frontiers | Research Topics

Machine learning and data science in heart failure and stroke

Edited by
Leonardo Roever and Qingpeng Zhang

Coordinated by
Sharen Lee

Published in
Frontiers in Cardiovascular Medicine

FRONTIERS EBOOK COPYRIGHT STATEMENT

The copyright in the text of individual articles in this ebook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this ebook is the property of Frontiers.

Each article within this ebook, and the ebook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this ebook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or ebook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-8325-3338-3
DOI 10.3389/978-2-8325-3338-3

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Machine learning and data science in heart failure and stroke

Topic editors
Leonardo Roever — Federal University of Uberlandia, Brazil
Qingpeng Zhang — City University of Hong Kong, Hong Kong, SAR China

Topic coordinator
Sharen Lee — The Chinese University of Hong Kong, China

Citation
Roever, L., Zhang, Q., Lee, S., eds. (2023). *Machine learning and data science in heart failure and stroke*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-3338-3

Table of contents

frontiers | Frontiers in Cardiovascular Medicine

TYPE Editorial
PUBLISHED 22 August 2023
DOI 10.3389/fcvm.2023.1275739

Check for updates

OPEN ACCESS

EDITED AND REVIEWED BY
Matteo Cameli,
University of Siena, Italy

*CORRESPONDENCE
Leonardo Roever
leonardoroever@hotmail.com

RECEIVED 10 August 2023
ACCEPTED 14 August 2023
PUBLISHED 22 August 2023

CITATION
Roever L, Lee S and Zhang Q (2023) Editorial: Machine learning and data science in heart failure and stroke.
Front. Cardiovasc. Med. 10:1275739.
doi: 10.3389/fcvm.2023.1275739

COPYRIGHT
© 2023 Roever, Lee and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Editorial: Machine learning and data science in heart failure and stroke

Leonardo Roever[1,2]*, Sharen Lee[3] and Qingpeng Zhang[4]

[1]Department of Clinical Research, Federal University of Uberlândia, Uberlândia, Brazil, [2]Gilbert and Rose-Marie Chagoury School of Medicine, Lebanese American University, Beirut, Lebanon, [3]Li Ka Shing Institute of Health Sciences, Faculty of Medicine, The Chinese University of Hong Kong, Hong Kong, Hong Kong SAR, China, [4]School of Data Science, City University of Hong Kong, Hong Kong, Hong Kong SAR, China

KEYWORDS

machine learning, data science, heart failure, stroke, algorithms

Editorial on the Research Topic
Machine learning and data science in heart failure and stroke

In this Topic on Machine Learning and Data Science in Heart Failure and Stroke, an international collection of manuscripts is presented that aims to contribute to the advancement of understanding of evidence-based approaches to the prevention, prognosis, diagnosis, and treatment of cardiovascular disease. In conclusion, the high-quality clinical case contributions presented in this Research Topic have significantly boosted knowledge, diagnosis, and treatment of cardiovascular disease in complex cases.

Liu et al. presented a retrospective study that aimed to build a machine learning (ML) model to predict the occurrence of AKI in patients with HF. The Medical Information Mart for Intensive Care-IV (MIMIC-IV) database was used. the ML model was established to predict AKI development using decision tree, random forest (RF), support vector machine (SVM), K-Nearest Neighbor (KNN) and logistic regression (LR) algorithms. Demographic, clinical, and treatment characteristics were used to establish the model. A total of 2,678 patients with HF were analyzed, 919 developed AKI. Among the 5 ML algorithms, the RF algorithm exhibited the highest performance with an AUROC of 0.96 Liu et al.

The study by Tu and collaborators aimed to identify central HF-related genes and regulatory networks using bioinformatics and validation assays. Using four sets of RNA-seq data in the Gene Expression Omnibus (GEO) database, the authors screened for HF differentially expressed genes (DEGs) using Removal of Undesired Variation from RNA-seq Data (RUVSeq) and the Robust Classification Aggregation (RRA) method.: A total of 201 robust DEGs were identified in patients with HF and NFDs. In this study identified ASPN, COL1A1 and FMOD as potential diagnostic biomarkers for HF Tu et al.

Dao and colleagues performed a systematic review of 19 studies that analyzed 5,614 participants. The objective was to compare the sensitivity and specificity of the diagnosis between the third heart sound (S3) and the left ventricular ejection fraction (LVEF) in heart failure (HF). In the result of this research, it was observed that S3 alone presented lower sensitivity in the diagnosis of HF compared to LVEF, but it was useful in the early pathological evaluation Dao et al.

Cheng et al. in their article portrayed the relationship between blood pressure patterns and age, as well as the tendency towards high prevalence of BP over time in different age groups. A total of 71,468 participants aged over 18 years with complete information on weight, height, age, gender, glucose, triglycerides, total cholesterol, systolic (SBP) and diastolic (DBP) blood pressure were included for analysis. The risk of high SBP showed a continuous increase from 35 to 79 years of age and a concomitant early increase in the risk of high DBP; after the age of 50–65 years, the high risk of PAD decreased. High-risk SBP progresses more rapidly in early life in Chinese women compared to later life Cheng et al.

Sung and colleagues developed an electronic health record based machine learning model to assess the risk of newly detected atrial fibrillation (NDAF) at an early stage after stroke. The study population consisted of a training set of 4,064 and a temporal test set of 1,492 patients. At the median follow-up of 10.2 months, the incidence rate of NDAF was 87.0 per 1,000 person-years in the test set. On the test set, the model based on both structured and unstructured data achieved a C-score of 0.840, which was significantly higher than the AS5F and CHASE-LESS scores. More studies are needed to assess the clinical utility of the prediction model Sung et al.

Huang et al. presented a retrospective study including 4,570 Chinese adults with the aim of identifying independent risk factors for carotid atherosclerosis (CAS) and constructing and validating a CAS risk prediction model based on the Chinese population. Participants were randomly assigned to the training and validation sets in a 7:3 ratio. C-index curves and receiver operating characteristics, calibration plots, and decision curve analysis (DCA) were used to assess discrimination, calibration, and clinical applicability of the risk model. In the training, internal validation, and external validation sets, the risk model showed good discriminatory power with C-indices of 0.961 (0.953–0.969), 0.953 (0.939–0.967) and 0.930 (0.920–0.94 0), respectively, and excellent calibration. The development of risk models can contribute to the early identification and prevention of CAS Huang et al.

Burton and colleagues studied 396 patients using electromechanical (EM) waveforms to assess left ventricular end-diastolic pressure elevation (LVEDP). This analysis identified subgroups of patients with varying degrees of LVEDP elevation based on waveform characteristics Burton et al.

Susic et al. used data recorded on 37 patients using two types of electronic stethoscopes. This study demonstrated that, in patients with chronic heart failure, machine learning algorithms can outperform cardiologists in detecting episodes of decompensation based on heart sounds alone Susic et al.

In the research by Sabovčik and collaborators, data from 30,354 individuals from 6 cohorts were used. The predictive performance of increased survival gradient (GBS), CoxNet, the PCP-HF risk score, and a stacking method were evaluated. In the accuracy recall (PR) analysis for predicting 10-year HF risk, the stacking method, combining the SGB, CoxNet, Gaussian mixture, and PCP-HF models, outperformed other models with PR/AUC 0.804, while PCP-HF achieved only 0.551. Flexible ML algorithms can be used to capture these diverse distributions and produce more accurate prediction models Sabovčik et al.

Gtif and colleagues used data from 116 patients with heart failure with the pathogenesis of reduced ejection fraction (HFrEF). A generalized linear model (GLM), random forest, and extreme gradient augmentation models were developed to predict the risk of post-discharge mortality using clinical and laboratory data. The result obtained was a discriminatory power of 74.5% for post-mortality by the area under the curve (AUC) Gtif et al.

In conclusion, this research topic presented several machine learning models that can be used to improve diagnosis and treatment in patients with stroke and heart failure. More studies are needed to improve and validate these techniques nets patient type.

Author contributions

LR: Writing – original draft, Writing – review & editing. SL: Writing – original draft, Writing – review & editing. QZ: Writing – original draft, Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Cardiovascular Medicine

TYPE Original Research
PUBLISHED 29 July 2022
DOI 10.3389/fcvm.2022.941237

Check for updates

OPEN ACCESS

EDITED BY
Qingpeng Zhang,
City University of Hong Kong,
Hong Kong SAR, China

REVIEWED BY
Pradyumna Agasthi,
Mayo Clinic, United States
Mei Liu,
University of Kansas Medical Center,
United States

*CORRESPONDENCE
Pei-Ju Lee
pjlee@mis.ccu.edu.tw
Ya-Han Hu
yhhu@mgt.ncu.edu.tw

SPECIALTY SECTION
This article was submitted to Heart Failure and Transplantation, a section of the journal Frontiers in Cardiovascular Medicine

RECEIVED 11 May 2022
ACCEPTED 11 July 2022
PUBLISHED 29 July 2022

CITATION
Sung S-F, Sung K-L, Pan R-C, Lee P-J and Hu Y-H (2022) Automated risk assessment of newly detected atrial fibrillation poststroke from electronic health record data using machine learning and natural language processing.
Front. Cardiovasc. Med. 9:941237.
doi: 10.3389/fcvm.2022.941237

COPYRIGHT
© 2022 Sung, Sung, Pan, Lee and Hu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Automated risk assessment of newly detected atrial fibrillation poststroke from electronic health record data using machine learning and natural language processing

Sheng-Feng Sung[1,2], Kuan-Lin Sung[3], Ru-Chiou Pan[4], Pei-Ju Lee[5]* and Ya-Han Hu[6]*

[1]Division of Neurology, Department of Internal Medicine, Ditmanson Medical Foundation Chiayi Christian Hospital, Chiayi City, Taiwan, [2]Department of Nursing, Min-Hwei Junior College of Health Care Management, Tainan, Taiwan, [3]School of Medicine, National Taiwan University, Taipei, Taiwan, [4]Clinical Data Center, Department of Medical Research, Ditmanson Medical Foundation Chiayi Christian Hospital, Chiayi City, Taiwan, [5]Department of Information Management and Institute of Healthcare Information Management, National Chung Cheng University, Chiayi County, Taiwan, [6]Department of Information Management, National Central University, Taoyuan, Taiwan

Background: Timely detection of atrial fibrillation (AF) after stroke is highly clinically relevant, aiding decisions on the optimal strategies for secondary prevention of stroke. In the context of limited medical resources, it is crucial to set the right priorities of extended heart rhythm monitoring by stratifying patients into different risk groups likely to have newly detected AF (NDAF). This study aimed to develop an electronic health record (EHR)-based machine learning model to assess the risk of NDAF in an early stage after stroke.

Methods: Linked data between a hospital stroke registry and a deidentified research-based database including EHRs and administrative claims data was used. Demographic features, physiological measurements, routine laboratory results, and clinical free text were extracted from EHRs. The extreme gradient boosting algorithm was used to build the prediction model. The prediction performance was evaluated by the C-index and was compared to that of the AS5F and CHASE-LESS scores.

Results: The study population consisted of a training set of 4,064 and a temporal test set of 1,492 patients. During a median follow-up of 10.2 months, the incidence rate of NDAF was 87.0 per 1,000 person-year in the test set. On the test set, the model based on both structured and unstructured data achieved a C-index of 0.840, which was significantly higher than those of the AS5F (0.779, $p = 0.023$) and CHASE-LESS (0.768, $p = 0.005$) scores.

Conclusions: It is feasible to build a machine learning model to assess the risk of NDAF based on EHR data available at the time of hospital admission.

Inclusion of information derived from clinical free text can significantly improve the model performance and may outperform risk scores developed using traditional statistical methods. Further studies are needed to assess the clinical usefulness of the prediction model.

KEYWORDS

atrial fibrillation, electronic health records, ischemic stroke, natural language processing, prediction

Introduction

Ischemic stroke is associated with a substantial risk of recurrence with a one-year recurrence rate ranging from 6 to 18% (1–4). The risk of stroke recurrence depends on the subtypes of ischemic stroke. As compared to other stroke subtypes, the recurrence rate of cardioembolic stroke is relatively high (5, 6). Moreover, cardioembolic strokes are often followed by strokes of the same type (6, 7). Atrial fibrillation (AF) is the most common cause of cardioembolic stroke, and even embolic stroke of undetermined source (ESUS) may originate from subclinical AF (8). As the population ages, AF-related strokes have increased and may triple in the next few decades (9, 10). Fortunately, the advancement of non-vitamin K antagonist oral anticoagulant therapy has made great progress in preventing patients with AF from cardioembolic stroke (8). Nonetheless, since AF can be paroxysmal, it may go undetected and therefore undiagnosed in patients undergoing routine electrocardiography (ECG) examinations. In fact, for ischemic stroke patients with undiagnosed AF, delayed use of oral anticoagulants may double the risk of recurrent stroke or transient ischemic attack (TIA) (11). Considering the impact of anticoagulant therapy on the outcome, poststroke screening for AF is thus critical for preventing recurrent stroke in patients with acute ischemic stroke (AIS).

Approximately 30% of all ischemic strokes are without any apparent cause (12). Among these cryptogenic strokes, nearly two-thirds are considered to stem from embolism (12). A study points out that through a series of heart rhythm monitoring, AF can be detected in up to 24% of patients with AIS or TIA (13). In addition to 24-h or even 72-h Holter monitoring (14), numerous studies have established that extended ECG monitoring *via* either implantable or external devices increases the yield of AF detection in patients with AIS (15, 16). However, given the limited medical resources, setting the right priorities of extended ECG monitoring by stratifying patients into different risk groups likely to have newly detected AF (NDAF) is more crucial than implementing population-level screening (17).

To date, more than twenty risk scores have been proposed to assess the risk of poststroke NDAF (18, 19). These risk scores vary in their complexity, target population, outcome definition, predictor variables, and ease of implementation. Most risk scores were derived or validated in patients with AIS while some of them were derived from a specific population with cryptogenic stroke or ESUS (20, 21). The simplest risk score consists of only two predictor variables, that is, age and stroke severity as assessed using the National Institutes of Health Stroke Scale (NIHSS) (22). Nevertheless, many of the risk scores require additional diagnostic work-up or interpretation of examination results to obtain the necessary predictors, such as markers of blood, ECG, echocardiography, as well as brain and vascular imaging (18). Routine use of such risk scores may be impractical in the context of the extra time and cost required.

On the other hand, with the ubiquitous use of electronic health records (EHRs) and the advancement in computational power, it has become feasible to use EHRs for the creation, validation, and implementation of data-driven risk prediction models (23, 24). For example, a previous study developed and validated an EHR-based prediction tool for 5-year AF risk in the general population (25), demonstrating a simple and cost-conscious approach to AF screening. Furthermore, in addition to structured numerical and categorical data, EHRs accommodate a multitude of unstructured textual data such as narrative clinical notes. Combining information extracted from clinical free text through natural language processing with structured data has shown promising results in improving the performance of risk prediction models (26–28).

