Prediction of Cardiovascular Diseases Based on Pulse Wave Signals via Artificial Intelligence Methods

February 2025

WU DANDAN

Graduate School of Engineering

CHIBA UNIVERSITY

（千葉大学審査学位論文）

Prediction of Cardiovascular Diseases Based on Pulse Wave Signals via Artificial Intelligence Methods

人工知能方法を用いた脈拍信号による心疾患予測

February 2025

WU DANDAN

Graduate School of Engineering

CHIBA UNIVERSITY

Abstract

Cardiovascular disease (CVD), a major cause of death and disability globally, encompasses disorders impacting the heart and vascular system, including coronary heart disease, stroke, and heart failure (HF). HF, a serious complication of is defined by the heart's reduced capacity to effectively circulate blood, often resulting from other cardiovascular conditions. Patients with HF typically experience symptoms like shortness of breath, fatigue, and leg swelling. Given its chronic and progressive nature, early diagnosis and continuous monitoring are crucial for improving outcomes. Traditional HF management relies on clinical labels, which limits the identification of patient subgroups. Due to the variability in pulse wave morphology, understanding key features is essential for more accurate diagnosis. An unsupervised clustering algorithm can help detect HF subgroups without depending on labels, enhancing personalized care.

Left ventricular enlargement (LVE), a common issue in CVD, often arises from long-term pressure or volume overload. Observed in conditions like hypertension, HF, and coronary artery disease, untreated LVE can lead to worsening heart function and increased risk of complications. Early detection and monitoring of LVE progression are vital for improving HF patient outcomes and optimizing treatment strategies.

Driven by the swift developments in artificial intelligence (AI), including machine learning (ML), integrating physiological signals like pulse waves offers non-invasive early screening, personalized treatment, and long-term monitoring solutions. This has shown great potential in the diagnosis, prediction, and management of CVDs. Developing ML strategies based on pulse wave signals to identify subgroups among HF patients and detect LVE holds significant clinical potential, providing clinicians with an effective, non-invasive tool for diagnosing and managing HF and ventricular remodeling. Ultimately, this supports personalized treatment and reduces complications. Building on the foundation of this study context, the aspects of the thesis are consolidated as below:

Chapter I: This chapter introduces the background on CVDs emphasizing HF and LVE as

critical complications. It discusses the limitations of traditional diagnostic methods relying on clinical labels and highlights the potential of AI in non-invasive, accurate, and personalized diagnosis and monitoring through pulse wave analysis. The chapter sets the stage for developing AI-based strategies to improve CVD patient management.

Chapter II: We collected and standardized the pulse wave time series and clinical data from 380 HF patients. Using the K-means^{++} algorithm, we clustered the datasets, evaluated the clustering performance, and analyzed the clinical differences between the groups. An RF classifier was used to quantify the importance ratio of pulse wave time-frequency features in distinguishing between clusters. We successfully identified two patient subgroups, revealing significant differences in age, heart rate, and ejection time between the clusters. Additionally, the Upstroke time, mean value, fall time, and skewness were identified as key pulse wave morphological features. This provides initial evidence that non-invasive and easily accessible pulse wave signals can be used for stratified management of HF patients.

Chapter III: We established a dataset of 264 HF patients through strict data screening and divided the patients into LVE and non-LVE groups based on clinical labels. Fourier series analysis was used to verify the significant distinctions between the two groups, ensuring the validity of the dataset. We then proposed a weighted RF classification algorithm to distinguish between LVE and non-LVE patients and compared the classification results with the clinical labels. The findings indicated that the classification model demonstrated excellent performance, achieving an accuracy of 0.91 and an area under the receiver operating characteristic curve (AUC-ROC) of 0.93, effectively identifying LVE in HF patients. This study emphasizes the promise of ML models based on pulse wave signals for non-invasive identification of LVE patients.

Chapter IV: Building on the previous exploration, we further collected the left ventricular end-diastolic diameter index (LVDdI) from 264 HF patients as a key indicator for quantitatively predicting LVE. We proposed a regression algorithm based on a densely connected convolutional network to directly predict the specific values of the LVDdI. Subsequently, we used the Bland-Altman method to assess the consistency between the predicted values and the clinical measurements. The results showed that the regression model attained an accuracy of 0.88 and an

AUC-ROC of 0.89, demonstrating good predictive performance. In comparison to the classification algorithm in Chapter 3, this chapter focuses on directly predicting the specific values of LVE using the regression model, rather than simply classifying patients as LVE or non-LVE. This approach not only provides more precise quantitative predictions, enabling physicians to more accurately assess the extent of LVE, but also offers data support for further personalized treatment and long-term monitoring.

Chapter V: This thesis underscores the effectiveness of AI-driven analysis of pulse wave signals for diagnosing and predicting CVDs. By employing ML techniques, the research successfully identified key pulse wave features and demonstrated their potential for accurate, non-invasive detection of conditions such as HF and LVE. The findings suggest that these AI-based methods can enhance clinical precision, support personalized patient management, and streamline healthcare processes. Overall, the study points to broad prospects for integrating pulse wave analysis with AI to improve the diagnosis, monitoring, and treatment of CVDs.

Keywords： Cardiovascular disease; Heart failure; Deep learning; Machine learning; Pulse wave; Left ventricular enlargement; Importance ratio.

Table of Contents

List of Figures

List of Tables

Chapter I

General Introduction

1.1 Cardiovascular diseases (CVDs)

1.1.1 CVDs and heart failure (HF)

CVD encompasses a collection of disorders that affect the heart and vascular system, primarily comprising conditions like atherosclerosis, myocardial infarction, coronary artery disease, and stroke. These diseases typically result from the hardening, narrowing, or blockage of blood vessels, leading to insufficient blood supply to organs and subsequent dysfunction. HF is one of the serious consequences of CVD. It occurs when the heart deteriorates due to a pathological condition, rendering it unable to adequately circulate sufficient blood to fulfill the body's metabolic requirements.

1.1.2 Health Burden of HF

HF is recognized as a worldwide health crisis, affecting an estimated 64.3 million individuals globally in 2017 [1–3]. With improvements in patient survival rates following HF diagnoses, the widespread use of evidence-based treatment methods, and the increasing life expectancy of the global population, the prevalence of HF is projected to increase further.

The economic burden of HF on healthcare systems worldwide is also becoming increasingly severe. In 2012, the total healthcare expenditure related to HF in the United States was approximately $30.7 billion, with projections suggesting a 127% increase by 2030, reaching $69.8 billion, equivalent to around $244 per U.S. adult [4,5], as shown in Fig. 1-1.

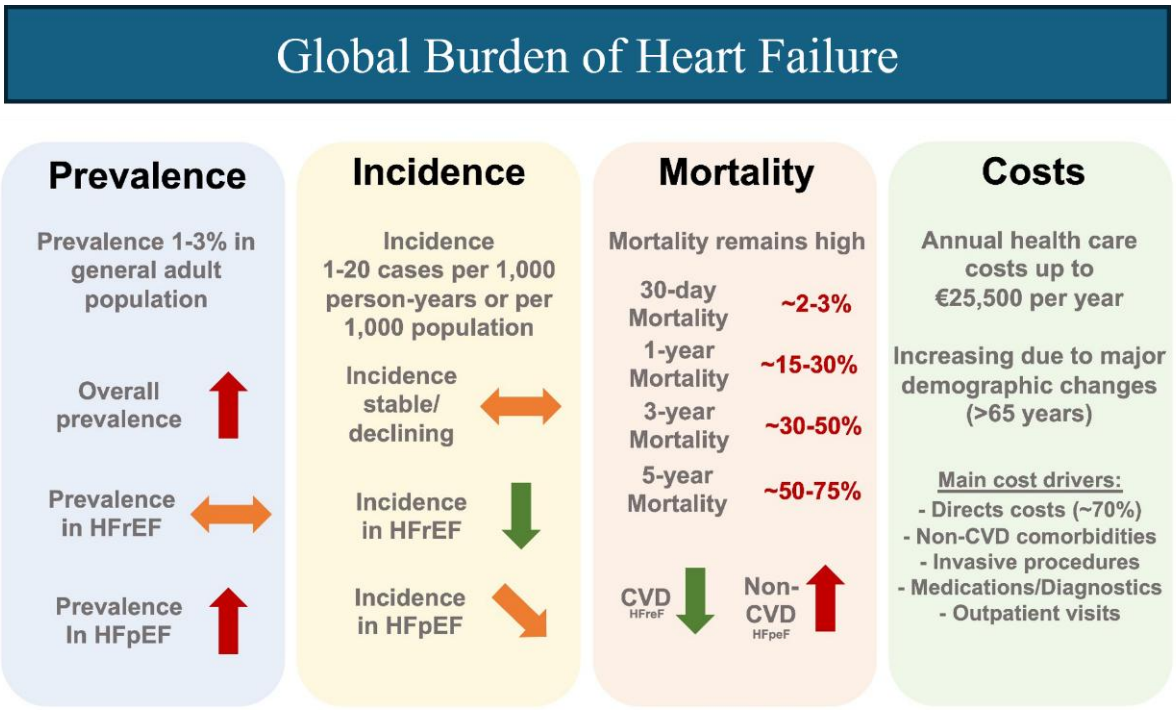


Fig. 1-1 Global Burden of HF. Prevalence: 1-3% in adults, overall increasing; Incidence: 1-20 cases per 1,000 people, stable or declining; Mortality: 30-day 2-3%, 1-year 15-30%, 5-year 50-75%; Costs: Annual healthcare costs up to €25,500, rising with aging population.

1.1.3 Diagnosis and management of HF and Left ventricular enlargement (LVE)

HF is a chronic, progressive condition typically linked to structural or functional abnormalities of the heart, resulting in the heart's failure to adequately pump blood to satisfy the body's demands. HF is characterized by high mortality [6], frequent recurrence [7], a significant decline in quality of life [8], and substantial economic burden on patients. As a result, the diagnosis and management of HF hold considerable research importance.

Early and accurate detection of HF is crucial for improving long-term prognosis, as it enables physicians to implement effective treatments promptly, thereby reducing complications and lowering mortality rates. HF can be diagnosed using various methods, typically including clinical

evaluation, laboratory testing, and clinical signal monitoring. Clinical evaluation involves detailed inquiries about the patient's symptoms and physical examination. Physicians usually assess whether the patient exhibits typical HF symptoms, like breathlessness, fatigue, or leg swelling, and use auscultation to detect any abnormal heart or lung sounds that may indicate fluid buildup or other issues [9]. Laboratory testing, particularly with biomarkers like brain natriuretic peptide (BNP) and N-terminal pro-brain natriuretic peptide (NT-proBNP), is vital for HF assessment [10,11]. When cardiac function is impaired, these biomarkers typically rise significantly, indicating increased heart pressure or volume overload. They are key indicators in the diagnosis of HF. Clinical signal monitoring methods include pulse wave analysis, electrocardiograms (ECG), and heart sounds, which are commonly used tools [12,13]. Pulse wave analysis, as a non-invasive hemodynamic detection technique, is widely applied for cardiovascular function evaluation owing to its efficiency, rapidity, and strong reliability [14]. Pulse waves originate from each heartbeat and travel along the arteries, reflecting at arterial branches. Characteristics such as pulse wave speed, shape, amplitude, and duration can reveal the functional state of the blood vessels [15,16]. Pathophysiological conditions like arterial stiffness or changes in vascular resistance can alter these waveforms. For instance, arterial stiffness may cause an increase in pulse wave velocity, an elevated main peak, or the early appearance of reflected waves. Thus, pulse wave analysis provides valuable insights into vascular health and cardiac function for physicians. Jin et al. explored the characteristics of carotid-femoral pulse wave velocity (cfPWV) in healthy individuals and its correlation with cardiovascular risk factors. The results indicated that cfPWV values differed by gender and age, with factors such as age, systolic blood pressure, and heart rate significantly affecting cfPWV in both males and females [17]. Valencia-Hernández et al [18]. investigated whether aortic pulse wave velocity (PWV) could predict major cardiovascular incidents and improve the efficacy of the American Heart Association/American College of Cardiology risk assessment score. The results showed that aortic stiffness, as measured by PWV, is a significant predictor of adverse cardiovascular incidents and can enhance risk assessment in the elderly population. In 1973, Dawber et al. categorized four distinct DVP shapes according to the features of the dicrotic notch, as shown in Fig. 1-2. These categories range from Class 1, which features a

clearly visible and well-defined dicrotic notch, to Class 4, where the dicrotic notch is not visible. Other methods for identifying typical DVP waveforms have also been explored: Sherebrin and Sherebrin [19] classified DVP into three categories by age using frequency analysis; Tigges et al. [20] applied ML and deep learning (DL) techniques based on handcrafted features to classify DVP shapes into the four categories proposed by Dawber et al.; Wang et al. [21] introduced a multi-Gaussian fitting approach, classifying DVP into five categories. This shows that utilizing non-invasive and easily obtainable pulse wave signals to explore the classification and subtype detection of CVDs is feasible and promising.

CVDs are commonly associated with the presence of multiple overlapping risk factors, while pulse waveforms are shaped by variables such as age, gender, and lifestyle. This complexity poses challenges in isolating pulse wave parameters that are directly linked to particular cardiovascular risk factors. Thus, it becomes vital to identify and evaluate characteristic pulse waveforms influenced by distinct risk factors to improve the accuracy of diagnosing and treating cardiovascular conditions.

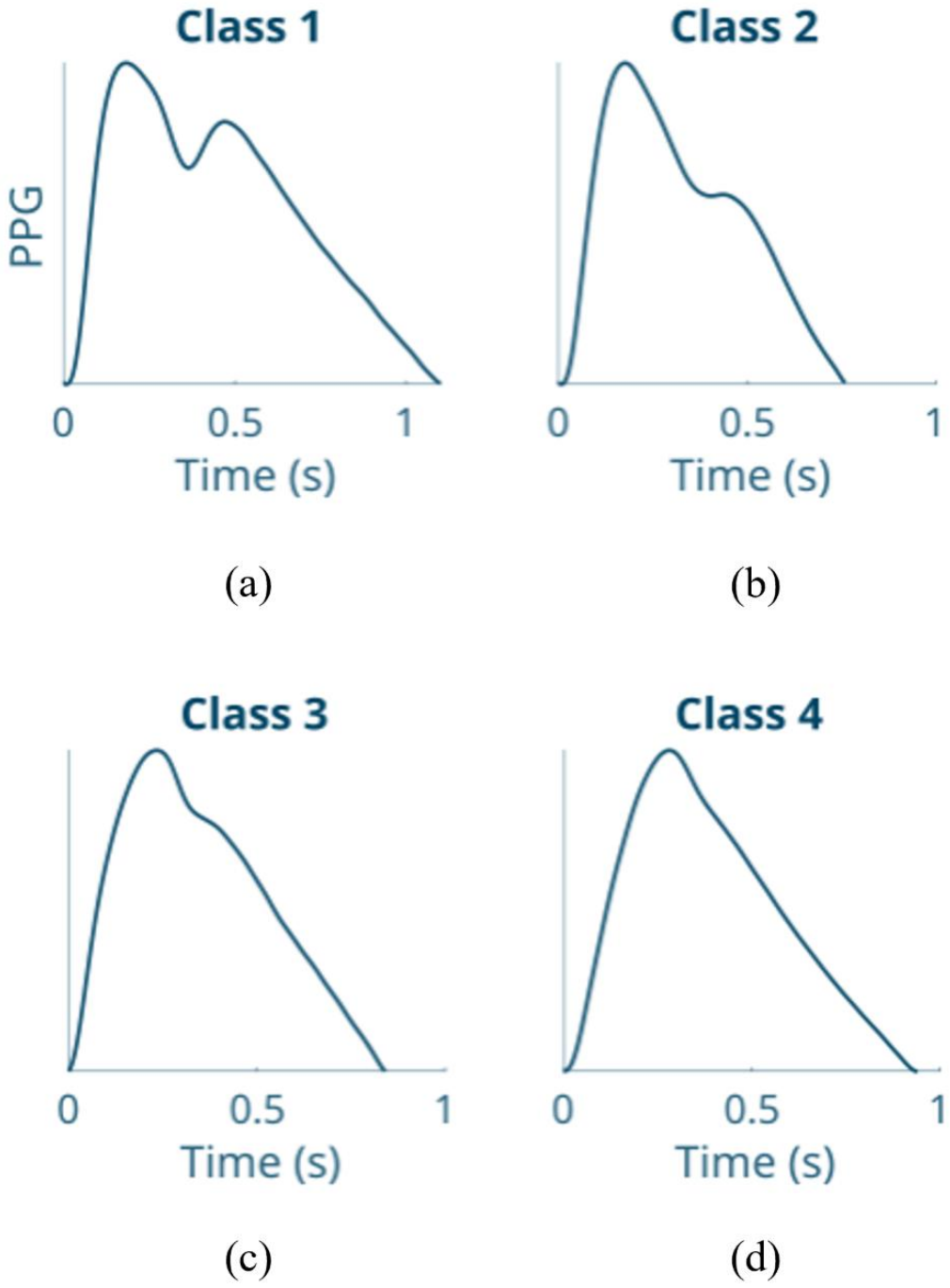


Fig. 1-2 Four pulse waveform classes based on the features of the dicrotic notch. (a) Class 1: clearly visible and well-defined dicrotic notch. (b) Class 2: dicrotic notch is visible but less distinct than

in class 1. (c) Class 3: weak dicrotic notch, only slightly visible. (d) Class 4: dicrotic notch is not visible.

LVE refers to the abnormal enlargement or expansion of the left ventricle, as shown in Fig. 1-2, typically caused by prolonged pressure or volume overload, which prevents the myocardium from effectively pumping sufficient blood [22]. LVE is a common pathological feature in many heart diseases, such as HF, cardiomyopathy, and coronary artery disease. Its main characteristics include an enlarged left ventricular chamber and increased tension in the myocardial wall. When LVE occurs, the ventricular wall becomes thinner, reducing the heart's contractile function and pumping efficiency, resulting in insufficient cardiac output. The enlarged left ventricle may also increase the heart's oxygen demand, making it more prone to arrhythmia or further cardiac damage [23]. Therefore, prompt identification and intervention for LVE are crucial in preventing the progression of heart disease.

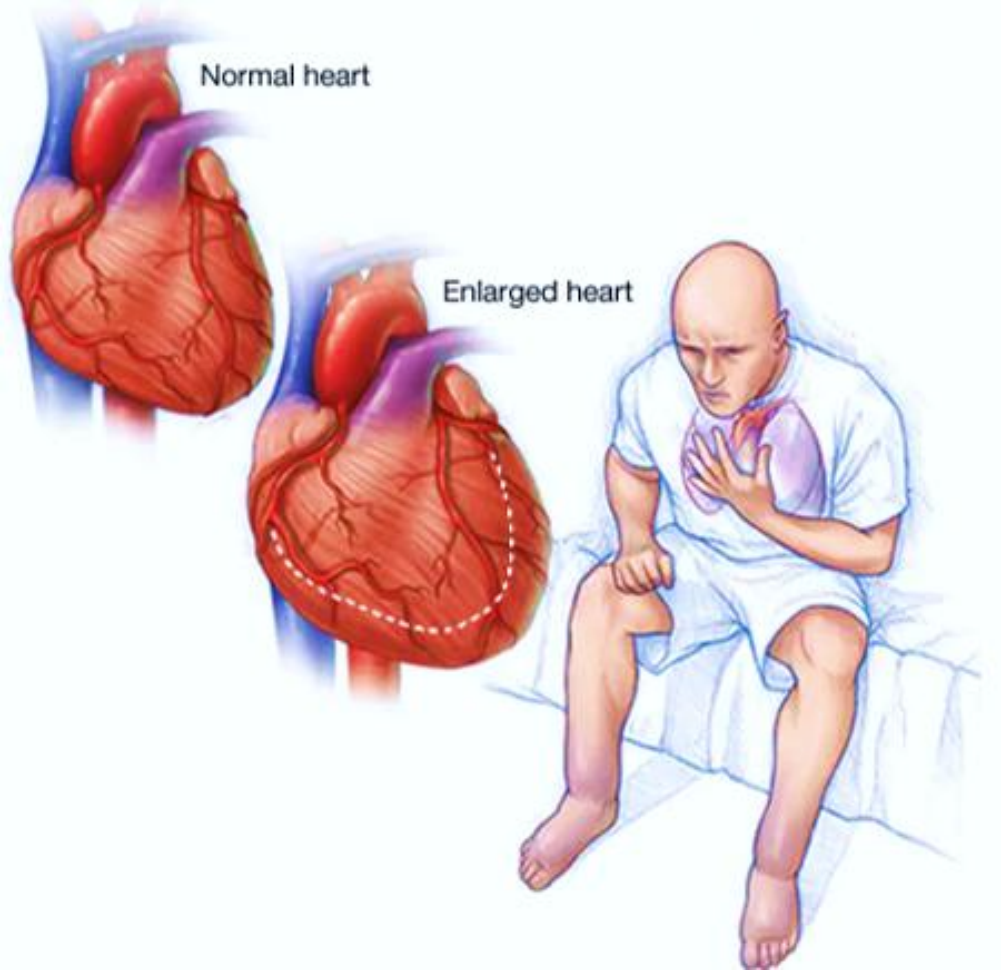


Fig. 1-3 Comparison of a normal heart and a heart with LVE.

The detection of LVE primarily relies on medical history evaluation and imaging techniques, such as transthoracic echocardiography (TTE) and cardiac magnetic resonance (CMR). CMR is considered the most precise approach for assessing the structure, size, and position of the heart [24]. However, due to its high cost, lengthy examination time, and magnetic field restrictions for patients with metal implants, CMR is not commonly used as a routine diagnostic tool [25]. In contrast, TTE is widely recognized for its non-invasive nature, real-time imaging capability, and patient-friendly approach, making it a popular choice for assessing and diagnosing cardiac function [26,27]. Nevertheless, the accessibility and convenience of TTE are somewhat limited by the need

for specialized equipment and skilled operators. Therefore, developing a non-invasive and easily accessible physiological signal tool for identifying LVE is of great importance. Pulse wave, as a physiological signal, has been widely applied in health monitoring and CVD prediction [28]. Pulse waves not only contain valuable information about the physiological and pathological state of the human cardiovascular system, but also provide key physiological data related to blood supply capacity and transport efficiency [29]. Nowadays, with various low-cost home electronic devices, pulse wave signals can be easily measured non-invasively, offering a low-cost, user-friendly solution for diagnosing CVD and its associated complications. The convenience and accessibility of this technology, especially for home monitoring, enhances its potential in early diagnosis and long-term health management, while also helping to reduce the burden on healthcare resources.

Currently, many studies have already investigated the application of pulse wave signals for detecting specific CVDs. For instance, research conducted by Ninna Hahn Tougaard et al. found that higher cfPWV among patients with type 1 diabetes is linked to a higher risk of developing kidney disease, cardiovascular events, and mortality. cfPWV could function as a useful mechanism for risk categorization in this population [30]. A study by Koivistoinen et al. found that PWV is an independent predictor of blood pressure progression and the incidence of hypertension in young individuals. Adding PWV to traditional cardiovascular risk factors significantly improved the prediction of hypertension onset [31]. Chou et al. proposed a method for identifying four life-threatening arrhythmias using pulse-to-pulse interval analysis. The Random Forest (RF) classifier demonstrated the best performance, with a Kappa coefficient of 98.86%, outperforming other classifiers, and showing great potential for detecting arrhythmias in mobile health applications [32]. Therefore, pulse wave, as a low-cost, patient-friendly physiological signal that can be non-invasively measured through home electronic devices, is a feasible method for detecting LVE. With the rapid development of ML and DL, many methods have been successfully applied to pulse wave signal analysis, showing great potential and feasibility in pulse wave pattern classification and cardiac function prediction.

1.2 Artificial intelligence (AI)

1.2.1 Definitions

AI refers to the development of computational systems and machines able to execute tasks that rely on cognitive abilities. Such tasks encompass learning, reasoning, and problem-solving, understanding natural language, pattern recognition, and adapting to new environments [33]. AI integrates fields such as computing, mathematical studies, and cognitive sciences, and neuroscience to create algorithms and models that enable machines to think and learn [34]. The core areas of AI include ML, DL, and natural language processing (NLP) [33,34]. These technologies allow AI systems to continuously improve their performance through data analysis, decision-making, and more natural interaction with humans. Today, AI is extensively utilized in domains like medical care, finance, autonomous driving, and mobile applications, significantly enhancing efficiency, accuracy, and innovation across industries [35].

1.2.2 Deep learning and machine learning

ML and DL are two important subfields of AI, with their core being the ability to enable machines to automatically learn and improve through data. ML is a technology that allows computational frameworks to extract knowledge from data without direct coding instructions [36]. It builds mathematical models, learns using training data, and makes forecasts or choices derived from new data. ML primarily includes three types: supervised learning, unsupervised learning, and reinforcement learning, as shown in Fig. 1-4 [37]. Supervised learning is trained using known data and outcomes, such as in classification and regression tasks. Unsupervised learning, on the other hand, deals with unlabeled data, aiming to discover hidden patterns, such as in clustering analysis [38]. Reinforcement learning learns through trial and error, optimizing decisions using a reward system, and is often used in complex decision-making problems [39].

DL, a branch of ML, is modeled after the neural architecture of the human brain. It utilizes multi-layer artificial neural networks (deep neural networks, DNN) to automatically learn and

extract complex patterns and features from data [40], as shown in Fig. 1-5. Compared to traditional ML, DL excels at handling large volumes of unstructured data, such as images, audio, and text. Its key features include the use of multi-layer neural networks to automatically extract higher-level features, strong adaptive learning capabilities, and outstanding performance for advanced applications such as visual recognition, audio interpretation, and language understanding.

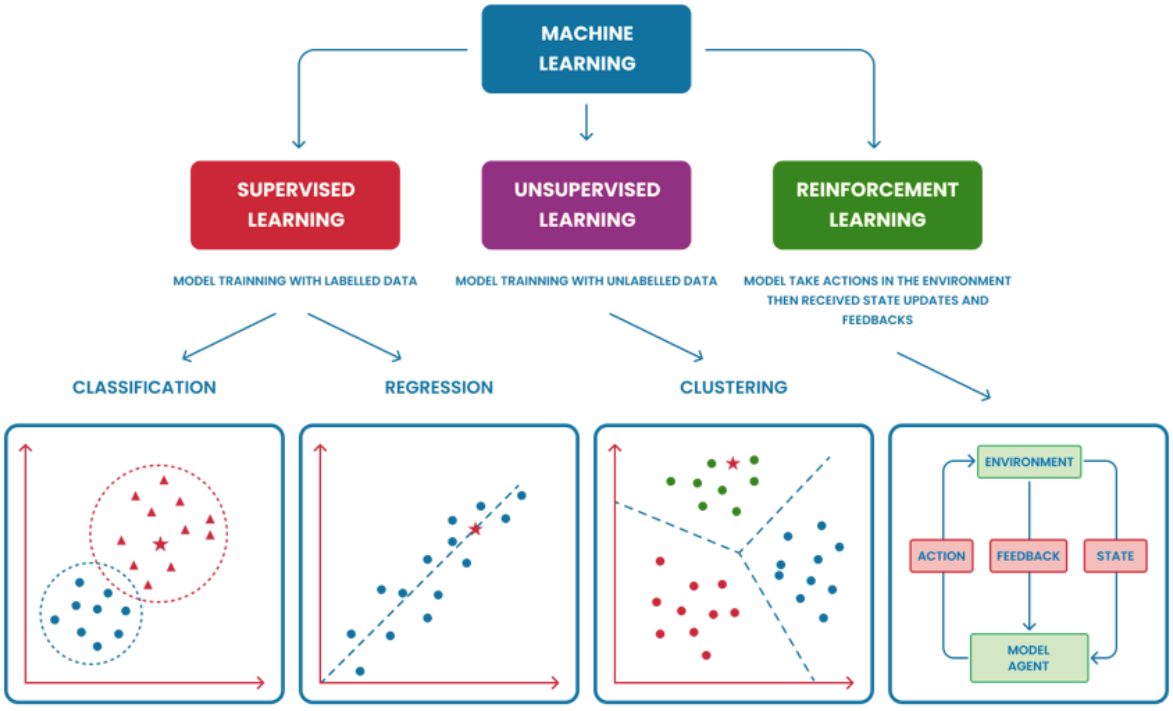


Fig. 1-4 The three types of ML: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning trains models using labeled data, with typical applications in classification and regression. Unsupervised learning trains models using unlabeled data, commonly applied in clustering analysis. Reinforcement learning optimizes decision-making processes through interaction with the environment and feedback mechanisms.

