1h, 2h, 3h, 4h, 6h, 8h, and 24h. In this study, the time scales were selected by following reasons: 5min and 24h are two of the most classical time scales used by former research.

8h is a close time scales to whole night sleep. 4h, 6h, 8h are selected with echelon distribution. 1h, 2h, 3h, and 4h were also selected with echelon distribution. As described in Section 6.3, the optimal time scale was 2-h in 3-level risk assessment according to statistical analysis. This result was achieved according to the criteria in this study at the Section 6.3 based on stability, separability, and availability. The optimal time scale of data for risk assessment of CHF should show better ability in differentiating different risk groups and low-cost. While comparing different time scales, the 24-h data were segmented into various scales, so the segment number is different at each time scale. To decrease the influence of segment number to statistical analysis, eight random selection of segment was done to equal while comparing data at different time scales. In addition, we compared data at different time scales from small to large step by step, instead of comparison between every two time scales to get faster processing. Here, 5-min data showed its characteristic as discussed above, the usability of 5-min HRV in 3-level risk assessment was proved inferior [146] and was not included. While searching for researches focused on comparing different time scales of data, not much work has been found. Maria et al. found that 8-min HRV from bedside ECG provided additional important prognosis information during controlled breathing than 5-min data [62]. This work discussed the influence of breath to short-term data, Besides, Jong et al. [148] explored optimal timing in screening CHF patients from healthy persons and found 7 PM to 9 PM showed superlative screening performance. The optimal timing is a 2-h duration, which is helpful to identify suitable measurement in non-invasive monitoring and assessment of CHF in the clinical practice.

After applying statistical analysis, the 2-h HRV measures were then applied into the SVM classifier. The exhaustive search feature selection helped to improve the classification model. In this study, the classification accuracy of CHF detection using 2-h data is 91.49% (shown in Table 6.5 NR vs. (LR&HR)). The results were as good as other work using 5-min and 24-h data. Furthermore, sensitivity and specificity of 90.91% and 65% in classifying LR from HR were achieved applying 2-h HRV in SVM (shown in Table 6.6: LR vs. HR). This result indicated that in CHF quantification, the performance of our work is as good as Melillo's work, but with a shorter time scale. Thus, we concluded that the optimal time scale - 2h - in this study had a comparative

ability in risk assessment of CHF to former work and can be served as a robust and reliable alternative in the future.

In clinical diagnosis, it is difficult to diagnose CHF patients at lower risk (NYHA I-II). One reason is that there is no obvious symptom in the early stage of CHF [8]. This results in less chance to diagnose early. The other reason is that even under examination in the hospital, an accurate diagnosis is not simple for a cardiologist. Clinical diagnoses of CHF include three stages: physical examination, history information, and diagnostic tests. According to the hospital examination description, several problems exist. First, hospital resources are generally limited. During hospital examination, no single test can diagnose heart failure, which means the diagnosis process described above need multiple departments cooperation and multitool detection. These often result in time delay and high cost. Thus, in-hospital examinations bring heavy financial burden and time-consuming [137]. Second, the diagnosis by a cardiologist requires rich experience and a rigorous attitude. This means possible human error during diagnosis. Finally, some examinations are invasive and could bring potential pain to patients. These facts resulted in a decrease possibility to early diagnose lower risk patients in the hospital. In this study, a non-invasive, cheap and robust method was proposed using only 2-hour ECG data in CHF prognosis and achieving precisions in healthy persons (NR), lower risk patients (LR) and higher risk patients (HR) as 91.89%, 65%, and 90.91% respectively. Difficulty in prognosis lower-risk patients also helps to explain that only 13 out of 20 lower-risk patients were truly recognized during classification.

Though the performance of the 2-h data in this study showed a good result, there exist some limitations. First, the subject size in this research is still limited, especially for CHF patients at a lower risk. An abundant data collection should be obtained for a more comprehensive analysis. Besides, only SVM classifiers with Gaussian kernel function were used for 3-level risk assessment of CHF. In the future, classifiers that can handle big data samples should be included. The seven time scales were not uniformly distributed, and a finer time scale might be required. Heartbeats in a time scale are personally different, which might be another research aspect in the future. Finally, the application of our method in actual home monitoring should be tested using on wearable devices with high-quality ECG signals.

6.5 Conclusion

In this work, we applied statistical analysis to optimal time scale analysis and then used a support vector machine (SVM) techniques to evaluate 3-level risk assessments for CHF. A feature selection method was based on an exhaustive search process. All possible binary classifications were searched during evaluating the performance of data at optimal time scale. The data used in this study included 116 samples of ECG records from the MIT/BIH database, and 10 classical HRV measurements were calculated. According to the study, we reached several conclusions:

1) In this study, we applied HRV measurements at the optimal time scale (2-h) to construct a 3-level risk assessment model of CHF. The model achieved good precisions in healthy persons (NR), lower risk patients (LR) and higher risk patients (HR) as 91.89%, 65% and 90.91% (see Table 6.7).

2) We found the optimal time scale is 2-h while using HRV for 3-level risk assessment, based on significant analyses among data at seven different time scales according to The Criteria. The sample size was balanced using a random selection for comparability of different time scales.

3) Comparing to previous research, the 2-h data showed comparable performance both in CHF detection and quantification, comparing with both 5-min and 24-h data.

Considering these advantages, HRV techniques at the optimal time scale (2-h) is robust and reliable for CHF risk assessments in clinical practice, especially in home-monitoring in the future.

Chapter 7

An Automatic Method to Differentiate Sleep Apnea Patients with Congestive Heart Failure Using HRV

This Chapter focused on the classification between sleep apnea and congestive heart failure using HRV, to build a more comprehensive cardiopulmonary assessment system.

7.1 Introduction

Over the past decades, numerous studies have explored cardiovascular variation in electrocardiograph (ECG) recording of cardiopulmonary patients, especially in sleep apnea [64] and congestive heart failure (CHF) [50]. For patients with obstructive sleep apnea, breath disorder condition went worse resulting to heart pumping failure [65]. While patients with CHF and those with sleep breathing disorders shared common symptoms of obstructive and central apnea. Even worse, both diseases are associated with a high risk of death and financial loss [65].

As the in diagnosing apnea, polysomnography (PSG) has been widely used in monitoring sleep breathing and the gold standard for the diagnosis of sleep apnea

[149]. PSG is a comprehensive sleep study recording multiple physiological signals, including electroencephalograph (EEG), Electrocardiograph (ECG), respiratory airflow, electrooculogram (EOG) and electromyogram (EMG). In all these signals, ECG showed the heart rhythm of the human body during sleep. The variation in the time intervals between heart beats (beat-to-beat interval) is heart rate variability (HRV), which can be derived from the ECG signal. There are many reports using HRV for the detection of sleep apnea and CHF [49, 64, 66, 67].

Technically, a majority of researchers have applied three approaches to determining HRV: the time domain analysis, frequency domain analysis, and nonlinear analysis [66]. Time and frequency domain analyses address variability degrees and the underlying rhythms, respectively [49], while nonlinear analysis processes the dynamic variation features in HRV generated from physiological systems [50], e.g. complexity.

Statistical analysis based on these three approaches of HRV analysis showed a significant difference between samples from non-CHF people and CHF patients and classifier based automatic detection method for CHF had achieved the accuracy over 90%. Decreased time and frequency domain HRV measurement was a marker of sympathoexcitation [54, 55]; this measurement is of clinical significance. Besides, the nonlinear analysis provides an improved decomposition of cycle fluctuation of HRV in a period and achieves significant discrimination power for CHF detection [59].

In sleep apnea detection using heart rate variability, Ravelo et al. (2015) used the permutation entropy of HRV and quadratic discriminant analysis and reached an AUC of 91.7% [64]. Li et al. (2018) used the classic HRV and deep neural network and hidden Markov model to detect sleep apnea, achieving an accuracy of 85% based on 1-min ECG data [67]. Take together, these researches showed that there was the potential to apply HRV for detection of sleep apnea and CHF. However, currently, there are no autonomic methods for distinguishing between sleep apnea patients with and without CHF.

In this study, an autonomic model was constructed to distinguish between sleep apnea patients with and without CHF. Twenty whole night PSG records were downloaded and preprocessed to obtain suitable ECG data. Then the ECG data were segmented into multiple 5-minute epochs to calculate six classical HRV measures.

Comparison of data was analyzed using one-way ANOVA. Next, a support vector machine classifier was applied for optimal model construction. The exhaustive search was used to improve the performance of the final model. Finally, the classification performance was evaluated using classical measures and cross-validation.

7.2 Methods

7.2.1 Sample Data

The data used in this work were from the Sleep Heart Health Study database implemented by The National Sleep Research Resource (NSRR) [150]. The NSRR was supported by Grant Number HL114473 from the National Heart, Lung, and Blood Institute, NIH.

Whole night Polysomnography (PSG) data of sleep apnea patients were obtained. The whole reports provided accounts of patients with and without congestive heart failure (CHF). the apnea/hypopnea index (AHI) was used to evaluate the severity of apnea. Twenty sleep apnea patients (with AHI between 5 and 30) were included in this study, 10 of whom had CHF (aged 72.3±6.60 years) and 10 without CHF (aged 72.2±6.64 years). CHF information was recorded in cardiovascular outcome report file of the dataset (https://sleepdata.org/datasets/shhs).