AF-related strokes tend to be more severe and may manifest with different clinical features than other subtypes of ischemic strokes (29, 30). A higher risk of NDAF has been observed in patients with greater stroke severity (22, 31). Previous studies have shown that information extracted from clinical text can be used to represent patients' stroke severity (28, 32). Furthermore, stroke patients with AF have a higher prevalence of heart diseases and experience more cardiac events than those without AF (29–31). Symptoms, signs, or examinations related to heart diseases are typically documented in clinical notes. However, such information may not be captured or routinely collected as structured data in the EHR system. We thereby hypothesized that clinical text contains information that can discriminate between strokes stemming from AF and those not stemming from AF. In this study, we aimed to develop an EHR-based

machine learning (ML) model to assess the risk of NDAF. To this end, we investigated various ML models using structured data, unstructured textual data, or a combination of both. In addition, the prediction performance of the developed ML models was compared to that of two traditional risk scores on a temporal test set of patients hospitalized for AIS.

Materials and methods

Data sources

The study data was obtained from the stroke registry of the Ditmanson Medical Foundation Chia-Yi Christian Hospital and the Ditmanson Research Database (DRD), a deidentified database comprising both EHR data and administrative claims data for research purposes. The DRD currently holds clinical information of over 1.4 million patients. The hospital stroke registry has prospectively enrolled consecutive hospitalized stroke patients since 2007 conforming to the design of the nationwide Taiwan Stroke Registry (33). To create the dataset for this study, we linked the stroke registry to the DRD using a unique encrypted patient identifier. Information regarding risk factors and stroke severity as assessed using the NIHSS was obtained from the stroke registry. Billing information and medical records from 2 years before to 1 year after the index stroke were extracted from the DRD.

The study protocol was approved by the Ditmanson Medical Foundation Chia-Yi Christian Hospital Institutional Review Board (IRB2020135). The requirement for informed consent was waived because of the retrospective design. The study protocol conforms to the ethical guidelines of the 1975 Declaration of Helsinki.

Study population

The study population selection is shown in Supplementary Figure 1. The stroke registry was queried for all hospitalizations for AIS between Oct 2007 and Sep 2020. Only the first hospitalization was included for each patient. Patients who suffered an in-hospital stroke or whose records could not be linked were excluded. The study population was split into a training set (patients admitted before the end of 2016) and a temporal test set (those admitted from 2017 onwards). All patients were traced in the DRD until AF was detected, death, the last visit within 1 year after the index stroke, or February 28, 2021, whichever came first.

Predictor and outcome variables

The class label (outcome) was AF, which was defined according to an AF ascertainment algorithm detailed in the Supplementary Methods in the Supplementary Material. According to the time sequence between AF detection and the index stroke (13), AF was further categorized as known AF before the index stroke, AF detected on admission, AF detected during the index stroke hospitalization, and AF detected after discharge (Figure 1). During the training phase, we trained ML models to predict which stroke is likely to stem from AF. Therefore, patients with all kinds of AF were retained in the training set. Because the study purpose was to build an ML model to assess the risk of NDAF poststroke, i.e., AF detected during the index stroke hospitalization and AF detected after discharge (Figure 1), patients who had known AF before the index stroke or AF detected on admission (34) were further excluded from the test set.

A total of 20 structured predictor variables (Supplementary Table 3), including age, sex, body mass index (BMI), vital signs, and results of routine blood tests, were chosen because they are readily available from EHRs upon admission. Missing values were imputed as mean values for continuous variables. Besides these structured variables, the free text extracted from the History of Present Illness section of the admission note was preprocessed through the following steps: spell checking, abbreviation expansion, removal of non-word symbols, removal of words suggestive of AF ("paroxysmal", "atrial", "fibrillation"), lowercase conversion, lemmatization, marking of negated words with the suffix "_NEG" using the Natural Language Toolkit mark_negation function, and stop-word removal.

The preprocessed text was then vectorized using the bag-of-words (BOW) approach with three different types of feature representation (Figure 2). We built a document-term matrix in which each column represents each unique feature (word) from the text corpus while the rows represent each document (present illness for each patient). The cells represent the counts of each word within each document (term frequency), the absence or presence of each word within each document (binary representation), or the term frequency with inverse document frequency (TF-IDF) weighting (35). Because medical terms are commonly comprised of two words or even more, we further experimented with adding word bigram features (two-word phrases) to the basic BOW model. To reduce noises such as redundant and less informative features as well as to improve training efficiency (36), we performed feature selection by filtering out words that appeared in <5% of all documents in the training set, followed by performing a penalized logistic regression with 10-fold cross-validation to identify the most predictive words (37).

Baseline models

For comparison with ML models, we only considered traditional risk scores that are based on variables available from

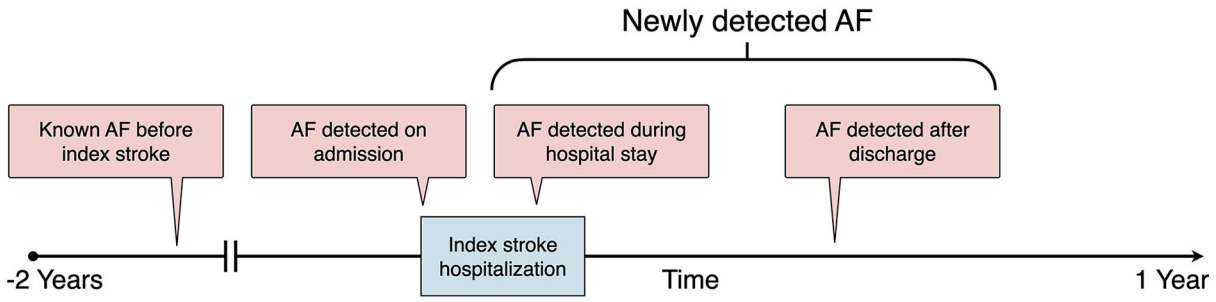


FIGURE 1
Definition of AF categories according to the time sequence between AF detection and the index stroke. AF, atrial fibrillation.

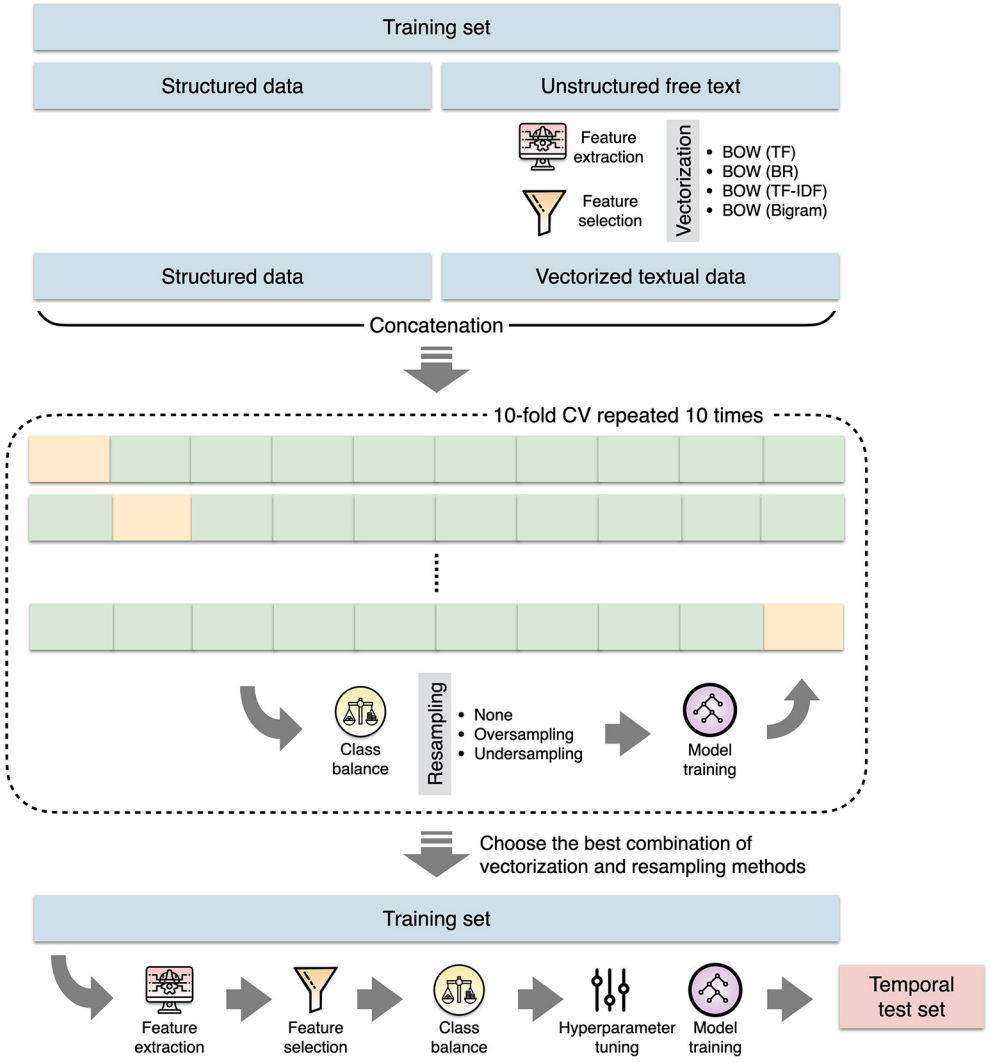


FIGURE 2
The process of machine learning model construction. BOW, bag-of-words; BR, binary representation; CV, cross validation; TF, term frequency; TF-IDF, term frequency with inverse document frequency.

EHRs upon admission. According to a validation study that evaluated eight such risk scores, two risk scores performed better than the others, demonstrating adequate discrimination and calibration (19). These two risk scores were thus used as the baseline models. The AS5F score, composed by age and NIHSS, was developed and validated in cohorts of patients who underwent extended Holter monitoring after AIS or TIA (22). The CHASE-LESS score was constructed from patients hospitalized for AIS in a claims database (31). It comprises seven components, including age, NIHSS, as well as the presence of coronary artery disease, congestive heart failure, hyperlipidemia, diabetes, and prior stroke or TIA.

Machine learning models

ML models were constructed by using structured data, vectorized textural data, or a combination of both (Figure 2). Because class imbalance might influence the classification performance, we experimented with resampling methods to maintain the ratio of majority and minority classes as 1:1, 2:1, or 3:1 (38). The extreme gradient boosting (XGB) algorithm was used to build classifiers. The XGB classifier trains a series of classification and regression trees where each successive tree attempts to correct the errors of the preceding trees.

During the training process, we first evaluated a suite of different combinations of text vectorization techniques and resampling methods without hyperparameter tuning. We repeated 10-fold cross-validation 10 times to obtain the performance estimates. The area under the receiver operating characteristic curve (AUC) was used as the evaluation metric because both positive and negative classes are important. After the optimal combination of text vectorization and resampling methods was determined, ML models were trained from the full training set through feature extraction, feature selection, class balancing, followed by hyperparameter tuning. Hyperparameter optimization for each model was performed by repeating 10-fold cross-validation 10 times. Model error was minimized in terms of AUC. We performed a grid search to find optimal hyperparameters following steps proposed in a prior study (39). After building the XGB classifiers, we used Shapley additive explanations (40) to interpret the output of the XGB classifiers. The experiments were carried out by using scikit-learn, XGBoost, imbalanced-learn, and SHAP libraries within Python 3.7 environment.

Statistical analysis

Categorical variables were reported with counts and percentages. Continuous variables were presented as means with standard deviations or medians and interquartile ranges. Differences between groups were tested by Chi-square tests for categorical variables and *t* tests or Mann-Whitney U tests for continuous variables, as appropriate.

The incidence rate of NDAF was expressed as events per 1,000 person-years. To assess the prediction performance of each prediction model, Cox proportional hazard regression analyses were performed by entering each risk score or the predicted probability output by each ML model as a continuous variable. Harrell's concordance index (C-index) was calculated to evaluate and compare model performance. The C-index ranges from 0.5 to 1.0, with 0.5 indicating random guess and 1 indicating perfect model discrimination. A model with a C-index value above 0.7 is considered acceptable for clinical use (41).

All statistical analyses were performed using Stata 15.1 (StataCorp, College Station, Texas) and R version 4.1.1 (R Foundation for Statistical Computing, Vienna, Austria). Two-tailed p values were considered statistically significant at <0.05.

Results

Characteristics of the study population

A total of 6,321 patients were eligible for this study (Supplementary Figure 1). The training set consisted of 4,604 patients who were admitted before the end of 2016. Among patients in the training set, 422 (9.2%) had known AF, 265 (5.6%) were diagnosed with AF on admission, and 232 (5.0%) developed NDAF during follow-up. Among 1,717 patients who were admitted from 2017 onwards, 122 and 103 were excluded because of having known AF before the index stroke and being diagnosed with AF on admission, respectively. Therefore, the temporal test set consisted of 1,492 patients. During a median follow-up of 10.2 months, 87 (5.8%) patients in the temporal test set were identified as having NDAF. Each patient had an average of 3.1 hospital visits per month during the follow-up period. The incidence rate of NDAF was 87.0 per 1,000 person-year. Table 1 lists the characteristics of the patients. Patients in the training set were older, more likely to be female, less likely to have diabetes mellitus, and tended to have hypertension, coronary artery disease, congestive heart failure, as well as prior stroke or TIA. They also had significantly higher NIHSS, AS5F, and CHASE-LESS scores.

Performance of prediction models

According to the estimates of AUC obtained from the 10 times of 10-fold cross-validation (Figure 3), ML models using a combination of both structured and unstructured data achieved higher AUCs than those using structured or unstructured data alone. Data resampling did not improve the performance of models. Text vectorization using BOW with TF-IDF weighting generally performed higher than the other

TABLE 1 Characteristics of the study population.

Characteristic	Training set ($N = 4,604$)	Temporal test set ($N = 1,492$)	*P*
Age, mean (SD)	69.2 (12.3)	68.0 (13.5)	0.002
Female	1,896 (41.2)	531 (35.6)	<0.001
Hypertension	3,705 (80.5)	1,119 (75.0)	<0.001
Diabetes mellitus	1,958 (42.5)	683 (45.8)	0.028
Hyperlipidemia	2,670 (58.0)	852 (57.1)	0.546
Coronary artery disease	560 (12.2)	103 (6.9)	<0.001
Congestive heart failure	228 (5.0)	25 (1.7)	<0.001
Prior stroke or TIA	1,143 (24.8)	274 (18.4)	<0.001
NIHSS, median (IQR)	5 (3-10)	5 (2-8)	<0.001
AS5F, median (IQR)	67.4 (59.2–76.5)	65.8 (56.9–74.2)	<0.001
CHASE-LESS, median (IQR)	6 (5-8)	6 (4-7)	<0.001

Data are numbers (percentage) unless specified otherwise.
IQR, interquartile range; NIHSS, National Institutes of Health Stroke Scale; SD, standard deviation; TIA, transient ischemic attack.

feature value representation methods. Therefore, we used the full original training set to build three ML models, that is, a model based on structured data (model A), a model based on textual data vectorized using BOW with TF-IDF weighting (model B), and a model based on both structured data and unstructured textual data vectorized using BOW with TF-IDF weighting (model C).

Table 2 lists the performance of prediction models. All the prediction models significantly predicted the risk of NDAF. Among the ML models, model C had the highest C-index (0.840), which was significantly higher than those of model A (0.791, $p = 0.009$) and model B (0.738, $p < 0.001$). Model C outperformed the AS5F (0.779, $p = 0.023$) and CHASE-LESS (0.768, $p = 0.005$) scores. The C-index of model A was comparable to those of AS5F ($p = 0.715$) and CHASE-LESS ($p = 0.487$) scores. Although model B attained the lowest C-index, its performance was also comparable to the AS5F ($p = 0.163$) and CHASE-LESS ($p = 0.282$) scores.