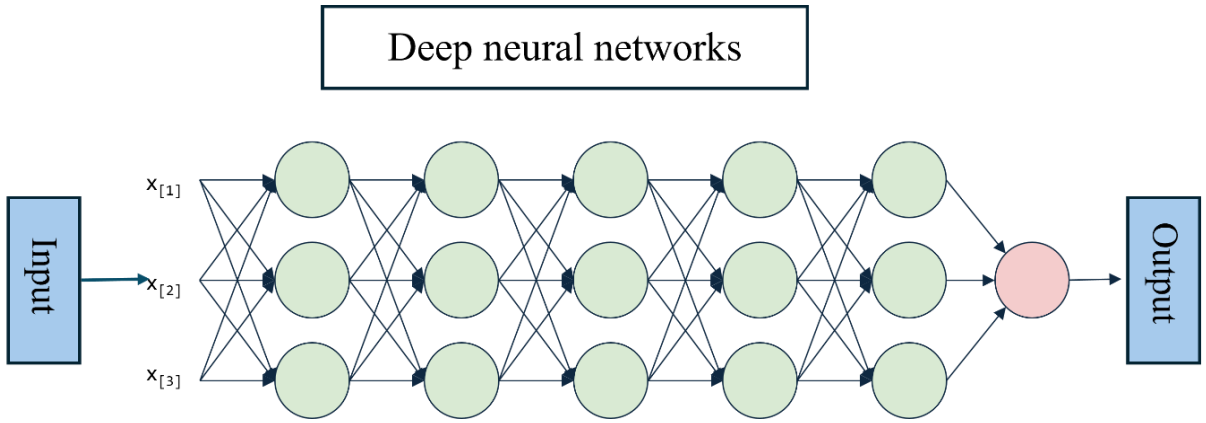


Fig. 1-5 Schematic Diagram of DNN Structure. The basic structure of a DNN consisting of an input layer, multiple hidden layers in the middle, and an output layer. Input data is processed through neurons in the hidden layers, with each layer's neurons interconnected. Features are extracted layer by layer, allowing the network to learn complex patterns, and the final prediction is generated in the output layer.

1.2.3 CVDs detection via AI

The combination of physiological signals and ML has become an effective method for detecting CVDs. Physiological signals such as pulse waves and ECG provide rich data regarding the circulatory system, including blood pressure, heart rate, and hemodynamic parameters. By integrating ML techniques, the analysis of these signal features enables accurate prediction of cardiovascular function parameters and early detection of CVDs. In recent years, ML models, such as DL and neural networks, have been widely used in pulse wave signal classification and cardiac function prediction, demonstrating significant potential in CVD diagnosis. The minimally invasive characteristic and cost-effectiveness of this approach make it a patient-friendly tool for early diagnosis and monitoring. Han et al. proposed a new method that combines a multi-lead residual neural network and feature fusion, utilizing 12-lead ECG data for myocardial infarction detection and localization, achieving high accuracy and outperforming traditional methods [41]. Mousavi et al. proposed HAN-ECG, an interpretable two-way recurrent neural network for detecting atrial

fibrillation using ECG signals. The model leverages a hierarchical attention mechanism to enhance interpretability and achieve superior performance compared to existing algorithms, providing clinically meaningful insights into its decision-making process [42]. Wang et al. explored the feasibility of using ML- driven pulse wave analysis for the early identification of abdominal aortic aneurysms (AAA) through a synthetic pulse wave dataset. By training a ML model on silico data, the study achieved a sensitivity of 86.8% and a specificity of 86.3% for detecting AAA from photoplethysmogram pulse wave signals [43]. Li et al. developed a CNN model to classify pulse waveforms associated with five conditions, achieving a 95% success rate [44].

These studies have laid a solid foundation for detecting specific CVDs using physiological signals combined with ML, demonstrating the great potential of ML methods in cardiovascular health monitoring. However, most research has focused on the general population, with few studies specifically targeting HF patients. For this patient group, existing ML models and detection methods have yet to fully address their unique pathological characteristics. Additionally, although cardiac remodeling, especially LVE, is a key physiological manifestation of HF, achieving its non-invasive detection and long-term monitoring remains a significant challenge. Current detection methods mainly rely on imaging techniques such as CMR or TTE, which, although accurate, have limitations in terms of cost, availability, and patient comfort. Therefore, developing non-invasive detection technologies based on physiological signals like pulse waves, combined with the powerful feature extraction and pattern recognition capabilities of ML, could provide a more cost-effective and convenient solution for detecting and managing LVE in HF patients, which remains an important issue to be addressed.

1.2.4 Unsupervised learning for pulse waves in CVDs

Unsupervised learning is a ML method that does not rely on predefined labels or classifications during the training process [45]. Instead, it identifies potential relationships and features by analyzing and exploring the structure and patterns within the data. It is mainly used for handling unlabeled data, with common applications including clustering, dimensionality reduction, and

anomaly detection. Through unsupervised learning, models can discover hidden patterns in the data, helping to uncover unknown knowledge or information, which is particularly valuable in complex datasets [46].

The application of unsupervised learning in CVDs primarily includes patient subgroup discovery, anomaly detection, feature extraction and dimensionality reduction, as well as pattern recognition and risk stratification. Using unsupervised learning algorithms (such as clustering analysis), patients with CVDs can be grouped into different subgroups based on physiological signals like pulse waves and ECGs, aiding in personalized treatment. Additionally, unsupervised learning can detect abnormal heart rhythms in ECGs, enabling early identification of cardiovascular events for timely warnings. Unsupervised learning also enhances the processing efficiency and accuracy of complex physiological data through feature extraction and dimensionality reduction techniques, identifies key patterns associated with increased heart attack risk, supporting more precise diagnosis and risk stratification. Through these applications, unsupervised learning plays a significant role in the accurate diagnosis and personalized treatment of CVDs. Flores A et al. used unsupervised ML to identify four coronary artery disease subgroups with distinct clinical outcomes and risk characteristics based on clinical and genetic data. Compared to traditional methods, the identified subgroups provided better assessment of major cardiovascular incidents and mortality, highlighting the potential of unsupervised algorithms in processing diverse patient data for precision population health [47]. Grigorios M Karageorgos et al. proposed an unsupervised DL-based method to enhance arterial wall motion estimation for imaging vascular elasticity. The model demonstrated high accuracy in generating localized pulse wave velocity maps, with low error rates and strong correlation. The results indicate that this novel approach could significantly improve the quality of arterial stiffness assessments in cardiovascular diagnostics [48]. Rousseau-Portalis et al. used an unsupervised clustering method to evaluate the association between cfPWV and Coronary Calcium Score (CCS). The results showed that cfPWV is closely correlated with CCS, with both measures progressing more rapidly in the high-risk group. This suggests that cfPWV could enhance cardiovascular risk stratification and serve as a gatekeeper for CCS testing [49]. These studies demonstrate that unsupervised learning holds great potential

and promise in the discovery of subgroups, early diagnosis, personalized treatment, and long-term monitoring of CVDs.

Compared to supervised learning, which relies on clinical labels, unsupervised learning can detect undefined or abnormal subgroups. Although there have been some clustering studies on pulse wave morphology, such as [20] et al. using ML and DL methods trained on handcrafted features to classify digital volume pulse waveforms into four categories proposed by Dawber et al., these studies often focus on broad populations. HF, as a complex syndrome with diverse causes and clinical characteristics, presents a challenge. Using unsupervised learning to detect subtypes of HF patients based on pulse wave signals can provide valuable support for clinical stratification and personalized treatment. However, achieving this in such a complex context remains a significant challenge.

1.3 Objectives

Building on the aforementioned research background, this study gathered pulse wave signals and in-hospital clinical information from 380 HF patients. The research objectives are divided into the following three parts:

(1) Subgroup detection and pulse waveform importance ratio analysis in HF patients:
- K-means++ algorithm was used to cluster the pulse wave time series of HF patients, aiming to detect different subgroups.
- Statistical analysis was conducted to explore the clinical significance of differences in clinical parameters between subgroups.
- Time-frequency features were extracted, and their importance ratios in distinguishing clinical differences were quantified using a RF classifier.

(2) Identification of LVE in HF patients through classification algorithms:
- HF patients were grouped based on clinical labels, and statistical differences between groups were verified using Fourier series.

- Multiple classification algorithms were compared to evaluate their performance.
- A weighted RF model was proposed to effectively classify HF patients with or without LVE.

(3) Prediction of left ventricular end-diastolic diameter index (LVDdI) in HF patients:

- LVDdI was selected as a key indicator for the quantitative prediction of LVE.
- A regression algorithm based on densely connected convolutional networks was proposed to directly predict the LVDdI.
- The performance of classification and regression algorithms in predicting LVE was compared.

1.4 Thesis contents

The content of this thesis is outlined as follows:

- Chapter I: a brief introduction to CVDs and AI methods, presenting the research objectives of this thesis.
- Chapter II: a study on detecting subgroups of HF patients through pulse wave signals using unsupervised learning and evaluating the importance ratio of pulse wave time-frequency features.
- Chapter III: a study on identifying LVE in HF patients through pulse wave signals using classification algorithms.
- Chapter IV: a study on quantitatively predicting left ventricular end-diastolic diameter index using pulse wave signals and regression models.
- Chapter V: a summary of this study and perspectives extending beyond the scope of this thesis.

Chapter II

Pulse Wave Time Series Unsupervised Clustering with Importance Ratios for HF Subgroup Detection

2.1 Introduction

Heart failure (HF) is a severe cardiovascular condition that represents a significant global public health threat, with its prevalence and mortality rates continuing to rise [50–52]. As of 2017, an estimated 64.3 million individuals worldwide were affected by HF, with over 8 million new cases diagnosed annually [53,54]. Therefore, condition monitoring and deterioration prediction of HF patients are clinically significant.

Pulse wave, a carrier of cardiovascular information, serves as a crucial physiological signal for assessing cardiovascular health, enabling the calculation of parameters like heart rate (HR), pulse wave velocity, and blood pressure; moreover, its waveforms, which exhibit strong correlations with early arterial pathologies such as atherosclerosis and endothelial dysfunction, are utilized for diverse preliminary cardiovascular assessments—including blood pressure estimation [55–57], diabetes mellitus detection [58,59], and vascular ageing assessment [60–63]. Given the non-invasive and cost-effective nature of pulse waveform acquisition that integrates pulse monitoring, flexible sensing, and the Internet of Medical Things, it is positioned at the forefront of CVD prevention and diagnosis [64]. However, due to pulse waves varying significantly across individuals and pathologies, as well as with age, blood pressure, and conditions like arrhythmias and coronary artery disease, there is a necessity to develop methods to analyze pulse waveforms accurately. ML methods, for their powerful feature extraction capabilities, have prompted the adoption of pulse waves-based quantitative and qualitative analysis on CVDs. For instance, Tigges et al. [65], by manually extracting features in conjunction with ML and DL (DL) methods, classified distinct pulse waveforms into the following four categories proposed by Dawber et al. [66]: 1) Class I: A clear incisura is marked on the descending slope of the pulse wave. 2) Class II: An incisura does not form, but the descending line transitions to a horizontal trajectory. 3) Class III: A notch is absent, but a distinct alteration in the descent angle is evident. 4) Class IV: No indication of a notch is observed. Li et al. further investigated the potential of convolutional neural networks for pulse waveform classification and implemented a DL mapping from pulse waveforms to different CVD risk factor categories, including hypertension, hyperlipidemia, atherosclerosis, and type 2 diabetes [44]. Additionally, our previous research has demonstrated that ML models, such as fully connected networks and Dense Net, are highly effective in accurately predicting various cardiac function parameters, identifying left ventricular dilation, conducting

comprehensive quantitative analyses of cardiac function status in both patients with CVD and healthy individuals, and assessing the cardiac blood supply capacity in HF patients [67–69].

Despite the success of these ML-based classification and regression studies in pulse wave feature extraction and risk prediction, they typically rely on predefined clinical labels. However, obtaining high-quality and consistent clinical labels is challenging and often depends on the judgment of medical personnel. When dealing with patients with multiple comorbidities, such as those with HF, studies often focus on data from patients with a single disease label to ensure the accuracy of detection tasks. This approach requires large datasets and overlooks the interactions between multiple diseases, limiting the model's applicability in complex clinical scenarios. To overcome these limitations, we turned to cluster analysis [70,71], an unsupervised learning method that does not depend on predefined category labels. Cluster analysis can identify unknown clinical subgroups with distinct characteristics, making it particularly suitable for complex clinical cases. This method helps us develop more precise and differentiated treatment and management strategies for different patient subgroups, thereby improving treatment outcomes and management quality. Currently, Zanelli et al. have used unsupervised clustering (ML method without using predefined labels or categories) to categorize digital volume pulse waveforms and investigate their correlation with clinical data, demonstrating how different clustering methods can reveal associations between pulse waveforms and clinical characteristics such as age and pulse transit time [72]. Garmaev et al. demonstrated a clustering analysis method based on the time parameters of pulse signals, classifying the pulse signals of healthy young individuals into two, three, and four categories to analyze significant differences between them [73]. This method can accurately assess cardiac cycle time parameters and aid in creating decision support systems for recording and analyzing cardiac signals. Although previous studies on pulse wave clustering have made significant progress, they typically focus on broader populations and rarely specifically target HF patients. Additionally, while these studies have interpreted the clinical significance of the clusters, they have not clearly elucidated how pulse wave morphology features influence clustering outcomes.

Given the limitations, our study is the first to perform pulse wave clustering analysis specifically on HF patients to identify potential subgroups with significant clinical characteristics. We not only explored the key morphological features of pulse waves in these HF patient subgroups but also quantified the importance ratios (IR) of these features to better understand their role in

subgroup identification. As shown in Fig. 2-1, we collected pulse wave time series data from 380 HF patients, applied the K-means++ algorithm for clustering, and evaluated the clustering performance as well as the clinical differences between clusters. Subsequently, we extracted the time-frequency features of the pulse waves and quantified their IRs using an RF model. This approach enabled us to identify potential subgroups with impaired cardiac function and to provide clinical recommendations on which pulse wave features to focus on during the diagnosis and monitoring of HF patients, thereby enhancing personalized management of HF, improving patient outcomes, and reducing hospitalization rates and enhancing quality of life.

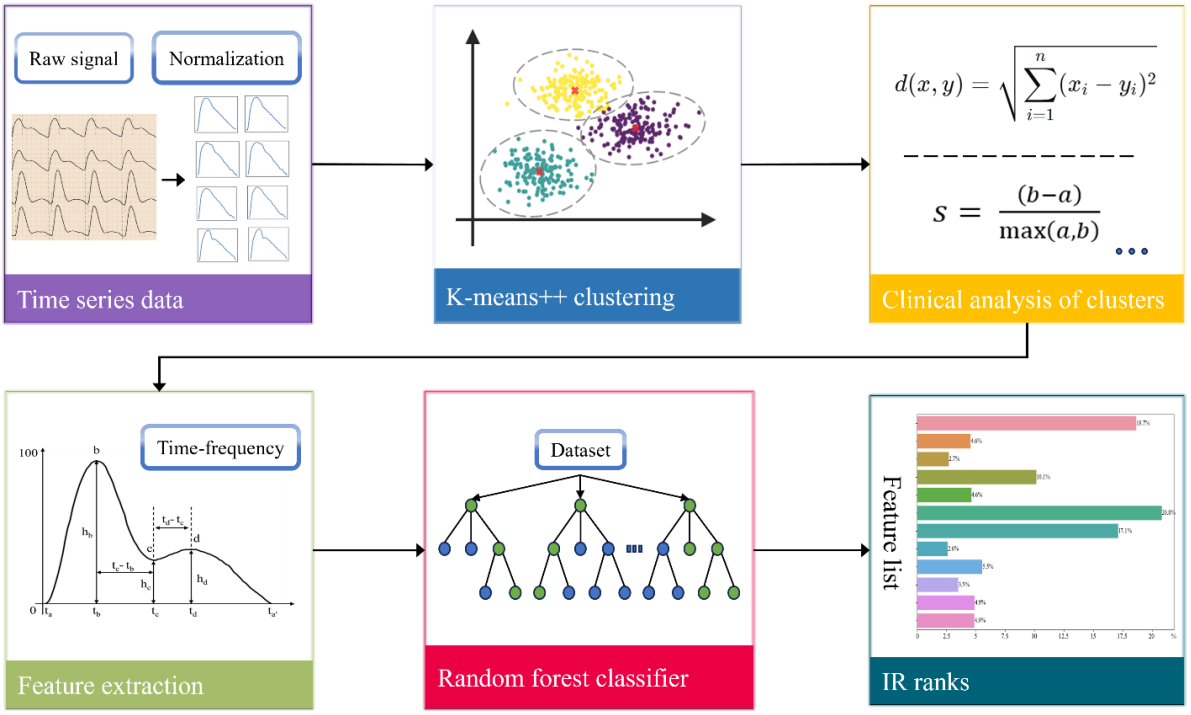


Fig. 2-1 Flowchart of pulse wave cluster analysis and feature importance ratio evaluation in HF patients.

2.2 Methods

2.2.1 Clinical data

The data collection encompassed unprocessed extremity wave signals along with clinical, physiological, and pathological information from 380 HF patients, in accordance with the Framingham HF diagnostic criteria [74]. This cohort originated from patient admissions at Chiba

University Hospital between January 2019 and December 2022. Following stabilization from acute HF episodes, measurements of pulse waves and blood pressure were conducted using Omron 203RPEIII apparatus, as shown in Fig. 2-2. Additionally, transthoracic echocardiography employing a Vivid E9 system (GE Healthcare, Horten, Norway) was performed within a week of the pulse wave assessments.

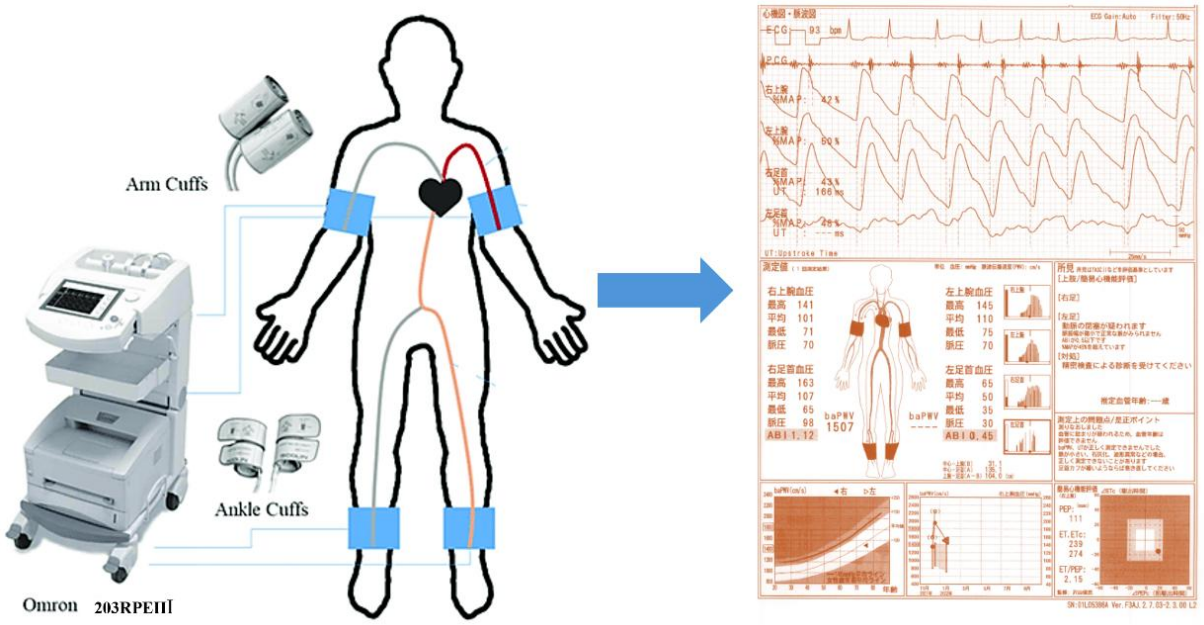


Fig. 2-2 Schematic diagram of limb blood pressure pulse detection device and clinical report. The device shown is an Omron 203RPEIII, which is connected to the cuffs placed on the patient's limbs. On the right, the clinical report displays pulse waveforms, heart rate, and various blood pressure measurements, including systolic and diastolic pressures at different limb locations. The diagram also presents analysis results such as pulse wave velocity (baPWV) and ankle-brachial index (ABI), providing a comprehensive overview of the patient's cardiovascular health.

Table 2-1 details the clinical characteristics of the included patients. Body mass index is calculated using the formula BMI = weight / height², indicating a person's weight relative to their height. The heart rate is obtained by measuring the number of heart beats per minute, typically using an electrocardiogram (ECG) or a fingertip pulse oximeter. Systolic blood pressure and diastolic blood pressure are measured with a blood pressure monitor, with the former indicating arterial pressure during heart contraction and the latter during heart relaxation. N-terminal pro B-type natriuretic peptide (NTproBNP) and B-type natriuretic peptide (BNP) are cardiac biomarkers

measured through blood tests, used to evaluate the risk of HF. Left ventricular ejection fraction is measured by echocardiography or magnetic resonance imaging, representing the percentage of blood ejected from the left ventricle during each heartbeat. Ejection time is measured via echocardiography or ECG combined with carotid pulse waveforms, indicating the duration of left ventricular ejection. These parameters are collectively used to evaluate cardiac and overall health.

Table 2-1: Clinical characteristics of the included patients. Mean, standard deviation, minimum, and maximum values of the clinical data related to the pulse wave dataset used.

Feature	Mean (± std)	Min	Max
Age[years]	67 ± 16	16	95
Weight[kg]	64 ± 20	34	215
Height[cm]	166 ± 73	125	185
BMI [kg/m2]	24 ± 6	14	76.18
HR [beats per minute]	90 ± 26	33	188
BPs[mmHg]	129 ± 33	71	266
BPd[mmHg]	80 ± 22	37	175
NTproBNP [pg/mL]	5691 ± 7924	203	45210
BNP [pg/mL]	531 ± 772	8.4	8323
LVEF [%]	39 ± 15	10	72.4
ET [ms]	251 ± 39	137	399

Note: *std* standard deviation, *BMI* body mass index, *HR* heart rate, *BPs* systolic blood pressure, *BPd* diastolic blood pressure, *BNP* B-type natriuretic peptide, *NTproBNP* N-terminal pro B-type natriuretic peptide, *LVEF* left ventricular ejection fraction, *ET* ejection time.

2.2.2 Time series dataset

The data set used in this study includes pulse wave signals obtained from 380 HF patients. The pulse wave signals were collected from the left upper arms of the HF patients and subsequently underwent denoising and normalization processes as described in previous studies [75,76]. Initially, wavelet transform decomposition was utilized to eliminate noise present in the signals [77]. To avert distortion of the pulse wave signals, the number of sampling points for each pulse wave cycle

was standardized to 100, in adherence to the Nyquist theorem and the actual sampling frequency [78,79]. Given that our research focuses on variations in the model, we normalized the amplitude of the pulse wave within each cycle to a range from 0 to 100, constructing a time series dataset.

2.2.3 t-SNE dimensionality reduction

A nonlinear dimensionality reduction method known as t-distributed stochastic neighbor embedding (t-SNE) was used to visualize the clustering results [80,81]. T-SNE is commonly utilized to depict high-dimensional data by projecting the distribution of data points from high-dimensional space to a low-dimensional space, making the clustering structure of the data more intuitively observable [82,83]. Specifically, the method models the data points in high-dimensional space using a probability distribution, where the distances between neighboring points are converted into conditional probabilities to reflect their similarity. Then, t-SNE finds a matching probability distribution into a lower-dimensional space and optimizes this matching process to generate a low-dimensional representation that preserves both the global and local structures of the original data [84]. Through this process, t-SNE effectively captures the complex structure within high-dimensional data, ensuring that the data points displayed in low-dimensional space retain their relative distribution and neighborhood relationships from the original dataset, thus providing a powerful tool for interpreting and analyzing clustering results.

2.2.4 Clustering algorithm

The K-means++ algorithm was employed to facilitate the clustering of pulse waves. The K-means++ algorithm represents an enhancement of the traditional K-means algorithm, introduced by David Arthur in 2006 [85–87]. It improves both the efficiency and clustering accuracy of K-means by strategically selecting initial centroids. As illustrated in Fig. 2-2, the algorithm begins by randomly selecting the first centroid from the dataset. Thereafter, it calculates the distance between each data point and the nearest selected centroid, using these distances to probabilistically choose additional centroids, with a preference for points farther from the existing centroids. This methodical selection of initial centroids is pivotal in enhancing the overall clustering quality. The process is repeated until k centroids (corresponding to the specified number of clusters) are chosen. Following this, the standard K-means algorithm is applied, where data points are assigned to the

nearest centroid, and centroids are iteratively updated to minimize within-cluster variance. The algorithm continues to iterate, refining the centroid positions, until convergence is achieved, indicated by the stabilization of centroid locations.

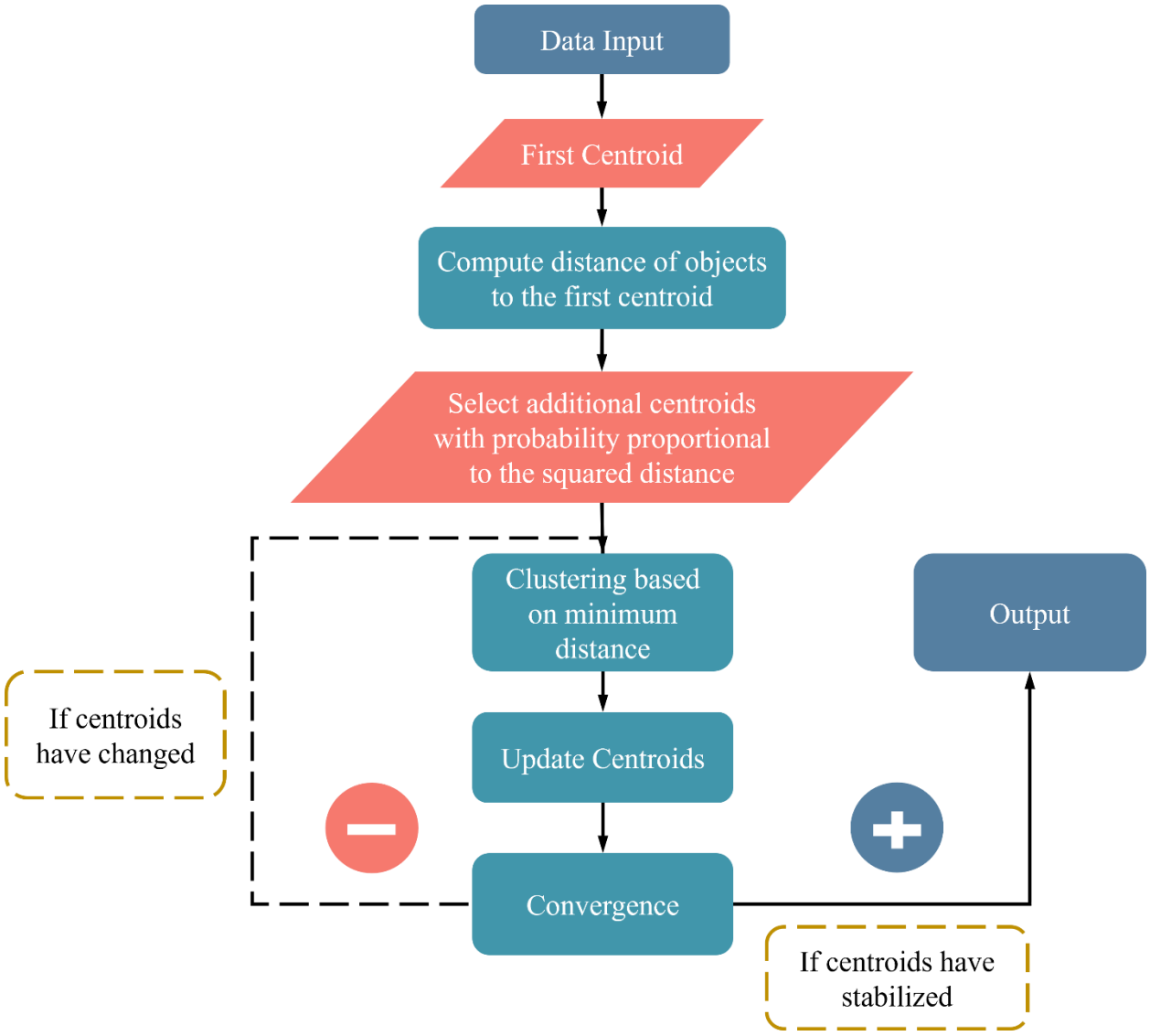


Fig. 2-2 K-means++ clustering algorithm process flowchart.