The whole night PSG data was downloaded to get sleep apnea information and cardiovascular information. In this study, the raw ECG signal was processed for measure calculation and model construction, detailed shown in Fig.7.1. To calculate heart rate variability (HRV) measures, the Pan-Tompkins algorithm (PT) was applied to detect R peaks [151]. All R-peak data were then manually checked after PT detection to remove possible false detection and sensory loss. These processes guaranteed the accuracy of R peak data. Then RR interval data (difference between adjacent R peaks) were calculated. Here, RR intervals longer than 3 seconds were excluded according to the ACC/AHA Guideline [152]. These steps were performed in case of unstable measurement conditions and artificial error.

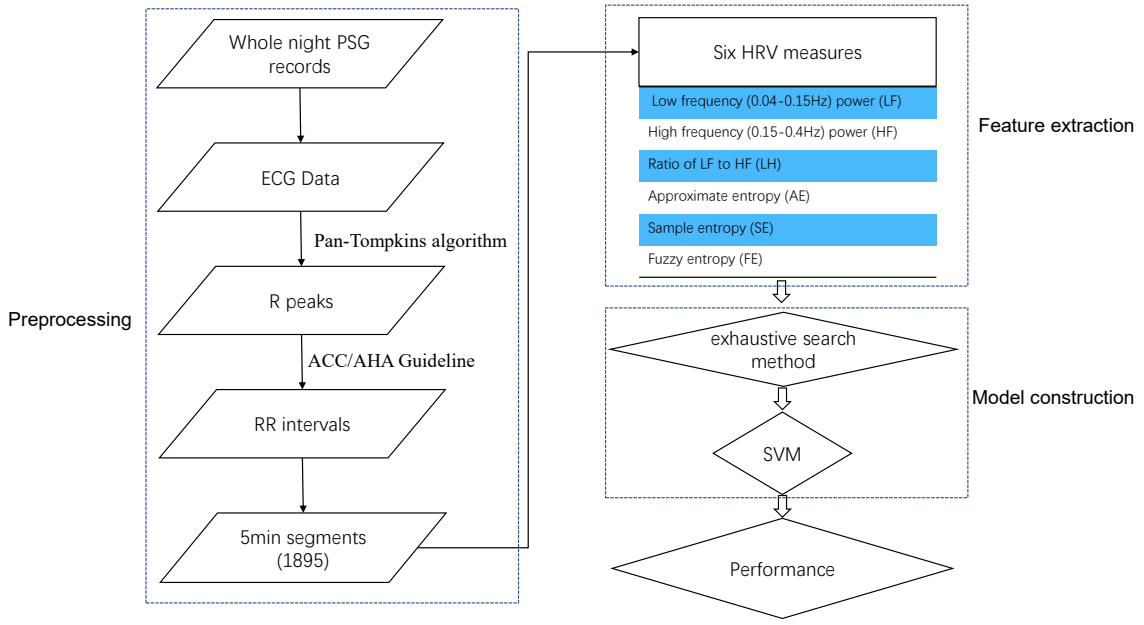

Fig. 7.1 Workflow.

7.2.2 HRV Calculation

After preprocessing, whole night RR interval data of sleep apnea patients with and without congestive heart failure (CHF) were obtained. The whole night RR interval data were segmented into multiple 5-minute segments in sequence. In this research, a total of 1895 epochs from 20 PSG records were deemed suitable for HRV analysis. Six classical HRV measures of all these epochs were then calculated to evaluate differences in the cardiovascular autonomic nervous system.

The six calculated classical HRV measures included: (1) frequency measures: low frequency (0.04-0.15Hz) power of RR intervals (LF); high frequency (0.15-0.4Hz) power of RR intervals (HF); ratio of LF to HF (LH) [153]; (2) complexity measures: approximate entropy of RR intervals (AE); sample entropy of RR intervals (SE) [45]; fuzzy entropy of RR intervals (FE) [114]. Fast Fourier transform (FFT) based power spectral density measurement was performed for different spectral components. All settings for entropy calculation were the same. All these measures were described in Ref. [152].

In this study, the six HRV measures were obtained to analyze the cardiovascular difference between sleep apnea patients with and without CHF. All these measures addressed two perspectives: statistical level and complexity to track dynamic change trend of apnea patients in 5-minute cycles. The statistical level difference was shown via the frequency domain measures, while the complexity was reflected by the three entropy measures.

All measures are presented as mean±standard deviation (see Table 7.1). Differences between groups are assessed using one-way analysis of variance (one-way ANOVA). Statistical tests are conducted with a significant level set at 0.05. A p-value <0.05 is considered statistically significant and <0.001 is considered extremely significant. All analysis were performed using the SPSS software (version 19, SPSS Inc., Chicago, IL, USA).

7.2.3 Model Construction

In this study, support vector machine classifier was applied into the automatic model construction to distinguish sleep apnea patients with congestive heart failure CHF from those without CHFs.

Support vector machine (SVM) is a widely used classifier in machine learning. The advantage of SVM is in binary classification assigning new samples to one category or the other [80]. SVM algorithm will separate input features by a hyperplane to reach the largest margin between sleep apnea patients with and without CHF, also called maximum-margin hyperplane [81].

The hyperplane was calculated based on the function:

$$f(x) = W^T x + b, \begin{cases} f(x) > 0, then & x \in class \quad 1 \\ f(x) < 0, then & x \in class \quad 2 \end{cases} \tag{7.1}$$

Here, W and b are the weight vector and bias that maximize the margin, respectively. x is the feature vectors of one sample. The W and b were then decided by the Lagrangian function.

As mentioned in the Subsection 7.2.1, there are two different groups: sleep apnea patients with CHF (CHF) and without CHF (NoCHF). There were six HRV measures calculated for each 5-minute epochs. All measures were done with statistical difference analysis to evaluate their ability in distinguishing the two groups.

During model construction procedure, train set and test set were randomly selected in a ratio of 9:1. All these six HRV measures with significant difference between the two groups were applied into the classifier. In order to improve the performance of the classifier, an exhaustive search method (ESM) - also called brute-force search method - was applied to get the optimal feature subset. The ESM is a very general method that consist of all possible subset combinations for classification between sleep apnea patients between with CHFs and without CHFs.

As mentioned above, there are six features obtained. Thus, there were 63 feature subsets to be evaluated. Among these feature subsets, each feature subset was inputted into an SVM classifier as a Wrapper to evaluate the classification performance. Classification performance was verified to avoid over-training.

To evaluate the performance of a constructed model, some widely used parameters were defined. Let TP is the number of true positive, TN is the number of true negatives, FP is the number of false positive, and FN is the number of false negatives. Then, the accuracy, sensitivity, and specificity can be calculated:

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \tag{7.2}$$

$$Sensitivity = TP/(TP + FN) \tag{7.3}$$

$$Specificity = \ TP/(TN + FP) \tag{7.4}$$

Hence, an autonomic model for detection of sleep apnea patients with and without CHF was constructed. The optimal model is the one with high accuracy and balanced sensitivity and specificity. All calculations were performed using the software MATLAB 9.3.0.71359 (version R2017b, The MathWorks, Inc., Natick, MA, USA).

7.3 Results and Discussion

To construct an autonomic model to distinguish sleep apnea patients with and without Congestive heart failure (CHF), 1895 short term (5-minute) Electrocardiograph (ECG) data from whole night Polysomnography (PSG) of twenty subjects were analyzed. First, the Pan-Tompkins algorithm was applied to get accurate RR interval data with a manual check. Then, six HRV measures were calculated and statistically analyzed. Finally, support vector machine classifier and exhaustive search method were applied to build the optimal model.

In this work, RR interval data was achieved from the raw ECG signal of PSG. Pan-Tompkins algorithm is one of the widely used methods in detecting R-peak in ECG [151]. Here, Pan-Tompkins algorithm application and manually check help guarantee the accuracy and reliability of RR interval data.

With the R-peaks data, RR intervals were calculated. Any RR intervals > 3 seconds were deleted, as described in Ref [152], RR intervals longer than 2.5 seconds indicating an unusual heart situation and heart attack happened. Thus, RR intervals longer than 3 seconds were considered as unreliable data in this study.

All twenty subjects had AHI over 5, indicating that they suffered from the sleep apnea syndrome according to the updated 2012 American Academy of Sleep Medicine (AASM) respiratory scoring rules. The AHI information was given on SHHS dataset with sleep specialist notation. In the future, sleep apnea patients with different AHI level should be compared. Besides, all subjects in the two groups had a close age distribution as described in the Subsection 7.2.1. Thus, age on the cardiovascular system was not a factor for consideration in this study.

Table 7.1 shows the mean and standard deviation of HRV measures of sleep apnea patients with and without CHF (CHF vs. NoCHF). More subjects should be included in future work.

Table 7.2 shows the p-value of HRV measures between sleep apnea patients with and without CHF. According to Table 7.2, all measures showed significance ($p<0.001$), except for sample entropy ($p=0.003$), suggesting that all measures showed good ability to distinguishing sleep apnea patients with CHF from those without CHFs.

Table 7.1 Mean and standard deviation (SD) of heart rate variability (HRV) measures of sleep apnea patients with and without congestive heart failure

Measures	Apnea patients	Mean	Standard deviation
Low-frequency power	**NoCHF**	0.0016	0.0020
	CHF	0.0023	0.0043
High-frequency power	**NoCHF**	0.0008	0.0013
	CHF	0.002	0.0044
Ratio of low to high-frequency power	**NoCHF**	1.725	2.3012
	CHF	1.0538	0.9542
Approximate entropy	**NoCHF**	-0.2098	0.0615
	CHF	-0.1952	0.0618
Sample entropy	**NoCHF**	0.4607	0.2194
	CHF	0.4288	0.2432
Fuzzy entropy	**NoCHF**	0.0809	0.0442
	CHF	0.0963	0.0782

N: normal samples; P: CHF patients, in which 1 is of NYHA I-II, 2 is of NYHA III, 3 is of NYHA III-IV; SI: static indices; DI: dynamic indices; * represents that significance level of features were under 0.1; C-SVM: classical SVM.