Model interpretation

Figure 4A shows the top 20 most important features in model C ordered by the mean absolute Shapley value, which indicates the global importance of each feature on the model output. Figure 4B presents the beeswarm plot depicting the Shapley value for every patient across these features, demonstrating each feature's contribution to the model output. According to the magnitude and direction of the Shapley value, patients who were female and those with increased age, high heart rate, elevated creatinine, elevated blood urea nitrogen, and high BMI were more likely to have NDAF. Patients with high triglyceride, platelet count, and pulse pressure were less likely to have NDAF. Words associated with an increased risk of NDAF included "unit", "middle", "cardiovascular", "heart", "electrocardiogram", and "family", whereas those associated with a decreased risk were "numbness", "diabetes", "day", "visit", and "ago".

The top 20 most important features in model A and model B are shown in Supplementary Figures 2, 3, respectively. The important structured and unstructured predictors identified in model C were generally consistent with those identified separately in model A (structured data) and model B (unstructured textual data).

Discussion

We found that prediction of NDAF using routinely collected variables from EHRs was feasible. ML models performed better than or were comparable to existing risk scores. The ML model based on both structured variables and text had higher discriminability than those of AS5F and CHASE-LESS scores. Furthermore, by using the Shapley value to reveal the significance of features, we identified important predictors of NDAF that may help gain insight into clinical practice for stroke prevention.

Important predictors of newly detected atrial fibrillation

Many studies have investigated prediction models for NDAF in the general population (25, 42, 43) or in selected patient groups such as those with stroke or TIA (18, 21, 22, 31, 44, 45). Owing to the different characteristics of at-risk populations, it is arguable whether the relationships between the predictors and

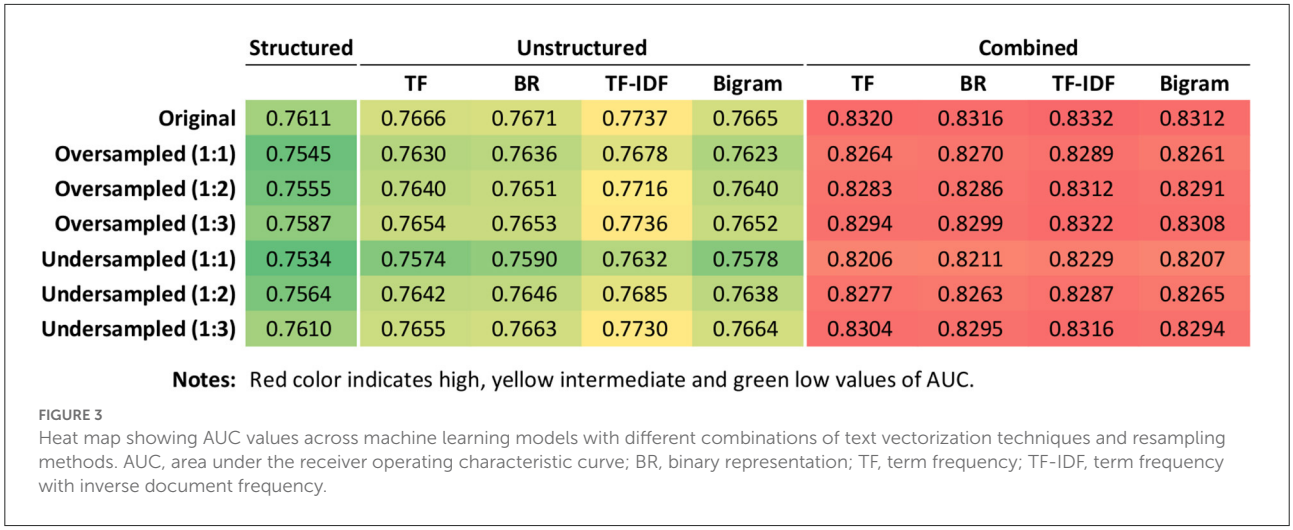


FIGURE 3
Heat map showing AUC values across machine learning models with different combinations of text vectorization techniques and resampling methods. AUC, area under the receiver operating characteristic curve; BR, binary representation; TF, term frequency; TF-IDF, term frequency with inverse document frequency.

TABLE 2 Performance of prediction models for predicting newly detected atrial fibrillation.

Risk score	HR (95% CI)	*P*	Schoenfeld's global test	C-index (95% CI)
AS5F	1.10 (1.08–1.13)	<0.001	0.062	0.779 (0.734–0.825)
CHASE-LESS	1.49 (1.38–1.60)	<0.001	0.296	0.768 (0.721–0.816)
Model A (structured)	1.05 (1.04–1.06)	<0.001	0.764	0.791 (0.745–0.836)
Model B (unstructured)	1.04 (1.03–1.05)	<0.001	0.060	0.738 (0.688–0.788)
Model C (combined)	1.05 (1.04–1.06)	<0.001	0.600	0.840 (0.803–0.876)

CI, confidence interval; HR, hazard ratio.

NDAF are similar across patient groups. Among the identified structured predictor variables, some of them such as age and BMI were common to the general population and patients with stroke (18, 42), others are known predictors in the general population but have seldom been used to predict poststroke NDAF, while still others are controversial predictors that warrant further study. For example, chronic kidney disease is a positive predictor whereas hyperlipidemia is a negative predictor of NDAF in the general population (25, 43). This study echoes those findings by showing positive associations of NDAF with elevated creatinine, elevated blood urea nitrogen, as well as decreased triglyceride level (Figure 4B). On the other hand, the evidence on the relationship between heart rate and NDAF is conflicting (46).

The central hypothesis of this study is that clinical free text contains information that may be used to predict NDAF. We indeed identified several words that could help make predictions. The reason why some of these words were associated with the risk of NDAF may be obscure at first glance but could be revealed by examining each word in its context. For example, the word "unit" from the term "intensive care unit" and the word "middle" from the term "middle cerebral artery infarction" typically imply severe stroke, which is a known predictor of NDAF (22, 31). These results demonstrate that useful and informative predictors could be derived from unstructured text in EHRs without intervening human curation. Despite this, since clinicians may use different terms to describe the same condition in clinical text, the relationship between such terms might not be accurately represented. Concept-based feature extraction using specialized medical ontologies can be explored in future research (35).

Advantages of EHR-based machine learning models

Traditional prediction models used in clinical practice are generally built on limited predefined variables using logistic regression. Although such models have reasonable prediction performance, whether they are applicable in routine clinical practice and relevant to a specific context is yet to be determined (47). First, logistic regression models necessitate the assumptions of linear and additive relationships among predictors being fulfilled, while ML algorithms, especially tree-based models, are more effective in capturing potential nonlinear relationships and handling complex interactions between the predictor and outcome variables (48). Second,

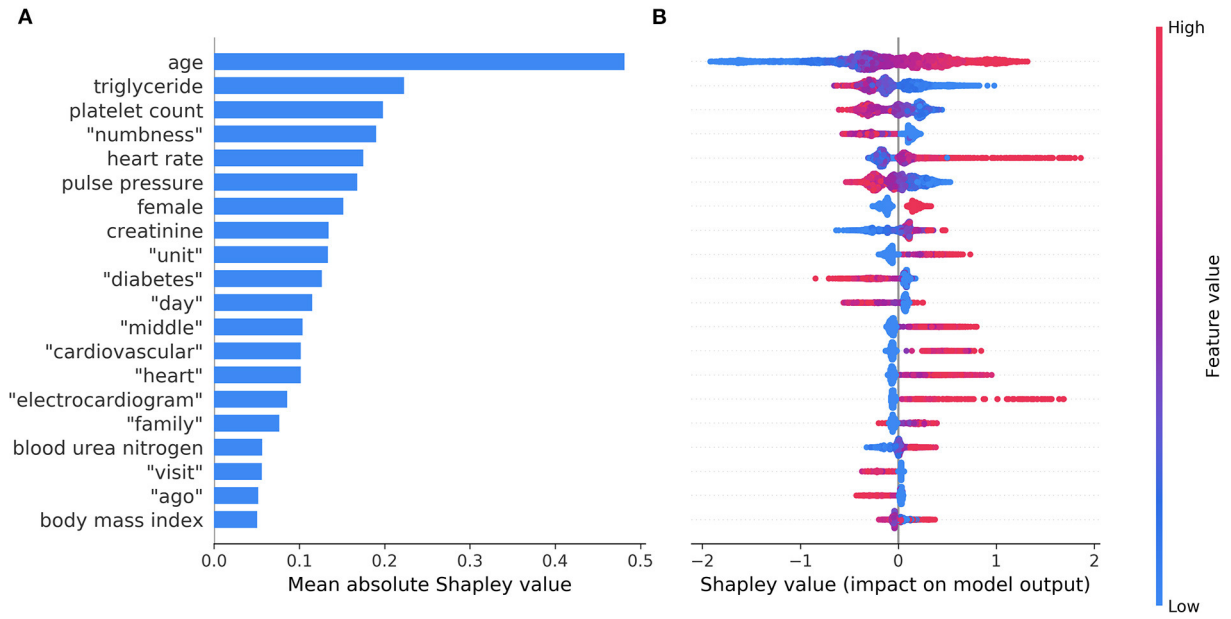


FIGURE 4
The top 20 most important features identified by the model based on both structured data and unstructured textual data. The mean absolute Shapley values that indicate the average impact on model output are shown in a bar chart **(A)**. The individual Shapley values for these features for each patient are depicted in a beeswarm plot **(B)**, where a dot's position on the x-axis denotes each feature's contribution to the model prediction for that patient. The color of the dot specifies the relative value of the corresponding feature.

considering the wide variety of data in EHRs, data-driven prediction modeling may allow identifying novel predictors in the context of insufficient prior knowledge of the real system (49). In this respect, ML is suitable for building complex models and analyzing noisy data such as that stored in EHRs (50). ML techniques were also applied to predict cardioembolic vs. non-cardioembolic stroke mechanism in patients with ESUS (51). Recently, deep learning techniques have been introduced to predict new-onset AF in the general population using structured primary care data or unstructured 12-lead ECG traces (52, 53).

Clinical applications and significance

Poststroke AF screening is essential for choosing the optimal strategy for secondary stroke prevention. However, to be resource efficient, extended ECG monitoring should be prioritized for patients at a high risk of NDAF. The developed ML model will be suited for assessing the risk of individual patients and assisting in personalized clinical decisions. Moreover, locally constructed prediction models may be more suitable for real-world clinical use than externally developed risk models (25). Since the prediction model was derived from EHRs, it is ideal to implement this model in the EHR as a decision support tool. With this tool, the calculation of risk estimates and the flagging of high-risk patients can be automated within the EHR, streamlining the process of risk stratification for poststroke AF screening.

Limitations

This study has several limitations. First, patients were traced through EHRs. Because patients might be diagnosed with AF outside the study hospital, some outcome misclassification was inevitable. Nevertheless, the frequent visits to the study hospital observed in this stroke population (>3 visits per month) might have alleviated this problem. Second, the diagnosis of AF was made in usual-care settings, where AF was detected almost exclusively by 12-lead ECG or 24-h Holter ECG. Advanced ECG monitoring *via* either implantable or external devices to detect subclinical or low-burden AF was not used. Consequently, the study findings are valid for relatively high-burden AF (54). Third, although data-driven ML modeling has its own advantages, the predictor-outcome relationships discovered from data does not mean causality. In other words, prediction accuracy should not be equated to causal validity (55). Fourth, as this is a single-site study, the generalizability of the study findings may be restricted. Variations in the terminology used in clinical documentation are to be expected across healthcare settings. However, the methods used here may allow other healthcare systems to develop their own customized versions of prediction models.

Conclusions

It is feasible to build an ML model to predict NDAF based on EHR data available at the time of hospital admission. Inclusion of information derived from clinical free text can significantly improve the model performance and may outperform risk scores developed using traditional statistical methods. These improvements may be due to both the modeling approach to delineate nonlinear decision boundaries and the use of textual features that help characterize nuances of disease presentation across patients. Despite these findings, further studies are required to confirm the approach's generalizability and the clinical usefulness of the prediction model.

Data availability statement

The data used in this study cannot be made available because of restrictions regarding the use of EMR data. Requests to access these datasets should be directed to Y-HH, yhhu@mgt.ncu.edu.tw.

Ethics statement

The studies involving human participants were reviewed and approved by Ditmanson Medical Foundation Chia-Yi Christian Hospital Institutional Review Board. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

Study concept and design: S-FS and Y-HH. Acquisition of data: S-FS and R-CP. Drafting of the manuscript: S-FS and K-LS. Study supervision: P-JL and Y-HH. Analysis and interpretation of data and critical revision of the manuscript for important intellectual content: all authors. All authors had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Funding

This research was supported in part by the Ditmanson Medical Foundation Chia-Yi Christian Hospital-National Chung Cheng University Joint Research Program (Grant number CYCH-CCU-2022-03).

Acknowledgments

The authors thank the help from the Clinical Data Center, Ditmanson Medical Foundation Chia-Yi Christian Hospital for providing administrative and technical support. This study is based in part on data from the Ditmanson Research Database (DRD) provided by Ditmanson Medical Foundation Chia-Yi Christian Hospital. The interpretation and conclusions contained herein do not represent the position of Ditmanson Medical Foundation Chia-Yi Christian Hospital. The authors also thank Ms. Li-Ying Sung for English language editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcvm.2022.941237/full#supplementary-material

References

1. Wang Y, Xu J, Zhao X, Wang D, Wang C, Liu L, et al. Association of hypertension with stroke recurrence depends on ischemic stroke subtype. *Stroke.* (2013) 44:1232–7. doi: 10.1161/strokeaha.111.000302

2. Kang K, Park TH, Kim N, Jang MU, Park S-S, Park J-M, et al. Recurrent stroke, myocardial infarction, and major vascular events during the first year after acute ischemic stroke: the multicenter prospective observational study about recurrence and its determinants after acute ischemic stroke I. *J Stroke Cerebrovasc Dis.* (2016) 25:656–64. doi: 10.1016/j.jstrokecerebrovasdis.2015.11.036

3. Hsieh C-Y, Wu DP, Sung S-F. Trends in vascular risk factors, stroke performance measures, and outcomes in patients with first-ever ischemic stroke in Taiwan between 2000 and 2012. *J Neurol Sci.* (2017) 378:80–4. doi: 10.1016/j.jns.2017.05.002

4. Lin B, Zhang Z, Mei Y, Wang C, Xu H, Liu L, et al. Cumulative risk of stroke recurrence over the last 10 years: a systematic review and meta-analysis. *Neurol Sci.* (2021) 42:61–71. doi: 10.1007/s10072-020-04797-5

5. Rücker V, Heuschmann PU, O'Flaherty M, Weingärtner M, Hess M, Sedlak C, et al. Twenty-year time trends in long-term case-fatality and recurrence

rates after ischemic stroke stratified by etiology. *Stroke.* (2020) 51:2778–85. doi: 10.1161/strokeaha.120.029972

6. Kolmos M, Christoffersen L, Kruuse C. Recurrent ischemic stroke – a systematic review and meta-analysis. *J Stroke Cerebrovasc Dis.* (2021) 30:105935. doi: 10.1016/j.jstrokecerebrovasdis.2021.105935