2.2.5 Clustering performance evaluation

We evaluated the clustering performance by calculating the Silhouette coefficient, Calinski-Harabasz (CH) index, and Davies-Bouldin (DB) index. The silhouette coefficient [88] is calculated

by considering the mean intra-cluster distance *a* and the distance *b* between a sample and the nearest cluster to which the sample does not belong. The formula for calculating the silhouette coefficient is as follows:

$$s = \frac{(b-a)}{max(a,b)} \quad (1)$$

The CH Index [89], also known as the Variance Ratio Criterion, evaluates the dispersion of clusters. It represents the ratio of between-cluster dispersion to within-cluster dispersion. A higher value indicates better-defined clusters. It is calculated as follows:

$$CH = \frac{B_k}{W_k} \times \frac{N-k}{k-1} \quad (2)$$

Where: *CH* is the CH index; Bk is the between-cluster dispersion matrix; W_k is the within-cluster dispersion matrix; *N* is the total number of data points; *k* is the number of clusters.

The DB Index [90] evaluates the average similarity of each cluster relative to its closest neighboring cluster. A lower value indicates better clustering. It is calculated as follows:

$$DB = \frac{1}{k}\sum_{i=1}^{k} \max_{j \neq i} \left(\frac{S_i+S_J}{M_{ij}}\right) \quad (3)$$

Where: *DB* is the DB index; S_i the average distance of all points in cluster i to the centroid of cluster i; S_j the average distance of all points in cluster j to the centroid of cluster j; M_{ij} is the distance between the centroids of clusters i and $\max_{j \neq i} \left(\frac{S_i+S_J}{M_{ij}}\right)$ identifies the cluster j that maximizes the ratio for each cluster i; *k* is the number of clusters.

2.2.6 Clinical relevance analysis of the obtained clusters

We first calculated the meaning clinical characteristics of each cluster to compare the differences between them. Then, we used a t-test to evaluate whether the differences in clinical parameters between clusters were statistically significant. The t-test is a widely utilized statistical technique for comparing whether there is a significant difference between the averages of two independent samples. The formula for the independent samples t-test is as follows:

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}} \quad (4)$$

Using this method, we can determine whether the differences in key clinical characteristics across clusters are significant enough, thereby revealing potential clinical stratifications between the groups. To determine if a significant difference exists between the two groups, we typically rely on the p-value obtained from the t-test results. The p-value represents the probability of observing the current data under the assumption that no difference exists between the two groups. If the p-value falls below the predefined significance threshold (typically 0.05) [91,92], we consider the difference statistically significant. This indicates that there is a meaningful difference between the two groups, indicating that the observed variations in clinical characteristics are not due to random factors but may reflect distinct clinical conditions. Through this analysis, we can identify which clinical characteristics exhibit statistically significant differences across the clusters, providing further insights into the clinical distinctions of these subgroups and supporting future clinical research or personalized treatment.

2.2.7 Importance ratio of time-frequency features

To systematically evaluate the importance ratios (IRs) of different pulse wave features in the clustering model, we first extracted a range of key time-domain and frequency-domain features to comprehensively capture various aspects of the pulse wave signals. In the time domain, the extracted features included mean, peak value, peak count, skewness, kurtosis, upstroke time, and downstroke time, which help describe the overall shape and trends of the signal over time [93]. In the frequency domain, the selected features comprise dominant frequency, mean power spectral density, frequency center, spectral entropy, and total harmonic distortion, primarily reflecting the energy distribution and complexity of the signal in the frequency dimension [94,95].

Next, to quantify the contribution of these features in distinguishing between different groups within the clustering model, we employed an RF classifier as the primary analysis tool. RF is capable of capturing feature importance by constructing multiple decision trees, offering both robustness and accuracy. In our experiment, we set the model to use 100 estimators (decision trees) and introduced a random seed to guarantee the consistency of the results. This process allowed us to reliably assess the weight of each time-domain and frequency-domain feature in the model's

decision-making process, helping us identify which features had the greatest impact on the clustering.

Furthermore, to enhance the intuitive understanding of the model's decision-making process, we selected and visualized a single decision tree from the RF model. This decision tree clearly illustrates how the model makes branching decisions based on various features, highlighting the key role certain features play in distinguishing between different clusters. This visualization helps explain the underlying decision logic of the model, simplifying the understanding of potential applications of pulse wave signal features in clinical stratification management. As shown in Figure 2-3, the decision path and the importance of each feature are displayed. This approach enables a deeper analysis of the role pulse wave features play in different clinical scenarios, providing a valuable reference for future personalized treatment and diagnosis.

This decision tree demonstrates the clustering classification based on various pulse wave parameters. The root node starts with "peak count," where binary decisions are made to split the samples. As the tree branches out, the samples are progressively split by thresholds on spectral features such as "upstroke time," "downstroke time," "frequency center," "mean psd," and "skewness." Each node in the tree corresponds to a decision made using a single feature. The node color indicates the predicted cluster (blue for Cluster 0, orange for Cluster 1), and the color depth is related to the purity of the node, with deeper colors indicating more confident classifications. Each node provides the Gini index (a measure of impurity), the number of samples classified at that node, and the distribution between the two clusters (Cluster 0 and Cluster 1). By following the decision path from the root down to the leaves, one can observe how each feature contributes to classifying the data into the two clusters.

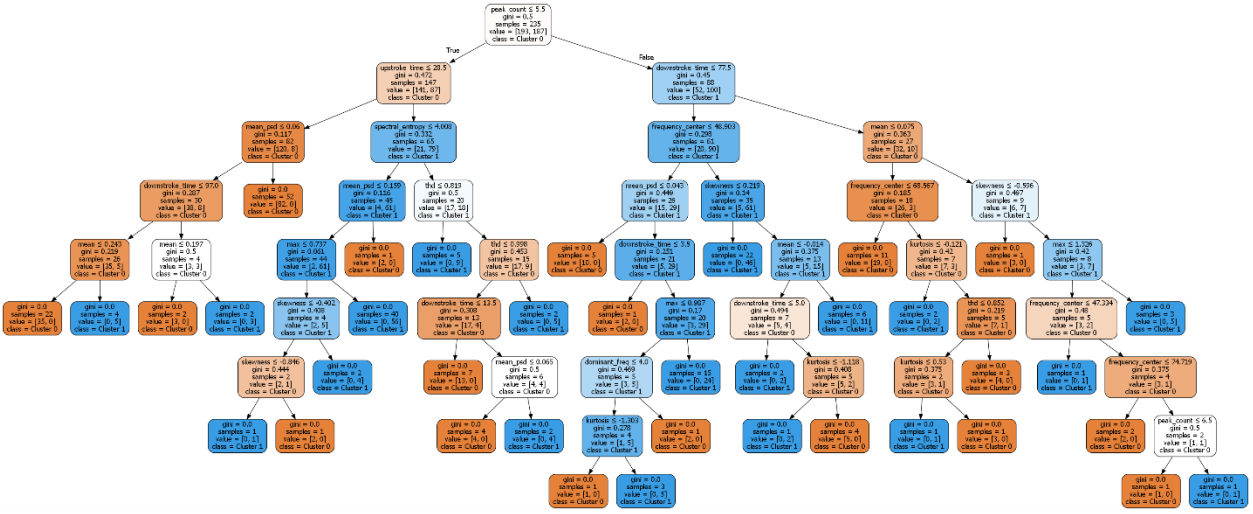


Fig. 2-3 Visualization of a single decision tree in the RF model: the key role of pulse wave features in the clustering model decision pathway.

2.3 Results

2.3.1 Clustering performance evaluation

We used the silhouette coefficient to determine the optimal number of clusters. When the number of clusters is two, we obtained cluster 0 and cluster 1. The silhouette coefficient for clustering based on time series is 0.74, indicating this is the optimal value. We believe that optimal clustering with fewer clusters is due to our dataset consisting only of HF patients with similar characteristics. This allows for a more refined differentiation of the pulse waveforms.

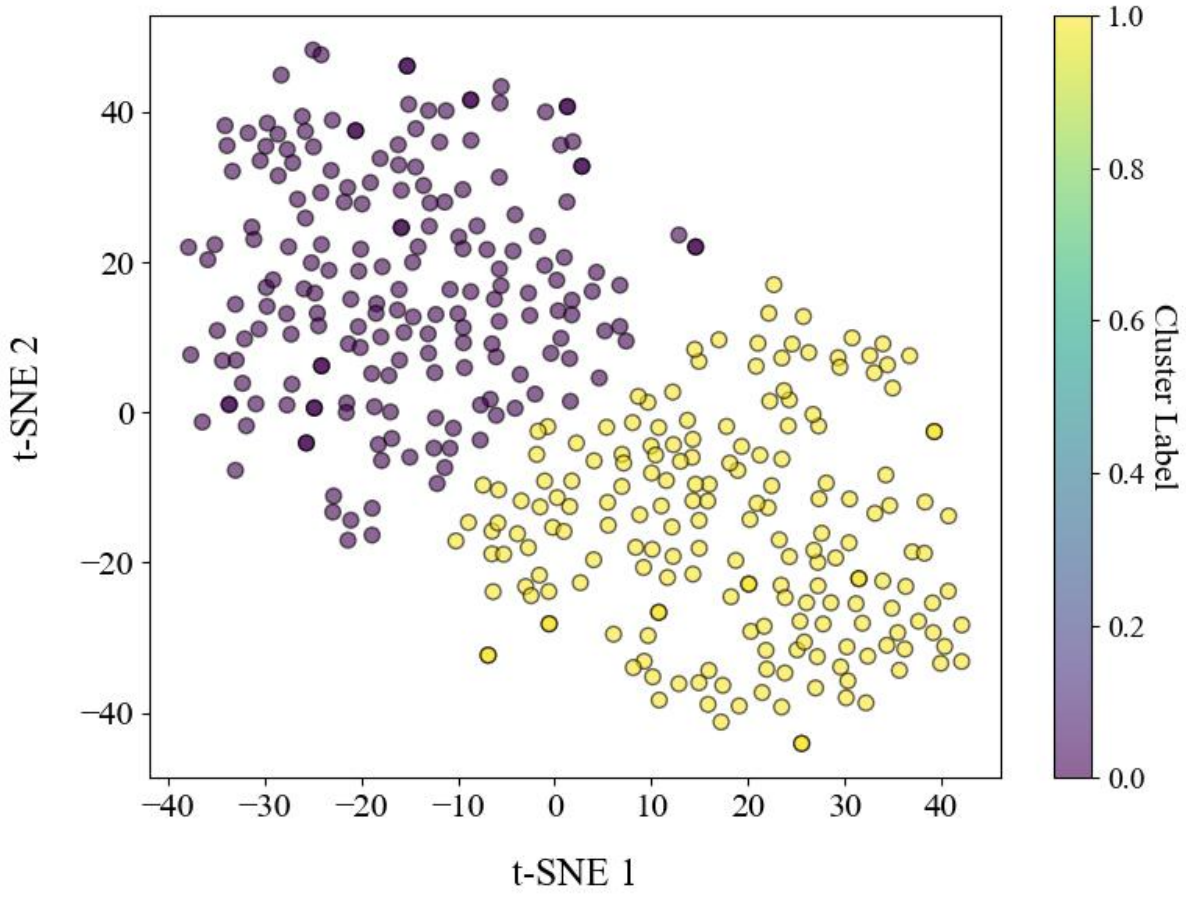


Fig. 2-4 T-SNE visualization of clustering approaches. Purple dots represent Cluster 0, and yellow dots represent Cluster 1. Each point represents a high-dimensional data point in a two-dimensional space. The t-SNE1 and t-SNE2 axes represent the two main directions after dimensionality reduction, with the values indicating the relative positions and distances between the data points.

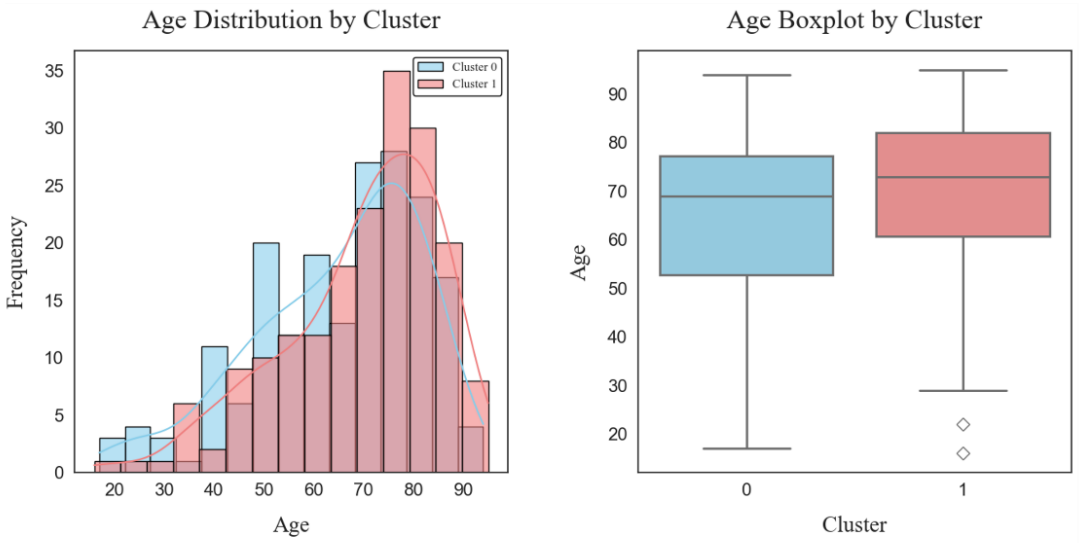
From Fig. 2-4, it is evident that the boundaries between Cluster 0 and Cluster 1 are clear and well-separated. To quantitatively evaluate the clustering effectiveness, we calculated the Silhouette coefficient, CH index, and DB index, as shown in Table 3. The Silhouette coefficient is 0.74, which is close to 1, signifying that the data points are effectively grouped within the clusters and clearly separated between clusters, suggesting an ideal clustering effect. The CH index is 585, indicating tight cohesion within clusters. The DB index is low at 0.75, further indicating low dispersion within clusters and high separation between clusters. These indices demonstrate that our clustering performance is high and reasonable. However, further analysis with clinical parameters is needed to explain the clinical relevance of the clusters.

2.3.2 Clinical relevance analysis of the obtained clusters

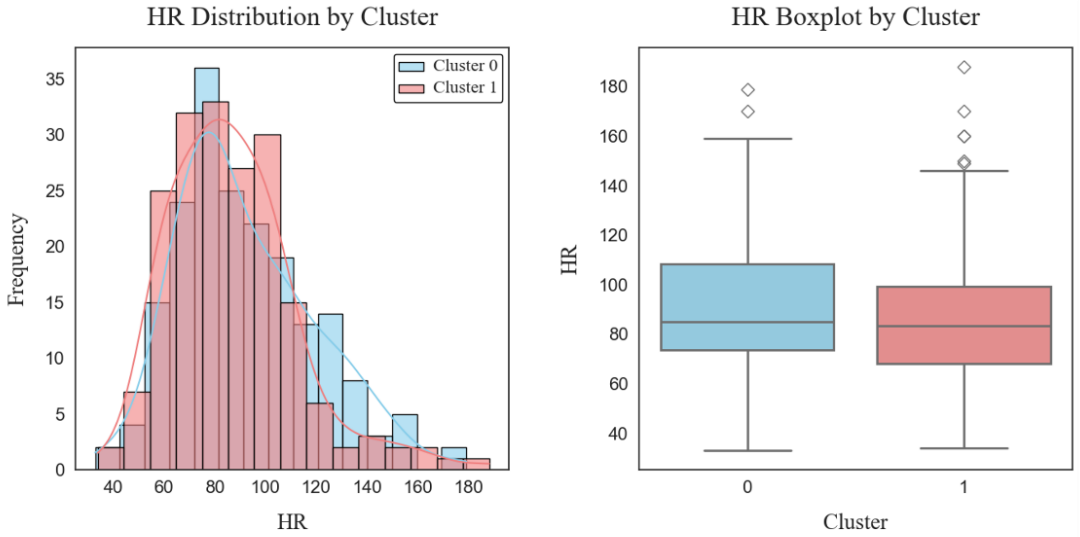
We calculated the mean and standard deviation of clinical characteristics for Cluster 0 and Cluster 1, followed by performing t-tests to assess statistical significance, as illustrated in Table 2-2. The t-test results demonstrated significant differences in age, HR, and ET between Cluster 0 and Cluster 1. Cluster 0 exhibited a lower proportion of elderly patients but a higher proportion of patients with elevated HRs and shorter ETs compared to Cluster 1.

To further visualize and analyze the differences in clinical characteristics between the two clusters, we plotted histograms of the distributions of HR, and ET for Cluster 0 and Cluster 1, as depicted in Fig. 2-5. Specifically, Fig. 2-5(a) shows that Cluster 0 has a wider age range and a more heterogeneous age composition, with an overall mean age lower than that of Cluster 1, where the ages of patients are concentrated between 70 and 90 years. The proportion of elderly patients (aged over 70 years) in Cluster 0 is 47%, compared to 57% in Cluster 1. Despite the higher proportion of elderly patients in Cluster 1, it has significantly fewer patients with high HRs compared to Cluster 0. As shown in Fig. 2-5(b), the proportion of patients with tachycardia (HR > 100 bpm) in Cluster 0 is 34%, compared to 23% in Cluster 1. Tachycardia in HF patients is associated with poorer prognosis [96,97], suggesting that despite the older age of Cluster 1 patients, their cardiac function is more stable. Furthermore, as shown in Fig. 2-5(c), the proportion of patients with ET less than 250 milliseconds in Cluster 0 is 54%, significantly higher than the 46% in Cluster 1, further indicating that the cardiac ejection efficiency of Cluster 1 patients is superior.

(a)


Age Distribution by Cluster
Cluster 0
Cluster 1
Frequency
Age
Age Boxplot by Cluster
Age
Cluster

(b)


HR Distribution by Cluster
Cluster 0
Cluster 1
Frequency
HR
HR Boxplot by Cluster
HR
Cluster

(c)

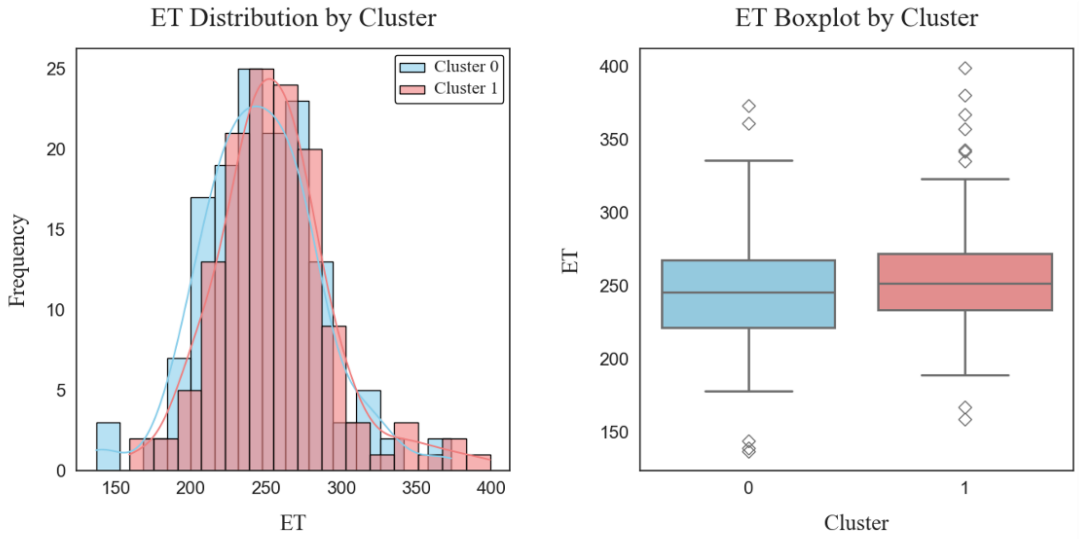

ET Distribution by Cluster
Cluster 0
Cluster 1
Frequency
ET
ET Boxplot by Cluster
ET
Cluster

Fig. 2-5 Histogram and boxplot for clusters 0 and 1. (a) Age (b) HR (c) ET. Each parameter is presented with a histogram (left) and a boxplot (right). The histograms display the frequency distribution, while the boxplots indicate the median, interquartile range, and outliers.

Table 2-2: t-test of clinical relevance for clustering results based on time series.

Feature	Cluster 0 Mean (± std)	Cluster 1 Mean (± std)	p
Age [years]	65 ± 17	70 ± 16	0.004 < 0.05
Weight [kg]	66 ± 19	63 ± 20	0.156
Height [cm]	163 ± 10	162 ± 10	0.714
BMI [kg/m2]	25 ± 6	24 ± 7	0.159
HR [beats per minute]	92 ± 27	87 ± 25	0.039 < 0.05
BPs [mmHg]	129 ± 34	130 ± 33	0.610
BP_d [mmHg]	81 ± 23	80 ± 22	0.553
NTproBNP [pg/mL]	5332 ± 8065	6050 ± 7835	0.644
BNP [pg/mL]	514 ± 728	547 ± 817	0.727
LVEF [%]	38 ± 15	39 ± 15	0.618
ET [ms]	247 ± 39	356 ± 39	0.049 < 0.05

These findings challenge the traditional perspective that older HF patients inherently have poorer prognoses. They underscore the transformative potential of time-series clustering analysis in clinical management. Our study demonstrates that detailed clustering analysis can identify subgroups of elderly patients with comparatively stable cardiac function, thereby providing new avenues for personalized treatment and prognosis optimization. This methodological approach emphasizes the critical importance of exploring latent pulse wave patterns without relying on predefined labels. Consequently, it offers more precise and individualized strategies for clinical management and prognostic evaluation of HF patients, enhancing patient outcomes and advancing the field of cardiology.

2.3.3 Importance ratio of time-frequency features

Based on the analysis presented, we identified two distinct subgroups within the HF patient population: Cluster 0 was characterized by fewer elderly patients, higher HR, and shorter ET,

compared to Cluster 1. To further elucidate the pulse wave features contributing to the differentiation of HF patients, we extracted time-frequency features from the pulse wave signals, as illustrated in Fig. 2-6 and Table 2-3.

Table 2-3: Mean and t-test results of time-frequency features for clusters.

Features	Cluster 0	Cluster 1	p
Mean	45 ± 6	50 ± 4	< 0.001
Peak value	94 ± 11	95 ± 5	0.172
Peak count	1.40 ± 0.58	1.26 ± 0.52	0.013
Skewness	0.24 ± 0.23	0.01 ± 0.22	< 0.001
Kurtosis	-1.04 ± 0.24	-1.11 ± 0.17	0.001
Upstroke time	20 ± 3	27 ± 5	< 0.001
Downstroke time	78 ± 3	72 ± 5	< 0.001
Mean psd	3.30 ± 0.43	3.40 ± 0.30	0.069
Frequency center	12 ± 3	11 ± 3	0.156
Spectral entropy	2.12 ± 0.17	1.93 ± 0.14	< 0.001
THD	0.58 ± 0.17	0.46 ± 0.12	< 0.001

Note: Mean psd means power spectral density, THD total harmonic distortion.

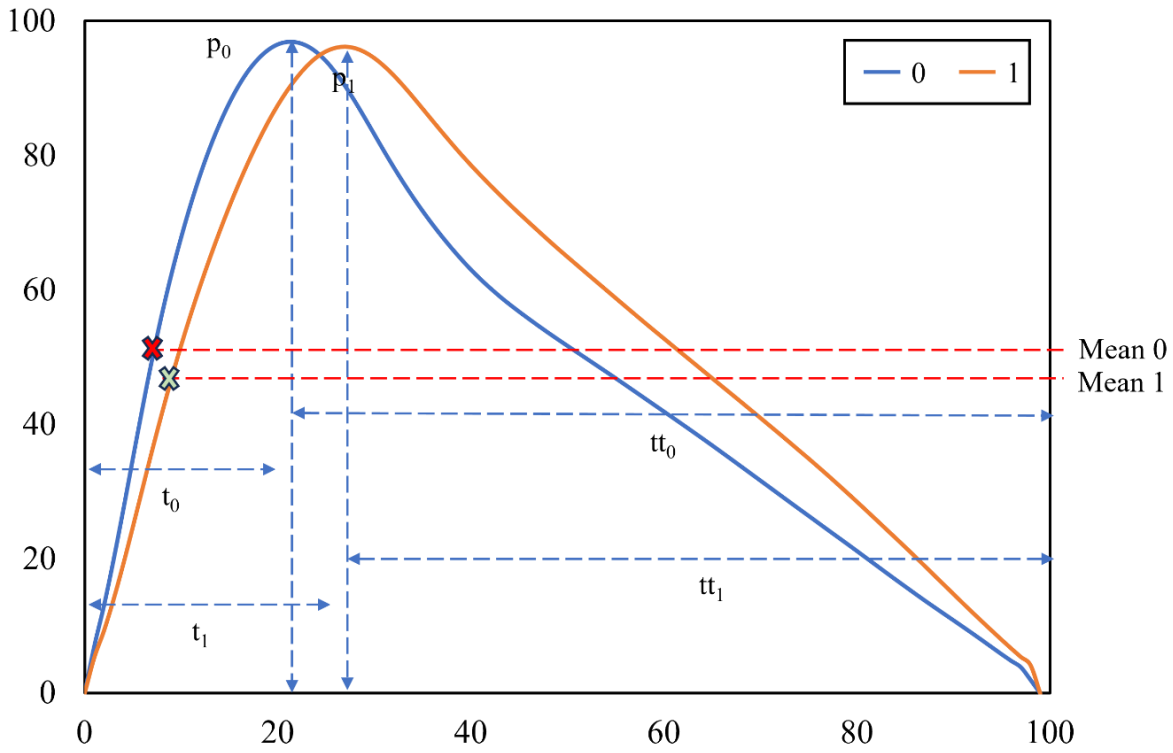
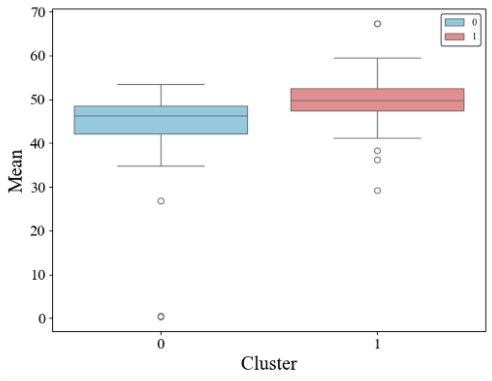
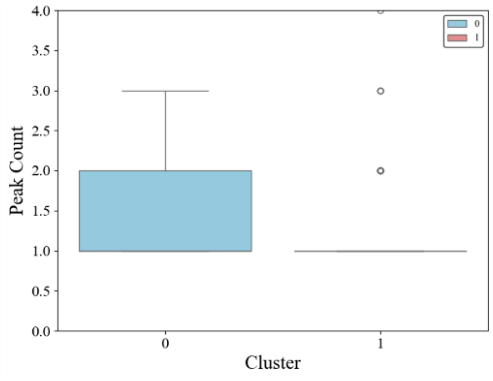
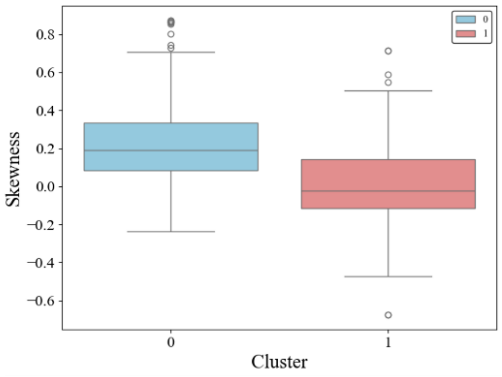


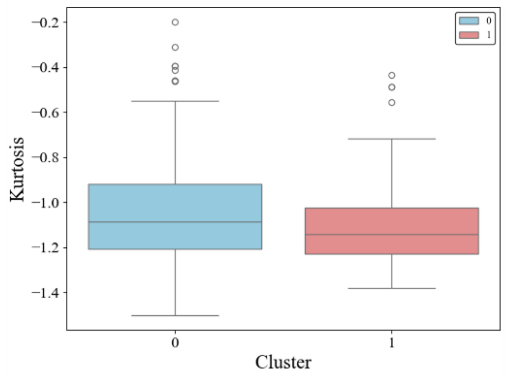
Fig. 2-6 Centroid waveforms and selected features of the Clusters. p0 and p1 represent the peak values of centroid waveforms 0 and 1, respectively. t0 and t1 represent the upstroke times of centroid waveform 0 and centroid waveform 1, respectively, while tt0 and tt1 represent the downstroke times of centroid waveform 0 and centroid waveform 1, respectively.