Table 7.2 *p*-value of heart rate variability (HRV) measures between sleep apnea patients with and without congestive heart failure

Measures	Low-frequency power	High-frequency power	Ratio of low to high-frequency power	Approximate entropy	Sample entropy	Fuzzy entropy
P value	<0.001	<0.001	<0.001	<0.001	0.003	<0.001

N: normal samples; P: CHF patients, in which 1 is of NYHA I-II, 2 is of NYHA III, 3 is of NYHA III-IV; SI: static indices; DI: dynamic indices; * represents that significance level of features were under 0.1; C-SVM: classical SVM.

The measures used in this study include three frequency domain measures and three entropy measures. The frequency domain measures reflected interactions of sympathetic and parasympathetic autonomic modulation [120], and marks sympathovagal unbalance levels. According to Table 7.2, it shows an increase mean value at low- and high-frequency power, while a decrease at ratio of low to high-frequency power. This indicated that the unbalance level of the cardiovascular autonomic nervous system increased. Besides, the entropy of HRV measures reflected the cardiovascular system autonomic balance in different aspects [56]. Entropy can be used to assess the temporal regularity of short-term time series with its anti-interference performance. Thus, the increasing trend of all three entropy measures of HRV could be used to explain cardiovascular system dynamic regularity decrease in cardiovascular autonomic regulation (as shown in Table 7.1). Although the standard deviation is a little high among these measures,

the statistical significant differences indicated that HRV could be a potential indicator for sleep apnea with CHF prognosis in clinical studies.

According to Table 7.2, all measures were with a significant difference. Therefore, all measures were applied as the input of SVM classifier. In order to improve the classifier performance, the exhaustive search method was used with 63 feature subsets inputted. According to the results, the optimal feature subset was high-frequency power, the ratio of low to high-frequency power, approximate entropy, and fuzzy entropy with an accuracy increase of 9.23%. With this input set, an accuracy of 81.68% was achieved. Fig.7.2 is the confusion matrix of the final model. According to Fig.7.2, 78 out of 103 sleep apnea patients without CHF were truly recognized, and 78 out of 88 sleep apnea patients with CHF were truly recognized. Thus, the sensitivity was 88.64% and specificity was 75.73%. The difference between sensitivity and specificity is 12.93%, indicating that the performance of this model is robust.

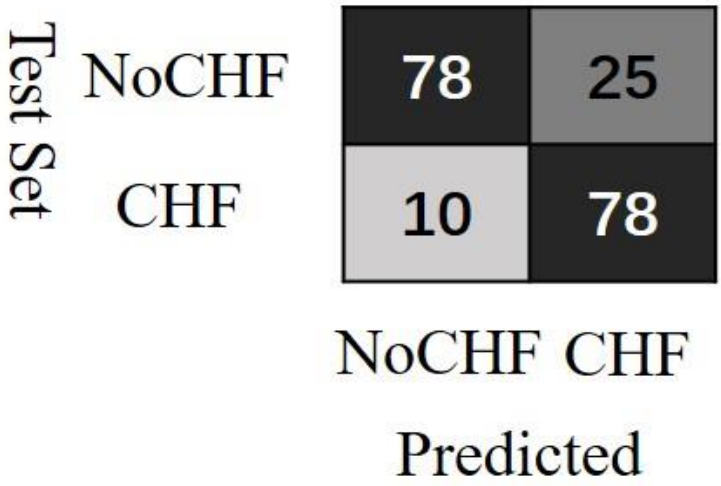

Fig. 7.2 Confusion matrix.

Thus, an autonomic model was successfully constructed to distinguish sleep apnea patients with CHF from without CHFs. The performance showed that 5-minute HRV measures from overnight PSG could potentially be used for CHF prognosis in sleep apnea patients.

The classification performance in this study was good for prognosis for CHF and sleep apnea. In future work, different classification algorithms and parameter setting optimization methods application will be our goal to improve the performance. Besides, 5-minute epochs were analyzed to extract classic HRV measures in this study. In recent years, there are different methods proposed in HRV analysis, thus more analysis in HRV to show the fluctuation of whole night change will be another direction.

7.4 Conclusion

In this study, an autonomic model was constructed to distinguish between sleep apnea patients with and without congestive heart failure and achieved an accuracy of 81.68%. The result indicated that 5-minute ECG signals had the potential in distinguishing apnea patients with and without CHF. The frequency and entropy features showed a clear significant difference between the two groups, and the classification performance showed that the model constructed could be used clinically. For future work, more subjects should be included to improve the application.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

This dissertation examines the risk assessment of congestive heart failure (CHF) using the ECG signal and intelligent algorithm. First, a novel dynamic heart rate variability (HRV) was proposed in autonomic unbalance for congestive heart failure stratification. Based on these features, a novel and effective method for CHF detection and quantification was proposed using dynamic heart rate variability measurement and 24-h data. Then an automatic CHF detection method was based on an unsupervised sparse auto encoder based deep learning algorithm with 5-min RR intervals. Next, an automatic risk assessment method of CHF using ECG at optimal timescale - 2h - was performed. Finally, an automatic method to differentiate sleep apnea patients with and without CHF using HRV from Polysomnography records was attempted for early warning of the comprehensive cardiopulmonary system.

The five parts of this dissertation have been achieved. Conclusions are detailed below.

1. In the first part, we aimed to explore dynamic indices performance in risk stratification of samples of CHF and normal people. A total of 116 samples were analysed, including 72 no risk (NR), 12 low risk (LR) and 32 high risk (HR) samples. Ten widely used classical HRV measures and 30 dynamic indices were

obtained from nominal 24-hour recordings in three perspectives: time domain, frequency domain and complexity to quantify risk levels. Significance difference analysis was performed among groups with these indices. The results of this study showed that dynamic indices found in this study can be considered as a marker in risk stratification of CHF patients with the following characteristics:

(1) Three dynamic indices (MLH, SDM and FELH) had a monotonic and significance different reduction along with severity, which indicated good discrimination power in quantifying three risk groups;

(2) Dynamic indices showed good power in differentiating LR and HR, while classical 24h indices did not;

(3) Three quarters indices showed clear difference among three risk levels.

The characteristics above suggested that the dynamic indices studied here could be useful for risk stratification and quantitatively evaluating autonomic nervous system (ANS) dysfunction. They might potentially be applied in clinical settings as a useful complement to clinical scales.

2. In the second part, to construct a multistage risk assessment model of CHF, we applied dynamic HRV measurements to CHF detection and quantification. The decision-tree based support vector machine (DT-SVM) algorithm and feature selection method based on backward elimination and significance analysis were included. The data used for this objective consisted of 126 dynamic indices (DI) and 54 static indices (SI) of HRV obtained from 116 samples of 24-h ECG records from the MIT/BIH database. According to the study, we reached several conclusions:

(1) In this part, we built a 4-level risk assessment model for CHF detection and quantification based on the DT-SVM algorithm. The model succeeded with a total accuracy of 96.61 %, in risk assessment among individuals of N (no risk), P1 (mild risk), P2 (moderate risk), and P3 (severe risk).

(2) We creatively proposed dynamic indices for the severity evaluation of CHF. In CHF quantification, the DI is obviously superior to the SI, increasing the accuracy from 73.91% to 91.30%.

(3) The DT-SVM-based multistage risk assessment model proposed in this work significantly improved discrimination power from 76.27 % to 96.61 % when compared to C-SVM. The performance of our classifier improved based on the combination of the BE algorithm and significance analysis.

(4) According to the analysis of SI, it was clear that SI of HRV was more powerful in disease detection than DI with an accuracy of 98.31 %, which was about 20% higher. This is consistent with the results of prior research regarding disease detection.

In light of these advantages, the stratifying CHF risk assessment model will be a reliable and objective prognostic marker for routine clinical application (especially daily health nursing) in the future.

3. In the third part, to develop a risk assessment model using 5-min RR intervals in CHF detection, a two-layer deep neural network model was constructed based on the sparse-auto-encoder (SAE)-based deep learning (DL) algorithm. The results indicated that RR intervals have the potential for CHF detection but are sensitive to body condition. Though the classification accuracy for the overall segments only reached 72.41%, further analysis demonstrated the efficiency of the proposed method and interpreted the rationality of the achieved classification results in CHF detection. It is certain that RR intervals contain information about cardiovascular condition. Actually, when analysing classification results of segments of each person, it showed that less than 30 % segments of 90.28 % (65/72) healthy persons were classified as negative, and higher than 30 % segments of 77.27 % (34/44) CHF patients were classified as negative. This supports the position that RR intervals can be used as a CHF detection indicator, but this is sensitive to test condition variations. This also demonstrates the dynamic change with body condition might not be fully reflected if we only analyse short/long term RR variability. Thus, the proposed assessment method is reliable for CHF detection and prognosis.

4. In the fourth part, we applied statistical analysis to optimal time scale analysis and then used support vector machine (SVM) techniques to evaluate 3-level risk assessments for CHF. A feature selection method was based on an exhaustive

search process. All possible binary classifications were searched during the evaluation of the performance data at optimal time scale. The data used in this study included 116 samples of ECG records from the MIT/BIH database, and 10 classical HRV measurements were calculated. According to the study, we reached several conclusions:

(1) In this study, we applied HRV measurements at the optimal time scale (2-h) to construct a 3-level risk assessment model of CHF. The model achieved good precisions in healthy persons (NR), lower risk patients (LR) and higher risk patients (HR) as 91.89%, 65% and 90.91% (see Table 6.7).