7. Flach C, Muruet W, Wolfe CDA, Bhalla A, Douiri A. Risk and secondary prevention of stroke recurrence. *Stroke.* (2020) 51:2435–44. doi: 10.1161/strokeaha.120.028992

8. Kamel H, Healey JS. Cardioembolic stroke. *Circ Res.* (2017) 120:514–26. doi: 10.1161/circresaha.116.308407

9. Go AS, Hylek EM, Phillips KA, Chang Y, Henault LE, Selby JV, et al. Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the anticoagulation and risk factors in atrial fibrillation (ATRIA) study. *JAMA.* (2001) 285:2370–5. doi: 10.1001/jama.285.18.2370

10. Yiin GSC, Howard DPJ, Paul NLM Li L, Luengo-Fernandez R, Bull LM, et al. Age-specific incidence, outcome, cost, and projected future burden of atrial fibrillation–related embolic vascular events. *Circulation.* (2014) 130:1236–44. doi: 10.1161/circulationaha.114.010942

11. Lip GYH, Hunter TD, Quiroz ME, Ziegler PD, Turakhia MP. Atrial fibrillation diagnosis timing, ambulatory ecg monitoring utilization, and risk of recurrent stroke. *Circ Cardiovasc Qual Outcomes.* (2017) 10:e002864. doi: 10.1161/circoutcomes.116.002864

12. Yaghi S, Bernstein RA, Passman R, Okin PM, Furie KL. Cryptogenic stroke. *Circ Res.* (2017) 120:527–40. doi: 10.1161/circresaha.116.308447

13. Sposato LA, Cipriano LE, Saposnik G, Vargas ER, Riccio PM, Hachinski V. Diagnosis of atrial fibrillation after stroke and transient ischaemic attack: a systematic review and meta-analysis. *Lancet Neurol.* (2015) 14:377–87. doi: 10.1016/s1474-4422(15)70027-x

14. Grond M, Jauss M, Hamann G, Stark E, Veltkamp R, Nabavi D, et al. Improved detection of silent atrial fibrillation using 72-hour holter ecg in patients with ischemic stroke. *Stroke.* (2013) 44:3357–64. doi: 10.1161/strokeaha.113.001884

15. Buck BH, Hill MD, Quinn FR, Butcher KS, Menon BK, Gulamhusein S, et al. Effect of implantable vs prolonged external electrocardiographic monitoring on atrial fibrillation detection in patients with ischemic stroke. *JAMA.* (2021) 325:2160–8. doi: 10.1001/jama.2021.6128

16. Noubiap JJ, Agbaedeng TA, Kamtchum-Tatuene J, Fitzgerald JL, Middeldorp ME, Kleinig T, et al. Rhythm monitoring strategies for atrial fibrillation detection in patients with cryptogenic stroke: A systematic review and meta-analysis. *Int J Cardiol Hear Vasc.* (2021) 34:100780. doi: 10.1016/j.ijcha.2021.100780

17. Jones NR, Taylor CJ, Hobbs FDR, Bowman L, Casadei B. Screening for atrial fibrillation: a call for evidence. *Eur Heart J.* (2019) 41:1075–85. doi: 10.1093/eurheartj/ehz834

18. Kishore AK, Hossain MJ, Cameron A, Dawson J, Vail A, Smith CJ. Use of risk scores for predicting new atrial fibrillation after ischemic stroke or transient ischemic attack—a systematic review. *Int J Stroke.* (2021) 174749302110458. doi: 10.1177/17474930211045880

19. Hsieh C-Y, Kao H-M, Sung K-L, Sposato LA, Sung S-F, Lin S-J. Validation of risk scores for predicting atrial fibrillation detected after stroke based on an electronic medical record algorithm: a registry-claims-electronic medical record linked data study. *Front Cardiovasc Med.* (2022) 9:888240. doi: 10.3389/fcvm.2022.888240

20. Ntaios G, Perlepe K, Lambrou D, Sirimarco G, Strambo D, Eskandari A, et al. Identification of patients with embolic stroke of undetermined source and low risk of new incident atrial fibrillation: The AF-ESUS score. *Int J Stroke.* (2020) 16:29–38. doi: 10.1177/1747493020925281

21. Muscari A, Barone P, Faccioli L, Ghinelli M, Trossello MP, Puddu GM, et al. Usefulness of the ACTEL score to predict atrial fibrillation in patients with cryptogenic stroke. *Cardiology.* (2020) 145:168–77. doi: 10.1159/000505262

22. Uphaus T, Weber-Krüger M, Grond M, Toenges G, Jahn-Eimermacher A, Jauss M, et al. Development and validation of a score to detect paroxysmal atrial fibrillation after stroke. *Neurology.* (2019) 92:e115–24. doi: 10.1212/wnl.0000000000006727

23. Healey JS, Wong JA. Pre-screening for atrial fibrillation using the electronic health record. *JACC Clin Electrophysiol.* (2019) 5:1342–3. doi: 10.1016/j.jacep.2019.08.019

24. Ding L, Liu C, Li Z, Wang Y. Incorporating artificial intelligence into stroke care and research. *Stroke.* (2020) 51:e351–4. doi: 10.1161/strokeaha.120.031295

25. Hulme OL, Khurshid S, Weng L-C, Anderson CD, Wang EY, Ashburner JM, et al. Development and validation of a prediction model for atrial fibrillation using electronic health records. *JACC Clin Electrophysiol.* (2019) 5:1331–41. doi: 10.1016/j.jacep.2019.07.016

26. Horng S, Sontag DA, Halpern Y, Jernite Y, Shapiro NI, Nathanson LA. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PLoS ONE.* (2017) 12:e0174708. doi: 10.1371/journal.pone.0174708

27. Weissman GE, Hubbard RA, Ungar LH, Harhay MO, Greene CS, Himes BE, et al. Inclusion of unstructured clinical text improves early prediction of death or prolonged ICU stay. *Crit Care Med.* (2018) 46:1125–32. doi: 10.1097/ccm.0000000000003148

28. Sung S, Chen C, Pan R, Hu Y, Jeng J. Natural language processing enhances prediction of functional outcome after acute ischemic stroke. *J Am Heart Assoc.* (2021) 10:e023486. doi: 10.1161/jaha.121.023486

29. Marini C, Santis FD, Sacco S, Russo T, Olivieri L, Totaro R, et al. Contribution of atrial fibrillation to incidence and outcome of ischemic stroke. *Stroke.* (2005) 36:1115–9. doi: 10.1161/01.str.0000166053.83476.4a

30. Arboix A, García-Eroles L, Massons JB, Oliveres M, Pujades R, Targa C. Atrial fibrillation and stroke: clinical presentation of cardioembolic versus atherothrombotic infarction. *Int J Cardiol.* (2000) 73:33–42.

31. Hsieh C-Y, Lee C-H, Sung S-F. Development of a novel score to predict newly diagnosed atrial fibrillation after ischemic stroke: The CHASE-LESS score. *Atherosclerosis.* (2020) 295:1–7. doi: 10.1016/j.atherosclerosis.2020.01.003

32. Sung S-F, Hsieh C-Y, Hu Y-H. Early prediction of functional outcomes after acute ischemic stroke using unstructured clinical text: retrospective cohort study. *JMIR Med Inform.* (2022) 10:e29806. doi: 10.2196/29806

33. Hsieh F-I, Lien L-M, Chen S-T, Bai C-H, Sun M-C, Tseng H-P, et al. Get with the guidelines-stroke performance indicators: surveillance of stroke care in the taiwan stroke registry. *Circulation.* (2010) 122:1116–23. doi: 10.1161/circulationaha.110.936526

34. Sposato LA, Chaturvedi S, Hsieh C-Y, Morillo CA, Kamel H. Atrial fibrillation detected after stroke and transient ischemic attack: a novel clinical concept challenging current views. *Stroke.* (2022) 53:e94–103. doi: 10.1161/strokeaha.121.034777

35. Mujtaba G, Shuib L, Idris N, Hoo WL, Raj RG, Khowaja K, et al. Clinical text classification research trends: systematic literature review and open issues. *Expert Syst Appl.* (2019) 116:494–520. doi: 10.1016/j.eswa.2018.09.034

36. Deng X, Li Y, Weng J, Zhang J. Feature selection for text classification: a review. *Multimed Tools Appl.* (2018) 78:3797–816. doi: 10.1007/s11042-018-6083-5

37. Ma S, Huang J. Penalized feature selection and classification in bioinformatics. *Brief Bioinform.* (2008) 9:392–403. doi: 10.1093/bib/bbn027

38. Branco P, Torgo L, Ribeiro RP. A Survey of predictive modeling on imbalanced domains. *ACM Comput Surv (CSUR).* (2016) 49:1–50. doi: 10.1145/2907070

39. Ogunleye AA, Qing-Guo W. XGBoost model for chronic kidney disease diagnosis. *IEEE/ACM Trans Comput Biol Bioinform.* (2019) 17:2131–40. doi: 10.1109/tcbb.2019.2911071

40. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. *Nat Mach Intell.* (2020) 2:56–67. doi: 10.1038/s42256-019-0138-9

41. LaValley MP. Logistic regression. *Circulation.* (2008) 117:2395–9. doi: 10.1161/circulationaha.106.682658

42. Schnabel RB, Sullivan LM, Levy D, Pencina MJ, Massaro JM, D'Agostino RB, et al. Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. *Lancet.* (2009) 373:739–45. doi: 10.1016/s0140-6736(09)60443-8

43. Liao J-N, Lim S-S, Chen T-J, Tuan T-C, Chen S-A, Chao T-F. Modified taiwan atrial fibrillation score for the prediction of incident atrial fibrillation. *Front Cardiovasc Med.* (2022) 8:805399. doi: 10.3389/fcvm.2021.805399

44. Chen Y-L, Wang H-T, Chen H-C, Liu W-H, Hsueh S, Chung W-J, et al. A risk stratification scoring system for new-onset atrial fibrillation after ischemic stroke. *Medicine.* (2020) 99:e20881. doi: 10.1097/md.0000000000020881

45. Ashburner JM, Wang X, Li X, Khurshid S, Ko D, Lipsanopoulos AT, et al. Re-CHARGE-AF: recalibration of the CHARGE-AF model for atrial fibrillation risk prediction in patients with acute stroke. *J Am Heart Assoc.* (2021) 10:e022363. doi: 10.1161/jaha.121.022363

46. Wang W, Alonso A, Soliman EZ, O'Neal WT, Calkins H, Chen LY, et al. Relation of resting heart rate to incident atrial fibrillation (From ARIC [atherosclerosis risk in communities] study). *Am J Cardiol.* (2018) 121:1169–76. doi: 10.1016/j.amjcard.2018.01.037

47. Drozdowska BA, Singh S, Quinn TJ. Thinking about the future: a review of prognostic scales used in acute stroke. *Front Neurol.* (2019) 10:274. doi: 10.3389/fneur.2019.00274

48. Orfanoudaki A, Chesley E, Cadisch C, Stein B, Nouh A, Alberts MJ, et al. Machine learning provides evidence that stroke risk is not linear: the non-linear Framingham stroke risk score. *PLoS ONE.* (2020) 15:e0232414. doi: 10.1371/journal.pone.0232414

49. Alaa AM, Bolton T, Angelantonio ED, Rudd JHF, Schaar M, van der. Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants. *PLoS ONE.* (2019) 14:e0213653. doi: 10.1371/journal.pone.0213653

50. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *New Engl J Med.* (2019) 380:1347–58. doi: 10.1056/nejmra1814259

51. Kamel H, Navi BB, Parikh NS, Merkler AE, Okin PM, Devereux RB, et al. Machine learning prediction of stroke mechanism in embolic strokes of undetermined source. *Stroke.* (2020) 51:e203–10. doi: 10.1161/strokeaha.120.029305

52. Raghunath S, Pfeifer JM, Ulloa-Cerna AE, Nemani A, Carbonati T, Jing L, et al. Deep neural networks can predict new-onset atrial fibrillation from the 12-lead ECG and help identify those at risk of atrial fibrillation–related stroke. *Circulation.* (2021) 143:1287–98. doi: 10.1161/circulationaha.120.047829

53. Nadarajah R, Wu J, Frangi AF, Hogg D, Cowan C, Gale C. Predicting patient-level new-onset atrial fibrillation from population-based nationwide electronic health records: protocol of FIND-AF for developing a precision medicine prediction model using artificial intelligence. *BMJ Open.* (2021) 11:e052887. doi: 10.1136/bmjopen-2021-052887

54. Aguilar M, Macle L, Deyell MW, Yao R, Hawkins N, Khairy P, et al. The influence of monitoring strategy on assessment of ablation success and post-ablation atrial fibrillation burden assessment: implications for practice and clinical trial design. *Circulation.* (2021) 145:21–30. doi: 10.1161/circulationaha.121.056109

55. Li J, Liu L, Le TD, Liu J. Accurate data-driven prediction does not mean high reproducibility. *Nat Mach Intell.* (2020) 2:13–5. doi: 10.1038/s42256-019-0140-2

frontiers | Frontiers in Cardiovascular Medicine

TYPE Original Research
PUBLISHED 02 August 2022
DOI 10.3389/fcvm.2022.946063

Check for updates

OPEN ACCESS

EDITED BY
Qingpeng Zihang,
City University of Hong Kong,
Hong Kong SAR, China

REVIEWED BY
Feng Jiang,
Fudan University, China
Zhiri Tang,
City University of Hong Kong,
Hong Kong SAR, China
Jun Lyu,
First Affiliated Hospital of Jinan University, China
Zetian Jia,
Second Hospital of Hebei Medical University, China

*CORRESPONDENCE
Yushan Mao
maoyushan@nbu.edu.cn

SPECIALTY SECTION
This article was submitted to General Cardiovascular Medicine, a section of the journal Frontiers in Cardiovascular Medicine

RECEIVED 17 May 2022
ACCEPTED 08 July 2022
PUBLISHED 02 August 2022

CITATION
Huang G, Jin Q, Tian X and Mao Y (2022) Development and validation of a carotid atherosclerosis risk prediction model based on a Chinese population.
Front. Cardiovasc. Med. 9:946063.
doi: 10.3389/fcvm.2022.946063

COPYRIGHT
© 2022 Huang, Jin, Tian and Mao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Development and validation of a carotid atherosclerosis risk prediction model based on a Chinese population

Guoqing Huang[1,2], Qiankai Jin[1,2], Xiaoqing Tian[1,2] and Yushan Mao[1]*

[1]Department of Endocrinology, The Affiliated Hospital of Medical School, Ningbo University, Ningbo, China, [2]School of Medicine, Ningbo University, Ningbo, China

Purpose: This study aimed to identify independent risk factors for carotid atherosclerosis (CAS) and construct and validate a CAS risk prediction model based on the Chinese population.

Methods: This retrospective study included 4,570 Chinese adults who underwent health checkups (including carotid ultrasound) at the Zhenhai Lianhua Hospital, Ningbo, China, in 2020. All the participants were randomly assigned to the training and validation sets at a ratio of 7:3. Independent risk factors associated with CAS were identified using multivariate logistic regression analysis. The least absolute shrinkage and selection operator combined with 10-fold cross-validation were screened for characteristic variables, and nomograms were plotted to demonstrate the risk prediction model. C-index and receiver operating characteristic curves, calibration plots, and decision curve analysis (DCA) were used to evaluate the risk model's discrimination, calibration, and clinical applicability.