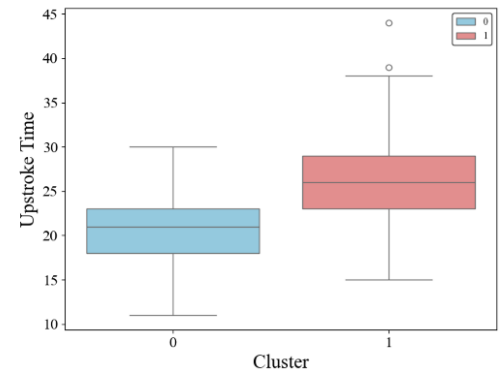
Figure 2-6 illustrates the centroid waveforms for Cluster 0 and Cluster 1. The centroid waveform, calculated as the meaning of all waveforms in the cluster at each time point, effectively encapsulates the predominant characteristics of each cluster. Distinct differences in pulse wave morphology, including mean, peak count, upstroke time, downstroke time, and skewness, are evident between the centroid waveforms of the two clusters. To rigorously quantify these differences, we conducted t-tests on the time-frequency features of the pulse waves. As detailed in Table 2-3, statistically significant differences were observed in mean, peak count, skewness, kurtosis, upstroke time, downstroke time, spectral entropy, and total harmonic distortion between the clusters, with p-values less than 0.05. To further elucidate these distinctions, we present boxplots of these eight parameters, highlighting the significant divergence between the clusters, as shown in Fig. 2-7.

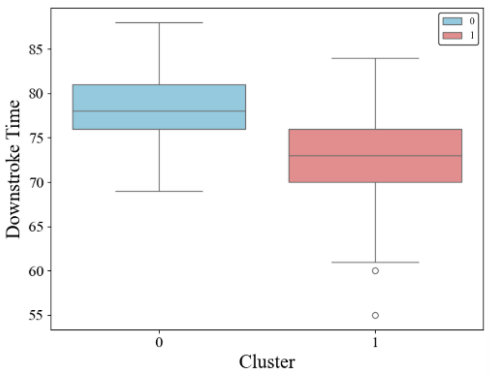

Mean
Cluster
0
1

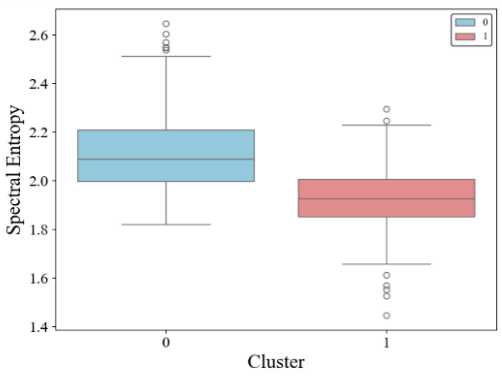

Peak Count
Cluster
0
1


Skewness
Cluster
0
1


Kurtosis
Cluster
0
1


Upstroke Time
Cluster
0
1


Downstroke Time
Cluster
0
1


Spectral Entropy
Cluster
0
1

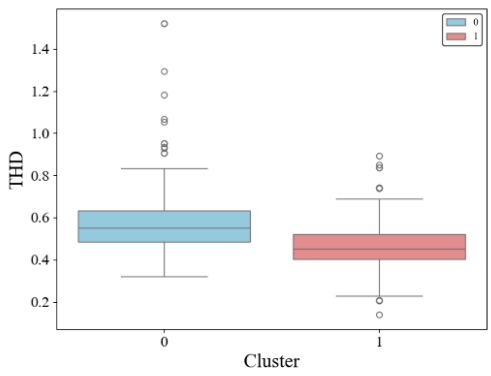

THD
Cluster
0
1

Fig. 2-7 Boxplots of key pulse wave features across clusters. Comparison of key pulse wave parameters between clusters 0 and 1. The boxplots illustrate differences in various pulse wave parameters for two distinct clusters, labeled as 0 (blue) and 1 (red). The central line in each box represents the median, while the boxes show the interquartile range, with whiskers extending to 1.5 times the interquartile range. Outliers are displayed as points outside this range, highlighting the variations in pulse wave characteristics between the two clusters.

We further quantified the relative importance of these pulse wave features using a RF classifier, calculating IR that indicates their contributions to the HF patient clustering model, as shown in Fig. 2-8. Upstroke time, mean, downstroke time, and skewness emerged as the four most critical features, with IRs of 20.8%, 18.7%, 17.1%, and 10.1%, respectively. This highlights their crucial role in distinguishing between HF patient clusters. It is recommended that clinicians pay particular attention to these differences in pulse wave features when monitoring and managing HF patients. As stated above, we found that upstroke time was the most important feature, explaining why Cluster 0 and Cluster 1 show significant differences in HR and ET. Cluster 0 exhibited significantly shorter upstroke times and longer downstroke times compared to Cluster 1, characteristics typically associated with higher HR and lower ET. This consistency with our clustering results further validates the rationale behind our findings.

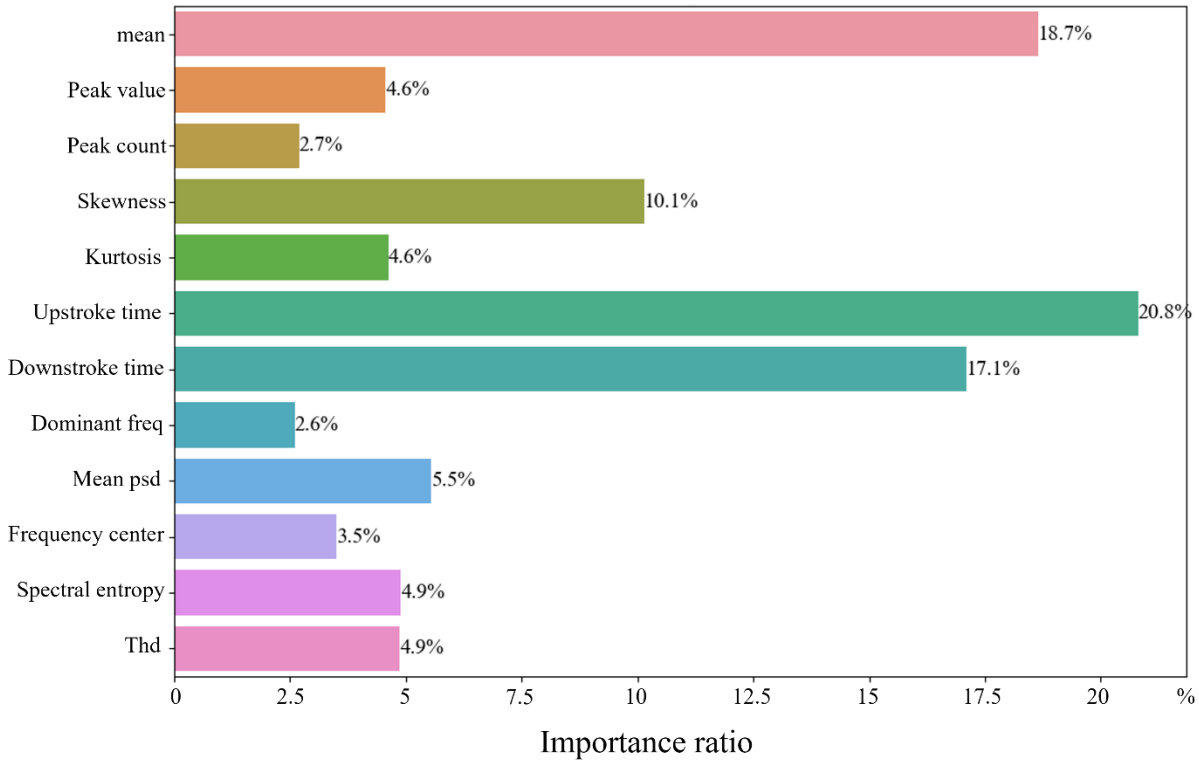


Fig. 2-8 Importance ratios of time-frequency features of pulse waves in clustering model. This horizontal bar chart shows the contribution of various features, with each bar representing a specific feature's percentage contribution to the overall model or analysis.

2.4 Discussion

In this study, we proposed a ML strategy that integrates unsupervised learning, pulse wave signal analysis, and feature importance evaluation. By employing the K-means++ algorithm, we conducted unsupervised clustering of pulse wave time series in HF patients to explore the clinical characteristic differences between clusters. Our results identified two distinct patient subgroups, with significant differences in age, HR, and ET. Additionally, for the first time, we introduced the concept of IR to perform an in-depth analysis of the time-frequency characteristics of pulse waves within the clustering model. These innovative approaches offer significant understanding of the stratification of HF patients and the potential for more personalized and precise management strategies.

Accurate diagnosis and effective treatment are crucial for improving the quality of life and prognosis of HF patients. Advanced feature extraction techniques, such as AI-driven analysis of clinical signals like pulse waveforms, have become the mainstream for predicting cardiovascular health. This study utilized the K-means++ unsupervised clustering algorithm to successfully classify the pulse wave time series of HF patients into two distinct subgroups with significant clinical differences. These subgroups exhibited marked variations in key clinical indicators such as age, heart HR, and ET. Specifically, 47% of patients in Cluster 0 were elderly, 34% had tachycardia, and 54% exhibited low ET; whereas in Cluster 1, 57% of the patients were elderly, 23% had tachycardia, and 46% had low ET. Notably, our study not only confirmed the age differences between subgroups consistent with the findings of Yousef et al. [98], Alty et al.[99], and Brillante et al. [100], but also, for the first time, revealed significant differences in HR and ET across these subgroups. Importantly, our findings challenge the traditional notion that elderly patients necessarily suffer from significantly deteriorated cardiac function. Instead, our research suggests that in HF patients, a higher age does not necessarily correlate with tachycardia and low ET. The successful identification of these two subgroups further validates the potential of pulse wave time series in identifying HF patient subgroups, optimizing management strategies, and enabling personalized monitoring. The choice of unsupervised learning methods was driven by the need to overcome the inherent limitations of traditional supervised learning approaches, which heavily depend on predefined labels, limiting their ability to handle data from patients with multiple comorbidities and reducing the dataset size. Through unsupervised learning, we were able to uncover deeper patterns in pulse waveforms, revealing latent characteristics within the HF patient population, thereby achieving more accurate patient stratification. Our study not only achieved exceptional clustering performance (Silhouette coefficient: 0.74; CH index: 585; DB index: 0.75), but also provided innovative methodological guidance for clinicians, enabling more efficient resource allocation in personalized medical management, reducing unnecessary waste, and improving patient outcomes.

To further elucidate and clarify the key factors involved in identifying HF patient subgroups through pulse wave analysis, we pioneeringly quantified the Importance Ratios (IR) of time-frequency features within the clustering model. The results revealed that upstroke time, mean value, downstroke time, and skewness were the four most influential features, with IRs of 20.8%, 18.7%, 17.1%, and 10.1%, respectively. These features played a crucial role in differentiating the patient

subgroups and provided valuable insights into underlying pathophysiological differences. Notably, the upstroke time was the most important key morphological feature in subgroup identification, which further explains the significant differences in HR and ET observed between Cluster 0 and Cluster 1. As HR increases, the speed of cardiac contraction accelerates, leading to a shortened upstroke time [101,102]. Cluster 0 exhibited a significantly shorter upstroke time compared to Cluster 1, a pattern consistent with the observed characteristics of higher HR and lower ET in Cluster 0, thereby validating the rationale behind our clustering results. By focusing on these critical features, clinicians can gain deeper insight into the cardiovascular status of patients, significantly enhancing the accuracy of HF diagnosis, monitoring, and prognostic evaluation. This highly targeted approach not only strengthens the scientific basis for HF management but also improves the precision of personalized prognostic predictions, leading to more favorable clinical outcomes for patients. Our study not only provides a new perspective on understanding the complex physiological mechanisms in HF patients but also paves the path for future progress in precision medicine and tailored treatment approaches.

Despite the significant findings of this study, it is important to acknowledge its limitations and explore potential directions for future improvements. First, HF is a complex and heterogeneous syndrome with a wide range of clinical presentations. While our sample is representative, it does not encompass all types of HF patients, which may limit the generalizability of the results. Future studies should focus on confirming these results in broader and more varied patient populations to ensure the broad applicability and robustness of the conclusions. Second, not all patients' data included complete clinical characteristics (e.g., BNP levels), which may have impacted the relevance of our analysis. Future studies should place greater emphasis on data completeness, particularly when incorporating multidimensional clinical data, to improve the precision and dependability of the analyses. Furthermore, the absence of a healthy control group in this study may have restricted the scope and depth of the clustering analysis. The lack of a control group could limit the identification of potential subgroups, thereby affecting our comprehensive understanding of the intrinsic heterogeneity within the HF patient population. Future studies should consider including a healthy control group to provide a more comprehensive baseline reference, thereby enhancing the robustness of the findings. To address these limitations, future research should not only expand the sample size and diversity but also incorporate long-term follow-up data to track patient prognosis and risk factors. Such longitudinal studies will provide

deeper insights into the progression and evolution of HF, offering a scientific basis for optimizing individualized treatment plans. By identifying high-risk patients and those with poor prognosis through pulse wave clustering techniques, this approach can make clinical management strategies more precise, improve the efficiency of resource allocation, and enhance clinical outcomes for HF patients. This research framework offers new perspectives to drive future progress in precision medicine and individualized healthcare, laying a robust basis for the advancement and optimization of clinical practice strategies.

2.5 Conclusion

This study is the first to apply the K-means++ algorithm for unsupervised clustering of pulse wave time series in HF patients, identifying two patient subgroups and revealing significant differences in age, HR, and ET between the clusters. Additionally, the study quantified the importance of pulse wave morphological features. The findings provide preliminary evidence for stratified management of HF patients using non-invasive and easily accessible pulse wave signals. Looking ahead, we plan to optimize clustering methods, expand the pulse wave dataset, and incorporate more relevant clinical information to further explore the hidden patterns in HF patients' pulse waves, thereby providing precise and personalized management strategies to enhance patient outcomes and overall quality of life.

Chapter III

Identification of Left Ventricular Enlargement in HF Patients through Classification Algorithms

3.1 Introduction

In recent years, HF (HF) has become an increasingly significant burden on public health [103–105], while ventricular structure remodeling has been recognized as a pivotal process in the development of CVDs, particularly the progression of HF [106,107]. The determinants affecting the left ventricular remodeling: (I) the intensity, longevity, and quickness of elevation in pressure burden; (II) the volume load [108]; (III) factors like age, racial/ethnic background, and gender [109]; (IV) accompanying conditions including coronary artery disease, diabetes, obesity, and valve-related heart issues [110]; (V) the hormonal environment within the nervous system; (VI) changes in the extracellular matrix [111]; and (VII) hereditary influences [112]. LVE, commonly reflecting ventricular remodeling, is strongly linked to symptoms like cardiac dysfunction, HF, and arrhythmias [113–115]. The increase in left ventricular size due to LVE exerts a severe impact on the clinical prognosis of patients with mild with or without severe HF [116]. Recent studies have identified LVE as a precursor to left ventricular dysfunction and clinical HF in asymptomatic individuals [117,118]. LVE is recognized as a risk factor for both the occurrence and mortality rates associated with CVDs (CVDs) [119], serving as a critical indicator of cardiac events, closely associated with deteriorating cardiac function and unfavorable prognosis [120].

LVE provides a crucial marker in cardiac pathology, and accurate evaluation of LVE facilitates the diagnosis and assessment of heart diseases and enables the prediction of cardiac events and prognosis, providing valuable guidance for medical treatments and management strategies. Thus, as illustrated in Fig. 3-1(a), the monitoring and evaluation of LVE are particularly important from a clinical perspective. Detection of LVE primarily relies on the patient's medical history and imaging tests, with cardiac magnetic resonance (CMR) often being the most precise method due to its ability to accurately assess the heart's morphology, size, and position [121–123]. CMR is not routinely available due to its high cost, time-consuming, and magnetic field limitations for patients with metal implants [124,125]. Consequently, transthoracic echocardiography (TTE), known for its noninvasive and real-time imaging, is widely regarded as an effective, patient-friendly method for evaluating and diagnosing cardiac functions [109,126,127]. However, the accessibility and convenience of TTE are limited due to the requirement for specialized equipment and skilled operators. This research leverages pulse waves as a key informational conduct within the cardiovascular system (CVS) to address these limitations and offer a noninvasive, cost-

effective solution. Building on our previous study, we demonstrate that pulse waves can effectively infer CVS conditions and aid in diagnosing CVDs [128]. By exploring the use of pulse waveforms, this research identifies them as a promising tool for rapidly identifying and evaluating LVE (Fig. 3-1(b)).

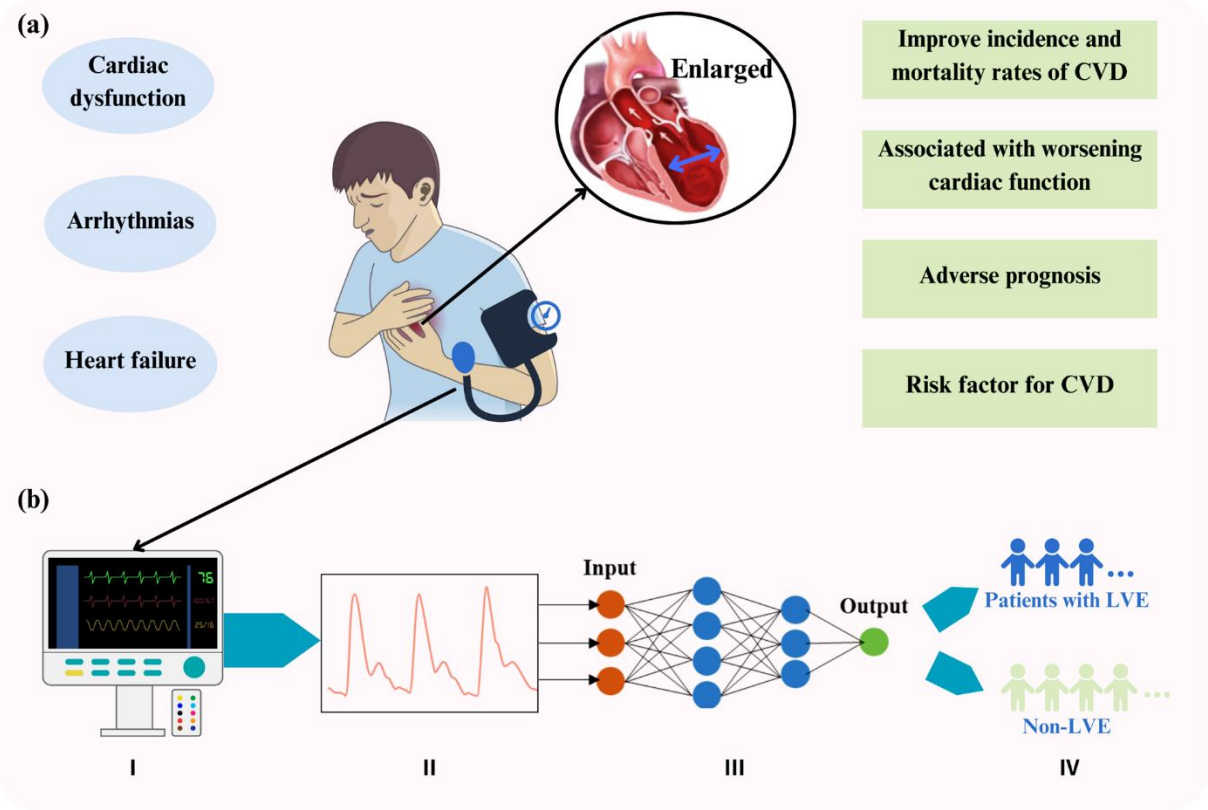


Fig. 3-1 Illustration of (a) LVE and representative symptoms in CVDs, data-driven LVE detection, and (b) ML-based strategy for predicting LVE using pulse-wave signals.

The physiological signal of pulse waves has been extensively utilized for health monitoring and CVD prediction [129–132]. Pulse waves contain valuable information on both physiological and pathological states of the human CVS while providing crucial physiological data related to blood delivery capacity and transport efficiency [128,133,134]. Non-invasive measurements of pulse wave signals can now be easily implemented using a range of affordable household electronic devices, providing helpful information for the low-cost and patient-friendly diagnosis of CVDs and relevant complications [135]. Recently, ML and DL methodologies have been employed for the analysis of pulse wave signals, demonstrating high potential and feasibility in terms of pulse-wave pattern classification and cardiac function prediction [57,136–138]. Wang et

al. successfully classified 407 datasets of pulse waveforms into five patterns by developing a Bayesian network based on six pulse-waveform parameters of depth, width, length, frequency, rhythm, and intensity, achieving classification with a success rate of 84% [139]. Xu et al. classified 320 datasets of pulse waveforms into 16 distinct patterns through a fuzzy neural network, identifying variations in pulse shape, width, position, and particular localized parameters, enabling a classification success rate of 90% [135]. Li et al. proposed a CNN model to classify pulse waveforms associated with five diseases, achieving a success rate of 95%. [140]. More recently, in our previous study [67], Wang et al. established an optimized ML strategy that enables quick and precise predictions of three cardiovascular functional parameters based on a pulse wave database of 412 patients, demonstrating the practicality and promise of ML-driven, pulse wave-based predictions of cardiac function. In this study, we aim to propose and establish a pulse wave signal-driven ML-based strategy for identifying and evaluating LVE in HF patients and to provide a clinically effective, patient-friendly, low-cost, and non-invasive method for early diagnosis and monitoring of CVDs.

3.2 Methods

Our ML framework consists of four parts: data processing, data screening, Fourier series calculation, and ML model analysis, as shown in Fig. 3-2:

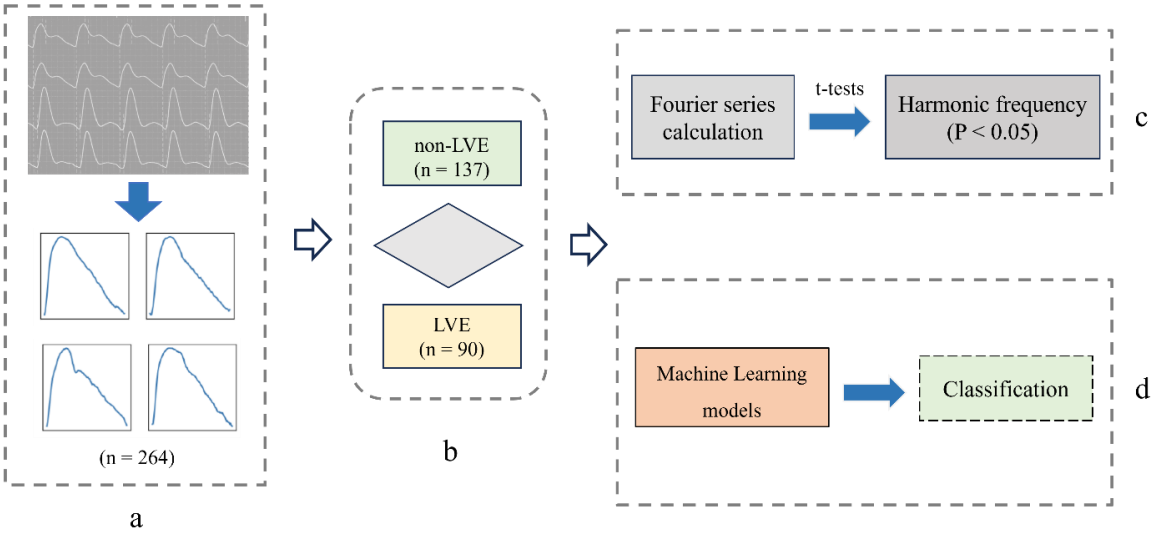


Fig. 3-2 Framework of the ML process: (a) Data processing. (b) Data screening. (c) Fourier series calculation. (d) ML models.

3.2.1 Ethics approval

This research received ethical approval from the Ethics Review Board of Chiba University Graduate School of Medicine in 2021 (Approval Number: M10089). The clinical data collection and analysis adhered to the applicable standards and rules.

3.2.2 Clinical data

The dataset comprised 264 HF patients, including unprocessed peripheral pulse wave data along with clinical. HF diagnosis was determined using the Framingham HF diagnostic criteria [141]. All patients were admitted to Chiba University Hospital between January 2019 and December 2022. Following stabilization of the acute HF condition, blood pressure and pulse waves were recorded using dedicated detection devices (Omron 203RPEIII). Additionally, all patients underwent TTE with the Vivid E9 system (GE Healthcare, Horten, Norway) within a week preceding or following the pulse wave measurements.

The observed parameters of relevant clinical information were also collected. Cardiovascular function parameters primarily include parameters related to cardiac pumping function, systolic and diastolic function, cardiac work, and vascular elasticity. The parameters of pumping function include cardiac output, stroke volume, stroke index, and ejection fraction. Cardiac output refers to the total volume of blood ejected by the ventricles in one minute and is an important indicator for evaluating the effectiveness of the cardiovascular circulatory system. In a resting state, the human body ejects approximately 4500 to 6000 milliliters of blood per minute [142]. Stroke volume, which represents the difference between the end-diastolic volume and the end-systolic volume, can evaluate the contractile ability of the myocardium. In a resting state, stroke volume is about 65 to 70 milliliters. Stroke index refers to the stroke volume per square meter in a resting state, approximately 40 to 60 milliliters per beat per square meter. Ejection fraction is the percentage of stroke volume relative to end-diastolic volume and can also assess myocardial contractility. In a resting state, it is around 55%-65%, with values below 50% considered indicative of HF [143,144].

Parameters of left ventricular systolic function include left ventricular function index, systolic index, and cardiac systolic function index, which are key parameters for analyzing cardiac systolic function.

Left ventricular diastolic function parameters include left ventricular end-diastolic volume and left ventricular end-diastolic pressure, which describe the volume and pressure within the left ventricle at the end of diastole. These are critical parameters for analyzing diastolic function.

Cardiac work parameters include left ventricular stroke work, left ventricular minute work, stroke work index, and cardiac function index. Left ventricular stroke work and minute work refer to the work done by the ventricle in a single contraction and over one minute, respectively. Vascular elasticity function parameters include vascular compliance and peripheral resistance.

3.2.3 Pulse waves

Pulse wave signals were collected from the left upper arms of HF patients, followed by applying denoising and normalization techniques grounded in methodologies from previous studies [145,146]. First, we utilized wavelet transform decomposition to remove noise from the signals[147]. Then, to avoid distortion of the pulse wave signals, we set the number of sampling points for each pulse wave cycle to 100, considering the Nyquist theorem and the actual sampling frequency[148,149]. Since our study focused on variations in the pulse wave model, we normalized the pulse wave amplitude within each cycle to a range of 0 to 100.

3.2.4 Datasets

Rigorous data screening was undertaken to ensure data quality. The 264 patients with HF were confirmed to satisfy the following screening criteria: 1) pulse wave data were obtained from the left upper arms of the patients; 2) for each patient, five or more accurate and reliable pulse wave cycles were captured; and 3) LV size, including LV diastolic diameter (LVDd) and LV systolic diameter (LVSd), was measured by transthoracic echocardiography. Our clinical data were used to evaluate the LVE of each patient based on the guidelines provided by the American Society of Echocardiography (ASE) [141]. LVE was defined as LVDd with an index, namely, LVDdI >36 mm/m2 for males and >37 mm/m2 for females[120].

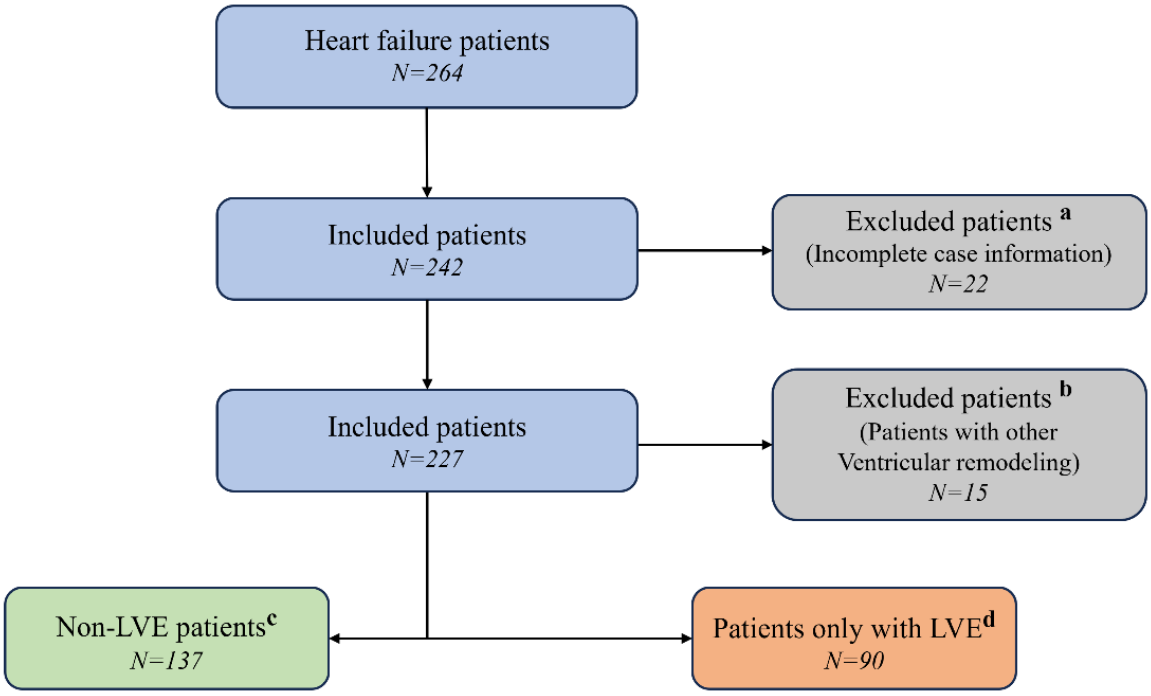


Fig. 3-3 Flow chart of patient screening under screening criteria a, b, c, and d: (1) exclusion of 22 subjects with incomplete information [a], (2) exclusion of 97 patients with other cardiac remodeling [b], (3) creation of a non-LVE patient group of 137 Subjects [c], and (4) creation of an LVE patient group of 90 Subjects [d].