(2) We found the optimal time scale is 2-h when using HRV for 3-level risk assessment, based on significance analyses among data at seven different time scales according to The Criteria. The sample size was balanced using a random selection for comparability of different time scales.

(3) Compared to previous research, the 2-h data showed comparable performance both in CHF detection and quantification using both 5-min and 24-h data.

Considering these advantages, HRV techniques at the optimal time scale (2-h) is robust and reliable for CHF risk assessments in clinical practice, especially in home-monitoring in the future.

5. In the fifth part, Twenty whole night PSG data from the Sleep Heart Health Study database were included in this study. The Pan-Tompkins algorithm was applied to the electrocardiograph signal to detect R peaks of the QRS complex. The whole night R peaks data were then manually checked and segmented into 1895 5-minute epochs to calculate three frequency domains and three nonlinear heart rate variability measures. All these measures were analysed for their statistical differences between groups (sleep apnea with and without CHF). Finally, a binary support vector machine classifier and extreme search method were performed to construct the model. An autonomic model was constructed to distinguish between sleep apnea patients with and without congestive heart failure and this achieved an accuracy of 81.68%. The result indicated that 5-minute ECG signals had the potential to distinguish apnea patients with and without CHF. The frequency and entropy features showed a clear significance difference between the

two groups, and the classification performance showed that the model constructed could be used clinically. For future work, more subjects should be included to improve the application.

In conclusion, dynamic indices of HRV is good at CHF stratification and 24-h data based multistage CHF quantification achieves a robust performance but is limited in terms of its accessibility. 5-min data with an unsupervised deep learning based CHF detection model proves the potential of short term data in CHF detection but is limited by its sensitivity to clinical condition changes. Based on these results, the optimal timescale analysis for risk assessment of CHF was performed and proved to have a comparable ability in 3-level assessment of CHF using 2-h data. This indicates our methods are applicable in robust, fast and convenient assessment for CHF, especially in home monitoring. Furthermore, due to the high prevalence between CHF and sleep apnea, we explored an automatic model for cardiopulmonary system risk assessment and achieved good performance. Thus, this project could be applied in wearable devices for convenient, robust and comprehensive early warning and assessment for CHF patients, especially in home monitoring.

8.2 Future Work

There are many areas that can still be explored based on our current project.

First, to achieve better clinical and industrial application, a larger and more complex balanced dataset with patients is needed, including different countries, age distributions, weight scales and genders.

Next, more involvement of cardiologists and sleep specialists should be sought to better understand clinical needs and improve this project.

Then, to embedded our model in industrial devices like an Apple watch, the algorithm efficiency and computing time should be improved. One of the important prospective areas in the risk assessment of CHF is to apply better artificial algorithms [154]. In Chapter 5, an unsupervised aprse quto encoder based deep learning algorithm was applied into over 30,000 5-min RR intervals segment and this proved the potential

of unsupervised CHF detection based on a deep learning algorithm. Recently, the long-short term memory (LSTM) neural network algorithm has attracted much attention in disease detection [155]. In the year of 2019, Wang et al applied the LSTM combining with an inception module from the GoogLeNet in CHF detection using different RR segment types[156]. The classification performance based on the blindfold validation showed an accuracy over 98%. Moreover, Mallya et al. claims that the RNN based LSTM algorithm could help in CHF onset prediction as it can achieve an AUCROC of 0.9147 [157]. These papers proved that short term HRV analysis can help clinicians in CHF prognosis, especially in healthcare monitoring. Accordingly, we could explore the LSTM algorithm performance in the monitoring of heart beat function to help improve early warning of cardiopulmonary system disease.

Another prospective area for exploration is multi signal fusion for CHF risk assessment. As we have mentioned in the introduction to this thesis, multiple ECG components have shown their power in CHF detection (see Subsection 2.6.3). The performance of short segment ECG signals has been proved in Ref. [158] with good results. Thus, there is good potential to help CHF prognosis and monitoring using more ECG information.

Finally, considering the clinical importance of heart function assessment, there are more biosignals which might help to improve the robustness and effectiveness of comprehensive assessment system for cardiopulmonary risk, including CHF assessment. As described in Chapter 7, CHF and sleep apnea discrimination based on whole night HRV from PSG records has achieved a good performance. For further heart function analysis of the cardiopulmonary system, the breath signal is one of the most important factors, like exhaled hydrogen [159, 160]. The medication information should also be explored in future to enhance the clinical explanation [161]. All of this information could be included in future analysis of CHF and sleep apnea discrimination based on results of this doctoral research.

References

[1] T. Kishi, "Heart failure as an autonomic nervous system dysfunction," *Journal of cardiology*, vol. 59, no. 2, pp. 117–122, 2012.

[2] M. Markaity, "Congestive heart failure: An'f'isn't an option," *Nursing made Incredibly Easy*, vol. 10, no. 2, pp. 13–23, 2012.

[3] W. H. Maisel and L. W. Stevenson, "Atrial fibrillation in heart failure: epidemiology, pathophysiology, and rationale for therapy," *The American journal of cardiology*, vol. 91, no. 6, pp. 2–8, 2003.

[4] M. Fikrle, T. Paleček, P. Kuchynka, E. Němeček, L. Bauerová, J. Straub, and R. Ryšavá, "Cardiac amyloidosis: A comprehensive review," *cor et vasa*, vol. 55, no. 1, pp. e60–e75, 2013.

[5] C. E. Skotzko, "Symptom perception in chf:(why mind matters)," *Heart failure reviews*, vol. 14, no. 1, pp. 29–34, 2009.

[6] D. J. Flavell, D. Boehm, A. Noss, S. Warnes, and S. U. Flavell, "Therapy of human t-cell acute lymphoblastic leukaemia with a combination of anti-cd7 and anti-cd38-saporin immunotoxins is significantly better than therapy with each individual immunotoxin," *British journal of cancer*, vol. 84, no. 4, pp. 571–578, 2001.

[7] I. Anand, J. J. McMurray, J. Whitmore, M. Warren, A. Pham, M. A. McCamish, and P. B. Burton, "Anemia and its relationship to clinical outcome in heart failure," *Circulation*, vol. 110, no. 2, pp. 149–154, 2004.

[8] L. W. S. Carol Flavell, "Take heart with heart failure," *Circulation*, vol. 104, no. 18, pp. e89–e91, 2001.

[9] A. L. Bui, T. B. Horwich, and G. C. Fonarow, "Epidemiology and risk profile of heart failure," *Nature Reviews Cardiology*, vol. 8, no. 1, p. 30, 2011.

[10] M. A. Pfeffer, B. Claggett, S. F. Assmann, R. Boineau, I. S. Anand, N. Clausell, A. S. Desai, R. Diaz, J. L. Fleg, I. Gordeev, *et al.*, "Regional variation in patients and outcomes in the treatment of preserved cardiac function heart failure with an aldosterone antagonist (topcat) trial," *Circulation*, vol. 131, no. 1, pp. 34–42, 2015.

[11] J. P. Saul, "Beat-to-beat variations of heart rate reflect modulation of cardiac autonomic outflow," *Physiology*, vol. 5, no. 1, pp. 32–37, 1990.

[12] A. G. Barnett, M. De Looper, and J. F. Fraser, "The seasonality in heart failure deaths and total cardiovascular deaths," *Australian and New Zealand journal of public health*, vol. 32, no. 5, pp. 408–413, 2008.

[13] F. H. S. Kuniyoshi, A. Garcia-Touchard, A. S. Gami, A. Romero-Corral, C. van der Walt, S. Pusalavidyasagar, T. Kara, S. M. Caples, G. S. Pressman, E. C. Vasquez, *et al.*, "Day–night variation of acute myocardial infarction in obstructive sleep apnea," *Journal of the American College of Cardiology*, vol. 52, no. 5, pp. 343–346, 2008.

[14] R. S. Leung and T. Douglas Bradley, "Sleep apnea and cardiovascular disease," *American journal of respiratory and critical care medicine*, vol. 164, no. 12, pp. 2147–2165, 2001.

[15] M. C. Khoo, T.-S. Kim, and R. B. Berry, "Spectral indices of cardiac autonomic function in obstructive sleep apnea," *Sleep*, vol. 22, no. 4, pp. 443–451, 1999.

[16] T. D. Bradley and J. S. Floras, "Sleep apnea and heart failure: Part ii: central sleep apnea," *Circulation*, vol. 107, no. 13, pp. 1822–1826, 2003.

[17] R. N. Aurora, S. P. Patil, and N. M. Punjabi, "Portable sleep monitoring for diagnosing sleep apnea in hospitalized patients with heart failure," *Chest*, vol. 154, no. 1, pp. 91–98, 2018.

[18] M. R. Cowie and A. M. Gallagher, "Sleep disordered breathing and heart failure: what does the future hold?," *JACC: Heart Failure*, vol. 5, no. 10, pp. 715–723, 2017.

[19] H. K. Khattak, F. Hayat, S. V. Pamboukian, H. S. Hahn, B. P. Schwartz, and P. K. Stein, "Obstructive sleep apnea in heart failure: review of prevalence, treatment with continuous positive airway pressure, and prognosis," *Texas Heart Institute Journal*, vol. 45, no. 3, pp. 151–161, 2018.

[20] L. M. Donovan, L. C. Feemster, E. M. Udris, M. F. Griffith, L. J. Spece, B. N. Palen, K. He, S. Parthasarathy, K. P. Strohl, V. K. Kapur, *et al.*, "Poor outcomes among patients with chronic obstructive pulmonary disease with higher risk for undiagnosed obstructive sleep apnea in the lott cohort," *Journal of Clinical Sleep Medicine*, vol. 15, no. 01, pp. 71–77, 2019.