Results: Age, body mass index, diastolic blood pressure, white blood cell count, mean platelet volume, alanine transaminase, aspartate transaminase, and gamma-glutamyl transferase were identified as independent risk factors for CAS. In the training, internal validation, and external validation sets, the risk model showed good discriminatory power with C-indices of 0.961 (0.953–0.969), 0.953 (0.939–0.967), and 0.930 (0.920–0.940), respectively, and excellent calibration. The results of DCA showed that the prediction model could be beneficial when the risk threshold probabilities were 1–100% in all sets. Finally, a network computer (dynamic nomogram) was developed to facilitate the physicians' clinical operations. The website is https://nbuhgq.shinyapps.io/DynNomapp/.

Conclusion: The development of risk models contributes to the early identification and prevention of CAS, which is important for preventing and reducing adverse cardiovascular and cerebrovascular events.

KEYWORDS

carotid atherosclerosis, independent risk factors, prediction model, early diagnosis, nomogram

Introduction

Atherosclerosis is a systemic atherosclerotic disease characterized by thickening, hardening, and loss of elasticity of the arterial walls and has become the pathological basis of many cardiovascular and cerebrovascular diseases, such as coronary heart disease and stroke (1). As part of systemic atherosclerosis, carotid atherosclerosis (CAS) has become an important window for observing systemic vascular health and the early risk of atherosclerosis due to its superficial location and ease of ultrasound manipulation (2, 3). In a global meta-analysis, approximately 28% of individuals in the general population (30–79 years) had carotid intimal thickening (4), potentially threatening people's health.

As the global aging process increases, the incidence of stroke has become a leading cause of morbidity and mortality (5). CAS is a major and potentially preventable cause of ischemic stroke. Some studies have shown that CAS is associated with 20–30% of strokes (6). Diagnostic examination modalities for CAS include carotid ultrasound, computed tomography, magnetic resonance imaging, and invasive angiography. Carotid ultrasound is the primary screening modality for CAS (6). Additionally, studies have shown that serum biomarkers such as interleukin, homocysteine, and adipokines contribute to the early diagnosis of CAS (7). The basic clinical treatment options for CAS are lifestyle modification, control of cardiovascular risk factors, antiplatelet aggregation (requiring adequate assessment of bleeding risk), and lipid-lowering (8, 9), whereas surgical intervention is required for severe carotid intimal thickening.

Reducing the risk factors associated with CAS is crucial in reducing adverse cardiovascular events occurrence. A meta-analysis by Ji et al. (10) showed that hyperlipidemia, hyperhomocysteinemia, hypertension (HTN), hyperuricemia, smoking history, metabolic syndrome, hypertriglyceridemia, diabetes mellitus (DM), and low-density lipoprotein (LDL) were significantly associated with CAS. In addition, a global meta-analysis showed that gender (male), smoking history, DM, HTN, and dyslipidemia are strongly associated with CAS (4). Identifying risk factors is a guide to targeting the prevention and control of CAS.

Although numerous risk factors associated with CAS are already known, few clinical risk prediction models related to CAS have been reported in Chinese populations. The nomogram is a common visual presentation tool for disease risk prediction models that is user-friendly and easy to understand. The current study aimed to develop and validate an analytical predictive model for CAS based on a Chinese population using statistical algorithms. This study will play an important role in the early identification and prevention of CAS in the Chinese population.

Materials and methods

Patients

A total of 4,738 adults (19–93 years) who underwent health checkups (including carotid ultrasound) at Zhenhai Lianhua Hospital, Ningbo, China, in 2020 were initially included in this study. Relevant information about the participants was obtained through the hospital's electronic medical record system. Those with serious missing information (exceeding 20% of the total) were excluded, and those with less missing information (less than 20% of the total) were filled by multiple interpolations. Ultimately, 4,570 participants were included in this study. CAS was defined as an increase in carotid intima-media thickness of $\geq$1 mm or plaque formation (11). CAS diagnosis was based on carotid ultrasound results, recorded independently by 2 ultrasound physicians. The external validation dataset (2,791) was obtained from the 2015 health checkups (different from 2020). This study was approved by the ethics committee of the Affiliated Hospital of Medical School, Ningbo University, Ningbo, China (KY20191114). A flowchart of the participants is shown in **Figure 1**.

Clinical baseline data

The methods and procedures for testing the clinical baseline data were based on previous studies (12–14). Gender, age, body mass index (BMI), systolic blood pressure (SBP), diastolic blood pressure (DBP), heart rate (HR), DM, HTN, Drinking and Smoking history, white blood cell count (WBC), neutrophil count, eosinophil count, basophil count, lymphocyte count, red blood cell count (RBC), hemoglobin, red blood cell distribution width, mean red blood cell volume, platelet count, platelet distribution width (PDW), mean platelet volume (MPV), alanine aminotransferase (ALT), aspartate aminotransferase (AST), total bilirubin, direct bilirubin, indirect bilirubin, total protein, albumin, globulin, gamma-glutamyl transpeptidase (GGT), total bile acids, blood urea nitrogen, serum creatinine, uric acid, fasting blood glucose, total cholesterol (TC), triglycerides, high-density lipoprotein, LDL, apolipoprotein A, apolipoprotein B (Apo-B), thyroid stimulating hormone, total triiodothyronine, total tetraiodothyronine, free triiodothyronine, and free tetraiodothyronine were obtained from the hospital electronic medical record system.

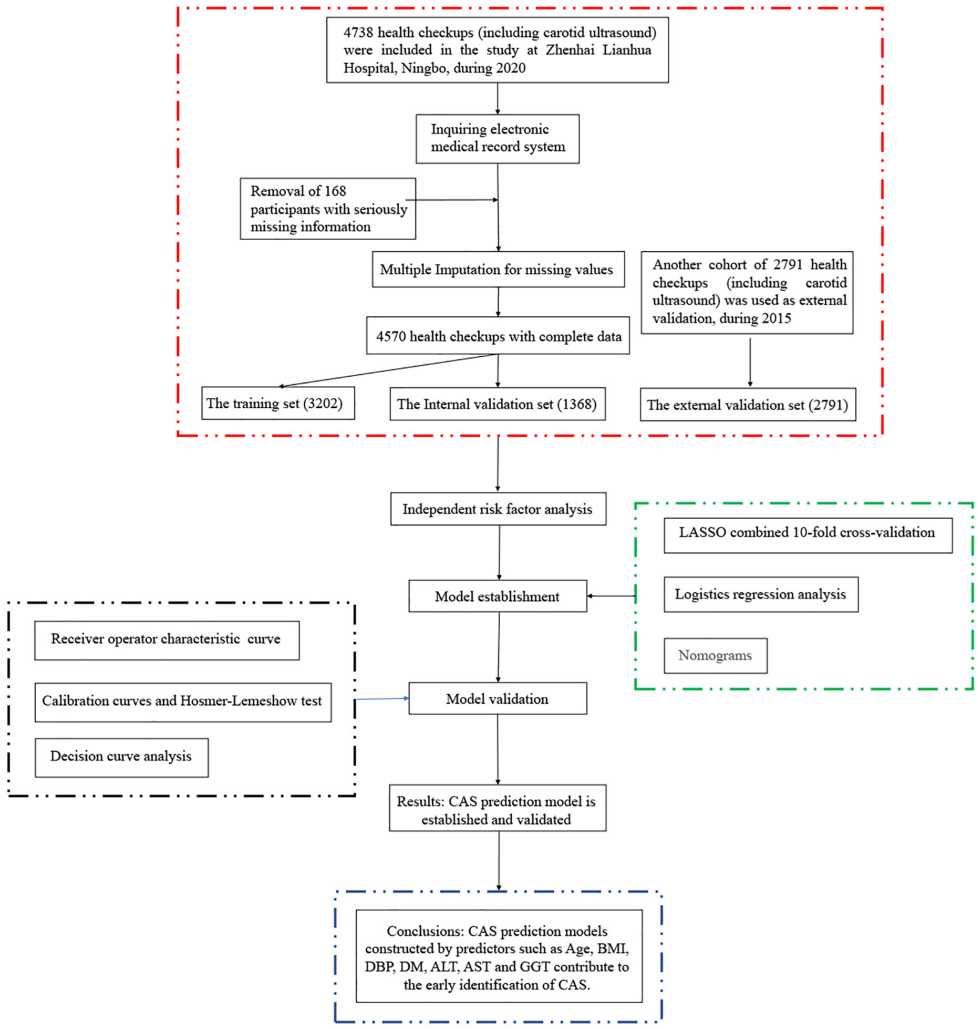


FIGURE 1
Flowchart of the participants.

Statistical analysis

This study enrolled 4,750 adults, and the measurement data were expressed as mean ± standard deviation; the count data were expressed as counts (%). Statistical analysis was performed with R software (version 4.1.2).[1] All tests were two-tailed, and a *P*-value <0.05 was considered statistically significant.

Participants were randomly assigned to the training and internal validation sets at a certain ratio (7:3) (15). Independent risk factors were identified using multivariate logistic regression analysis. The least absolute shrinkage and selection operator (LASSO) combined with 10-fold cross-validation was used to screen for characteristic variables associated with CAS. The visual presentation of the risk prediction model was displayed using a nomogram. Risk prediction models were evaluated in terms of discrimination (C-index and receiver operating characteristic (ROC) curves), calibration ability (Hosmer–Lemeshow test and calibration curves), and clinical applicability [decision curve analysis (DCA)] in the training, internal validation, and external validation sets, respectively.

1 https://www.R-project.org

Results

Characteristics of the study population

This study enrolled 4,570 participants (3,010 males and 1,560 females). The detection rate of CAS by carotid ultrasound in 2020 was 83.4% (3,813) compared with 73.1% (2,039) in

10.3389/fcvm.2022.946063

TABLE 1 Characteristics of participants in different cohorts.

	Training set	Internal validation set	External validation set
N	3202	1368	2791
Sex (male)	2116 (66.08)	894 (65.35)	1903 (68.18)
Age, years	60.30 (14.33)	60.46 (14.21)	59.90 (13.36)
BMI, kg/m^2	23.57 (3.05)	23.52 (2.97)	23.12 (2.83)
SBP, mmHg	136.06 (18.28)	136.62 (18.84)	131.40 (18.61)
DBP, mmHg	79.37 (11.16)	79.64 (11.56)	76.76 (11.32)
HR, times/min	78.53 (12.60)	78.87 (12.51)	78.88 (12.30)
DM	34 (1.06)	18 (1.32)	124 (4.44)
HTN	99 (3.09)	39 (2.95)	450 (16.12)
Drinking history	192 (6.00)	94 (6.87)	849 (30.42)
Smoking history	175 (5.47)	83 (6.07)	679 (24.33)
WBC, 10^9/L	5.93 (1.52)	5.97 (1.51)	5.98 (1.52)
NEC, 10^9/L	3.43 (1.14)	3.48 (1.15)	3.42 (1.14)
EOC, 10^9/L	0.15 (0.15)	0.14 (0.13)	0.15 (0.14)
BAC, 10^9/L	0.02 (0.01)	0.02 (0.01)	0.01 (0.01)
LYC, 10^9/L	1.96 (0.61)	1.96 (0.60)	2.05 (0.61)
RBC, 10^12/L	4.79 (0.50)	4.82 (0.50)	4.61 (0.45)
HGB, g/L	145.59 (14.83)	146.10 (15.04)	140.14 (14.16)
RDW, %	12.75 (0.92)	12.78 (1.00)	12.83 (0.87)
MCV, fl	92.89 (5.30)	92.55 (5.65)	92.94 (4.98)
PLT, 10^9/L	221.85 (58.54)	221.56 (55.23)	205.81 (52.34)
PDW, %	13.58 (2.33)	13.53 (2.30)	12.43 (1.99)
MPV, fl	11.00 (0.99)	10.97 (0.98)	10.40 (0.93)
ALT, U/L	24.18 (17.38)	24.52 (17.93)	19.01 (12.69)
AST, U/L	25.79 (12.66)	26.42 (14.49)	22.77 (10.01)
T-BIL, μmol/L	14.71 (6.12)	14.58 (6.06)	14.77 (6.34)
D-BIL, μmol/L	3.55 (1.77)	3.54 (1.81)	4.56 (1.60)
I-BIL, μmol/L	11.16 (4.84)	11.04 (4.73)	10.20 (4.94)
TP, g/L	74.21 (4.03)	74.24 (4.08)	72.52 (4.32)
ALB, g/L	44.79 (2.22)	44.76 (2.22)	44.99 (2.62)
GLOB, g/L	29.42 (3.74)	29.48 (3.56)	27.53 (4.18)
GGT, U/L	35.99 (46.92)	36.55 (44.48)	31.12 (35.03)
TBA, μmol/L	4.01 (3.93)	4.16 (5.17)	2.70 (2.90)
BUN, mmol/L	5.36 (1.54)	5.31 (1.55)	5.17 (1.52)
Scr, μmol/L	73.26 (36.99)	72.45 (36.39)	61.58 (18.63)
UA, μmol/L	356.77 (86.33)	357.98 (86.30)	326.81 (80.18)
FBG, mmol/L	5.90 (1.38)	5.90 (1.40)	5.37 (1.15)
TC, mmol/L	5.22 (1.12)	5.23 (1.09)	4.86 (0.98)
TG, mmol/L	1.57 (1.10)	1.56 (1.13)	1.35 (0.91)
HDL, mmol/L	1.28 (0.39)	1.31 (0.39)	1.52 (0.28)
LDL, mmol/L	2.99 (0.89)	2.97 (0.87)	2.70 (0.73)
Apo-A, g/L	1.45 (0.27)	1.47 (0.27)	1.54 (0.34)
Apo-B, g/L	0.99 (0.28)	0.98 (0.28)	0.70 (0.18)
TSH, mIU/L	2.11 (2.09)	2.14 (2.25)	2.09 (1.84)
TT3, nmol/L	1.59 (0.32)	1.60 (0.34)	1.64 (0.25)
TT4, nmol/L	116.70 (20.01)	117.75 (22.91)	113.37 (19.08)
FT3, pmol/L	5.15 (0.64)	5.17 (0.66)	4.56 (0.49)
FT4, pmol/L	11.19 (1.68)	11.24 (1.68)	11.10 (1.54)

BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; HR, heart rate; DM, diabetes mellitus; HTN, hypertension; WBC, white blood cell count; NET, neutrophil count; EOC, eosinophil count; BAC, basophil count; LYC, lymphocyte count; RBC, red blood cell count; HGB, hemoglobin; RDW, red blood cell distribution width; MCV, mean red blood cell volume; PLT, platelet count; PDW, platelet distribution width; MPV, mean platelet volume; ALT, alanine aminotransferase; AST, aspartate aminotransferase; T-BIL, total bilirubin; D-BIL, direct bilirubin; I-BIL, indirect bilirubin; TP, total protein; ALB, albumin; GLOB, globulin; GGT, gamma-glutamyl transpeptidase; TBA, total bile acids; BUN, blood urea nitrogen; Scr, serum creatinine; UA, uric acid; FBG, fasting blood glucose; TC, total cholesterol; TG, triglycerides; HDL, high-density lipoprotein; LDL, low-density lipoprotein; Apo-A, apolipoprotein -A; Apo-B, apolipoprotein-B; TSH, thyroid stimulating hormone; TT3, total triiodothyronine; TT4, total tetraiodothyronine; FT3, free triiodothyronine; FT4, free tetraiodothyronine.