The clinical data for 264 patients underwent classification, as depicted in Fig. 3-3. This dataset excluded 37 patients because of incomplete records. Among the remaining 227 qualified patients, 137 were identified as non-LVE patients, while 90 were diagnosed with LVE. The K-means clustering algorithm derived characteristic waveforms from these two datasets. These waveforms were then decomposed into 3rd-order Fourier series to explore the variances between the datasets, detailed as follows:

$$F(t) = A_0 + A_1\,cos(\omega t) + B_1\,sin(\omega t) + A_2\,cos(2\omega t) + B_2\,sin(2\omega t) + A_3\,cos(3\omega t) + B_3\,sin(3\omega t) \quad (5)$$

Where A_0 is the period-average value, A_1-A_3 and B_1-B_3 are the 1st-order to 3rd-order Fourier series coefficients, respectively, t is time, and ω is the angular frequency. Independent sample t-tests were then performed to statistically analyze the difference between the patient datasets with LVE and non-LVE.

3.2.5 Machine learning models

In this part, our objective was to address the classification problem between patients with LVE and non-LVE. We constructed a pulse wave dataset using data from 227 HF patients. The dataset was split into two subgroups, with roughly 80% allocated to the training dataset and the remaining 20% to the testing dataset. To select an appropriate ML model, we conducted numerous preliminary experiments to test and compare the predictive performance among different models, such as the weighted random forest (WRF) model, support vector machine (SVM) model, and fully convolutional neural network (FCNN) method.

3.2.5.1 SVM model

SVM is a commonly employed supervised learning technique for both classification and regression tasks [150]. It functions by identifying a separating hyperplane within the feature space that maximizes the margin between two classes, thus achieving classification. The idea of SVM is to construct a hyperplane that separates the samples of different classes as effectively as possible, while maximizing the distance between the hyperplane and the nearest data points from each class, known as support vectors, to improve the generalization ability of the classifier. In the case of linearly separable data, SVM can directly find a linear hyperplane, whereas for non-linearly separable data, SVM employs kernel functions (e.g., linear kernel, radial basis function, etc.) to transform the data into a higher-dimensional space, where a separating hyperplane can be identified. SVM excels at handling high-dimensional and small-sample datasets and is extensively used in areas such as text classification, image recognition, and gene expression data analysis.

3.2.5.2 Fully convolutional neural network

FCNN is a DL model specifically designed for dense prediction tasks, such as image segmentation and object detection, which require pixel-level predictions. Unlike traditional Convolutional Neural Networks (CNNs), FCNNs consist entirely of convolutional layers, eliminating fully connected layers, allowing them to produce outputs matching the spatial dimensions of the input, as depicted in Fig. 3-4. By using only convolutional layers, FCNNs maintain the spatial information of the input image, making them highly suitable for pixel-level tasks. Without fully connected layers, FCNNs is capable of processing input images with varying dimensions, providing greater flexibility. To recover the resolution lost during convolution and pooling processes, FCNNs utilize up sampling layers, such as transposed convolution or bilinear

interpolation, to ensure the output maintains the same spatial dimensions as the input image. Additionally, FCNNs use skip connections to combine features from different layers, allowing the fusion of abstract semantic information from deeper layers combined with detailed features from shallower layers, thereby improving prediction accuracy. FCNNs are widely applied in fields like semantic segmentation, object detection, medical image analysis, and optical flow estimation, excelling in pixel-level tasks due to their ability to retain spatial information and support end-to-end training.

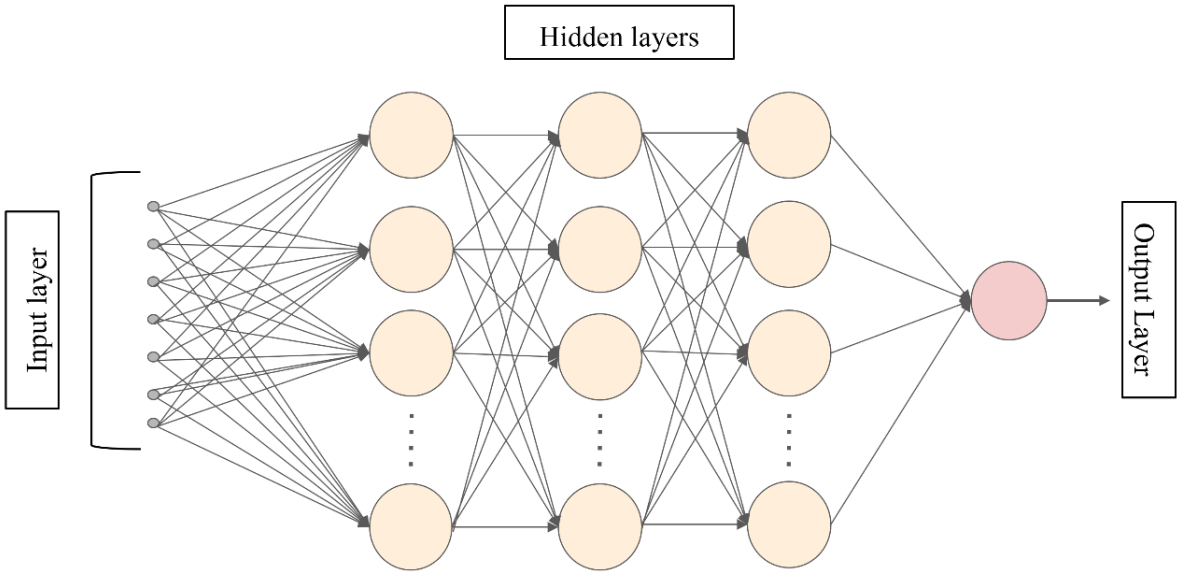


Fig. 3-4 Schematic Diagram of a FCNN Structure. In this network structure, each neuron in every layer is connected to all neurons in the previous layer, forming a fully connected architecture.

3.2.5.3 WRF model

The RF algorithm is a powerful an ensemble learning technique that merges the outputs of several decision trees to perform ML tasks such as classification and regression. Its fundamental principle involves generating various subsets of the dataset created using bootstrapping, with each decision tree being independently trained on its own subset. At each decision node, instead of using all available features, the RF randomly selects a subset of features and chooses the best one for splitting. This random selection of samples and features improves the model's generalization ability, reduces the correlation between decision trees, and effectively minimizes the risk of overfitting. In classification tasks, the RF algorithm determines the final classification by having all decision trees vote, and the class with the most votes is determined as the final output. This

majority voting mechanism ensures high accuracy and robustness in classification tasks. For regression tasks, the RF generates the final prediction by averaging the outputs of all the decision trees, ensuring stability and precision in predicting continuous variables. One of the key advantages of the RF algorithm is its ability to effectively reduce overfitting. By aggregating the decisions of multiple trees, it avoids the issue where a single decision tree might perform effectively on the training set but fails to generalize well to the test set. Additionally, RF handles high-dimensional data well. By randomly selecting features at each split, it reduces the computational complexity associated with processing large feature sets while maintaining sensitivity to the most important features.

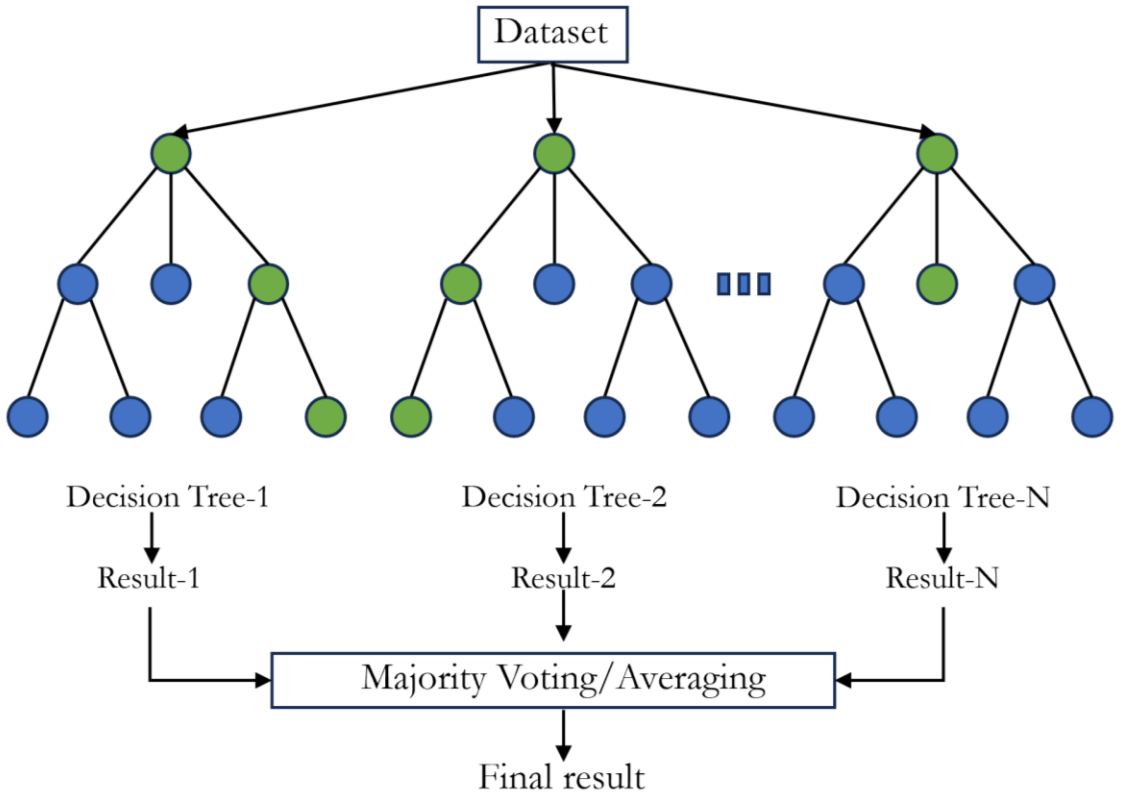


Fig. 3-5 Schematic Diagram of a RF Structure. The diagram shows the basic workflow of the RF model. A RF is composed of multiple decision trees, each of which is independently trained on the input dataset. The final prediction result is generated by combining the outputs of each decision tree through majority voting or averaging methods.

Due to the imbalance in the data from the two groups of patients, we employed the WRF method, where different weights were assigned to the two groups during the training process. We

utilized the class weight parameter of the RF classifier from the scikit-learn library to adjust the weights based on the class distribution, aiming to improve the performance of the minority class. We extensively investigated various weight settings and substantially determined weight parameters of 1 and 4 for non-LVE and LVE patients, respectively. To reduce the likelihood of overfitting, we used 5-fold cross-validation and recorded the average scores to evaluate the model's performance more accurately.

3.2.6 Performance evaluation

The performance of the weighted RF-based classification model was evaluated with standard classification metrics of accuracy, recall, F1 score, and AUC-ROC [151–153], as defined below.

$$Accuracy = \frac{TP+TN}{TP+PN+FP+FN} \tag{6}$$

$$Recal = \frac{TP}{TP+FN} \tag{7}$$

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

$$F1\ score = \frac{2\ Precision \times Recall}{Precision+Recall} \tag{9}$$

Here, *TP* and *TN* represent the total count of correctly detected positive and negative events, while *FP* and *FN* denote the total count of erroneously detected positive and negative events.

To assess the model's discrimination capability for the two datasets, we employed ROC analysis to derive the AUC. From the plots of the false-positive rate (FPR) on the x-axis and the true positive rate (TPR) on the y-axis, the ROC curve was obtained to calculate the AUC. TPR and FPR are given as follows:

$$TPR = \frac{TP}{TP+FN} \tag{10}$$

$$FPR = \frac{FP}{FP+TN} \tag{11}$$

The confusion matrix was further visualized to assess the model performance regarding the training and testing datasets.

3.3 Results

3.3.1 Fourier series calculation

Through data screening and pre-processing, we successfully generated two high-quality pulse wave data sets specifically tailored to LVE and non-LVE patients. As shown in Fig. 3-6, the representative waveforms of the two groups obtained based on K-means clustering show noticeable differences. Harmonic power decomposition (Fig. 3-7) was performed using the 1st to 3rd order Fourier coefficients to examine how the harmonics contribute differently to the waveforms of the two groups. The pulse waveform in LVE is mainly dominated by the 1st order Fourier coefficient, but less by the 2nd order: the LVE group has a significantly higher power, but a slightly lower power than the non-LVE group, and there is rarely a discrepancy in the 3rd order. This implies that the LVE-induced irregular movement of the LV exerts an essential impact by the 1st order on dominating the feature of the pulse waves.

We further analyzed the harmonic frequency dependency of the two pulse wave datasets through independent sample t-tests, summarized in Table 3-1. In the frequency domain, we observed a significant difference between the 1^{st-} and 2^{nd}-order harmonics ($p < 0.05$), while the influence of the 3rd-order harmonic was a margin ($p > 0.05$). This indicates that the datasets of the LVE and non-LVE groups are primarily affected by the 1st- and 2nd-order harmonics, while there is a notable frequency domain distinction between them. This provides further evidence of the validity associated with the dataset creation and an available guideline for data selection in subsequent classification and regression tasks, substantially enhancing the accuracy of pulse wave-driven identification of LVE.

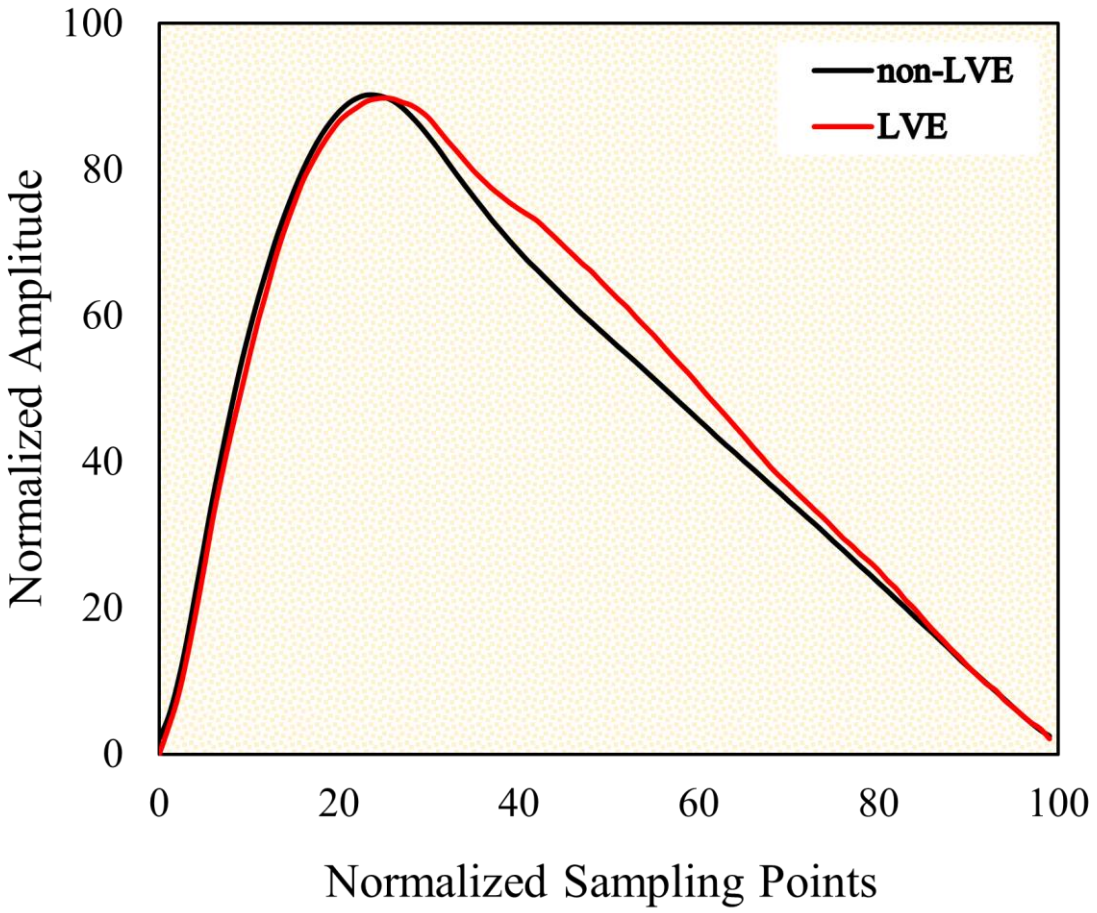


Fig. 3-6 Comparison of representative pulse waveforms associated with LVE and non-LVE patients. Normalized amplitudes vs. normalized sample points.

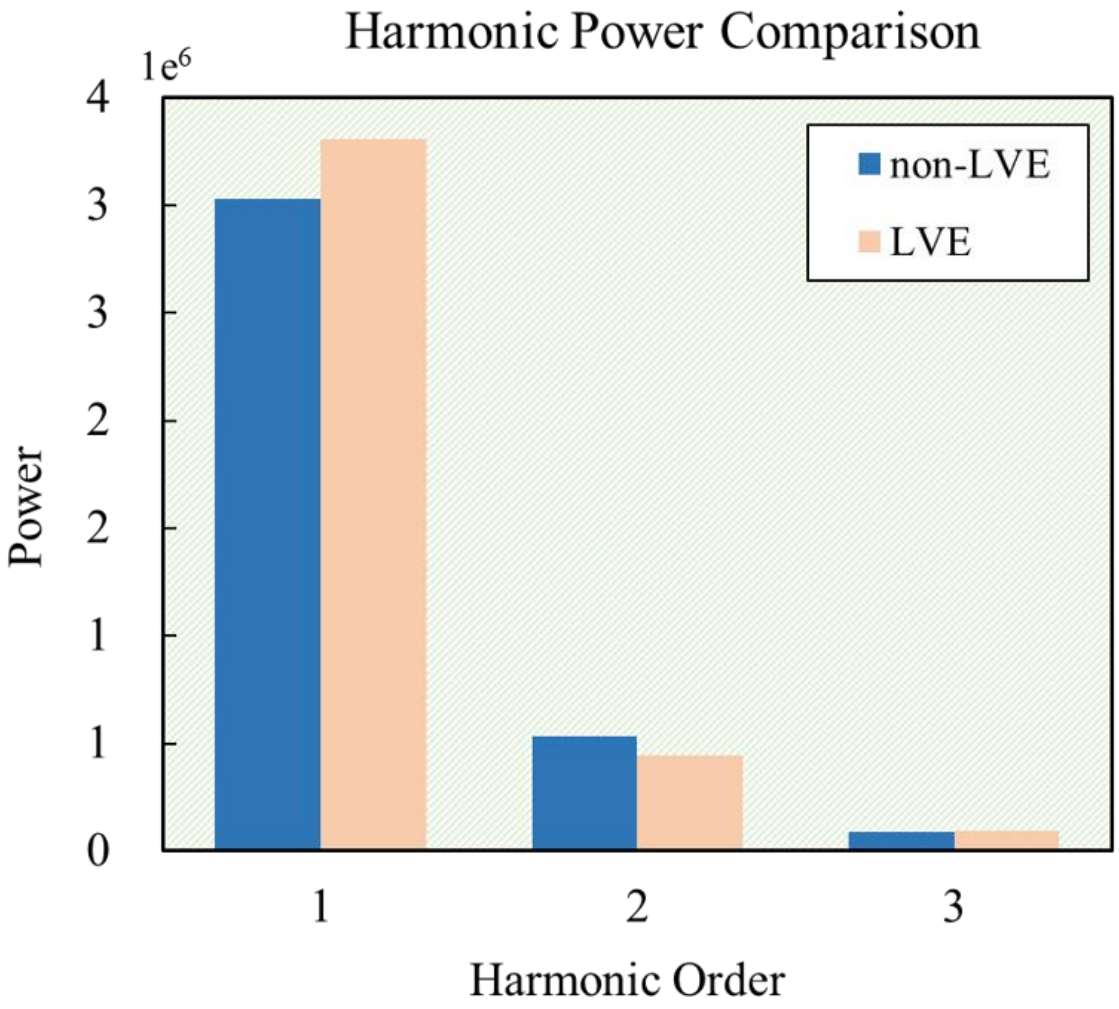


Fig. 3-7 Comparison of representative pulse waveforms associated with LVE and non-LVE patients. Harmonic powers vs. harmonic order.

Table 3-1: t-tests on the 1st-, 2nd-, and 3rd-order harmonics

Harmonic order	Non-LVE	LVE	p
1	3.11±0.65	2.92±0.74	0.048
2	0.61±0.21	0.69±0.35	0.027
3	0.13±0.10	0.12±0.09	0.363

Note: Data are presented as the mean ± SD; p values were calculated using the independent samples t-test, where $p<0.05$ represents significant differences.

3.3.2 Classification model

For a classification model performance test, the accuracy results of three models (WRF, SVM, and FCNN) are compared in Table 3-2. The WRF model shows the best classification performance with an accuracy of 91%: its specific classification metrics, as summarized in Table 3-3, for the LVE and non-LVE groups achieve an overall accuracy of 0.91, which is a remarkably high-accuracy classification, while the non-LVE group presents a slightly better performance (accuracy 0 = 0.93, recall H1 = 0.93, F1-score = 0.93) compared to the LVE group (accuracy 1 = 0.89, recall H2 = 0.89, F1-score = 0.89). This indicates that the current WRF classification model can achieve high prediction precisions with high recalls and F1 scores in identifying and predicting the most positive samples.

Table 3-2: Comparison of classification accuracy among the three methods.

ML-models	Number	Accuracy
WRF	227	0.91
SVM	227	0.81
FCNN	227	0.77

Table 3-3: Classification performance of the WRF model.

	Precision	Recall	F1-score	Support
Non-LVE	0.93	0.93	0.93	29
LVE	0.89	0.89	0.89	17
Accuracy		0.91		46

The classification performance was further visualized using the confusion matrix, which, as shown in Fig. 3-8, was employed to illustrate the classification outcomes of the LVE and non-LVE groups. While 2 out of 17 in the LVE group and 2 out of 29 in the non-LVE group were misclassified, the classification model overall demonstrated a capability to achieve the high-

accuracy prediction of 93% for the non-LVE group and 88% for 5 LVE patients, even for the limited number of patients. To assess the sensitivity and specificity of the classification model, we plotted the ROC curves in Fig. 3-9, where a scalar metric of the AUC was employed at an AUC of 0.93 for the WRF model. The results indicate that the current classification model enables excellent differentiation between the LVE and non-LVE groups in up to 93% of cases compared to the perfect case of AUC=1.

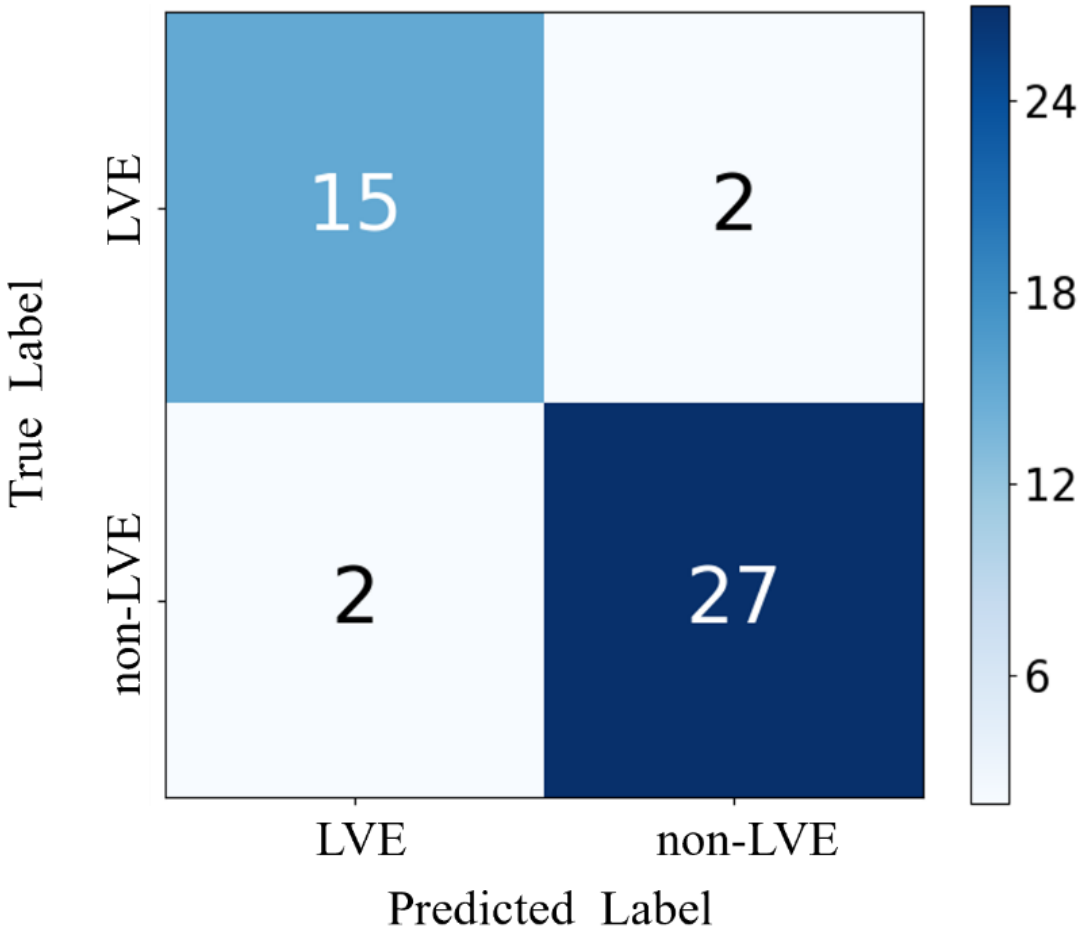


Fig. 3-8 Confusion matrices of classification model.

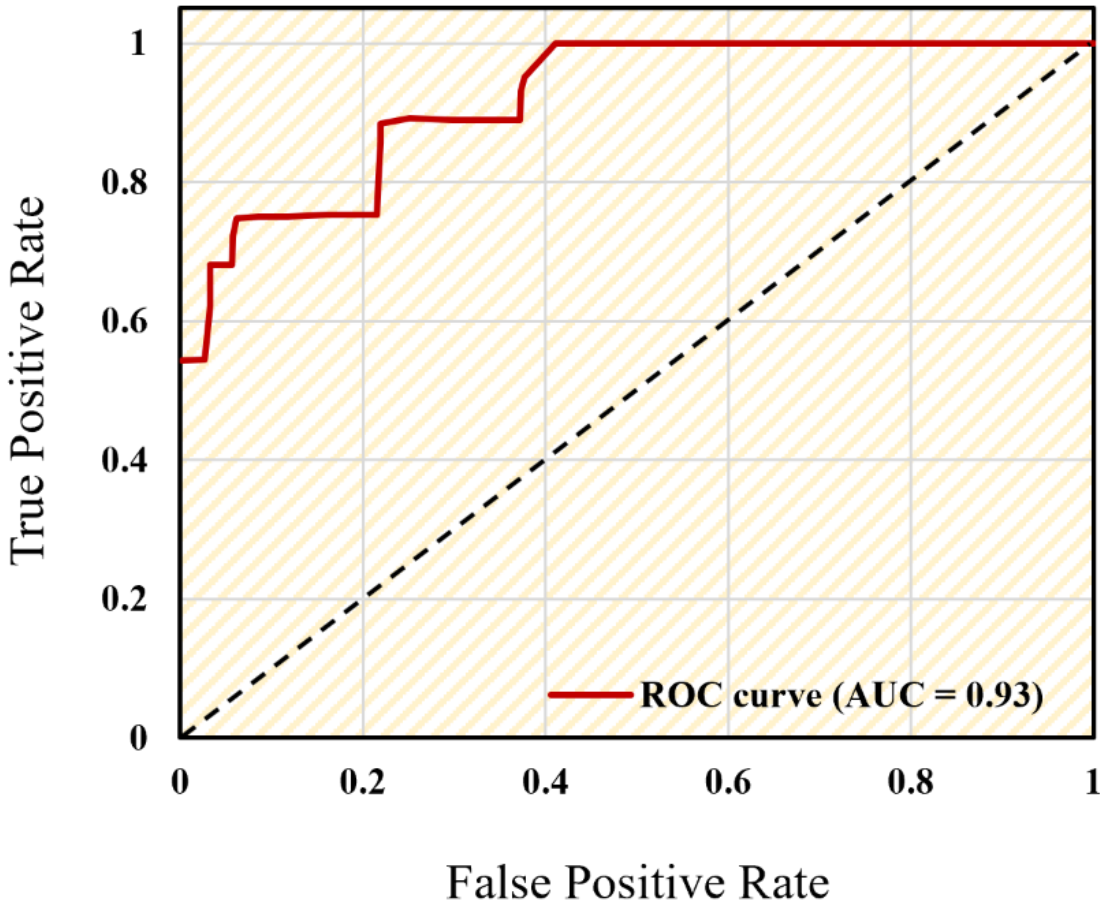


Fig. 3-9 ROC curve of classification model, AUC=0.93.