[21] L. Shen, P. S. Jhund, M. C. Petrie, B. L. Claggett, S. Barlera, J. G. Cleland, H. J. Dargie, C. B. Granger, J. Kjekshus, L. Køber, *et al.*, "Declining risk of sudden death in heart failure," *New England Journal of Medicine*, vol. 377, no. 1, pp. 41–51, 2017.

[22] S. Javaheri, F. Barbe, F. Campos-Rodriguez, J. A. Dempsey, R. Khayat, S. Javaheri, A. Malhotra, M. A. Martinez-Garcia, R. Mehra, A. I. Pack, *et al.*, "Sleep apnea: types, mechanisms, and clinical cardiovascular consequences," *Journal of the American College of Cardiology*, vol. 69, no. 7, pp. 841–858, 2017.

[23] J. Li, X. Gao, J. Han, G. Wang, L. Kang, and E. Ji, "Early cardiac injury in patients with obstructive sleep apnea," *Zhongguo ying yong sheng li xue za zhi= Zhongguo yingyong shenglixue zazhi= Chinese journal of applied physiology*, vol. 34, no. 5, pp. 457–461, 2018.

[24] J. Espiritu, "Health consequences of obstructive sleep apnea," *J Sleep Disord Ther*, vol. 8, no. 307, pp. 2167–0277, 2019.

[25] M. C. Iliou, S. Corone, B. Gellen, T. Denolle, F. Roche, A. Bellemain-Apaix, M. Bigot, M. E. Lopes, J.-L. Bussiere, and C. Darné, "0087: Sleep apneas treatment during cardiac rehabilitation can improve heart failure prognosis? satelit-hf study: Sleep apnea treatment during cardiac rehabilitation of chf patients," *Archives of Cardiovascular Diseases Supplements*, vol. 8, no. 1, p. 30, 2016.

[26] A. J. Sommerfeld, A. D. Althouse, J. Prince, and G. W. Hickey, "Obstructive sleep apnea is associated with increased readmissions in chf patients," *Journal of Cardiac Failure*, vol. 22, no. 8, p. S88, 2016.

[27] J.-L. Pepin, R. Tamisier, T. Damy, F. Goutorbe, A. Palot, P. Levy, J.-M. Davy, F. Lavergne, L. Morin, and M.-P. D'ortho, "Morbidity and mortality of chronic heart failure (chf) patients with central sleep apnoea (csa) treated by adaptive servoventilation (asv): Interim results of face cohort study," in *A19. SLEEP DISORDERED BREATHING, CARDIOVASCULAR DISEASE, AND MORTALITY*, pp. A1040–A1040, American Thoracic Society, 2018.

[28] J. L. Fleg, I. L. Piña, G. J. Balady, B. R. Chaitman, B. Fletcher, C. Lavie, M. C. Limacher, R. A. Stein, M. Williams, and T. Bazzarre, "Assessment of functional capacity in clinical and research applications: An advisory from the committee on exercise, rehabilitation, and prevention, council on clinical cardiology, american heart association," *Circulation*, vol. 102, no. 13, pp. 1591–1597, 2000.

[29] M. Senni, C. M. Tribouilloy, R. J. Rodeheffer, S. J. Jacobsen, J. M. Evans, K. R. Bailey, and M. M. Redfield, "Congestive heart failure in the community: a study of all incident cases in olmsted county, minnesota, in 1991," *Circulation*, vol. 98, no. 21, pp. 2282–2289, 1998.

[30] R. Levin, M. Dolgin, C. Fox, and R. Gorlin, "The criteria committee of the new york heart association: Nomenclature and criteria for diagnosis of diseases of the heart and great vessels," *LWW Handbooks*, vol. 9, p. 344, 1994.

[31] A. F. Members, J. J. McMurray, S. Adamopoulos, S. D. Anker, A. Auricchio, M. Böhm, K. Dickstein, V. Falk, G. Filippatos, C. Fonseca, *et al.*, "Esc guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The task force for the diagnosis and treatment of acute and chronic heart failure 2012 of the european society of cardiology. developed in collaboration with the heart failure association (hfa) of the esc," *European journal of heart failure*, vol. 14, no. 8, pp. 803–869, 2012.

[32] H. M. Azzazy and R. H. Christenson, "B-type natriuretic peptide: physiologic role and assay characteristics," *Heart failure reviews*, vol. 8, no. 4, pp. 315–320, 2003.

[33] K. M. Fussell, J. A. Awad, and L. B. Ware, "Case of fulminant hepatic failure due to unrecognized peripartum cardiomyopathy," *Critical care medicine*, vol. 33, no. 4, pp. 891–893, 2005.

[34] S. Kennedy, B. Simon, H. J. Alter, and P. Cheung, "Ability of physicians to diagnose congestive heart failure based on chest x-ray," *The Journal of emergency medicine*, vol. 40, no. 1, pp. 47–52, 2011.

[35] A. Bellone, A. Barbieri, C. Ricci, E. Iori, M. Donateo, M. Massobrio, and S. Bendinelli, "Acute effects of non-invasive ventilatory support on functional mitral regurgitation in patients with exacerbation of congestive heart failure," *Intensive care medicine*, vol. 28, no. 9, pp. 1348–1350, 2002.

[36] B. A. Lutomsky, T. Rostock, A. Koops, D. Steven, K. Müllerleile, H. Servatius, I. Drewitz, D. Ueberschär, T. Plagemann, R. Ventura, *et al.*, "Catheter ablation of paroxysmal atrial fibrillation improves cardiac function: a prospective study on the impact of atrial fibrillation ablation on left ventricular function assessed by magnetic resonance imaging," *Europace*, vol. 10, no. 5, pp. 593–599, 2008.

[37] T. F. Robinson, S. M. Factor, and E. H. Sonnenblick, "The heart as a suction pump," *Scientific American*, vol. 254, no. 6, pp. 84–91, 1986.

[38] K.-L. He, D. Burkhoff, W.-X. Leng, Z.-R. Liang, L. Fan, J. Wang, and M. S. Maurer, "Comparison of ventricular structure and function in chinese patients with heart failure and ejection fractions> 55% versus 40% to 55% versus< 40%," *The American journal of cardiology*, vol. 103, no. 6, pp. 845–851, 2009.

[39] K. O. Healy, C. A. Waksmonski, R. K. Altman, P. D. Stetson, A. Reyentovich, and M. S. Maurer, "Perioperative outcome and long-term mortality for heart failure patients undergoing intermediate-and high-risk noncardiac surgery: Impact of left ventricular ejection fraction," *Congestive heart failure*, vol. 16, no. 2, pp. 45–49, 2010.

[40] E. H. Conn, R. S. Williams, and A. G. Wallace, "Exercise responses before and after physical conditioning in patients with severely depressed left ventricular function," *The American journal of cardiology*, vol. 49, no. 2, pp. 296–300, 1982.

[41] X. Zheng and X. Song, "The economic development of china's demographic transition, and chronic growth," *Journal of social science in colleges and universities in China*, vol. 4.

[42] I. R. Legarreta, P. S. Addison, M. Reed, N. Grubb, G. R. Clegg, C. E. Robertson, and J. N. Watson, "Continuous wavelet transform modulus maxima analysis of the electrocardiogram: Beat characterisation and beat-to-beat measurement," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 3, no. 01, pp. 19–42, 2005.

[43] R. Dhingra, M. J. Pencina, T. J. Wang, B.-H. Nam, E. J. Benjamin, D. Levy, M. G. Larson, W. B. Kannel, R. B. D'Agostino Sr, and R. S. Vasan, "Electrocardiographic qrs duration and the risk of congestive heart failure: the framingham heart study," *Hypertension*, vol. 47, no. 5, pp. 861–867, 2006.

[44] L. G. Tereshchenko, I. Cygankiewicz, S. McNitt, R. Vazquez, A. Bayes-Genis, L. Han, S. Sur, J.-P. Couderc, R. D. Berger, A. B. De Luna, *et al.*, "Predictive value of beat-to-beat qt variability index across the continuum of left ventricular dysfunction: competing risks of noncardiac or cardiovascular death and sudden or nonsudden cardiac death," *Circulation: Arrhythmia and Electrophysiology*, vol. 5, no. 4, pp. 719–727, 2012.

[45] W. Chen, L. Zheng, K. Li, Q. Wang, G. Liu, and Q. Jiang, "A novel and effective method for congestive heart failure detection and quantification using dynamic heart rate variability measurement," *PloS one*, vol. 11, no. 11, p. e0165304, 2016.

[46] A. J. Camm, M. Malik, J. T. Bigger, G. Breithardt, S. Cerutti, R. J. Cohen, P. Coumel, E. L. Fallen, H. L. Kennedy, R. Kleiger, *et al.*, "Heart rate variability: standards of measurement, physiological interpretation and clinical use. task force of the european society of cardiology and the north american society of pacing and electrophysiology," 1996.

[47] G. Liu, L. Wang, Q. Wang, G. Zhou, Y. Wang, and Q. Jiang, "A new approach to detect congestive heart failure using short-term heart rate variability measures," *PloS one*, vol. 9, no. 4, p. e93399, 2014.

[48] P. Melillo, R. Fusco, M. Sansone, M. Bracale, and L. Pecchia, "Discrimination power of long-term heart rate variability measures for chronic heart failure detection," *Medical & biological engineering & computing*, vol. 49, no. 1, pp. 67–74, 2011.