10.3389/fcvm.2022.946063

TABLE 2 Univariate analysis of carotid atherosclerosis.

	Overall	HC	CAS	*P*-value
N	4570	757	3813	
Sex (male)	3010 (65.86)	459 (60.63)	2551 (66.90)	0.001
Age, years	60.35 (14.29)	40.08 (10.91)	64.37 (11.11)	<0.001
BMI, kg/m^2	23.55 (3.03)	21.84 (2.54)	23.89 (3.00)	<0.001
SBP, mmHg	136.22 (18.45)	120.52 (13.91)	139.34 (17.63)	<0.001
DBP, mmHg	79.45 (11.28)	73.82 (9.99)	80.57 (11.19)	<0.001
HR, times/min	78.63 (12.57)	79.50 (12.48)	78.46 (12.58)	0.036
DM	52 (1.13)	0 (0.0)	52 (1.36)	<0.001
HTN	138 (3.01)	3 (0.40)	135 (3.54)	<0.001
Drinking history	286 (6.26)	83 (10.96)	203 (5.32)	<0.001
Smoking history	258 (5.64)	69 (9.11)	189 (4.96)	<0.001
WBC, 10^9/L	5.94 (1.52)	5.69 (1.40)	5.99 (1.54)	<0.001
NEC, 10^9/L	3.45 (1.15)	3.24 (1.07)	3.49 (1.16)	<0.001
EOC, 10^9/L	0.15 (0.14)	0.14 (0.11)	0.15 (0.14)	0.023
BAC, 10^9/L	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)	0.545
LYC, 10^9/L	1.96 (0.61)	1.95 (0.55)	1.96 (0.62)	0.536
RBC, 10^12/L	4.80 (0.50)	4.88 (0.49)	4.79 (0.50)	<0.001
HGB, g/L	145.75 (14.90)	145.41 (15.45)	145.81 (14.78)	0.496
RDW, %	12.76 (0.95)	12.65 (1.09)	12.78 (0.91)	<0.001
MCV, fl	92.79 (5.41)	91.08 (5.49)	93.13 (5.33)	<0.001
PLT, 10^9/L	221.76 (57.56)	235.50 (54.53)	219.03 (57.76)	<0.001
PDW, %	13.56 (2.32)	13.80 (2.46)	13.51 (2.29)	0.002
MPV, fl	10.99 (0.99)	11.11 (1.02)	10.97 (0.98)	<0.001
ALT, U/L	24.28 (17.55)	19.60 (13.17)	25.21 (18.15)	<0.001
AST, U/L	25.98 (13.24)	22.08 (7.91)	26.75 (13.93)	<0.001
T-BIL, μmol/L	14.67 (6.11)	14.36 (6.16)	14.73 (6.09)	0.122
D-BIL, μmol/L	3.54 (1.78)	3.51 (1.74)	3.55 (1.79)	0.581
I-BIL, μmol/L	11.13 (4.80)	10.85 (4.77)	11.18 (4.81)	0.078
TP, g/L	74.22 (4.04)	74.16 (3.68)	74.23 (4.11)	0.68
ALB, g/L	44.78 (2.22)	45.76 (2.10)	44.59 (2.19)	<0.001
GLOB, g/L	29.44 (3.69)	28.41 (3.20)	29.64 (3.74)	<0.001
GGT, U/L	36.16 (46.20)	24.43 (22.53)	38.48 (49.24)	<0.001
TBA, μmol/L	4.06 (4.34)	3.34 (2.82)	4.20 (4.57)	<0.001
BUN, mmol/L	5.34 (1.54)	4.87 (1.16)	5.44 (1.59)	<0.001
Scr, μmol/L	73.01 (36.81)	68.14 (12.39)	73.98 (39.85)	<0.001
UA, μmol/L	357.13 (86.31)	332.13 (82.22)	362.10 (86.25)	<0.001
FBG, mmol/L	5.90 (1.39)	5.21 (0.76)	6.04 (1.44)	<0.001
TC, mmol/L	5.22 (1.11)	5.01 (0.98)	5.26 (1.13)	<0.001
TG, mmol/L	1.56 (1.11)	1.23 (1.03)	1.63 (1.11)	<0.001
HDL, mmol/L	1.29 (0.39)	1.38 (0.38)	1.28 (0.39)	<0.001
LDL, mmol/L	2.98 (0.89)	2.83 (0.77)	3.01 (0.91)	<0.001
Apo-A, g/L	1.46 (0.27)	1.46 (0.25)	1.46 (0.27)	0.794
Apo-B, g/L	0.98 (0.28)	0.89 (0.24)	1.00 (0.28)	<0.001
TSH, mIU/L	2.12 (2.14)	1.88 (1.03)	2.16 (2.30)	0.001
TT3, nmol/L	1.59 (0.32)	1.63 (0.34)	1.58 (0.32)	<0.001
TT4, nmol/L	117.01 (20.92)	111.20 (18.72)	118.17 (21.15)	<0.001
FT3, pmol/L	5.16 (0.65)	5.35 (0.69)	5.12 (0.63)	<0.001
FT4, pmol/L	11.21 (1.68)	11.28 (1.51)	11.19 (1.71)	0.218

BMI, body mass index; SBP, systolic blood pressure; DBP, diastolic blood pressure; HR, heart rate; DM, diabetes mellitus; HTN, hypertension; WBC, white blood cell count; NET, neutrophil count; EOC, eosinophil count; BAC, basophil count; LYC, lymphocyte count; RBC, red blood cell count; HGB, hemoglobin; RDW, red blood cell distribution width; MCV, mean red blood cell volume; PLT, platelet count; PDW, platelet distribution width; MPV, mean platelet volume; ALT, alanine aminotransferase; AST, aspartate aminotransferase; T-BIL, total bilirubin; D-BIL, direct bilirubin; I-BIL, indirect bilirubin; TP, total protein; ALB, albumin; GLOB, globulin; GGT, gamma-glutamyl transpeptidase; TBA, total bile acids; BUN, blood urea nitrogen; Scr, serum creatinine; UA, uric acid; FBG, fasting blood glucose; TC, total cholesterol; TG, triglycerides; HDL, high-density lipoprotein; LDL, low-density lipoprotein; Apo-A, apolipoprotein -A; Apo-B, apolipoprotein-B; TSH, thyroid stimulating hormone; TT3, total triiodothyronine; TT4, total tetraiodothyronine; FT3, free triiodothyronine; FT4, free tetraiodothyronine.

10.3389/fcvm.2022.946063

TABLE 3 Multivariate logistic regression analysis.

Variable	Coefficients	Odds ratio (95% CI)	*P*-value
Age	0.222	1.249 (1.225–1.274)	<0.001
BMI	0.200	1.221 (1.156–1.291)	<0.001
HR	−0.007	0.993 (0.982–1.004)	0.22786
SBP	0.007	1.007 (0.994–1.021)	0.31200
DBP	0.027	1.029 (1.008–1.048)	<0.01
Drinking history	−0.412	0.662 (0.393–1.117)	0.12159
Smoking history	0.305	1.357 (0.782–2.390)	0.28334
WBC	0.114	1.121 (1.018–1.236)	<0.05
RBC	−0.009	0.991 (0.684–1.435)	0.96104
MCV	0.006	1.006 (0.978–1.035)	0.67506
MPV	−0.170	0.844 (0.738–0.966)	<0.05
ALT	0.036	1.036 (1.020–1.053)	<0.001
AST	−0.037	0.964 (0.947–0.984)	<0.001
I-BIL	−0.002	0.998 (0.970–1.028)	0.908005
GLOB	0.024	1.024 (0.982–1.069)	0.26725
GGT	0.007	1.007 (1.003–1.013)	<0.05
FBG	0.098	1.112 (0.979–1.289)	0.12646
LDL	−0.200	0.765 (0.503–1.325)	0.41725
Apo-B	1.531	4.623 (0.922–23.683)	0.06452
TT3	0.198	1.219 (0.782–1.915)	0.38678

BMI, body mass index; HR, heart rate; SBP, systolic blood pressure; DBP, diastolic blood pressure; WBC, white blood cell count; RBC, red blood cell count; MCV, mean red blood cell volume; MPV, mean platelet volume; ALT, alanine aminotransferase; AST, aspartate aminotransferase; I-BIL, indirect bilirubin; GLOB, globulin; GGT, gamma-glutamyl transpeptidase; FBG, fasting blood glucose; LDL, low-density lipoprotein; Apo-B, apolipoprotein-B; TT3, total triiodothyronine.

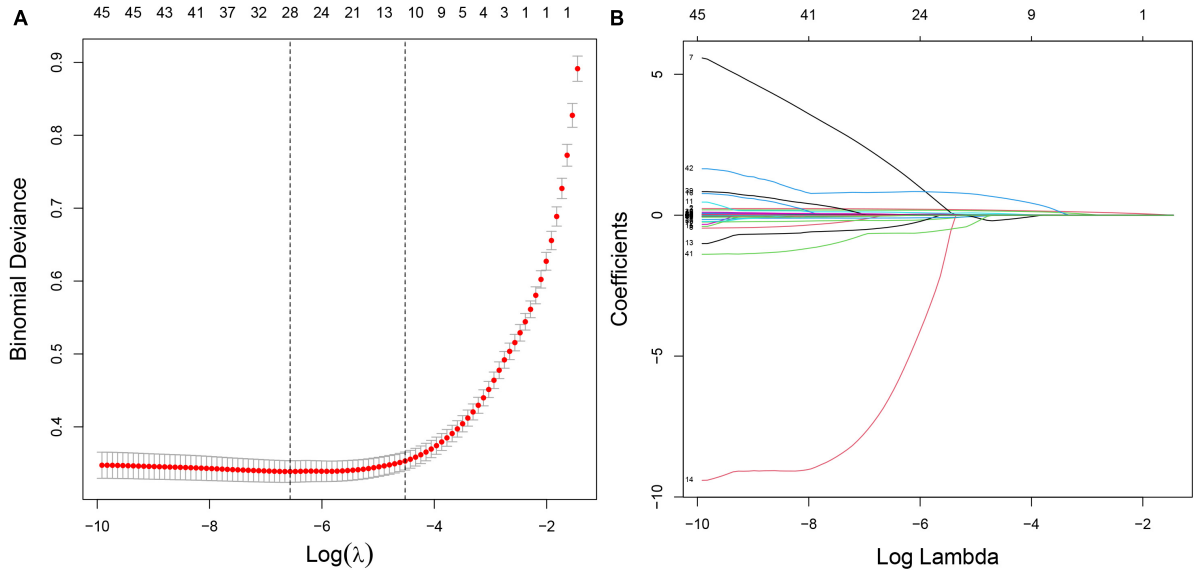


FIGURE 2
Screening of characteristic variables using the LASSO regression analysis. **(A)** The selection of the best parameter (lambda) in the LASSO model uses 10-fold cross-validation with the lowest standard. The relationship curve between partial likelihood deviation (binomial deviation) and log (lambda) was plotted. Dotted vertical lines were drawn at the optimal values by using the minimum criteria and the 1 SE of the minimum criteria (the 1 SE criteria). **(B)** LASSO coefficient profiles of the 28 characteristic variables. A coefficient profile plot was produced against the log (lambda) sequence. LASSO, least absolute shrinkage and selection operator; SE, standard error.

2015 (external validation set) at Zhenhai Lianhua Hospital, Ningbo, China. The total study population was divided into training (3,202) and internal validation (1,358) sets according to a 7:3 ratio, while an external validation set (2,791) was introduced to ensure the stability of the model (Table 1). Univariate analysis showed a higher proportion of males in the CAS group than in the healthy control (HC) group (66.90% vs. 60.63%, $P < 0.001$). The mean age of the HC and CAS groups was 40.08 and 64.37 years, respectively. The CAS group had higher proportions of DM, HTN, smoking history, and drinking history than the HC group. In addition, there were differences between the 2 groups in terms of routine blood tests, liver function, lipids, glucose, and thyroid function. The baseline information of the study cohort is shown in Table 2.

Independent risk factors

Based on univariate analysis (Table 1), we selected candidate variables with $P < 0.1$ to screen independent risk factors. Covariance analysis among candidate variables showed variance inflation factors (VIF) > 10 for PDW, MPV, TC, LDL, and Apo-B. When the Akaike information criterion was −12180.32, the stepwise backward logistic regression analysis found the optimal model to include 20 variables such as age, BMI, HR, SBP, and DBP (VIF < 10). Furthermore, we constructed logistic regression equations.

$$\log\left[\frac{P\left(\widehat{\text{Carotid\_atherosclerosis}}=1\right)}{1-P\left(\widehat{\text{Carotid\_atherosclerosis}}=1\right)}\right] \quad (1)$$

$$\begin{aligned}&-19.05+0.22\,(\text{Age})+0.2\,(\text{BMI})-\\&0.01\,(\text{HR})+0.01\,(\text{SBP})+0.03\,(\text{DBP})-\\&0.41\left(\text{Drinking\_history}_{\text{Yes}}\right)+\\&0.31\left(\text{Smoking\_history}_{\text{Yes}}\right)+0.11\,(\text{WBC})-\\&0.01\,(\text{RBC})+0.01\,(\text{MRBCV})-0.17\,(\text{MPV})+\\&0.04\,(\text{ALT})-0.04\,(\text{AST})+0\,(\text{I\_BIL})+\\&0.02\,(\text{GLOB})+0.01\,(\text{GGT})+0.11\,(\text{FBG})-\\&0.2\,(\text{LDL})+1.53\,(\text{Apo\_B})+0.2\,(\text{TT3})\end{aligned}$$

Where BMI, body mass index; HR, heart rate; SBP, systolic blood pressure; DBP, diastolic blood pressure; WBC, white blood cell count; RBC, red blood cell count; MRBCV, mean red blood cell velocities; MPV, mean platelet volume; ALT, alanine transaminase; AST, aspartate transaminase; I_BIL, indirect bilirubin; GLOB, globulin; GGT, gamma-glutamyl transpeptidase; FBG, fasting blood glucose; LDL, low-density lipoprotein; Apo_B, apolipoprotein B; TT3, total triiodothyronine.

TABLE 4 Coefficients and lambda.min value of the LASSO regression.

Variables	Coefficients	Lambda.min
Age	0.2196	0.0014
BMI	0.1777	
SBP	0.0097	
DBP	0.0293	
HR	−0.0040	
DM	1.8408	
Drinking history	−0.1880	
WBC	0.1084	
NEC	0.0220	
EOC	−0.4282	
BAC	−6.5259	
PLT	0.0014	
MPV	−0.0999	
ALT	0.0241	
AST	−0.0176	
I-BIL	0.0032	
ALB	−0.0096	
GLOB	0.0248	
GGT	0.0051	
Scr	0.0013	
UA	0.0020	
FBG	0.0118	
TG	0.0242	
Apo-A	−0.6405	
Apo-B	0.8163	
TT3	0.0375	
TT4	0.0082	
FT4	−0.0283	

BMI, body mass index; HR, heart rate; SBP, systolic blood pressure; DBP, diastolic blood pressure; DM, diabetes mellitus; WBC, white blood cell count; NET, neutrophil count; EOC, eosinophil count; BAC, basophil count; PLT, platelet count; MPV, mean platelet volume; ALT, alanine aminotransferase; AST, aspartate aminotransferase; I-BIL, indirect bilirubin; ALB, albumin; GLOB, globulin; GGT, gamma-glutamyl transpeptidase; Scr, serum creatinine; FBG, fasting blood glucose; TG, triglycerides; LDL, low-density lipoprotein; Apo-A, apolipoprotein -A; Apo-B, apolipoprotein-B; TT3, total triiodothyronine; TT4, total tetraiodothyronine; FT4, free tetraiodothyronine.