3.4 Discussion

LVE plays a pivotal role in the clinical evaluation of HF. In this study, we demonstrate that employing pulse wave signals combined with advanced ML techniques can provide a promising avenue for quantitative identification with LVE and non-LVE in HF patients.

Based on a high-quality pulse wave dataset consisting of 264 HF patients from both LVE and non-LVE groups, we proposed and implemented an ML-based classification strategy, utilizing a WRF model to classify patients with LVE. To assess the accuracy of the model, we compared the classification results with clinical labels. The findings demonstrated that the model effectively and accurately detected LVE in HF patients. The established WRF model significantly distinguished between the LVE and non-LVE groups, achieving a classification accuracy of up to 93%. These

findings indicate that the proposed classification strategy has strong feasibility and application potential, showcasing the superior performance of ML-driven approaches in LVE classification and identification among HF patients using pulse wave signals. This method provides an effective, non-invasive, data-driven solution for cardiac health management and holds promise for future use in clinical diagnostics and personalized treatment.

From a clinical perspective, the current pulse wave-driven, ML-based identification methodology can provide a noninvasive and low-cost tool for evaluating and diagnosing LVE in HF patients. As a noninvasive, real-time cardiac imaging technique to detect LVE, transthoracic echocardiography (TTE) is widely utilized. However, it has certain limitations regarding accessibility and convenience, which may lead to delayed medical treatments. This issue becomes even more crucial with expensive, high-sensitivity medical equipment such as non-enhanced multilayer spiral CT [154]. Recently, some attempts to utilize AI methods have been conducted. Nam et al. [155] developed a way to detect left atrial enlargement (LAE) and LVE on chest X-rays using DL algorithms, but the results for LVE detection were all $p>0.05$ compared to those for LAE, suggesting a need for inclusion of more images. The current ML-based methods demonstrate a patient-friendly and feasible tool for the first time to effectively differentiate LVE patients from the LVE and non-LVE groups using pulse wave signals. With the rapid advances in portable electronics and wearable devices such as smartphones and smartwatches, the noninvasive measurement of such physiological data has become more convenient and cost-effective. Pulse waves can thus be utilized for LVE detection, holding immense potential in assisting clinical management and treatment.

There are limitations in this study in terms of the relatively small sample size of data, the limited data scope, the singular data source, and a lack of pertinent clinical information. Although we adopted a rigorous approach in data selection and successfully created a high-quality classification dataset, the validation based on preliminary experiments still requires further improvement; data sourced from a single institution may not fully represent the diverse population of HF patients. Additionally, the lack of healthy subjects and patients with other CVDs may bring a potential bias, as it fails to encompass the entire HF demographic. To augment the generalizability of the ML models, there is an urgent need to expand the dataset by including a more diverse patient cohort. Further improvements in predictive accuracy and robust feasibility of

the ML-based methodology to meet practical clinical applications may be accomplished by integrating multiple data sources, such as TTE and clinical biochemical markers, which will be our future endeavors. More efforts will improve the precision and reliability in training and testing with larger datasets, amalgamating relevant clinical information, and attempting patient classification based on LVE severity.

3.5 Conclusion

The ML-based strategy successfully identified the patients with LVE in HF from those without LVE in HF. The proposed ML methods have been validated to provide effective classification with outstanding performance in identifying LVE in HF patients. This points to the potential and superiority of identifying patients with LVE based on pulse waves. Our study thus underscores the significance of the ML-based methodology for clinical practice, offering a reliable method for identifying and treating cardiac remodeling.

Chapter IV

Prediction of Left Ventricular End-Diastolic Diameter index in HF Patients

4.1 Introduction

HF is a common and severe CVD defined by the heart's failure to adequately circulate enough blood to satisfy the body's demands [156,157]. LVE, particularly the increase in left ventricular end-diastolic diameter (LVDd), is one of the key indicators of HF progression and is closely associated with worsening cardiac function. LVE was defined as LVDd indexed (LVDdI) as >36 mm/m² for males and >37 mm/m² for females according to the American Society of Echocardiography [158]. Monitoring LVDdI provides a quantitative understanding of left ventricular remodeling, which plays a crucial role in the management and treatment of HF patients. Accurate and timely prediction of LVDdI can help clinicians make personalized treatment decisions for patients, thereby improving outcomes and reducing mortality.

In recent years, the non-invasive analysis of physiological signals, such as pulse wave signals, has gained significant attention in the field of cardiovascular research [159–161]. Pulse wave signals convey critical insights into cardiovascular dynamics and have the potential to predict various cardiac parameters [162,163]. Advances in AI, particularly ML models, have provided new opportunities for leveraging pulse wave data to predict clinical indicators like LVDd. Among various ML approaches, DL models such as Densely Connected Convolutional Networks (DenseNet) [164] have demonstrated impressive performance in extracting meaningful patterns from complex physiological signals due to their deep feature extraction capability and efficient layer connectivity.

DenseNet is a type of neural network that introduces direct connections between each layer and every other layer in a feed-forward manner [165,166]. This architecture allows the network to reuse learned features and enhances the flow of information, making DenseNet especially suitable for tasks requiring the processing of high-dimensional input data, such as pulse wave signals. Alghamdi et al. introduced a new method that combines autoencoder and DenseNet architectures to predict heart disease, achieving a high mean accuracy of 99.67% on the Heart Disease UCI Cleveland dataset. The model demonstrated excellent performance, with a test accuracy of 99.99% after further testing, along with outstanding macro precision, recall, and F1 scores [59]. Hadaate Ullah et al. proposed an automatic end-to-end 2D CNN with DenseNet for arrhythmia recognition, achieving impressive results on the MIT-BIH and INCART datasets with accuracy of 99.80% and

99.63%, respectively [167]. The model surpasses current state-of-the-art approaches, demonstrating high precision and F1-scores, and is further validated through transfer learning on class-imbalanced datasets, showing its potential for real-life applications in arrhythmia diagnosis. This paper proposes a method for accurate aortic dissection diagnosis at the patient level using CT angiography (CTA) images. Xiong et al. presents a novel multi-lead myocardial infarction (MI) localization approach using DenseNet, achieving superior performance with 99.87% accuracy, 99.84% sensitivity, and 99.98% specificity, demonstrating its potential for clinical application in MI diagnosis [168]. Krishnan et al. proposed a transfer learning-based model using DenseNet121 for CVD prediction, enhanced by a hybrid Lion-based Butterfly Optimization Algorithm for feature selection. The model demonstrated significant improvements in accuracy, precision, recall, and F1-Score, while reducing training time by focusing on a smaller, more relevant feature set [169].

In this study, we proposed a DenseNet network-based approach to predict the LVDdI from pulse wave signals. By training the model on pulse wave data collected from HF patients, we aim to demonstrate that DenseNet can reliably predict LVDdI and provide insights into left ventricular function. This approach has potential for early diagnosis and monitoring of HF progression, providing a non-invasive, efficient, and accurate approach for clinical assessment.

4.2 Methods

4.2.1 Ethics approval

The approval number for this study is the same as that in section 3.2.1. The collection and analysis of clinical data were conducted in compliance with relevant guidelines and regulations.

4.2.2 Clinical data

The dataset of this study is consistent with that in section 3.2.2. The diagnosis of HF was based on the Framingham HF diagnostic criteria [170]. After the stabilization of acute HF, pulse wave and blood pressure were measured and recorded using a blood pressure/pulse wave monitoring device (Omron 203RPEIII). Additionally, all patients underwent TTE within a week preceding or following the pulse wave assessment. Relevant clinical parameters were also recorded.

4.2.3 Pulse waves

We first obtained pulse wave signals from the left upper arm of HF patients and processed the data using denoising and normalization techniques based on previous studies. During the denoising process, wavelet transformation was used to decompose the signals and eliminate noise [171]. Then, to prevent distortion of the pulse wave signals, and in accordance with the Nyquist theorem and the real sampling frequency, the number of sampling points per pulse wave cycle was set to 100 [172]. Since this study focuses on variations in the pulse wave model, the amplitude of each pulse wave cycle was normalized to a range of 0 to 100.

4.2.4 Datasets

Rigorous data screening was undertaken to ensure data quality. The 264 patients with HF were confirmed to satisfy the following screening criteria: 1) pulse wave data were collected from the left upper arms of the patients; 2) for each patient, five or more consistent and usable pulse wave cycles were documented; and 3) LV size, including LVDd and LV systolic diameter, was measured by transthoracic echocardiography. Our clinical data were used to evaluate the LVE of each patient based on the guidelines provided by the American Society of Echocardiography (ASE) [173]. LVE was defined as LVDd with an index, namely, LVDdI >36 mm/m2 for males and >37 mm/m2 for females [158].

4.2.5 Regression model

The densely connected convolutional network (Dense Net) was employed for regression prediction of LVDdI, a recently established, innovative network architecture capable of excelling in efficient feature extraction and regression prediction tasks [174]. In the context of our limited dataset, the dense net is verified, enabling the effective employment of data features, and achieving high performance while mitigating overfitting. The relationship between the input and output of the (n+1)-th layer (feature map) associated with the dense block module is given as:

$$Output^{n+1} = feature\ map = G^n(Output^1, Output^2, \dots, Output^n) \quad (11)$$

G represents several operations, including the rectified linear unit (ReLU), batch normalization, and convolution.

This diagram illustrates the architecture of a DL model with several key components. The model begins with an input layer, followed by a convolutional layer for feature extraction, and then passes through multiple dense blocks (Dense Block 1, Dense Block 2, and Dense Block 3), which include convolution and max-pooling operations (C+M). After each dense block, information is passed to additional modules depicted in separate boxes, suggesting lateral connections or auxiliary processing. Max-pooling is applied before the fully connected (FC) layer, and finally, the output layer produces the model's prediction.

As shown in Fig. 4-1, the specific network architecture includes two hidden layers: the 1st hidden layer with 128 neurons and the 2nd hidden layer with 32 neurons. The output layer consists of a single neuron dedicated to regression prediction. The activation function used in both hidden layers is ReLU, while dropout layers are introduced between the 1st and 2nd hidden layers to mitigate overfitting. This architecture highlights the sequential nature of the DL process, incorporating dense blocks and supplementary modules to enhance feature processing and improve model generalization.

The model was trained using the Adam optimizer under the following circumstances: learning rate = 0.001, ε = 0.001, ρ_1= 0.9, ρ_2 = 0.999, and δ = 1E−8 [175] while using the MSE as the loss function. The ML networks were trained with TensorFlow (v2.0.0rc, Python 3.7) on an NVIDIA Quadro K4000 GPU, encompassing 100 epochs, with each batch having 128 samples. The model performance was evaluated as validation based on 10% of the training dataset.

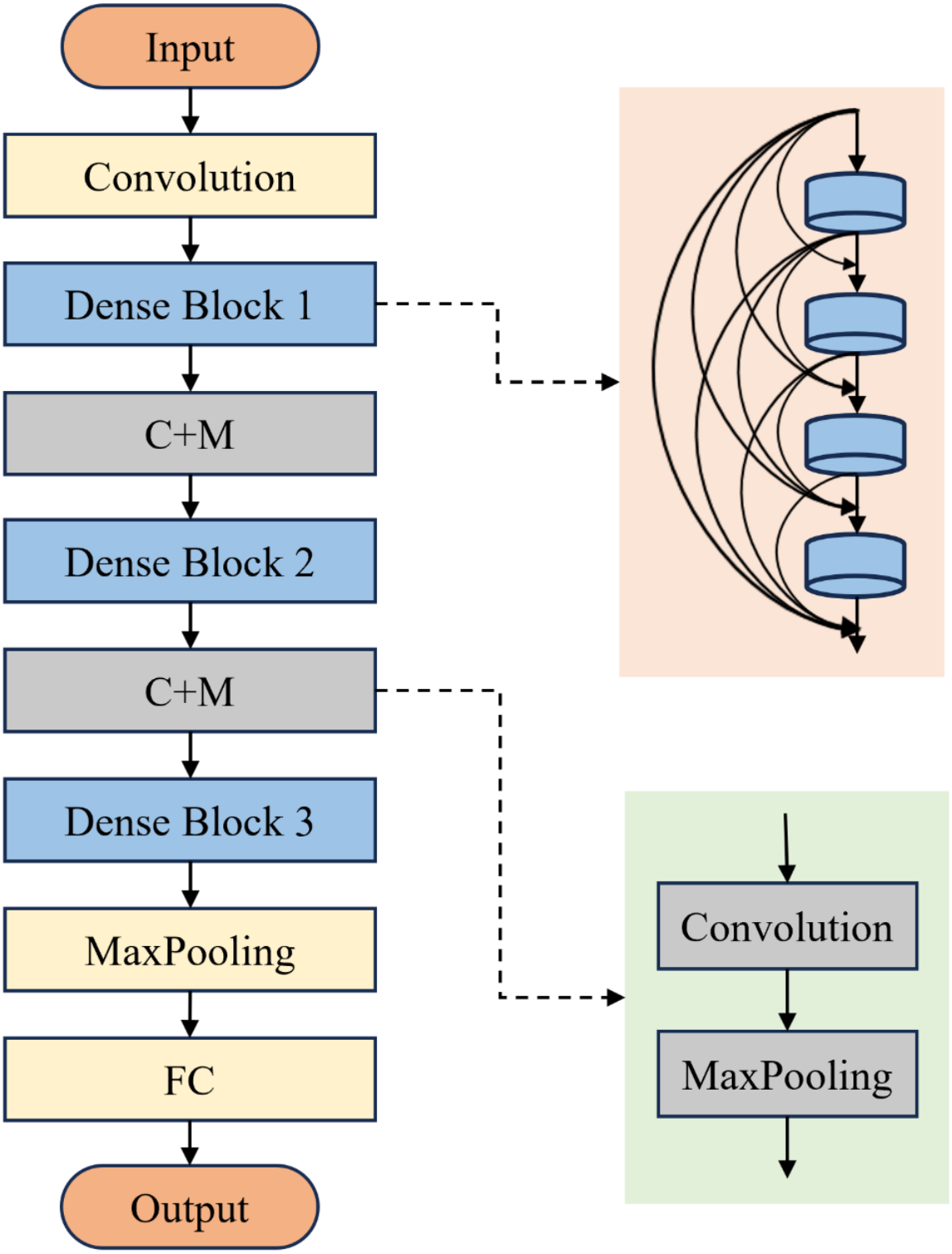


Fig. 4-1 Schematics of densely connected convolutional networks (Dense Nets) for regression modeling.

4.2.6 Performance evaluation

The confusion matrix was further visualized to assess the model performance regarding the training and testing datasets.

For the regression model, we used the metrics of accuracy and AUC-ROC to evaluate the performance. Given that the Mean absolute percentage error (MAPE) is defined as an error function [176,177]:

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}-y}{y}\right| \tag{12}$$

where y and ŷ denote the clinically observed values and the ML-predicted value, respectively; n is the quantity of the test dataset; thus, the accuracy is given as:

$$\mathrm{Accuracy}=1\text{-}\mathrm{MAPE} \tag{13}$$

Moreover, we conducted consistency analysis between clinical measurements and ML-based predictions using the Bland–Altman method. The Bland-Altman analysis is a statistical method used to evaluate the agreement between two different measurement techniques or instruments, particularly in medical research when comparing a new method to an established one. The analysis involves visualizing the differences between the two methods on the Y-axis versus the average of the two methods on the X-axis. The mean difference, or bias, indicates any systematic error between the methods, while the limits of agreement, determined as the mean difference ± 1.96 times the standard deviation of the discrepancies, represent the range within which 95% of the differences are expected to lie. If most data points fall within these limits, the two methods are considered to have good agreement. However, if the differences are consistently large or the limits are wide, the new method may not be suitable for use as a substitute for the reference method. Additionally, trends in the differences, such as increasing or decreasing variability with larger averages, may indicate systematic disagreement. Bland-Altman analysis is commonly used in clinical and biomedical studies to assess the reliability of new diagnostic tools or measurement devices and to determine whether they can be used interchangeably with existing methods. The visual representation provided by a Bland-Altman plot highlights any systematic bias or variability between the two methods, offering critical insights into their agreement. The Bland-Atman analysis reveals the trends, clustering patterns, and correlations of parameters between the clinical

measurement datasets and ML-based prediction datasets. When the parameters fall within an acceptable range, good agreement is obtained between the two datasets, and then both methods can be used interchangeably [178].

4.3 Results

A Dense Net was utilized to predict the LVDdI to optimize the ML network during the training process. The mean squared error (MSE), adopted as a loss function, displays a rapid and monotonic decline during the first several epochs (Fig. 4-2), converging quickly to a stable and minimum level after 100 training epochs. The relevant parameters and weight gained for the optimized model were then employed for the testing and ML-based predictions. As a result, the optimized regression model can achieve a prediction accuracy of 0.88.

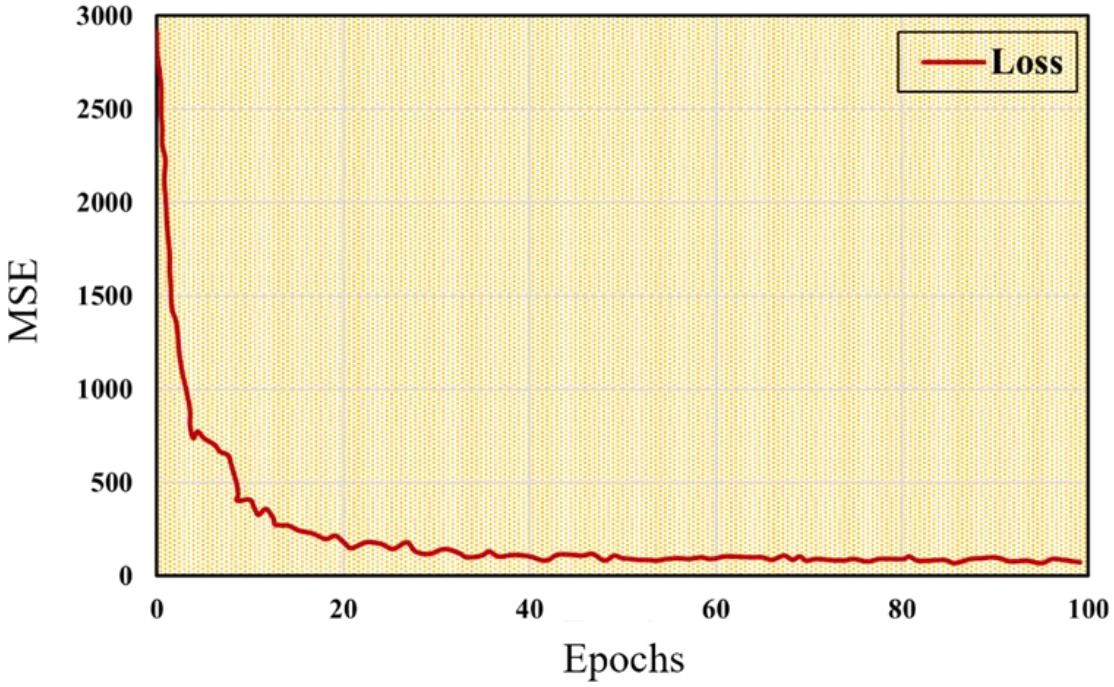


Fig. 4-2 Learning the MSE curve of the DenseNet model.

To evaluate the consistency between the Dense Net predictions and clinical measurements, we applied the Bland–Altman method to examine the mean values and differences, illustrated in Fig. 4-3, with an interval of 95% confidence. The Dense Net predictions are in good agreement with the clinical measurements, with most of the predicted LVDdI plots scattered within the 95%

confidence interval. We presented the confusion matrix of the regression model based on the predicted LVDdI for a clearer comparison with the classification model, as shown in Figure 4-4. In this model, 3 cases from the LVE group and 3 cases from the non-LVE group were misclassified, which is slightly higher than the classification model. However, the regression model still demonstrated a high overall prediction accuracy of 88%. In addition, as shown in Fig. 4-5, the ROC curve of the regression model, which estimates the LVDdI as a binary classification of the LVE and non-LVE groups, presents an AUC-ROC of 0.89, slightly lower than that of the classification model.

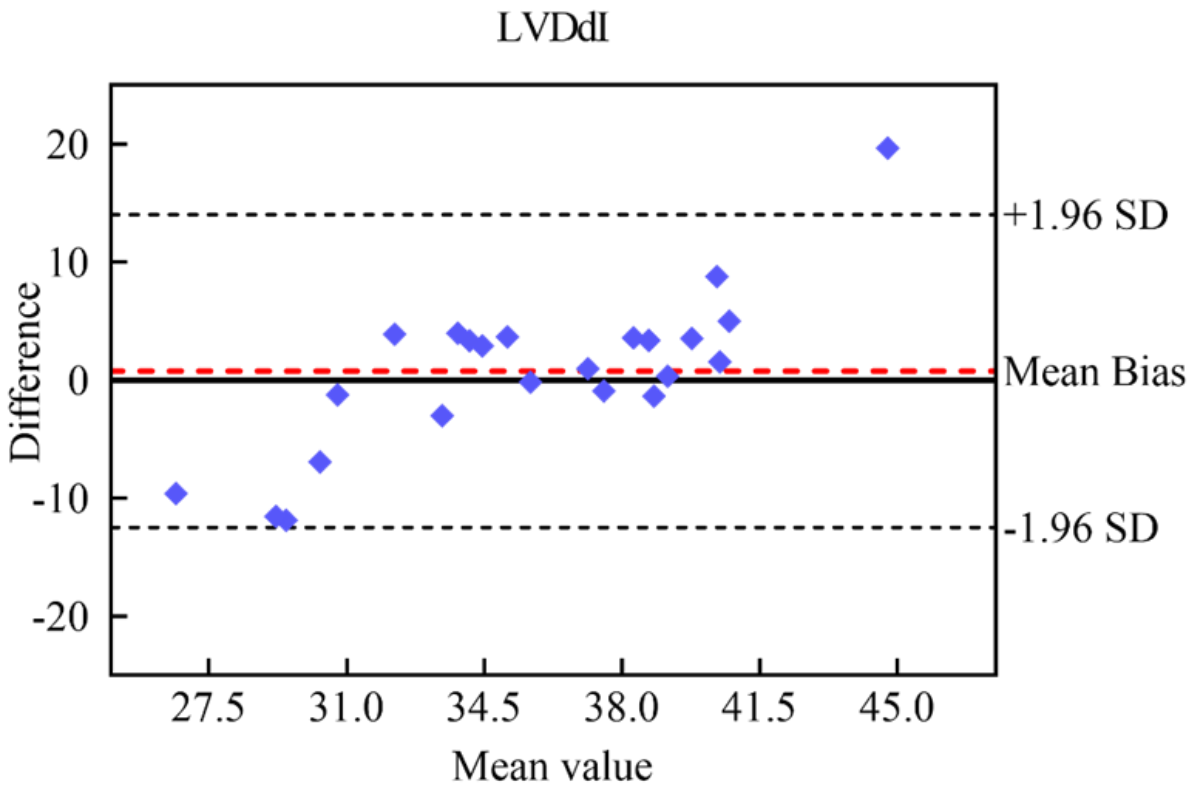


Fig. 4-3 Comparison between Dense Net predictions and clinical measurements based on Bland–Altman analyses for LVDdI.

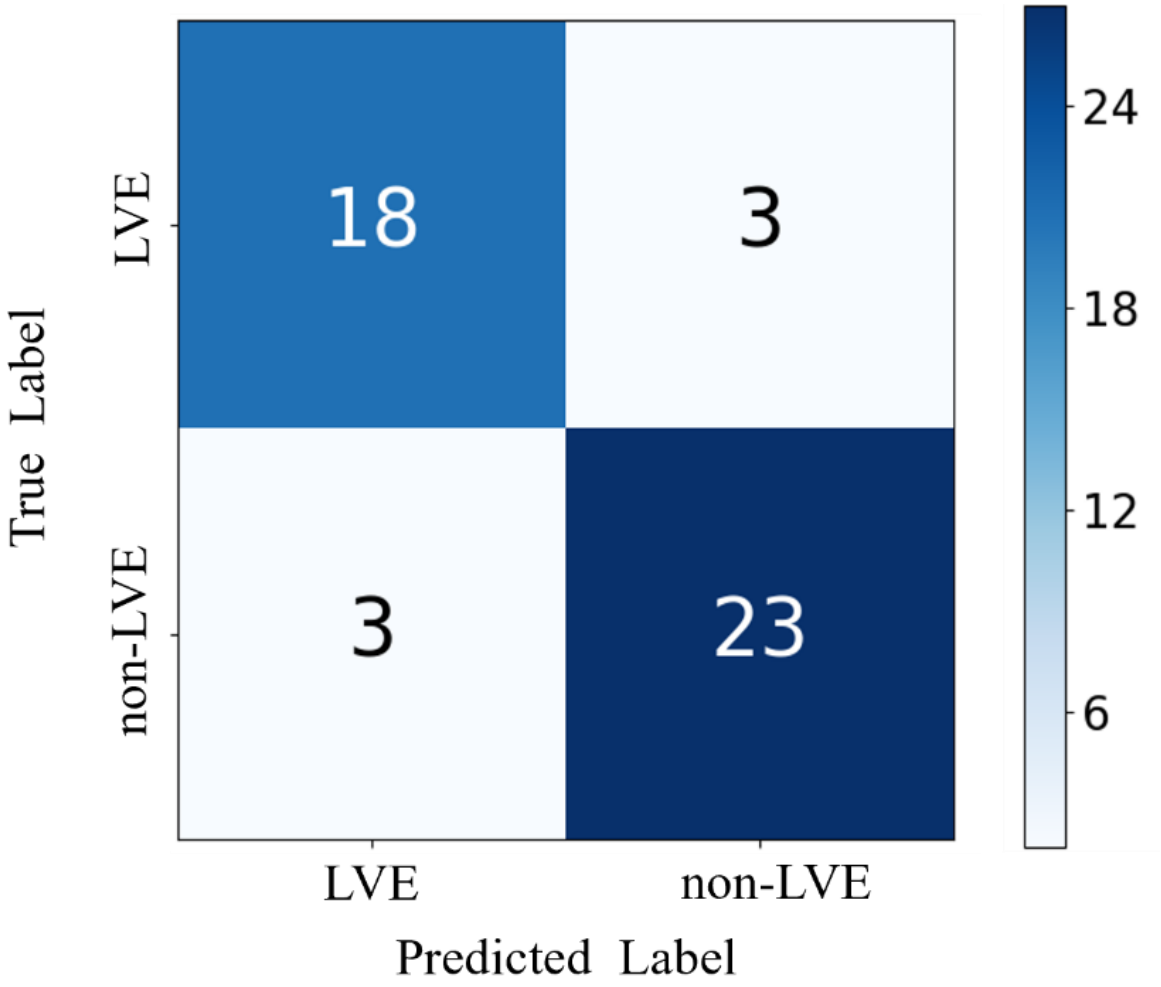


Fig. 4-4 Confusion matrices of the DenseNet model.