[49] P. K. Stein, M. S. Bosner, R. E. Kleiger, and B. M. Conger, "Heart rate variability: a measure of cardiac autonomic tone," *American heart journal*, vol. 127, no. 5, pp. 1376–1381, 1994.

[50] J.-L. Chen, Y.-J. Tseng, H.-W. Chiu, T.-C. Hsiao, and W.-C. Chu, "Nonlinear analysis of heart rate dynamics in hyperthyroidism," *Physiological measurement*, vol. 28, no. 4, p. 427, 2007.

[51] A. Rao, H. C. Ryoo, A. Akin, and H. H. Sun, "Classification of heart rate variability (hrv) parameters by receiver operating characteristics (roc)," in *Proceedings of the IEEE 28th Annual Northeast Bioengineering Conference (IEEE Cat. No. 02CH37342)*, pp. 167–168, IEEE, 2002.

[52] G. D'Addio, A. Porta, C. Picone, E. Ranaudo, G. Furgi, and F. Rengo, "Reproducibility of the heart rate variability regularity index in chronic heart failure patients," in *Computers in Cardiology, 2003*, pp. 449–452, IEEE, 2003.

[53] R. Maestri, G. D. Pinna, A. Accardo, P. Allegrini, R. Balocchi, G. D'ADDIO, M. Ferrario, D. Menicucci, A. Porta, R. Sassi, *et al.*, "Nonlinear indices of heart rate variability in chronic heart failure patients: redundancy and comparative

clinical value," *Journal of cardiovascular electrophysiology*, vol. 18, no. 4, pp. 425–433, 2007.

[54] M. G. Kienzle, D. W. Ferguson, C. L. Birkett, G. A. Myers, W. J. Berg, and D. J. Mariano, "Clinical, hemodynamic and sympathetic neural correlates of heart rate variability in congestive heart failure," *The American journal of cardiology*, vol. 69, no. 8, pp. 761–767, 1992.

[55] T. B. Kuo, T. Lin, C. C. Yang, C.-L. Li, C.-F. Chen, and P. Chou, "Effect of aging on gender differences in neural control of heart rate," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 277, no. 6, pp. H2233–H2239, 1999.

[56] M. M. J. De Jong and D. C. Randall, "Heart rate variability analysis in the assessment of autonomic function in heart failure," *Journal of Cardiovascular Nursing*, vol. 20, no. 3, pp. 186–195, 2005.

[57] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[58] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical review letters*, vol. 89, no. 6, p. 068102, 2002.

[59] Y.-C. Lin, Y.-H. Lin, M.-T. Lo, C.-K. Peng, N. E. Huang, C. C. Yang, and T. B. Kuo, "Novel application of multi dynamic trend analysis as a sensitive tool for detecting the effects of aging and congestive heart failure on heart rate variability," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 26, no. 2, p. 023109, 2016.

[60] C.-K. Peng, M. Costa, and A. L. Goldberger, "Adaptive data analysis of complex fluctuations in physiologic time series," *Advances in adaptive data analysis*, vol. 1, no. 01, pp. 61–70, 2009.

[61] L. Pecchia, P. Melillo, M. Sansone, and M. Bracale, "Discrimination power of short-term heart rate variability measures for chf assessment," *IEEE Transactions on Information Technology in biomedicine*, vol. 15, no. 1, pp. 40–46, 2010.

[62] S. Miyamoto, M. Fujita, K. Tambara, H. Sekiguchi, S. Eiho, K. Hasegawa, and S.-i. Tamaki, "Circadian variation of cardiac autonomic nervous activity is well preserved in patients with mild to moderate chronic heart failure: effect of patient position," *International journal of cardiology*, vol. 93, no. 2-3, pp. 247–252, 2004.

[63] M. Soliński, J. Gierałtowski, and J. Żebrowski, "Modeling heart rate variability including the effect of sleep stages," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 26, no. 2, p. 023101, 2016.

[64] A. Ravelo-García, J. Navarro-Mesa, U. Casanova-Blancas, S. Martin-Gonzalez, P. Quintana-Morales, I. Guerra-Moreno, J. Canino-Rodríguez, and E. Hernández-Pérez, "Application of the permutation entropy over the heart rate variability

for the improvement of electrocardiogram-based sleep breathing pause detection," *Entropy*, vol. 17, no. 3, pp. 914–927, 2015.

[65] T. D. Bradley and J. S. Floras, "Sleep apnea and heart failure: Part i: obstructive sleep apnea," *Circulation*, vol. 107, no. 12, pp. 1671–1678, 2003.

[66] T. F. o. t. E. S. o. C. t. N. A. S. o. P. Electrophysiology, "Heart rate variability: standards of measurement, physiological interpretation, and clinical use," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[67] K. Li, W. Pan, Y. Li, Q. Jiang, and G. Liu, "A method to detect sleep apnea based on deep neural network and hidden markov model using single-lead ecg signal," *Neurocomputing*, vol. 294, pp. 94–101, 2018.

[68] Y. İşler and M. Kuntalp, "Combining classical hrv indices with wavelet entropy measures improves to performance in diagnosing congestive heart failure," *Computers in biology and medicine*, vol. 37, no. 10, pp. 1502–1510, 2007.

[69] H. H. Bafroui and A. Ohadi, "Application of wavelet energy and shannon entropy for feature extraction in gearbox fault detection under varying speed conditions," *Neurocomputing*, vol. 133, pp. 437–445, 2014.

[70] J. K. Kanters, N.-H. HOLSTEIN-RATHLOU, and E. Agner, "Lack of evidence for low-dimensional chaos in heart rate variability," *Journal of cardiovascular electrophysiology*, vol. 5, no. 7, pp. 591–601, 1994.

[71] N. M. De Souza, L. C. M. Vanderlei, and D. M. Garner, "Risk evaluation of diabetes mellitus by relation of chaotic globals to hrv," *Complexity*, vol. 20, no. 3, pp. 84–92, 2015.

[72] M. Brennan, M. Palaniswami, and P. Kamen, "Do existing measures of poincare plot geometry reflect nonlinear features of heart rate variability?," *IEEE transactions on biomedical engineering*, vol. 48, no. 11, pp. 1342–1347, 2001.

[73] J. W. Kantelhardt, E. Koscielny-Bunde, H. H. Rego, S. Havlin, and A. Bunde, "Detecting long-range correlations with detrended fluctuation analysis," *Physica A: Statistical Mechanics and its Applications*, vol. 295, no. 3-4, pp. 441–454, 2001.

[74] A. H. Shirazi, M. R. Raoufy, H. Ebadi, M. De Rui, S. Schiff, R. Mazloom, S. Hajizadeh, S. Gharibzadeh, A. R. Dehpour, P. Amodio, *et al.*, "Quantifying memory in complex physiological time-series," *PLoS One*, vol. 8, no. 9, p. e72854, 2013.

[75] R. J. Storella, H. W. Wood, K. M. Mills, J. K. Kanters, M. V. Højgaard, and N.-H. Holstein-Rathlou, "Approximate entropy and point correlation dimension of heart rate variability in healthy subjects," *Integrative physiological and behavioral science*, vol. 33, no. 4, pp. 315–320, 1998.

[76] B. P. Kovatchev, L. S. Farhy, H. Cao, M. P. Griffin, D. E. Lake, and J. R. Moorman, "Sample asymmetry analysis of heart rate characteristics with application to neonatal sepsis and systemic inflammatory response syndrome," *Pediatric research*, vol. 54, no. 6, p. 892, 2003.

[77] F. Bailly, G. Longo, and M. Montevil, "A 2-dimensional geometry for biological time," *Progress in Biophysics and Molecular Biology*, vol. 106, no. 3, pp. 474–484, 2011.

[78] R. O. Duda and P. E. Hart, "Dg stork pattern classification," *Wiley-I nterscience, Ne w Y or k,*, vol. 200, no. 1, p. 1, 2001.

[79] K. Z. Mao, "Orthogonal forward selection and backward elimination algorithms for feature subset selection," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 1, pp. 629–634, 2004.

[80] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[81] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *et al.*, "Scikit-learn: Machine learning in python," *Journal of machine learning research*, vol. 12, no. Oct, pp. 2825–2830, 2011.

[82] F. Takahashi and S. Abe, "Decision-tree-based multiclass support vector machines," in *Proceedings of the 9th International Conference on Neural Information Processing, 2002. ICONIP'02.*, vol. 3, pp. 1418–1422, IEEE, 2002.

[83] G. Madzarov, D. Gjorgjevikj, and I. Chorbev, "A multi-class svm classifier utilizing binary decision tree," *Informatica*, vol. 33, no. 2, 2009.

[84] W. C. Levy, D. Mozaffarian, D. T. Linker, S. Sutradhar, S. Anker, A. Cropp, I. Anand, A. Maggioni, P. Burton, M. Sullivan, *et al.*, "The seattle heart failure model," *Circulation*, vol. 113, no. 11, pp. 1424–1433, 2006.

[85] J.-Y. Tabet, D. Logeart, C. Geyer, C. Guiti, P. Ennezat, M. Dahan, and A. Cohen-Solal, "Comparison of the prognostic value of left ventricular filling and peak oxygen uptake in patients with systolic heart failure," *European heart journal*, vol. 21, no. 22, pp. 1864–1871, 2000.

[86] S. A. Rezaeieh, A. Zamani, K. Bialkowski, and A. Abbosh, "Foam embedded wideband antenna array for early congestive heart failure detection with tests using artificial phantom with animal organs," *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 11, pp. 5138–5143, 2015.