Independent risk factors associated with CAS were also identified, including age, BMI, WBC count, MPV, ALT, AST, and GGT (Table 3).

Construction of predictive models

In the training set, 28 non-zero characteristic variables were screened using LASSO regression analysis (Figure 2 and Table 4). Low-weight variables (points < 20) were removed from the risk prediction model. Finally, we selected age, BMI, DBP, DM, ALT, AST, and GGT for model construction (Figure 3). In addition, we developed a web version of the

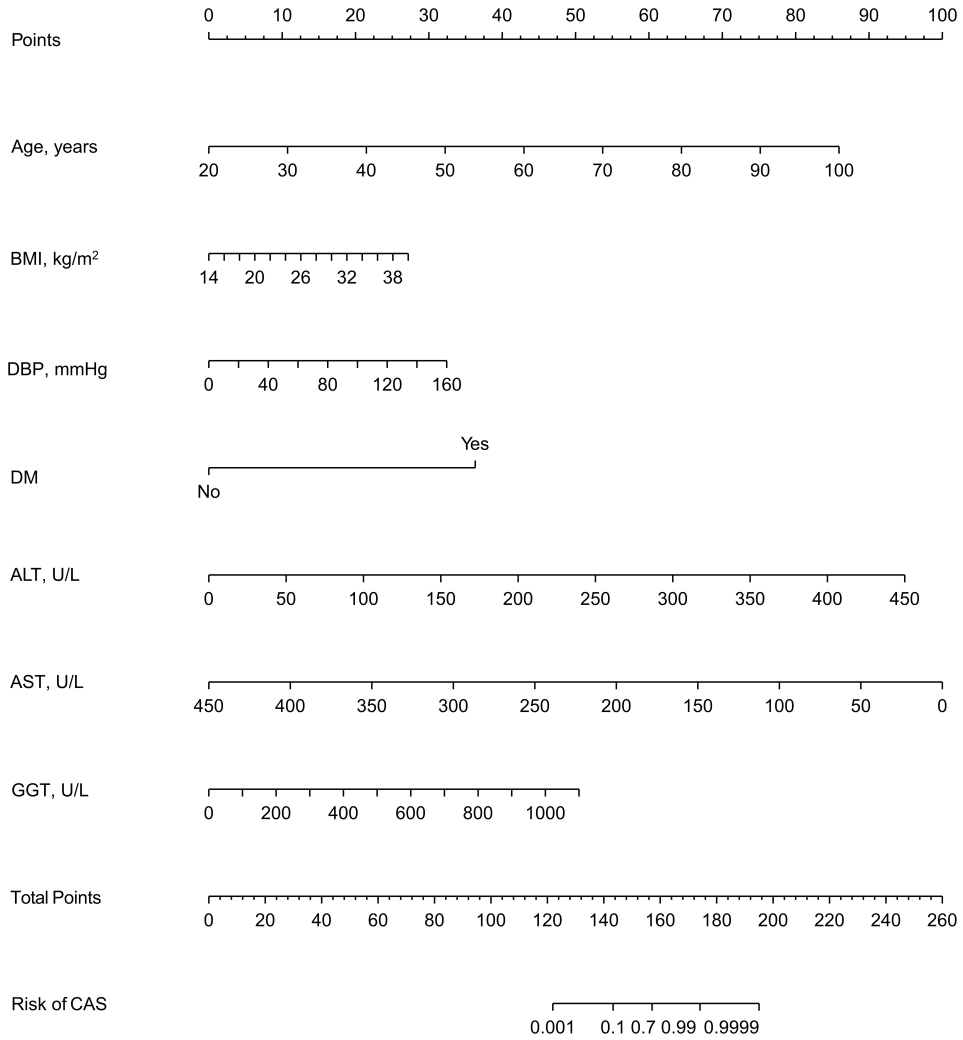


FIGURE 3
A nomogram for predicting the probability of the development of CAS. The nomogram is used by scoring each variable on its corresponding score scale. The scores for all variables are then summed up to obtain the total score, and a vertical line is drawn from the total point row to indicate the estimated probability of the development of CAS. carotid atherosclerosis, CAS; body mass index, BMI; systolic blood pressure, SBP; diastolic blood pressure, DBP; diabetes mellitus, DM; alanine aminotransferase, ALT; aspartate aminotransferase, AST; gamma-glutamyl transpeptidase, GGT.

dynamic nomogram (**Figure 4**) for ease of daily use. The URL is https://nbuhgq.shinyapps.io/DynNomapp/.

Validation of predictive models

The C-index and area under the ROC curve (AUC) were used to assess the discriminatory ability of the risk model. The C-index was 0.961 (0.953–0.969), 0.953 (0.939–0.967), and 0.930 (0.920–0.940) in the training, internal validation, and external validation sets, respectively (**Table 5**), whereas the AUC was 0.961, 0.953, and 0.930, respectively (**Figure 5**).

From the calibration curves, we observed that the predicted values were very close to the theoretical values in the training, internal validation, and external validation sets, showing a very good fit (**Figure 6**), which was further confirmed by the Hosmer-Lemeshow test ($P > 0.05$) (**Table 6**).

Decision curve analysis is often used to assess the clinical applicability of risk-prediction models. **Figure 7** shows that the risk threshold probabilities for the training, internal validation, and external validation sets were 1–100%, suggesting that the risk prediction model is beneficial in this range.

Discussion

With the popularity of ultrasound in health checkups, CAS detection rate is gradually increasing. The carotid ultrasound

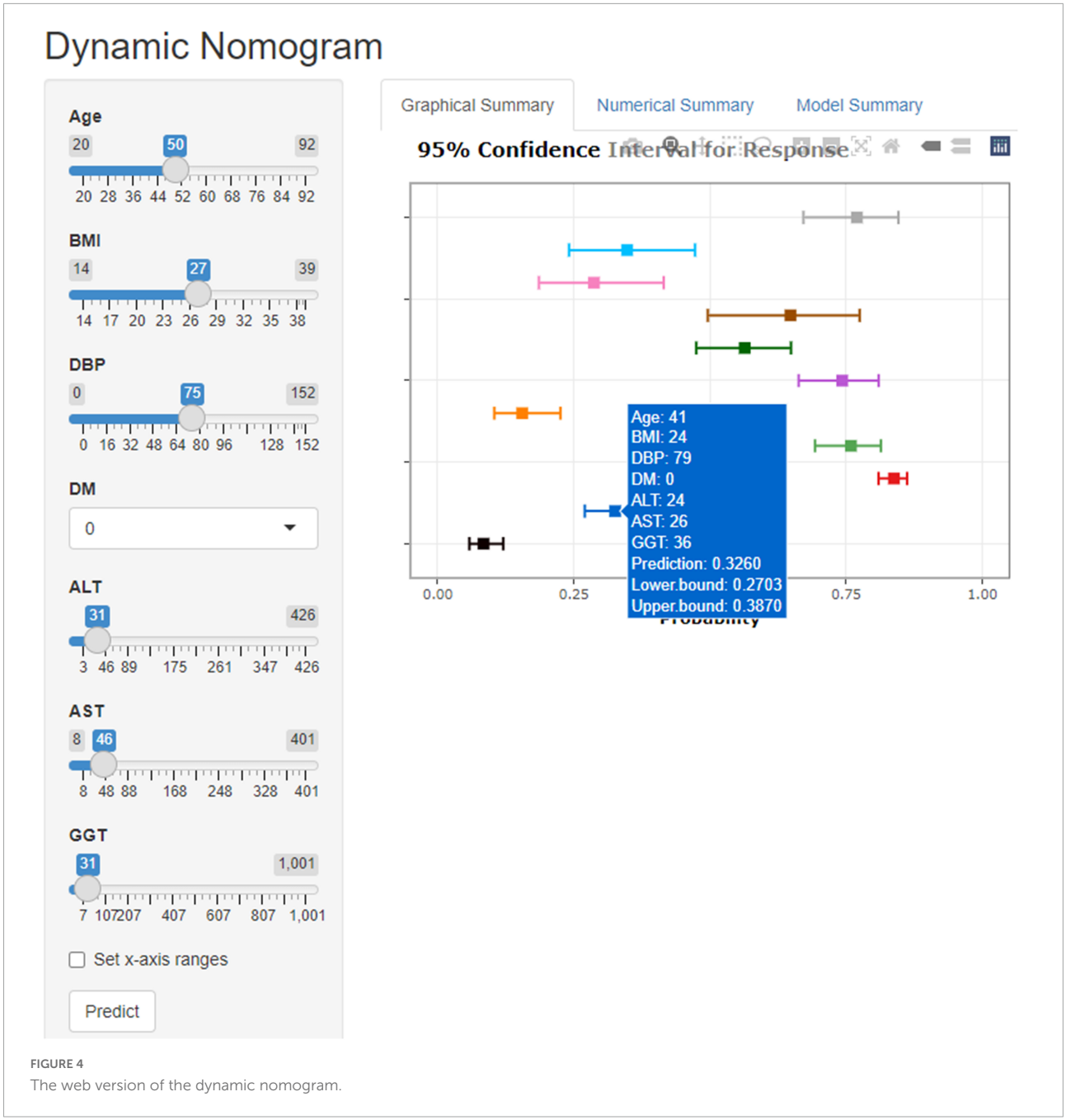


FIGURE 4
The web version of the dynamic nomogram.

TABLE 5 C-index in the study cohort.

	C-index (95%CI)	Dxy	aDxy	SD	Z	*P*-value
Training set	0.961 (0.953–0.969)	0.921	0.921	0.008	112.93	0
Internal validation set	0.953 (0.939–0.967)	0.905	0.905	0.028	23.15	0
External validation set	0.930 (0.920–0.940)	0.860	0.860	0.010	84.51	0

CAS detection rate in 2020 was 83.4% (3,813) compared to 73.1% (2,039) in 2015 (external validation set) at Zhenhai Lianhua Hospital in Ningbo, China. Age (odds ratio [OR]: 1.249 [1.225–1.274], $P < 0.001$), BMI (OR: 1.221 [1.156–1.291], $P < 0.001$), DBP (OR: 1.029 [1.008–1.048], $P < 0.01$), WBC (OR: 1.121 [1.018–1.236], $P < 0.05$), MPV (OR: 0.844 [0.738–0.966],

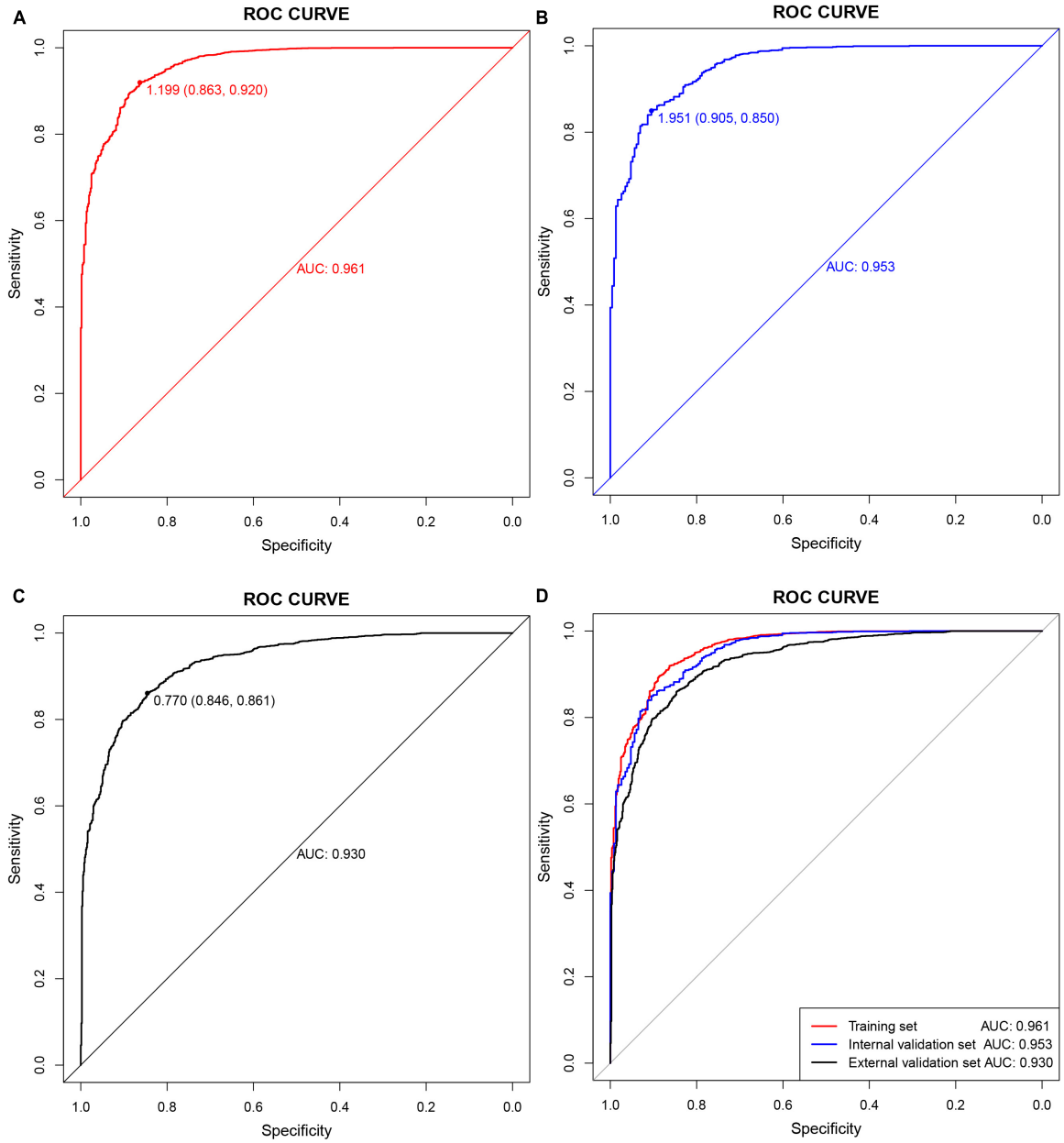


FIGURE 5
Receiver operating characteristic (ROC) curves. **(A)** Training set; **(B)** internal validation set; **(C)** external validation set; **(D)** all sets.

$P < 0.05$), ALT (OR: 1.036 [1.020–1.053], $P < 0.001$), AST (OR: 0.964 [0.947–0.984], $P < 0.001$), GGT (OR: 1.007 [1.003–1.013], $P < 0.05$) were identified as independent risk factors for CAS. After removing the low-weight variables, age, BMI, DBP, DM, ALT, AST, and GGT were filtered out to design risk prediction models. The prediction model showed excellent clinical differentiation with C-indexes of 0.961 (0.953–0.969), 0.953 (0.939–0.967), and 0.930 (0.920–0.940) and degrees of fit in the training, internal validation, and external validation sets, respectively. In addition, DCA showed that the prediction model could benefit people with a risk threshold of 1%–100%.