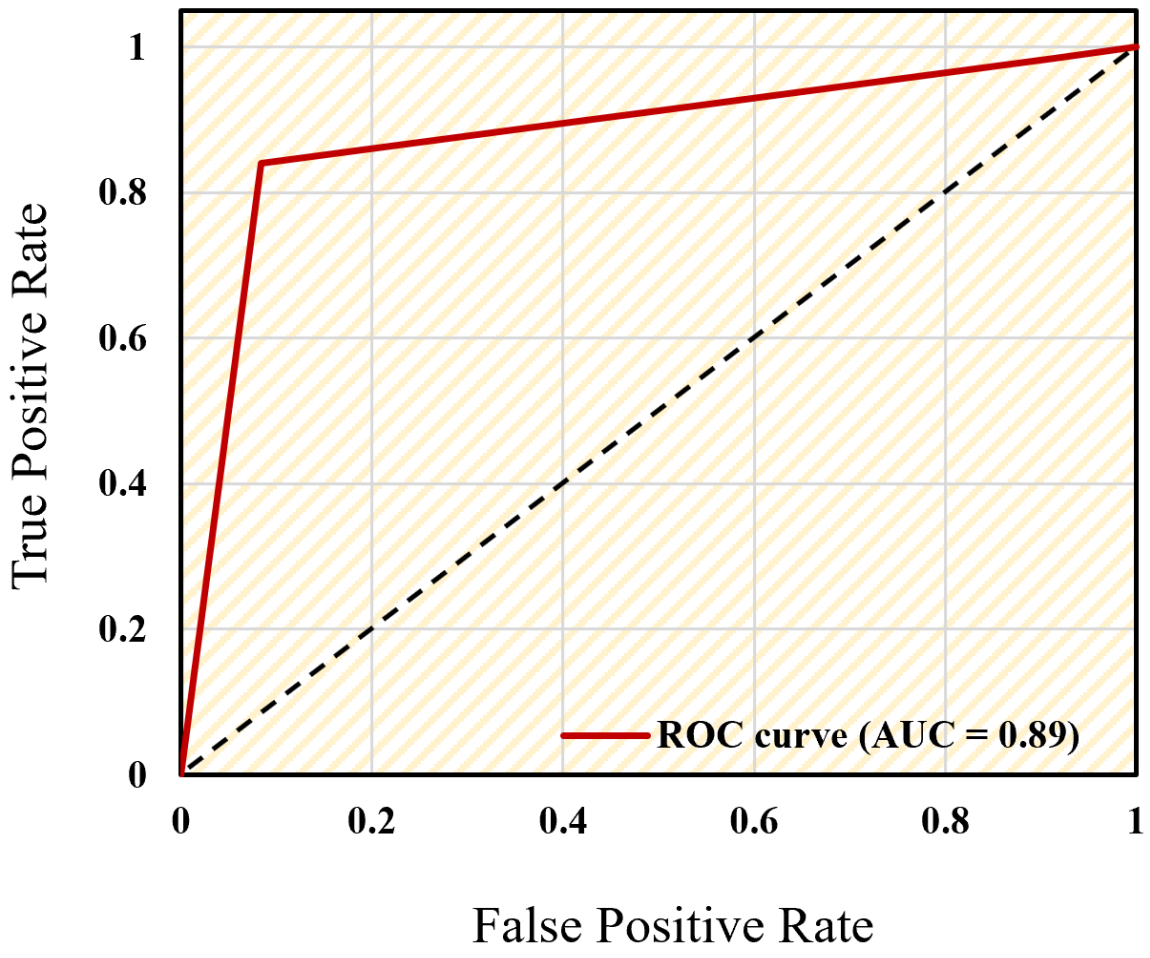


Fig. 4-5 ROC curve of regression model, AUC=0.89.

4.4 Discussion

LVE is one of the key precursors of HF; as the ventricular volume increases, the heart's pumping ability gradually decreases, contributing to the development and advancement of HF. In this study, we utilized pulse wave signals combined with the DenseNet network to quantitatively detect the LVDdI in HF patients, providing an accurate and non-invasive approach for clinical assessment of ventricular remodeling.

We collected a well-curated pulse wave dataset obtained from 264 HF patients and proposed a ML-based regression model that utilizes the DenseNet network to quantitatively predict the LVDdI value. Based on the standards of the American Society of Echocardiography, HF patients were then classified into LVE and non-LVE groups. To validate the model's accuracy, we compared

the classification labels from the predicted results with the clinical labels. The results showed that the predicted outcomes were highly consistent with the clinical labels, achieving a prediction accuracy of 93% with an AUC of 0.89. These findings indicate that the proposed regression strategy can effectively identify LVE and accurately predict LVDdI values, demonstrating strong feasibility and application potential. This method offers an effective, non-invasive, and precise diagnostic tool for detecting LVE in clinical settings and holds promise for future clinical diagnosis and personalized treatment.

In Chapter 3, the classification model we proposed for identifying LVE achieved an accuracy of 0.91 and an AUC of 0.93. In contrast, the regression model used in this chapter reached an accuracy of 0.88 and an AUC-ROC of 0.89. Although slightly lower than the classification model, it still demonstrates strong predictive performance. Unlike the classification approach in Chapter 3, this chapter emphasizes directly predicting the specific value of LVE using a regression model, rather than merely categorizing patients as LVE or non-LVE. This regression approach not only offers more precise quantitative predictions, allowing clinicians to more accurately assess the degree of LVE, but also provides valuable data support for personalized treatment and long-term health monitoring. This more detailed assessment enables clinicians to formulate more precise intervention strategies and follow-up plans in clinical practice, further improving treatment outcomes and prognosis for patients with HF.

Although we achieved accurate prediction of LVDdI in HF patients using pulse wave signals and the DenseNet model, there are still some limitations in this study. First, our dataset primarily focuses on HF patients, lacking data from healthy subjects, which necessitates further expansion of the dataset to include individuals with varying health conditions. Second, most of the HF patients had advanced disease; incorporating data from patients with mild HF could further validate the model's accuracy in detecting LVE at earlier stages. Additionally, despite our rigorous data selection process and the creation of a high-quality dataset, the experimental validation is based on preliminary results and requires further optimization. Moreover, since the data comes from a single institution, it may not fully represent HF patients from diverse backgrounds and varying demographics. Finally, pulse wave signals are not continuously monitored, so while we can accurately predict LVDdI values, real-time monitoring has not yet been achieved. To enable continuous monitoring of LVDdI, future research will require more comprehensive monitoring and

follow-up data, potentially integrating these into wearable devices. This would facilitate early detection of ventricular remodeling and improve prognostic management, meeting the demands of real-world clinical applications. Achieving these goals will be the focus of our future work.

4.5 Conclusion

We successfully proposed and implemented an ML-based method using pulse wave signals and the DenseNet network to predict the LVDdI in HF patients. Our model demonstrated strong predictive performance, providing a non-invasive, accurate, and efficient method for assessing left ventricular remodeling. The results highlight the potential of utilizing ML techniques to improve clinical decision-making, particularly in the early detection and management of HF. Future research will focus on expanding the dataset, optimizing the model, and integrating continuous monitoring capabilities into wearable devices to offer more comprehensive, personalized care.

Chapter V

Conclusions

5.1 Summary of contributions

This thesis introduces strategies based on ML and DL for the diagnosis and management of CVD. These ML and DL methods overcome the limitations of traditional approaches, offering the possibility of exploring unknown HF subgroups and demonstrating the feasibility of detecting specific cardiovascular conditions, such as LVE. Through these technologies, this work provides new tools for non-invasive and convenient CVD detection and management, further advancing personalized treatment.

In Chapter I, we introduce the risks, diagnosis, and treatment of CVDs, along with the classification and development of AI. We also review AI-based analytical methods and applications in CVD management.

In Chapter II, we collected and standardized the pulse wave time series and clinical data from 380 HF patients. The data was then clustered using the K-means++ algorithm, and the clustering performance was evaluated, with an in-depth analysis of the clinical differences between the groups. A RF classifier was applied to quantify the importance ratio of pulse wave time-frequency features in distinguishing between the clusters. Ultimately, two patient subgroups were identified, showing significant differences in age, heart rate, and ejection time. Additionally, rise time, mean value, fall time, and skewness were identified as key pulse wave morphological features. This finding provides preliminary evidence that pulse wave signals can be used for stratified management of HF patients, offering a non-invasive and convenient approach.

In Chapter III: After rigorous data screening, we constructed a dataset consisting of 264 HF patients and divided them into LVE and non-LVE groups based on clinical labels. Statistical differences between the two groups were verified using Fourier series analysis, ensuring the reliability of the dataset. We then proposed a weighted RF classification algorithm to differentiate between LVE and non-LVE patients, comparing the classification results with clinical labels. The results showed that the classification model achieved an accuracy of 0.91 and an AUC-ROC of 0.93, effectively identifying LVE in HF patients. This study highlights the potential of ML models based on pulse wave signals for non-invasive identification of LVE patients.

In Chapter IV: Building on the work from the previous chapter, we further collected the left ventricular end-diastolic diameter index (LVDdI) from 264 HF patients as a key indicator for

quantitatively predicting LVE. We proposed a regression algorithm based on a densely connected convolutional network to directly predict the specific values of LVDdI. The consistency between the predicted values and clinical measurements was assessed using the Bland-Altman method. The results demonstrated that the regression model achieved an accuracy of 0.88 and an AUC-ROC of 0.89, indicating good predictive performance. Unlike the classification algorithm in Chapter 3, this chapter highlights predicting the specific values of LVE through the regression model, rather than simply classifying patients into LVE or non-LVE groups. This approach provides more precise quantitative predictions, enabling clinicians to more accurately assess the extent of LVE, and offers valuable data support for personalized treatment and long-term monitoring.

5.2 Outlook:

In this thesis, AI methods, including ML and DL, have been demonstrated to play a critical role in the diagnosis and management of CVDs, contributing to health monitoring and management outcomes. Our future work will primarily focus on expanding the dataset and optimizing the network architecture.

In Chapter II, we plan to further refine key clinical indicators, such as BNP. Although BNP is already included in our dataset, its coverage is not yet comprehensive, which may affect the results of correlation analysis. Therefore, we intend to enhance this indicator by supplementing the dataset. Additionally, including both healthy individuals and HF patients in comparative testing will help improve the model's applicability. By improving the unsupervised learning model and expanding the range of clinical parameters, we aim not only to more accurately identify subgroups of HF patients but also to detect groups with potential HF risk, leading to more effective prevention and treatment strategies.

In Chapters III and IV, our primary goal is to expand the dataset by including a wider variety of HF patients. We will collaborate with hospitals to collect cardiac magnetic resonance data, which is considered the gold standard for detecting LVE, to enhance the reliability of the model’s training and validation. Additionally, in the future, we plan to collect longer pulse wave signals and conduct patient follow-ups to enable real-time detection and monitoring of LVE, as well as to predict patient recurrence rates and adverse events. These improvements will enhance the model’s

accuracy and clinical relevance, providing stronger support for the long-term management of HF patients.

Overall, AI-based strategies hold great potential for applications in clinical diagnosis, personalized treatment, and health monitoring. In clinical diagnosis, AI algorithms can rapidly process large volumes of pulse wave data, identifying subtle patterns and abnormalities, leading to more precise disease detection. For personalized treatment, AI-driven pulse wave analysis helps doctors develop tailored treatment plans based on the cardiovascular condition of patients, enabling precise drug adjustments and targeted interventions by identifying specific features, thus improving treatment effectiveness, reducing side effects, and effectively managing comorbidities. In health monitoring and long-term care, AI systems can be integrated into wearable devices, providing real-time monitoring of cardiovascular health, detecting changes in heart function early, preventing disease progression, and helping doctors adjust care plans based on actual data. This is particularly effective in managing HF, where it can significantly reduce complications and hospitalization rates.

However, the application of AI still faces challenges related to data quality and standardization, clinical integration, privacy and ethical issues. Ensuring data consistency and representativeness, seamless integration into clinical workflows, and protecting data privacy while complying with relevant regulations are essential for the widespread adoption of AI. Additionally, transparent and interpretable AI models are crucial for gaining the trust of clinicians and patients. Despite these challenges, there are still significant opportunities for innovation in the diagnosis and treatment of CVDs. In the future, combining multi-modal data (such as integrating pulse waves with ECG and CMR) to improve diagnostic accuracy, and fostering collaboration between AI developers and clinicians to advance real-time monitoring systems and early warning mechanisms, will pave new pathways for the prevention and treatment of CVDs.

Bibliography

[1] N.L. Bragazzi, W. Zhong, J. Shu, A. Abu Much, D. Lotan, A. Grupper, A. Younis, H. Dai, Burden of HF and underlying causes in 195 countries and territories from 1990 to 2017, Eur J Prev Cardiol 28 (2021) 1682–1690.

[2] M. Sokolski, K. Reszka, T. Suchocki, B. Adamik, A. Doroszko, J. Drobnik, J. Gorka-Dynysiewicz, M. Jedrzejczyk, K. Kaliszewski, K. Kilis-Pstrusinska, History of HF in patients hospitalized due to COVID-19: relevant factor of in-hospital complications and all-cause mortality up to six months, J Clin Med 11 (2022) 241.

[3] H. Dai, A.A. Much, E. Maor, E. Asher, A. Younis, Y. Xu, Y. Lu, X. Liu, J. Shu, N.L. Bragazzi, Global, regional, and national burden of ischaemic heart disease and its attributable risk factors, 1990–2017: results from the Global Burden of Disease Study 2017, European Heart Journal-Quality of Care and Clinical Outcomes 8 (2022) 50–60.

[4] J.B. Echouffo-Tcheugui, K.G. Bishu, G.C. Fonarow, L.E. Egede, Trends in health care expenditure among US adults with HF: The Medical Expenditure Panel Survey 2002-2011, Am Heart J 186 (2017) 63–72.

[5] J. Patel, HF population health considerations., American Journal of Managed Care 27 (2021).

[6] W.C. Lee, Y.E. Chavez, T. Baker, B.R. Luce, Economic burden of HF: a summary of recent literature, Heart & Lung 33 (2004) 362–371.

[7] V.L. Roger, Epidemiology of HF: a contemporary perspective, Circ Res 128 (2021) 1421–1434.

[8] L.A. Allen, M. Gheorghiade, K.J. Reid, S.M. Dunlay, P.S. Chan, P.J. Hauptman, F. Zannad, M.A. Konstam, J.A. Spertus, Identifying patients hospitalized with HF at risk for unfavorable future quality of life, Circ Cardiovasc Qual Outcomes 4 (2011) 389–398.

[9] N.S. Kakouros, S.N. Kakouros, Clinical assessment in acute HF, Hellenic J Cardiol 56 (2015) 285–301.

[10] G. Alcidi, G. Goffredo, M. Correale, N.D. Brunetti, M. Iacoviello, Brain natriuretic peptide biomarkers in current clinical and therapeutic scenarios of HF, J Clin Med 11 (2022) 3192.

[11] A. Palazzuoli, M. Gallotta, I. Quatrini, R. Nuti, Natriuretic peptides (BNP and NT-proBNP): measurement and relevance in HF, Vasc Health Risk Manag (2010) 411–418.

[12] N. Mallegni, G. Molinari, C. Ricci, A. Lazzeri, D. La Rosa, A. Crivello, M. Milazzo, Sensing devices for detecting and processing acoustic signals in healthcare, Biosensors (Basel) 12 (2022) 835.

[13] P.K. Jain, A.K. Tiwari, Heart monitoring systems—A review, Comput Biol Med 54 (2014) 1–13.

[14] P.M. Nabeel, V.R. Kiran, J. Joseph, V. V Abhidev, M. Sivaprakasam, Local pulse wave velocity: theory, methods, advancements, and clinical applications, IEEE Rev Biomed Eng 13 (2019) 74–112.

[15] P. Salvi, Pulse waves, How Vascular Hemodynamics Affects Blood Pressure (2012).

[16] P.M. Nabeel, V.R. Kiran, J. Joseph, V. V Abhidev, M. Sivaprakasam, Local pulse wave velocity: theory, methods, advancements, and clinical applications, IEEE Rev Biomed Eng 13 (2019) 74–112.

[17] J. Jin, T. Qi, M. Wang, X. Yu, H. Sun, X. Jin, P. Ju, Characteristics of pulse wave velocity and its correlation with cardiovascular disease risk factors in healthy individuals, Polish Heart Journal (Kardiologia Polska) (2024).

[18] C.A. Valencia-Hernández, J. V Lindbohm, M.J. Shipley, I.B. Wilkinson, C.M. McEniery, S. Ahmadi-Abhari, A. Singh-Manoux, M. Kivimäki, E.J. Brunner, Aortic pulse wave velocity as adjunct risk marker for assessing cardiovascular disease risk: prospective study, Hypertension 79 (2022) 836–843.

[19] M.H. Sherebrin, R.Z. Sherebrin, Frequency analysis of the peripheral pulse wave detected in the finger with a photoplethysmograph, IEEE Trans Biomed Eng 37 (1990) 313–317. https://doi.org/10.1109/10.52332.

[20] T. Tigges, Z. Music, A. Pielmus, M. Klum, A. Feldheiser, O. Hunsicker, R. Orglmeister, Classification of morphologic changes in photoplethysmographic waveforms, 2 (2016) 203–207. https://doi.org/doi:10.1515/cdbme-2016-0046.

[21] L. Wang, L. Xu, S. Feng, M.Q.-H. Meng, K. Wang, Multi-Gaussian fitting for pulse waveform using Weighted Least Squares and multi-criteria decision making method, Comput Biol Med 43 (2013) 1661–1672. https://doi.org/https://doi.org/10.1016/j.compbiomed.2013.08.004.

[22] J. Kockskaemper, F. Pluteanu, Left atrial myocardium in arterial hypertension, Cells 11 (2022) 3157.

[23] F. Liang, Y. Wang, Coronary heart disease and atrial fibrillation: a vicious cycle, American Journal of Physiology-Heart and Circulatory Physiology 320 (2021) H1–H12.

[24] E.-S.H. Ibrahim, L. Frank, D. Baruah, V.E. Arpinar, A.S. Nencka, K.M. Koch, L.T. Muftuler, O. Unal, J. Stojanovska, J.C. Rubenstein, Value CMR: Towards a Comprehensive, Rapid, Cost-Effective Cardiovascular Magnetic Resonance Imaging, Int J Biomed Imaging 2021 (2021) 8851958.

[25] A. Barison, F. Ricci, A.G. Pavon, G. Muscogiuri, G. Bisaccia, G. Camastra, M. De Lazzari, C. Lanzillo, M. Raguso, L. Monti, Cardiovascular magnetic resonance in patients with cardiac electronic devices: evidence from a multicenter study, J Clin Med 12 (2023) 6673.

[26] X. Wang, H. Zhu, Artificial Intelligence in Image-based Cardiovascular Disease Analysis: A Comprehensive Survey and Future Outlook, ArXiv Preprint ArXiv:2402.03394 (2024).

[27] F. Sana, E.M. Isselbacher, J.P. Singh, E.K. Heist, B. Pathik, A.A. Armoundas, Wearable devices for ambulatory cardiac monitoring: JACC state-of-the-art review, J Am Coll Cardiol 75 (2020) 1582–1592.

[28] J.-D. Huang, J. Wang, E. Ramsey, G. Leavey, T.J.A. Chico, J. Condell, Applying artificial intelligence to wearable sensor data to diagnose and predict cardiovascular disease: a review, Sensors 22 (2022) 8002.

[29] J.B. Park, J.E. Sharman, Y. Li, M. Munakata, K. Shirai, C.-H. Chen, S.Y. Jae, H. Tomiyama, H. Kosuge, R.M. Bruno, Expert consensus on the clinical use of pulse wave velocity in Asia, Pulse 10 (2022) 1–18.

[30] N.H. Tougaard, S. Theilade, S.A. Winther, N. Tofte, T.S. Ahluwalia, T.W. Hansen, P. Rossing, M. Frimodt-Møller, Carotid-femoral pulse wave velocity as a risk marker for development of complications in type 1 diabetes mellitus, J Am Heart Assoc 9 (2020) e017165.

[31] T. Koivistoinen, L.-P. Lyytikäinen, H. Aatola, T. Luukkaala, M. Juonala, J. Viikari, T. Lehtimäki, O.T. Raitakari, M. Kähönen, N. Hutri-Kähönen, Pulse wave velocity predicts the progression of blood pressure and development of hypertension in young adults, Hypertension 71 (2018) 451–456.

[32] L. Chou, K. Zhang, J. Liu, S. Gong, Y. Chou, Life-threatening arrhythmias recognition by pulse-to-pulse intervals analysis, International Journal of Computer Applications in Technology 67 (2021) 185–193.

[33] I.H. Sarker, AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems, SN Comput Sci 3 (2022) 158.

[34] N. Kriegeskorte, P.K. Douglas, Cognitive computational neuroscience, Nat Neurosci 21 (2018) 1148–1160.

[35] Y. Lu, Artificial intelligence: a survey on evolution, models, applications and future trends, Journal of Management Analytics 6 (2019) 1–29.

[36] I. El Naqa, M.J. Murphy, What is machine learning?, Springer, 2015.

[37] E.F. Morales, H.J. Escalante, A brief introduction to supervised, unsupervised, and reinforcement learning, in: Biosignal Processing and Classification Using Computational Learning and Intelligence, Elsevier, 2022: pp. 111–129.

[38] U.S. Shanthamallu, A. Spanias, Machine and deep learning algorithms and applications, Morgan & Claypool Publishers, 2021.

[39] M. Naeem, S.T.H. Rizvi, A. Coronato, A gentle introduction to reinforcement learning and its application in different fields, IEEE Access 8 (2020) 209320–209344.

[40] A. Shrestha, A. Mahmood, Review of deep learning algorithms and architectures, IEEE Access 7 (2019) 53040–53065.

[41] C. Han, L. Shi, ML–ResNet: A novel network to detect and locate myocardial infarction using 12 leads ECG, Comput Methods Programs Biomed 185 (2020) 105138.

[42] S. Mousavi, F. Afghah, U.R. Acharya, HAN-ECG: An interpretable atrial fibrillation detection model using hierarchical attention networks, Comput Biol Med 127 (2020) 104057.

[43] T. Wang, W. Jin, F. Liang, J. Alastruey, Machine learning-based pulse wave analysis for early detection of abdominal aortic aneurysms using in silico pulse waves, Symmetry (Basel) 13 (2021) 804.

[44] G. Li, K. Watanabe, H. Anzai, X. Song, A. Qiao, M. Ohta, Pulse-wave-pattern classification with a convolutional neural network, Sci Rep 9 (2019) 14930.

[45] G. James, D. Witten, T. Hastie, R. Tibshirani, J. Taylor, Unsupervised learning, in: An Introduction to Statistical Learning: With Applications in Python, Springer, 2023: pp. 503–556.

[46] M. Usama, J. Qadir, A. Raza, H. Arif, K.-L.A. Yau, Y. Elkhatib, A. Hussain, A. Al-Fuqaha, Unsupervised machine learning for networking: Techniques, applications and research challenges, IEEE Access 7 (2019) 65579–65615.

[47] A.M. Flores, A. Schuler, A.V. Eberhard, J.W. Olin, J.P. Cooke, N.J. Leeper, N.H. Shah, E.G. Ross, Unsupervised learning for automated detection of coronary artery disease subgroups, J Am Heart Assoc 10 (2021) e021976.

[48] G.M. Karageorgos, P. Liang, N. Mobadersany, P. Gami, E.E. Konofagou, Unsupervised deep learning-based displacement estimation for vascular elasticity imaging applications, Phys Med Biol 68 (2023) 155014.

[49] M. Rousseau-Portalis, L. Cymberknop, I. Farro, R. Armentano, Computational clustering reveals differentiated coronary artery calcium progression at prevalent levels of pulse wave velocity by classifying high-risk patients, Front Cardiovasc Med 10 (2023) 1161914.

[50] A. Groenewegen, F.H. Rutten, A. Mosterd, A.W. Hoes, Epidemiology of HF, Eur J Heart Fail 22 (2020) 1342–1356.

[51] V.L. Roger, Epidemiology of HF: a contemporary perspective, Circ Res 128 (2021) 1421–1434.

[52] J. Feng, Y. Zhang, J. Zhang, Epidemiology and burden of HF in Asia, JACC: Asia 4 (2024) 249–264.

[53] G. Savarese, P.M. Becher, L.H. Lund, P. Seferovic, G.M.C. Rosano, A.J.S. Coats, Global burden of HF: a comprehensive and updated review of epidemiology, Cardiovasc Res 118 (2022) 3272–3287.

[54] T. Zdrojewski, An Update on the Prevalence of HF, in: Hypertension and HF: Epidemiology, Mechanisms and Treatment, Springer, 2024: pp. 13–26.

[55] Y. Kurylyak, F. Lamonaca, D. Grimaldi, A Neural Network-based method for continuous blood pressure estimation from a PPG signal, in: 2013 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), IEEE, 2013: pp. 280–283.

[56] R. Li, K. Sughimoto, X. Zhang, S. Wang, H. Liu, Impacts of respiratory fluctuations on cerebral circulation: a machine-learning-integrated 0–1D multiscale hemodynamic model, Physiol Meas 44 (2023) 035013.

[57] R. Li, K. Sughimoto, X. Zhang, S. Wang, Y. Hiraki, H. Liu, Impact of Respiratory Fluctuation on Hemodynamics in Human Cardiovascular System: A 0-1D Multiscale Model, Fluids 7 (2022) 28.

[58] S. Zanelli, M. Ammi, M. Hallab, M.A. El Yacoubi, Diabetes detection and management through photoplethysmographic and electrocardiographic signals analysis: A systematic review, Sensors 22 (2022) 4890.

[59] N.S. Alghamdi, M. Zakariah, A. Shankar, W. Viriyasitavat, Heart disease prediction using autoencoder and DenseNet architecture, Egyptian Informatics Journal 28 (2024) 100559.

[60] P.H. Charlton, B. Paliakaitė, K. Pilt, M. Bachler, S. Zanelli, D. Kulin, J. Allen, M. Hallab, E. Bianchini, C.C. Mayer, Assessing hemodynamics from the photoplethysmogram to gain insights into vascular age: a review from VascAgeNet, American Journal of Physiology-Heart and Circulatory Physiology 322 (2022) H493–H522.

[61] C.M. McEniery, null Yasmin, I.R. Hall, A. Qasem, I.B. Wilkinson, J.R. Cockcroft, A. Investigators, Normal vascular aging: differential effects on wave reflection and aortic pulse wave velocity: the Anglo-Cardiff Collaborative Trial (ACCT), J Am Coll Cardiol 46 (2005) 1753–1760.

[62] L.A. Bortolotto, J. Blacher, T. Kondo, K. Takazawa, M.E. Safar, Assessment of vascular aging and atherosclerosis in hypertensive subjects: second derivative of photoplethysmogram versus pulse wave velocity, Am J Hypertens 13 (2000) 165–171.

[63] S. Laurent, L. Marais, P. Boutouyrie, The noninvasive assessment of vascular aging, Canadian Journal of Cardiology 32 (2016) 669–679.

[64] F. Qureshi, S. Krishnan, Wearable hardware design for the internet of medical things (IoMT), Sensors 18 (2018) 3812.

[65] T. Tigges, Z. Music, A. Pielmus, M. Klum, A. Feldheiser, O. Hunsicker, R. Orglmeister, Classification of morphologic changes in photoplethysmographic waveforms, Current Directions in Biomedical Engineering 2 (2016) 203–207.

[66] T.R. Dawber, H.E. THomas Jr, P.M. McNamara, Characteristics of the dicrotic notch of the arterial pulse wave in coronary heart disease, Angiology 24 (1973) 244–255.

[67] S. Wang, D. Wu, G. Li, X. Song, A. Qiao, R. Li, Y. Liu, H. Anzai, H. Liu, A machine learning strategy for fast prediction of cardiac function based on peripheral pulse wave, Comput Methods Programs Biomed 216 (2022) 106664.

[68] S. Wang, R. Ono, D. Wu, K. Aoki, H. Kato, T. Iwahana, S. Okada, Y. Kobayashi, H. Liu, Pulse wave-based evaluation of the blood-supply capability of patients with HF via machine learning, Biomed Eng Online 23 (2024) 7.

[69] D. Wu, R. Ono, S. Wang, Y. Kobayashi, K. Sughimoto, H. Liu, Pulse wave signal-driven machine learning for identifying left ventricular enlargement in HF patients, Biomed Eng Online 23 (2024) 60.

[70] J. Abonyi, B. Feil, Cluster analysis for data mining and system identification, Springer Science & Business Media, 2007.

[71] S.T. Wierzchoń, M.A. Kłopotek, Modern algorithms of cluster analysis, Springer, 2018.

[72] S. Zanelli, K. Eveilleau, P.H. Charlton, M. Ammi, M. Hallab, M.A. El Yacoubi, Clustered photoplethysmogram pulse wave shapes and their associations with clinical data, Front Physiol 14 (2023) 1176753.

[73] B.Z. Garmaev, V. V Boronoev, I. V Naguslaeva, V.D. Ompokov, Classification of pulse waves based on cluster analysis of time parameters, in: J Phys Conf Ser, IOP Publishing, 2019: p. 012048.

[74] P.A. McKee, W.P. Castelli, P.M. McNamara, W.B. Kannel, The natural history of congestive HF: the Framingham study, New England Journal of Medicine 285 (1971) 1441–1446.