[87] L. Zhao, S. Wei, C. Zhang, Y. Zhang, X. Jiang, F. Liu, and C. Liu, "Determination of sample entropy and fuzzy measure entropy parameters for distinguishing congestive heart failure from normal sinus rhythm subjects," *Entropy*, vol. 17, no. 9, pp. 6270–6288, 2015.

[88] P. F. Binkley, E. Nunziata, G. J. Haas, S. D. Nelson, and R. J. Cody, "Parasympathetic withdrawal is an integral component of autonomic imbalance in congestive heart failure: demonstration in human subjects and verification in a paced canine model of ventricular failure," *Journal of the American College of Cardiology*, vol. 18, no. 2, pp. 464–472, 1991.

[89] B. Takase, A. Kurita, M. Noritake, A. Uehata, T. Maruyama, H. Nagayoshi, T. Nishioka, K. Mizuno, and H. Nakamura, "Heart rate variability in patients with diabetes mellitus, ischemic heart disease, and congestive heart failure," *Journal of electrocardiology*, vol. 25, no. 2, pp. 79–88, 1992.

[90] M. Asyali, "Discrimination power of long-term heart rate variability measures," in *Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No. 03CH37439)*, vol. 1, pp. 200–203, IEEE, 2003.

[91] S.-N. Yu and M.-Y. Lee, "Bispectral analysis and genetic algorithm for congestive heart failure recognition based on heart rate variability," *Computers in Biology and Medicine*, vol. 42, no. 8, pp. 816–825, 2012.

[92] S. Janjarasjitt, "A spectral exponent-based feature of rr interval data for congestive heart failure discrimination using a wavelet-based approach," *Journal of Medical and Biological Engineering*, vol. 37, no. 2, pp. 276–287, 2017.

[93] A. Jovic, K. Brkic, and G. Krstacic, "Detection of congestive heart failure from short-term heart rate variability segments using hybrid feature selection approach," *Biomedical Signal Processing and Control*, vol. 53, p. 101583, 2019.

[94] R. K. Tripathy, M. R. Paternina, J. G. Arrieta, A. Zamora-Mendez, and G. R. Naik, "Automated detection of congestive heart failure from electrocardiogram signal using stockwell transform and hybrid classification scheme," *Computer methods and programs in biomedicine*, vol. 173, pp. 53–65, 2019.

[95] P. Melillo, N. De Luca, M. Bracale, and L. Pecchia, "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, pp. 727–733, 2013.

[96] F. Shahbazi and B. M. Asl, "Generalized discriminant analysis for congestive heart failure risk assessment based on long-term heart rate variability," *Computer methods and programs in biomedicine*, vol. 122, no. 2, pp. 191–198, 2015.

[97] U. Dixen, C. Joens, B. V. Rasmussen, J. Parner, and G. B. Jensen, "Signal-averaged p wave duration and the dimensions of the atria," *Annals of noninvasive electrocardiology*, vol. 9, no. 4, pp. 309–315, 2004.

[98] R. Proietti, A. Mafrici, and D. H. Spodick, "Dynamic variations of p-wave duration in a patient with acute decompensated congestive heart failure," *Cardiology journal*, vol. 19, no. 1, pp. 95–97, 2012.

[99] J. W. Magnani, N. Wang, K. P. Nelson, S. Connelly, R. Deo, N. Rodondi, E. B. Schelbert, M. E. Garcia, C. L. Phillips, M. G. Shlipak, *et al.*, "Electrocardiographic pr interval and adverse outcomes in older adults," 2012.

[100] H. J. Shenkman, V. Pampati, A. K. Khandelwal, J. McKinnon, D. Nori, S. Kaatz, K. R. Sandberg, and P. A. McCullough, "Congestive heart failure and qrs duration: establishing prognosis study," *Chest*, vol. 122, no. 2, pp. 528–534, 2002.

[101] J. E. Madias, "The importance of the p-waves in the differentiation of attenuation of the qrs voltage due to pericardial effusion versus due to peripheral edema," *Journal of cardiac failure*, vol. 14, no. 1, pp. 55–60, 2008.

[102] Y. Isler and M. Kuntalp, "Diagnosis of congestive heart failure patients using poincare measures derived from ecg signals," in *2009 14th National Biomedical Engineering Meeting*, pp. 1–4, IEEE, 2009.

[103] C. Kamath, "A new approach to detect congestive heart failure using sequential spectrum of electrocardiogram signals," *Medical engineering & physics*, vol. 34, no. 10, pp. 1503–1509, 2012.

[104] C. M. van Ravenswaaij-Arts, L. A. Kollee, J. C. Hopman, G. B. Stoelinga, and H. P. van Geijn, "Heart rate variability," *Annals of internal medicine*, vol. 118, no. 6, pp. 436–447, 1993.

[105] M. Kikuya, A. Hozawa, T. Ohokubo, I. Tsuji, M. Michimata, M. Matsubara, M. Ota, K. Nagai, T. Araki, H. Satoh, *et al.*, "Prognostic significance of blood pressure and heart rate variabilities: the ohasama study," *Hypertension*, vol. 36, no. 5, pp. 901–906, 2000.

[106] K. Jensen-Urstad, P. Reichard, and M. Jensen-Urstad, "Decreased heart rate variability in patients with type 1 diabetes mellitus is related to arterial wall stiffness," *Journal of internal medicine*, vol. 245, no. 1, pp. 57–61, 1999.

[107] A. M. Kiviniemi, M. P. Tulppo, D. Wichterle, A. J. Hautala, S. Tiinanen, T. Seppänen, T. H. Mäkikallio, and H. V. Huikuri, "Novel spectral indexes of heart rate variability as predictors of sudden and non-sudden cardiac death after an acute myocardial infarction," *Annals of medicine*, vol. 39, no. 1, pp. 54–62, 2007.

[108] H. Huikuri, "Heart rate variability in coronary artery disease," *Journal of internal medicine*, vol. 237, no. 4, pp. 349–357, 1995.

[109] H. A. Malave, A. A. Taylor, J. Nattama, A. Deswal, and D. L. Mann, "Circulating levels of tumor necrosis factor correlate with indexes of depressed heart rate variability: a study in patients with mild-to-moderate heart failure," *Chest*, vol. 123, no. 3, pp. 716–724, 2003.

[110] G.-Z. Liu, B.-Y. Huang, and L. Wang, "A wearable respiratory biofeedback system based on generalized body sensor network," *Telemedicine and e-Health*, vol. 17, no. 5, pp. 348–357, 2011.

[111] A. L. Goldberger, L. A. 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, 2000.

[112] D. S. Baim, W. S. Colucci, E. S. Monrad, H. S. Smith, R. F. Wright, A. Lanoue, D. F. Gauthier, B. J. Ransil, W. Grossman, and E. Braunwald, "Survival of patients with severe congestive heart failure treated with oral milrinone," *Journal of the American College of Cardiology*, vol. 7, no. 3, pp. 661–670, 1986.

[113] B. D. Bradley, G. Green, T. Ramsay, and A. J. Seely, "Impact of sedation and organ failure on continuous heart and respiratory rate variability monitoring in critically ill patients: a pilot study," *Critical care medicine*, vol. 41, no. 2, pp. 433–444, 2013.

[114] E. Cirugeda-Roldan, D. Cuesta-Frau, P. Miro-Martinez, and S. Oltra-Crespo, "Comparative study of entropy sensitivity to missing biosignal data," *Entropy*, vol. 16, no. 11, pp. 5901–5918, 2014.

[115] W. Chen, Z. Wang, H. Xie, and W. Yu, "Characterization of surface emg signal based on fuzzy entropy," *IEEE Transactions on neural systems and rehabilitation engineering*, vol. 15, no. 2, pp. 266–272, 2007.

[116] L. A. Zadeh, "Fuzzy sets," *Information and control*, vol. 8, no. 3, pp. 338–353, 1965.

[117] C. Notarius and J. Floras, "Limitations of the use of spectral analysis of heart rate variability for the estimation of cardiac sympathetic activity in heart failure," *EP Europace*, vol. 3, no. 1, pp. 29–38, 2001.

[118] P. J. Schwartz, M. T. La Rovere, G. M. De Ferrari, and D. L. Mann, "Autonomic modulation for the management of patients with chronic heart failure," *Circulation: Heart Failure*, vol. 8, no. 3, pp. 619–628, 2015.

[119] S. Guzzetti, R. Magatelli, E. Borroni, and S. Mezzetti, "Heart rate variability in chronic heart failure," *Autonomic neuroscience*, vol. 90, no. 1-2, pp. 102–105, 2001.

[120] A. Zygmunt and J. Stanczyk, "Methods of evaluation of autonomic nervous system function," *Archives of medical science: AMS*, vol. 6, no. 1, p. 11, 2010.

[121] J. M. Yentes, N. Hunt, K. K. Schmid, J. P. Kaipust, D. McGrath, and N. Stergiou, "The appropriate use of approximate entropy and sample entropy with short data sets," *Annals of biomedical engineering*, vol. 41, no. 2, pp. 349–365, 2013.

[122] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial intelligence*, vol. 97, no. 1-2, pp. 273–324, 1997.

[123] H. Xue, S. Chen, and Q. Yang, "Structural regularized support vector machine: a framework for structural large margin classifier," *IEEE Transactions on Neural Networks*, vol. 22, no. 4, pp. 573–587, 2011.

[124] F. Provost and R. Kohavi, "On applied research in machine learning," *MACHINE LEARNING-BOSTON-*, vol. 30, pp. 127–132, 1998.

[125] A. O. Verkerk, R. Wilders, R. Coronel, J. H. Ravesloot, and E. E. Verheijck, "Ionic remodeling of sinoatrial node cells by heart failure," *Circulation*, vol. 108, no. 6, pp. 760–766, 2003.