[75] X. Zhang, Y. Shang, D. Guo, T. Zhao, Q. Li, X. Wang, A more effective method of extracting the characteristic value of pulse wave signal based on wavelet transform, J Biomed Sci Eng 9 (2016) 9–19.

[76] H.-T. Wu, C.-H. Lee, C.-E. Chen, A.-B. Liu, Predicting arterial stiffness with the aid of ensemble empirical mode decomposition (EEMD) algorithm, in: 2010 IEEE International Conference on Wireless Communications, Networking and Information Security, IEEE, 2010: pp. 179–182.

[77] F. Chang, W. Hong, T. Zhang, J. Jing, X. Liu, Research on wavelet denoising for pulse signal based on improved wavelet thresholding, in: 2010 First International Conference on Pervasive Computing, Signal Processing and Applications, IEEE, 2010: pp. 564–567.

[78] L. Xu, M.Q.-H. Meng, K. Wang, W. Lu, N. Li, Pulse images recognition using fuzzy neural network, Expert Syst Appl 36 (2009) 3805–3811.

[79] X. Hu, H. Zhu, J. Xu, D. Xu, J. Dong, Wrist pulse signals analysis based on deep convolutional neural networks, in: 2014 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, IEEE, 2014: pp. 1–7.

[80] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE., Journal of Machine Learning Research 9 (2008).

[81] S. Arora, W. Hu, P.K. Kothari, An analysis of the t-sne algorithm for data visualization, in: Conference on Learning Theory, PMLR, 2018: pp. 1455–1462.

[82] G.C. Linderman, S. Steinerberger, Clustering with t-SNE, provably, SIAM J Math Data Sci 1 (2019) 313–332.

[83] T.T. Cai, R. Ma, Theoretical foundations of t-sne for visualizing high-dimensional clustered data, Journal of Machine Learning Research 23 (2022) 1–54.

[84] W. Li, J.E. Cerise, Y. Yang, H. Han, Application of t-SNE to human genetic data, J Bioinform Comput Biol 15 (2017) 1750017.

[85] D. Arthur, S. Vassilvitskii, k-means++: The advantages of careful seeding, Stanford, 2006.

[86] S. Daoudi, C.M. Anouar Zouaoui, M.C. El-Mezouar, N. Taleb, Parallelization of the K-Means++ Clustering Algorithm., Ingénierie Des Systèmes d'Information 26 (2021).

[87] D. Arthur, S. Vassilvitskii, k-means++: The advantages of careful seeding, Stanford, 2006.

[88] P.J. Rousseeuw, Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, J Comput Appl Math 20 (1987) 53–65.

[89] X. Wang, Y. Xu, An improved index for clustering validation based on Silhouette index and Calinski-Harabasz index, in: IOP Conf Ser Mater Sci Eng, IOP Publishing, 2019: p. 052024.

[90] F. Ros, R. Riad, S. Guillaume, PDBI: A partitioning Davies-Bouldin index for clustering evaluation, Neurocomputing 528 (2023) 178–199.

[91] S. Kwak, Are Only p-values less than 0.05 significant? A p-value greater than 0.05 is also significant!, J Lipid Atheroscler 12 (2023) 89.

[92] N.M. Neumann, A. Plastino, J.A.P. Junior, A.A. Freitas, Is p-value 0.05 enough? A study on the statistical evaluation of classifiers, Knowl Eng Rev 36 (2020) e1.

[93] H. Liu, D. Fujita, L. Zhang, A. Suzuki, Real-time pulse waveform profiling algorithm for wearable applications, IEEE Access 6 (2018) 59296–59306.

[94] M.E. Mollerus, C. Renier, M. Lipinski, Spectral methods to distinguish ventricular fibrillation from artefact in implantable cardioverter-defibrillators, Europace 13 (2011) 1346–1351.

[95] B. Boashash, Time-frequency signal analysis and processing: a comprehensive reference, Academic press, 2015.

[96] D. Dobre, J.S. Borer, K. Fox, K. Swedberg, K.F. Adams, J.G.F. Cleland, A. Cohen-Solal, M. Gheorghiade, F. Gueyffier, C.M. O'Connor, Heart rate: a prognostic factor and therapeutic target in chronic HF. The distinct roles of drugs with heart rate-lowering properties, Eur J Heart Fail 16 (2014) 76–85.

[97] S. Zeid, G. Buch, D. Velmeden, J. Söhne, A. Schulz, A. Schuch, S.-O. Tröbs, M.W. Heidorn, F. Müller, K. Strauch, Heart rate variability: Reference values and role for clinical profile and mortality in individuals with HF, Clinical Research in Cardiology (2023) 1–14.

[98] Q. Yousef, M.B.I. Reaz, M.A.M. Ali, The analysis of PPG morphology: investigating the effects of aging on arterial compliance, Measurement Science Review 12 (2012) 266–271.

[99] S.R. Alty, N. Angarita-Jaimes, S.C. Millasseau, P.J. Chowienczyk, Predicting arterial stiffness from the digital volume pulse waveform, IEEE Trans Biomed Eng 54 (2007) 2268–2275.

[100] D.G. Brillante, A.J. O'sullivan, L.G. Howes, Arterial stiffness indices in healthy volunteers using non-invasive digital photoplethysmography, Blood Press 17 (2008) 116–123.

[101] R.S. Gubner, M. Rodstein, H.E. Ungerleider, Ballistocardiography: An appraisal of technic, physiologic principles, and clinical value, Circulation 7 (1953) 268–286.

[102] M.E. Lough, S.J. Berger, A. Larsen, C.P. Sandoval, Cardiovascular Clinical Assessment and Diagnostic Procedures, Priorities in Critical Care Nursing-E-Book (2022) 105.

[103] G. Savarese, P.M. Becher, L.H. Lund, P. Seferovic, G.M.C. Rosano, A.J.S. Coats, Global burden of HF: a comprehensive and updated review of epidemiology, Cardiovasc Res 118 (2022) 3272–3287.

[104] K.A.A. Clark, S.W. Reinhardt, F. Chouairi, P.E. Miller, B. Kay, M. Fuery, A. Guha, T. Ahmad, N.R. Desai, Trends in HF hospitalizations in the US from 2008 to 2018, J Card Fail 28 (2022) 171–180.

[105] S. Wei, N. Le, J.W. Zhu, K. Breathett, S.J. Greene, M.A. Mamas, F. Zannad, H.G.C. Van Spall, Factors associated with racial and ethnic diversity among HF trial participants: a systematic bibliometric review, Circ Heart Fail 15 (2022) e008685.

[106] J.N. Cohn, R. Ferrari, N. Sharpe, Cardiac remodeling—concepts and clinical implications: a consensus paper from an international forum on cardiac remodeling, J Am Coll Cardiol 35 (2000) 569–582. https://doi.org/https://doi.org/10.1016/S0735-1097(99)00630-0.

[107] M.R. Zile, W.H. Gaasch, K. Patel, I.B. Aban, A. Ahmed, Adverse Left Ventricular Remodeling in Community-Dwelling Older Adults Predicts Incident HF and Mortality, JACC Heart Fail 2 (2014) 512–522. https://doi.org/https://doi.org/10.1016/j.jchf.2014.03.016.

[108] L.H. Opie, P.J. Commerford, B.J. Gersh, M.A. Pfeffer, Controversies in ventricular remodelling, The Lancet 367 (2006) 356–367.

[109] S. Cheng, V.R.S. Fernandes, D.A. Bluemke, R.L. McClelland, R.A. Kronmal, J.A.C. Lima, Age-related left ventricular remodeling and associated risk for cardiovascular outcomes: the Multi-Ethnic Study of Atherosclerosis, Circ Cardiovasc Imaging 2 (2009) 191–198.

[110] A. D'Andrea, A. Carbone, E. Agricola, L. Riegler, S. Sperlongano, G. Tocci, R. Scarafile, T. Formisano, C. Capogrosso, M.C. Bigazzi, Predictive value of left ventricular myocardial deformation for left ventricular remodeling in patients with classical low-flow, low-gradient aortic stenosis undergoing transcatheter aortic valve replacement, Journal of the American Society of Echocardiography 32 (2019) 730–736.

[111] G.L. Brower, J.D. Gardner, M.F. Forman, D.B. Murray, T. Voloshenyuk, S.P. Levick, J.S. Janicki, The relationship between myocardial extracellular matrix remodeling and ventricular function, European Journal of Cardio-Thoracic Surgery 30 (2006) 604–610.

[112] M.H. Drazner, The progression of hypertensive heart disease, Circulation 123 (2011) 327–334.

[113] L. Wu, L. Zhang, Z. Ai, L. Zou, Y. Zhu, Y. Bao, J. Li, S. Kang, H. Fan, D. Zhang, Association between risk factors and left ventricular remodeling in middle-aged and aged population: a community-based study, J Hypertens 30 (2012) 1862–1873.

[114] A.H. Gradman, F. Alfayoumi, From left ventricular hypertrophy to congestive HF: management of hypertensive heart disease, Prog Cardiovasc Dis 48 (2006) 326–341.

[115] S. Frantz, M.J. Hundertmark, J. Schulz-Menger, F.M. Bengel, J. Bauersachs, Left ventricular remodelling post-myocardial infarction: pathophysiology, imaging, and novel therapies, Eur Heart J 43 (2022) 2549–2561.

[116] R.S. Vasan, M.G. Larson, E.J. Benjamin, J.C. Evans, D. Levy, Left Ventricular Dilatation and the Risk of Congestive HF in People without Myocardial Infarction, New England Journal of Medicine 336 (1997) 1350–1355. https://doi.org/10.1056/NEJM199705083361903.

[117] P. Gaudron, C. Eilles, I. Kugler, G. Ertl, Progressive left ventricular dysfunction and remodeling after myocardial infarction. Potential mechanisms and early predictors., Circulation 87 (1993) 755–763.

[118] J. Yeboah, D.A. Bluemke, W.G. Hundley, C.J. Rodriguez, J.A.C. Lima, D.M. Herrington, Left ventricular dilation and incident congestive HF in asymptomatic adults without cardiovascular disease: multi-ethnic study of atherosclerosis (MESA), J Card Fail 20 (2014) 905–911.

[119] D. Levy, K.M. Anderson, D.D. Savage, S.A. Balkus, W.B. Kannel, W.P. Castelli, Risk of ventricular arrhythmias in left ventricular hypertrophy: The Framingham Heart Study, Am J Cardiol 60 (1987) 560–565. https://doi.org/https://doi.org/10.1016/0002-9149(87)90305-5.

[120] S. Yamanaka, Y. Sakata, K. Nochioka, M. Miura, S. Kasahara, M. Sato, H. Aoyanagi, T. Fujihashi, H. Hayashi, T. Shiroto, Prognostic impacts of dynamic cardiac structural changes in HF patients with preserved left ventricular ejection fraction, Eur J Heart Fail 22 (2020) 2258–2268.

[121] M. Salerno, C.M. Kramer, Advances in parametric mapping with CMR imaging, JACC Cardiovasc Imaging 6 (2013) 806–822.

[122] D.J. Pennell, Ventricular volume and mass by CMR, Journal of Cardiovascular Magnetic Resonance 4 (2002) 507–513.

[123] G. Vincenti, P. Monney, J. Chaptinel, T. Rutz, S. Coppo, M.O. Zenge, M. Schmidt, M.S. Nadar, D. Piccini, P. Chèvre, Compressed sensing single–breath-hold CMR for fast quantification of LV function, volumes, and mass, JACC Cardiovasc Imaging 7 (2014) 882–892.

[124] E.-S.H. Ibrahim, L. Frank, D. Baruah, V.E. Arpinar, A.S. Nencka, K.M. Koch, L.T. Muftuler, O. Unal, J. Stojanovska, J.C. Rubenstein, Value CMR: towards a comprehensive, rapid, cost-effective cardiovascular magnetic resonance imaging, Int J Biomed Imaging 2021 (2021) 1–12.

[125] C.M. Kramer, Potential for rapid and cost-effective cardiac magnetic resonance in the developing (and Developed) World, J Am Heart Assoc 7 (2018) e010435.

[126] L. Turvey, D.X. Augustine, S. Robinson, D. Oxborough, M. Stout, N. Smith, A. Harkness, L. Williams, R.P. Steeds, W. Bradlow, Transthoracic echocardiography of hypertrophic cardiomyopathy in adults: a practical guideline from the British Society of Echocardiography, Echo Res Pract 8 (2021) G61–G86.

[127] M.B. Jensen, E. Sloth, K.M. Larsen, M.B. Schmidt, Transthoracic echocardiography for cardiopulmonary monitoring in intensive care, Eur J Anaesthesiol 21 (2004) 700–707.

[128] S. Wang, D. Wu, G. Li, X. Song, A. Qiao, R. Li, Y. Liu, H. Anzai, H. Liu, A machine learning strategy for fast prediction of cardiac function based on peripheral pulse wave, Comput Methods Programs Biomed 216 (2022) 106664.

[129] K. Guk, G. Han, J. Lim, K. Jeong, T. Kang, E.-K. Lim, J. Jung, Evolution of wearable devices with real-time disease monitoring for personalized healthcare, Nanomaterials 9 (2019) 813.

[130] K. Sughimoto, K. Okauchi, D. Zannino, C.P. Brizard, F. Liang, M. Sugawara, H. Liu, K. Tsubota, Total cavopulmonary connection is superior to atriopulmonary connection Fontan in preventing thrombus formation: Computer simulation of flow-related blood coagulation, Pediatr Cardiol 36 (2015) 1436–1441.

[131] K. Sughimoto, J. Levman, F. Baig, D. Berger, Y. Oshima, H. Kurosawa, K. Aoki, Y. Seino, T. Ueda, H. Liu, Machine learning predicts blood lactate levels in children after cardiac surgery in paediatric ICU, Cardiol Young 33 (2023) 388–395.

[132] X. Song, Y. Liu, S. Wang, H. Zhang, A. Qiao, X. Wang, Non-invasive hemodynamic diagnosis based on non-linear pulse wave theory applied to four limbs, Front Bioeng Biotechnol 11 (2023) 1081447.

[133] I.B. Wilkinson, I.R. Hall, H. MacCallum, I.S. Mackenzie, C.M. McEniery, B.J. Van der Arend, Y.-E. Shu, L.S. MacKay, D.J. Webb, J.R. Cockcroft, Pulse-wave analysis: clinical evaluation of a noninvasive, widely applicable method for assessing endothelial function, Arterioscler Thromb Vasc Biol 22 (2002) 147–152.

[134] R.R. Townsend, H.R. Black, J.A. Chirinos, P.U. Feig, K.C. Ferdinand, M. Germain, C. Rosendorff, S.P. Steigerwalt, J.A. Stepanek, Clinical use of pulse wave analysis: proceedings from a symposium sponsored by North American Artery, The Journal of Clinical Hypertension 17 (2015) 503–513.

[135] O.T. Inan, P.F. Migeotte, P. Kwang-Suk, M. Etemadi, K. Tavakolian, R. Casanella, J. Zanetti, J. Tank, I. Funtova, G.K. Prisk, seismocardiography: a review of recent advances. Biomedical and Health Informatics, IEEE Journal Of 19 (2015) 1414–1427.

[136] R. Li, K. Sughimoto, X. Zhang, S. Wang, Y. Hiraki, H. Liu, Impact of Respiratory Fluctuation on Hemodynamics in Human Cardiovascular System: A 0-1D Multiscale Model, Fluids 7 (2022) 28.

[137] S. Wang, D. Wu, G. Li, Z. Zhang, W. Xiao, R. Li, A. Qiao, L. Jin, H. Liu, Deep learning-based hemodynamic prediction of carotid artery stenosis before and after surgical treatments, Front Physiol 13 (2023) 2674.

[138] K. Sughimoto, J. Levman, F. Baig, D. Berger, Y. Oshima, H. Kurosawa, K. Aoki, Y. Seino, T. Ueda, H. Liu, Machine learning predicts blood lactate levels in children after cardiac surgery in paediatric ICU, Cardiol Young 33 (2023) 388–395.

[139] H. Wang, Y. Cheng, A quantitative system for pulse diagnosis in traditional Chinese medicine, in: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, IEEE, 2006: pp. 5676–5679.

[140] G. Li, K. Watanabe, H. Anzai, X. Song, A. Qiao, M. Ohta, Pulse-wave-pattern classification with a convolutional neural network, Sci Rep 9 (2019) 14930.

[141] P.A. McKee, W.P. Castelli, P.M. McNamara, W.B. Kannel, The natural history of congestive HF: the Framingham study, New England Journal of Medicine 285 (1971) 1441–1446.

[142] F. Ashcroft, Life at the extremes: the science of survival, Univ of California Press, 2002.

[143] W.J. Paulus, C. Tschöpe, J.E. Sanderson, C. Rusconi, F.A. Flachskampf, F.E. Rademakers, P. Marino, O.A. Smiseth, G. De Keulenaer, A.F. Leite-Moreira, How to diagnose diastolic HF: a consensus statement on the diagnosis of HF with normal left ventricular ejection fraction by the HF and Echocardiography Associations of the European Society of Cardiology, Eur Heart J 28 (2007) 2539–2550.

[144] J.F. Polak, B.L. Holman, J. Wynne, W.S. Colucci, Right ventricular ejection fraction: an indicator of increased mortality in patients with congestive HF associated with coronary artery disease, J Am Coll Cardiol 2 (1983) 217–224.

[145] X. Zhang, Y. Shang, D. Guo, T. Zhao, Q. Li, X. Wang, A more effective method of extracting the characteristic value of pulse wave signal based on wavelet transform, J Biomed Sci Eng 9 (2016) 9–19.

[146] H.-T. Wu, C.-H. Lee, C.-E. Chen, A.-B. Liu, Predicting arterial stiffness with the aid of ensemble empirical mode decomposition (EEMD) algorithm, in: 2010 IEEE International

Conference on Wireless Communications, Networking and Information Security, IEEE, 2010: pp. 179–182.

[147] F. Chang, W. Hong, T. Zhang, J. Jing, X. Liu, Research on wavelet denoising for pulse signal based on improved wavelet thresholding, in: 2010 First International Conference on Pervasive Computing, Signal Processing and Applications, IEEE, 2010: pp. 564–567.

[148] L. Xu, M.Q.-H. Meng, K. Wang, W. Lu, N. Li, Pulse images recognition using fuzzy neural network, Expert Syst Appl 36 (2009) 3805–3811.

[149] X. Hu, H. Zhu, J. Xu, D. Xu, J. Dong, Wrist pulse signals analysis based on deep convolutional neural networks, in: 2014 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology, IEEE, 2014: pp. 1–7.

[150] S. Ghosh, A. Dasgupta, A. Swetapadma, A study on support vector machine based linear and non-linear pattern classification, in: 2019 International Conference on Intelligent Sustainable Systems (ICISS), IEEE, 2019: pp. 24–28.

[151] C. Goutte, E. Gaussier, A probabilistic interpretation of precision, recall and F-score, with implication for evaluation, in: European Conference on Information Retrieval, Springer, 2005: pp. 345–359.

[152] T. Saito, M. Rehmsmeier, The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets, PLoS One 10 (2015) e0118432.

[153] M. Grandini, E. Bagli, G. Visani, Metrics for multi-class classification: an overview, ArXiv Preprint ArXiv:2008.05756 (2020).

[154] N.N. Kathiria, Z. Devcic, J.S. Chen, D.M. Naeger, M.D. Hope, C.B. Higgins, K.G. Ordovas, Assessment of left ventricular enlargement at multidetector computed tomography, J Comput Assist Tomogr 39 (2015) 794–796.

[155] J.G. Nam, J. Kim, K. Noh, H. Choi, D.S. Kim, S.-J. Yoo, H.-L. Yang, E.J. Hwang, J.M. Goo, E.-A. Park, Automatic prediction of left cardiac chamber enlargement from chest radiographs using convolutional neural network, Eur Radiol 31 (2021) 8130–8140.

[156] N.C. Okwose, S.J. Charman, A.S. Fuller, G.A. MacGowan, N. Filipović, D.G. Jakovljevic, Heart Physiology and Heart Disease, in: In Silico Clinical Trials for Cardiovascular Disease: A Finite Element and Machine Learning Approach, Springer, 2024: pp. 47–76.

[157] M. Agarwal, S.N. Pakhare, Emerging therapeutic strategies in HF management: A narrative review of current evidence and future directions, IJCS 6 (2024) 32–38.

[158] Y. Eda, T. Nabeta, S. Iikura, Y. Takigami, T. Fujita, Y. Iida, Y. Ikeda, S. Ishii, J. Ako, Non-dilated left ventricular cardiomyopathy vs. dilated cardiomyopathy: clinical background and outcomes, ESC Heart Fail (2024).

[159] N. Pilz, A. Patzak, T.L. Bothe, Continuous cuffless and non-invasive measurement of arterial blood pressure—concepts and future perspectives, Blood Press 31 (2022) 254–269.

[160] S.N.A. Ismail, N.A. Nayan, R. Jaafar, Z. May, Recent advances in non-invasive blood pressure monitoring and prediction using a machine learning approach, Sensors 22 (2022) 6195.

[161] N. Pilz, V. Heinz, T. Ax, L. Fesseler, A. Patzak, T.L. Bothe, Pulse wave velocity: methodology, clinical applications, and interplay with heart rate variability, Rev Cardiovasc Med 25 (2024).

[162] F. Piccioli, A. Valiani, J. Alastruey, V. Caleffi, The effect of cardiac properties on arterial pulse waves: An in-silico study, Int J Numer Method Biomed Eng 38 (2022) e3658.

[163] F. Jiang, Y. Zhou, T. Ling, Y. Zhang, Z. Zhu, Recent research for unobtrusive atrial fibrillation detection methods based on cardiac dynamics signals: A survey, Sensors 21 (2021) 3814.

[164] T. Zhou, X. Ye, H. Lu, X. Zheng, S. Qiu, Y. Liu, Dense convolutional network and its application in medical image analysis, Biomed Res Int 2022 (2022) 2384830.

[165] A. Dhillon, G.K. Verma, Convolutional neural network: a review of models, methodologies and applications to object detection, Progress in Artificial Intelligence 9 (2020) 85–112.

[166] F. Fooladgar, S. Kasaei, Lightweight residual densely connected convolutional neural network, Multimed Tools Appl 79 (2020) 25571–25588.

[167] H. Ullah, M.B. Bin Heyat, F. Akhtar, Sumbul, A.Y. Muaad, M.S. Islam, Z. Abbas, T. Pan, M. Gao, Y. Lin, An End-to-End Cardiac Arrhythmia Recognition Method with an Effective DenseNet Model on Imbalanced Datasets Using ECG Signal, Comput Intell Neurosci 2022 (2022) 9475162.

[168] P. Xiong, Y. Xue, J. Zhang, M. Liu, H. Du, H. Zhang, Z. Hou, H. Wang, X. Liu, Localization of myocardial infarction with multi-lead ECG based on DenseNet, Comput Methods Programs Biomed 203 (2021) 106024.

[169] V.G. Krishnan, M.V.V. Saradhi, S.S. Kumar, G. Dhanalakshmi, P. Pushpa, V. Vijayaraja, Hybrid optimization based feature selection with DenseNet model for heart disease prediction, Int J Electr Electron Res 11 (2023) 253–261.

[170] S. Kenchaiah, R.S. Vasan, HF in women–insights from the Framingham Heart Study, Cardiovasc Drugs Ther 29 (2015) 377–390.

[171] D. Zhang, D. Zhang, Wavelet transform, Fundamentals of Image Data Mining: Analysis, Features, Classification and Retrieval (2019) 35–44.

[172] E. Por, M. van Kooten, V. Sarkovic, Nyquist–Shannon sampling theorem, Leiden University 1 (2019) 1–2.

[173] B.W. Eidem, The American Society of Echocardiography is Truly the Global Leader For All Users of Cardiovascular Ultrasound, Journal of the American Society of Echocardiography 37 (2024) A15–A16.

[174] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: pp. 4700–4708.

[175] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, ArXiv Preprint ArXiv:1412.6980 (2014).

[176] D.G. Mayer, D.G. Butler, Statistical validation, Ecol Modell 68 (1993) 21–32.

[177] A. De Myttenaere, B. Golden, B. Le Grand, F. Rossi, Mean absolute percentage error for regression models, Neurocomputing 192 (2016) 38–48.

[178] G.Y. Zou, Confidence interval estimation for the Bland–Altman limits of agreement with multiple observations per individual, Stat Methods Med Res 22 (2013) 630–642.

Acknowledgements

As my doctoral studies are near completion, I would like to take this opportunity to express my heartfelt gratitude to all those who have provided support and assistance throughout my academic and personal journey.

First and foremost, I would like to extend my sincerest thanks to my advisor, Professor Liu Hao. Four years ago, he graciously offered me the invaluable opportunity to pursue my PhD at Chiba University, which not only altered the trajectory of my academic career but also opened a new chapter of exploration and personal growth. During this time, he has provided me with substantial academic guidance, helping me to enhance my research abilities, and has also offered me endless encouragement and support in my personal life. Whether I was facing research challenges or personal struggles, he was always there with patience and wisdom to guide me through. I vividly recall the many discussions we had about my research, and his meticulous attention to detail and rigorous academic approach have profoundly shaped my way of thinking and research abilities. It is through his unwavering mentorship that I have been able to make continuous progress in my academic endeavors.

I would also like to express my heartfelt gratitude to the members of my defense committee. Their valuable insights into my research and their commitment to taking the time for my defense are truly appreciated. I would like to extend special thanks to Professors Liu Youjun and Qiao Aike, whose guidance and support were instrumental in my progress.

Meanwhile, I am deeply grateful to Prof. Kobayashi and Dr. Ono from Chiba University Hospital for providing the majority of the data used in this study and for all their contributions to the research.

During my time studying in Japan, I was fortunate to meet many exceptional friends and classmates, including Li Ruichen, Sun Jianwei, Rong Jiaxin, Xue Yujing, Wang Dongyue, Miki Toru, Koda Daiki, and Maeda Kenshu. Their help and companionship have made my doctoral journey filled with cherished memories. I will cherish our friendship and the wonderful memories in my heart.

Furthermore, I owe my deepest gratitude to my parents and family for their unwavering support and love, which have been the bedrock of my ability to complete my studies. Their constant encouragement, despite the distance, has been a source of strength that kept me grounded and

focused throughout the most challenging moments of this journey. Lastly, I want to express my heartfelt thanks to my boyfriend, Wang Sirui, whose unwavering support, understanding, and companionship helped me persevere through this journey. His presence has been a constant source of comfort and motivation, and I could not have reached this point without him by my side.

Through this acknowledgment, I hope to convey my sincere appreciation to everyone who has supported me throughout my academic path. The doctoral journey is full of challenges, but it is through the care and presence of so many people that I have been able to reach this point.

Publications and presentations

Publications in peer-reviewed journals:

1. Wu, D., Ono, R., Wang, S., Kobayashi, Y., Sughimoto, K., and Liu, H. (2024). Pulse wave signal-driven machine learning for identifying left ventricular enlargement in HF patients. BioMedical Engineering OnLine, 23(1), 60.
2. Wu, D., Ono, R., Wang, S., Kobayashi, Y., and Liu, H. (2024). Pulse Wave Time Series Unsupervised Clustering with Importance Ratios for HF Subgroups Detection. Journal of Biomechanical Science and Engineering, 24-00325.
3. Wang, S., Wu, D., Li, G., Song, X., Qiao, A., Li, R., Liu, Y., Anzai, H., and Liu, H. (2022). A machine learning strategy for fast prediction of cardiac function based on peripheral pulse wave. Computer Methods and Programs in Biomedicine, 216, 106664.
4. Wang, S., Wu, D., Li, G., Zhang, Z., Xiao, W., Li, R., Qiao, A., Jin, L., Liu, H. (2023). Deep learning-based hemodynamic prediction of carotid artery stenosis before and after surgical treatments. Frontiers in Physiology, 13, 2674.
5. Wang, S., Ryohei O., Wu D., Aoki K., Kato, H., Iwahana T., Okada S., Kobayashi Y., Liu H. (2024). Pulse wave-based evaluation of the blood-supply capability of patients with HF via machine learning. BioMedical Engineering OnLine, 23(1), 7.

Presentations:

1. Wu, D., Ono, R., Wang, S., and Liu, H. "脈波信号を用いた機械学習による心疾患左室肥大の予測." 日本機械学会第 34 回バイオフロンティア講演会, 山口, 2023/12/16.
2. Wu, D., and Liu, H. "脈波信号を利用した機械学習による心臓疾患の左室拡張予測." 日本機械学会第 36 回バイオエンジニアリング講演会, 名古屋, 2024/05/12.
3. Wu, D., and Liu, H. "Pulse wave Signal-Based Machine Learning Approaches for Identifying Left Ventricular Enlargement in HF Patients." 18th International Conference on Biomedical Engineering (ICBME 2024), Singapore, 2024/12/09.