[126] A. Mortara, P. Sleight, G. D. Pinna, R. Maestri, A. Prpa, M. T. La Rovere, F. Cobelli, and L. Tavazzi, "Abnormal awake respiratory patterns are common in chronic heart failure and may prevent evaluation of autonomic tone by measures of heart rate variability," *Circulation*, vol. 96, no. 1, pp. 246–252, 1997.

[127] S.-X. Zhou, J. Lei, C. Fang, Y.-L. Zhang, and J.-F. Wang, "Ventricular electrophysiology in congestive heart failure and its correlation with heart rate variability and baroreflex sensitivity: a canine model study," *Europace*, vol. 11, no. 2, pp. 245–251, 2009.

[128] Y. Zhu, X. Yang, Z. Wang, and Y. Peng, "An evaluating method for autonomic nerve activity by means of estimating the consistency of heart rate variability and qt variability," *IEEE transactions on biomedical engineering*, vol. 61, no. 3, pp. 938–945, 2013.

[129] L. Pecchia, P. Melillo, and M. Bracale, "Remote health monitoring of heart failure with data mining via cart method on hrv features," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 3, pp. 800–804, 2010.

[130] B. J. Bae, H. J. Kim, S. J. Kim, K. Y. Lee, W. D. Kim, and K. H. Yoo, "A case of yellow nail syndrome: Misdiagnosis as congestive heart failure," *Tuberculosis and Respiratory Diseases*, vol. 71, no. 1, pp. 46–49, 2011.

[131] C. Raphael, C. Briscoe, J. Davies, Z. I. Whinnett, C. Manisty, R. Sutton, J. Mayet, and D. P. Francis, "Limitations of the new york heart association functional classification system and self-reported walking distances in chronic heart failure," *Heart*, vol. 93, no. 4, pp. 476–482, 2007.

[132] A. M. Cheriyadat, "Unsupervised feature learning for aerial scene classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 439–451, 2013.

[133] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, pp. 1097–1105, 2012.

[134] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, p. 607, 1996.

[135] W. Sun, S. Shao, R. Zhao, R. Yan, X. Zhang, and X. Chen, "A sparse auto-encoder-based deep neural network approach for induction motor faults classification," *Measurement*, vol. 89, pp. 171–178, 2016.

[136] F. Tekiner, K. Gemici, B. Emrecan, E. Tekiner, and J. Jordan, "The efficacy and prognostic value of heart rate variability in 24-hour and short time recordings for determining cardiac autonomic dysfunction in congestive heart failure/kalp yetersizliginde uzun ve kisa sureli kayitlarda kalp hizi degiskenliginin kardiyak otonomik disfonksiyonu belirlemede etkinligi ve prognostik degeri," *The Anatolian Journal of Cardiology (Anadolu Kardiyoloji Dergisi)*, vol. 7, no. 2, pp. 118–124, 2007.

[137] T.-L. Jong, B. Chang, and C.-D. Kuo, "Optimal timing in screening patients with congestive heart failure and healthy subjects during circadian observation," *Annals of biomedical engineering*, vol. 39, no. 2, pp. 835–849, 2011.

[138] K. Swedberg, P. Eneroth, J. Kjekshus, and L. Wilhelmsen, "Hormones regulating cardiovascular function in patients with severe congestive heart failure and their relation to mortality. consensus trial study group.," *Circulation*, vol. 82, no. 5, pp. 1730–1736, 1990.

[139] G. Riegger, H. Bouzo, P. Petr, J. Munz, R. Spacek, H. Pethig, V. Von Behren, M. George, and H.-J. Arens, "Improvement in exercise tolerance and symptoms of congestive heart failure during treatment with candesartan cilexetil," *Circulation*, vol. 100, no. 22, pp. 2224–2230, 1999.

[140] J. Sztajzel *et al.*, "Heart rate variability: a noninvasive electrocardiographic method to measure the autonomic nervous system," *Swiss medical weekly*, vol. 134, no. 35-36, pp. 514–522, 2004.

[141] S. C. Matthews, M. P. Paulus, A. N. Simmons, R. A. Nelesen, and J. E. Dimsdale, "Functional subdivisions within anterior cingulate cortex and their relationship to autonomic nervous system function," *Neuroimage*, vol. 22, no. 3, pp. 1151–1156, 2004.

[142] M. Kumar, R. Pachori, and U. Acharya, "Use of accumulated entropies for automated detection of congestive heart failure in flexible analytic wavelet transform framework based on short-term hrv signals," *Entropy*, vol. 19, no. 3, p. 92, 2017.

[143] R. Castaldo, P. Melillo, U. Bracale, M. Caserta, M. Triassi, and L. Pecchia, "Acute mental stress assessment via short term hrv analysis in healthy adults: A systematic review with meta-analysis," *Biomedical Signal Processing and Control*, vol. 18, pp. 370–377, 2015.

[144] D. Rodrigues, Y. Tran, R. Guest, J. Middleton, and A. Craig, "Influence of neurological lesion level on heart rate variability and fatigue in adults with spinal cord injury," *Spinal cord*, vol. 54, no. 4, p. 292, 2016.

[145] A. Voss, R. Schroeder, A. Heitmann, A. Peters, and S. Perz, "Short-term heart rate variability-influence of gender and age in healthy subjects," *PloS one*, vol. 10, no. 3, p. e0118308, 2015.

[146] W. Chen, G. Liu, S. Su, Q. Jiang, and H. Nguyen, "A chf detection method based on deep learning with rr intervals," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 3369–3372, IEEE, 2017.

[147] M. Aiserman, E. M. Braverman, and L. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition," *Avtomat. i Telemeh*, vol. 25, pp. 917–936, 1964.

[148] M. T. La Rovere, G. D. Pinna, R. Maestri, A. Mortara, S. Capomolla, O. Febo, R. Ferrari, M. Franchini, M. Gnemmi, C. Opasich, *et al.*, "Short-term heart rate variability strongly predicts sudden cardiac death in chronic heart failure patients," *circulation*, vol. 107, no. 4, pp. 565–570, 2003.

[149] R. Ferber, R. Millman, M. Coppola, J. Fleetham, C. Friederich Murray, C. Iber, W. V. McCall, G. Nino-Murcia, M. Pressman, M. Sanders, *et al.*, "Portable recording in the assessment of obstructive sleep apnea," *Sleep*, 1994.

[150] G.-Q. Zhang, L. Cui, R. Mueller, S. Tao, M. Kim, M. Rueschman, S. Mariani, D. Mobley, and S. Redline, "The national sleep research resource: towards a sleep data commons," *Journal of the American Medical Informatics Association*, vol. 25, no. 10, pp. 1351–1358, 2018.

[151] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE Trans. Biomed. Eng*, vol. 32, no. 3, pp. 230–236, 1985.

[152] M. H. Crawford, S. J. Bernstein, P. C. Deedwania, J. P. DiMarco, K. J. Ferrick, A. Garson Jr, L. A. Green, H. L. Greene, M. J. Silka, P. H. Stone, *et al.*, "Acc/aha guidelines for ambulatory electrocardiography: executive summary and recommendations: a report of the american college of cardiology/american heart association task force on practice guidelines (committee to revise the guidelines for ambulatory electrocardiography) developed in collaboration with the north american society for pacing and electrophysiology," *Circulation*, vol. 100, no. 8, pp. 886–893, 1999.

[153] M. P. Tarvainen, P. O. Ranta-Aho, and P. A. Karjalainen, "An advanced detrending method with application to hrv analysis," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 2, pp. 172–175, 2002.

[154] Y. Zhang, Q. Yang, W. Pang, A. Argha, P. Xu, S. Su, and D. Yao, "Congestive heart failure detection via short-time electrocardiographic monitoring for fast reference advice in urgent medical conditions," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2256–2259, IEEE, 2018.

[155] L. Wang, W. Zhou, N. Liu, Y. Xing, and X. Zhou, "Chf detection with lstm neural network," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 514–517, IEEE, 2018.

[156] L. Wang and X. Zhou, "Detection of congestive heart failure based on lstm-based deep network via short-term rr intervals," *Sensors*, vol. 19, no. 7, p. 1502, 2019.

[157] S. Mallya, M. Overhage, N. Srivastava, T. Arai, and C. Erdman, "Effectiveness of lstms in predicting congestive heart failure onset," *arXiv preprint arXiv:1902.02443*, 2019.

[158] A. A. Bhurane, M. Sharma, R. San-Tan, and U. R. Acharya, "An efficient detection of congestive heart failure using frequency localized filter banks for the diagnosis with ecg signals," *Cognitive Systems Research*, vol. 55, pp. 82–94, 2019.

[159] A. Shibata, Y. Sugano, A. Shimouchi, T. Yokokawa, N. Jinno, H. Kanzaki, K. Ohta-Ogo, Y. Ikeda, H. Okada, T. Aiba, *et al.*, "Decrease in exhaled hydrogen as marker of congestive heart failure," *Open heart*, vol. 5, no. 2, p. e000814, 2018.

[160] E. Karasan, A. Abid, R. J. Mieloszyk, B. S. Krauss, T. Heldt, and G. C. Verghese, "An enhanced mechanistic model for capnography, with application to chf-copd discrimination," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5267–5272, IEEE, 2018.

[161] A. Maisel, "B-type natriuretic peptide measurements in diagnosing congestive heart failure in the dyspneic emergency department patient," *Reviews in Cardiovascular Medicine*, vol. 3, no. S4, pp. 10–17, 2019.