Our study showed that age and DBP were risk factors for CAS, consistent with previous studies (4, 16). Previous studies have shown that BMI and abnormal lipid metabolism are closely related (17, 18) and are an important cause of CAS. This study found BMI to be an independent risk factor for CAS, consistent with a national cross-sectional study in China (11). However, a study in Iceland showed a negative association between BMI and CAS, which

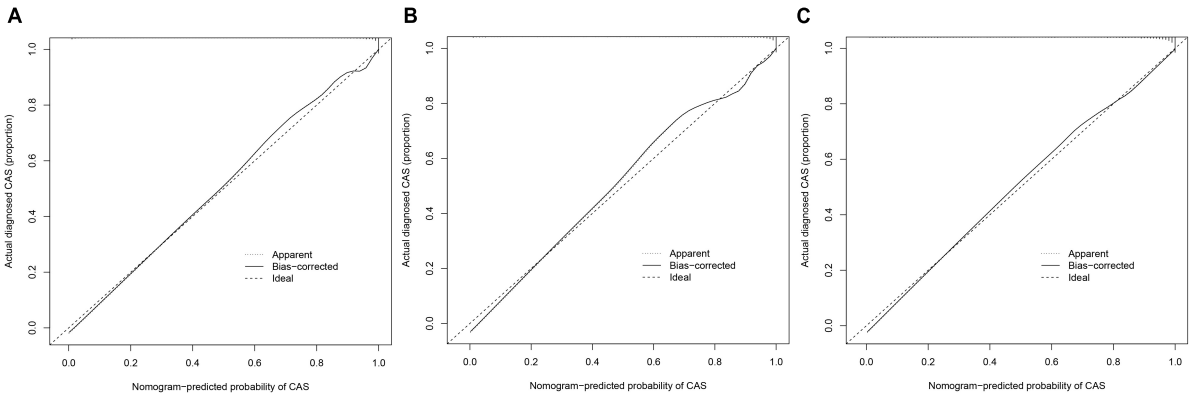


FIGURE 6
Calibration curves. The *x*-axis represents the predicted CAS risk. The *y*-axis represents the actual diagnosed CAS. The diagonal dotted line represents a perfect prediction by an ideal model. The solid line represents the performance of the nomogram, of which a closer fit to the diagonal dotted line represents a better prediction. **(A)** Training set; **(B)** internal validation set; **(C)** external validation set.

may be related to geographic location, ethnicity, diet, and lifestyle (19). Consistent with previous studies (20), we also found WBC count to be an independent risk factor for CAS in our study. CAS is gradually gaining acceptance as an inflammation-associated disease, and studies have shown that inflammatory markers such as interleukin 6 and C-reactive protein are serum markers of CAS (21). In addition, a randomized controlled trial has shown that anti-inflammatory therapy helps reduce the incidence of major cardiovascular adverse events (22). As the pathological basis of atherosclerosis, platelets promote the development and progression of atherosclerosis (23), which was also observed in our study. AST, ALT, and GGT levels are commonly used as indicators of liver dysfunction in clinical practice. In recent years, there has been an increasing number of studies on liver enzymes and cardiovascular risk, and a prospective meta-analysis revealed ALT and GGT levels in relation to cardiovascular disease risk (24). A study by Abdou et al. (25) based on a small sample of abdominally obese people (50) showed that AST was negatively associated with CAS; however, in our study, it was the opposite. A study in rural northeastern China showed that higher levels of education and income were associated with a lower risk of CAS ($P < 0.05$), although this relationship was absent after correction for confounders (26). The role of exercise in CAS remains controversial. A 6-year clinical trial showed that aerobic physical activity did not slow the progression of CAS (27) and might contribute to carotid endothelial injury (28). However, some studies have shown that lack of exercise is a risk factor for CAS (11, 26), and aerobic exercise could help combat carotid intima-media thickening in obese patients (29). In addition, a study in China showed that geographic location (rural areas) is associated with CAS (11).

TABLE 6 Hosmer–Lemeshow test.

	Training set	Internal validation set	External validation set
χ^2	12.9146	13.1931	16.7528
P-value	0.1665	0.1541	0.0527

Early identification and prevention of CAS are essential in reducing adverse cerebrovascular disease occurrence. The diagnosis of CAS in clinical practice relies mainly on carotid ultrasonography. However, there are still difficulties in the large-scale availability of ultrasound in health checkups owing to limited medical resources. The development of risk prediction models provides an alternative method for CAS detection that could benefit populations in less medically developed regions or countries. In addition, risk models could help physicians selectively perform further tests, which is beneficial in terms of saving health care resources. Xing et al. (30) constructed a risk prediction model for atherosclerosis in a systemic lupus erythematosus population based on RNA sequencing, and the model exhibited excellent clinical predictive value (AUC:0.922). However, this model has some limitations. First, the study's sample size was limited (67); furthermore, it was not validated with internal and external samples, and the cost of RNA sequencing in atherosclerosis diagnosis was too high. In addition, an atherosclerosis prediction was constructed based on operational research (31), but it is too obscure and clinically inoperative.

Data mining in big clinical data provides technical support for establishing risk-prediction models (32). Nomograms can generate individual probabilities of clinical events by integrating different outcome and predictor variables and

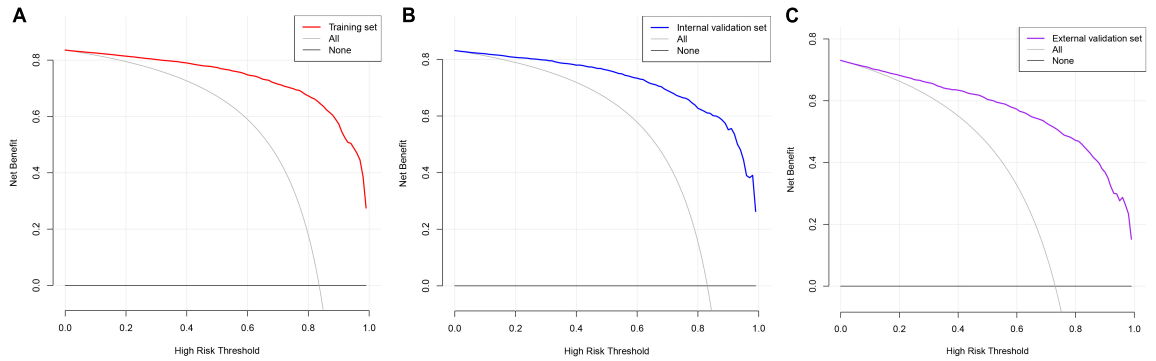


FIGURE 7
Decision curve analysis. The black straight line represents the net benefit when none of the participants are considered to develop CAS, while the light gray curve represents the net benefit when all participants are considered to develop CAS. The area between the color (red, blue, and purple) curve and the light gray curve in the model curve indicates the clinical utility of the model. **(A)** Training set; **(B)** internal validation set; **(C)** external validation set.

merging biological information with clinical prediction models (33–35). In recent years, it has been widely used as a prediction method in clinical settings (36, 37), and it plays a role in promoting personalized medicine (38, 39) and facilitates physicians in predicting disease risk (40). There have been many studies on CAS prevalence and risk factors (4, 41), but few studies on CAS risk-prediction models have been reported. LASSO is a widely used algorithm in machine learning to filter characteristics and interpretable predictors from a large number of potentially co-linear variables by constructing a penalty function (42, 43). LASSO combined with 10-fold cross-validation was used to screen for the characteristic variables associated with CAS. After several modeling attempts, we removed weakly weighted variables in the model (points $<$ 20). Ultimately, 7 clinically common indicators, including age, BMI, DBP, DM, ALT, AST, and GGT, were screened as predictors to construct a risk prediction model. A network computer (dynamic nomogram) was developed to increase the tractability of the model. Our risk prediction model showed a high predictive value and clinical applicability, which might be useful for CAS screening in developing countries, including China.

This study has some inevitable limitations. First, the inclusion of the study population was regional, which may have affected the extrapolation of the prediction model. Second, the collection of clinical baseline data was not sufficiently comprehensive, and potential clinical predictors may have been overlooked. Third, the prediction models were constructed based on cross-sectional studies, and the stability of the models must be tested in clinical practice. In future studies, we will cooperate with multiple centers to continuously test and revise the prediction model during clinical practice and further improve its extrapolation of the prediction model.

Conclusion

This study identified eight CAS-associated independent risk factors, including age, BMI, DBP, WBC, MPV, ALT, AST, and GGT, based on the Chinese population. Meanwhile, we developed risk models for identifying individuals at high risk of CAS, which is important for preventing and reducing adverse prognostic events.

Data availability statement

The original contributions presented in this study are included in the article, further inquiries can be directed to the corresponding author.

Ethics statement

The study was approved by the ethics committee of the Affiliated Hospital of Medical School, Ningbo University, Ningbo, China (KY20191114). The patients/participants provided their written informed consent to participate in this study.

Author contributions

GH conceived and designed the research and drafted the manuscript. GH and QJ did the statistical analysis. GH and XT took part in the discussion. YM revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Grant of Zhejiang Medicine and Health Technology Project, China (Nos. 2018ZH029 and 2020KY871), Major Project for Science and Technology Innovation 2025 in Ningbo, China (No. 2019B10035), Grant of Ningbo Social Development, China (No. 2019C50080), and NINGBO Medical and Health Leading Academic Discipline Project (No. 2022-F24).

Acknowledgments

We thank Zhongwei Zhu of Ningbo Zhenhai Lianhua Hospital for his long-term support of this study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Herrington W, Lacey B, Sherliker P, Armitage J, Lewington S. Epidemiology of atherosclerosis and the potential to reduce the global burden of atherothrombotic disease. *Circ Res.* (2016) 118:535–46. doi: 10.1161/CIRCRESAHA.115.307611

2. Spanos K, Petrocheilou G, Karathanos C, Labropoulos N, Mikhailidis D, Giannoukas A. Carotid bifurcation geometry and atherosclerosis. *Angiology.* (2017) 68:757–64. doi: 10.1177/0003319716678741

3. Meng XF, Liu GM, Wang XL, Xu Q. [Analysis of carotid atherosclerosis and related risk factors in a university physical examination population in Beijing]. *Zhonghua Yi Xue Za Zhi.* (2017) 97:2620–4. doi: 10.3760/cma.j.issn.0376-2491.2017.33.014

4. Song P, Fang Z, Wang H, Cai Y, Rahimi K, Zhu Y, et al. Global and regional prevalence, burden, and risk factors for carotid atherosclerosis: a systematic review, meta-analysis, and modelling study. *Lancet Glob Health.* (2020) 8:e721–9. doi: 10.1016/S2214-109X(20)30117-0

5. Paul S, Candelario-Jalil E. Emerging neuroprotective strategies for the treatment of ischemic stroke: an overview of clinical and preclinical studies. *Exp Neurol.* (2021) 335:113518. doi: 10.1016/j.expneurol.2020.113518

6. Gokaldas R, Singh M, Lal S, Benenstein RJ, Sahni R. Carotid stenosis: from diagnosis to management, where do we stand? *Curr Atheroscler Rep.* (2015) 17:480. doi: 10.1007/s11883-014-0480-7

7. Martinez E, Martorell J, Riambau V. Review of serum biomarkers in carotid atherosclerosis. *J Vasc Surg.* (2020) 71:329–41. doi: 10.1016/j.jvs.2019.04.488

8. Gresele P, Paciullo F, Migliacci R. Antithrombotic treatment of asymptomatic carotid atherosclerosis: a medical dilemma. *Intern Emerg Med.* (2020) 15:1169–81. doi: 10.1007/s11739-020-02347-7

9. Paraskevas KI, Veith FJ, Eckstein H-H, Ricco J-B, Mikhailidis DP. Cholesterol, carotid artery disease and stroke: what the vascular specialist needs to know. *Ann Transl Med.* (2020) 8:1265. doi: 10.21037/atm.2020.02.176

10. Ji X, Leng X-Y, Dong Y, Ma Y-H, Xu W, Cao X-P, et al. Modifiable risk factors for carotid atherosclerosis: a meta-analysis and systematic review. *Ann Transl Med.* (2019) 7:632. doi: 10.21037/atm.2019.10.115

11. Wang X, Li W, Song F, Wang L, Fu Q, Cao S, et al. Carotid atherosclerosis detected by ultrasonography: a national cross-sectional study. *J Am Heart Assoc.* (2018) 7:e008701. doi: 10.1161/JAHA.118.008701

12. Xu L, Xie J, Chen S, Chen Y, Yang H, Miao M, et al. Light-to-moderate alcohol consumption is associated with increased risk of Type 2 diabetes in individuals with nonalcoholic fatty liver disease: a nine-year cohort study. *Am J Gastroenterol.* (2020) 115:876–84. doi: 10.14309/ajg.0000000000000607

13. Huang H, Wang J, Xu L, Miao M, Xu C. Association between high-density lipoprotein cholesterol to apolipoprotein A-I ratio and nonalcoholic fatty liver disease: a cross-sectional study. *Int J Endocrinol.* (2021) 2021:6676526. doi: 10.1155/2021/6676526

14. Mao Y-S, Liu Z-M, Chen C-X, Zhu Z-W, Hong Z-L. Ningbo thyroid dysfunction prevalence study: a cross-sectional survey in an employees-cohort. *Chin Med J (Engl).* (2010) 123:1673–8.

15. Lyu J, Li Z, Wei H, Liu D, Chi X, Gong D-W, et al. A potent risk model for predicting new-onset acute coronary syndrome in patients with type 2 diabetes mellitus in Northwest China. *Acta Diabetol.* (2020) 57:705–13. doi: 10.1007/s00592-020-01484-x

16. de Weerd M, Greving JP, de Jong AWF, Buskens E, Bots ML. Prevalence of asymptomatic carotid artery stenosis according to age and sex: systematic review and metaregression analysis. *Stroke.* (2009) 40:1105–13. doi: 10.1161/STROKEAHA.108.532218

17. Mota dos Santos C, Sá Silva C, César de Araújo E, Kruze Grande de Arruda I, da Silva Diniz A, Coelho Cabral P. [Lipid and glucose profiles in outpatients and their correlation with anthropometric indices]. *Rev Port Cardiol.* (2013) 32:35–41. doi: 10.1016/j.repc.2012.05.017

18. Zaid M, Ameer F, Munir R, Rashid R, Farooq N, Hasnain S, et al. Anthropometric and metabolic indices in assessment of type and severity of dyslipidemia. *J Physiol Anthropol.* (2017) 36:19. doi: 10.1186/s40101-017-0134-x

19. Sturlaugsdottir R, Aspelund T, Bjornsdottir G, Sigurdsson S, Thorsson B, Eiriksdottir G, et al. Prevalence and determinants of carotid plaque in the cross-sectional REFINE-Reykjavik study. *BMJ Open.* (2016) 6:e012457. doi: 10.1136/bmjopen-2016-012457

20. Ortega E, Gilabert R, Nuñez I, Cofán M, Sala-Vila A, de Groot E, et al. White blood cell count is associated with carotid and femoral atherosclerosis. *Atherosclerosis.* (2012) 221:275–81. doi: 10.1016/j.atherosclerosis.2011.12.038

21. Ammirati E, Moroni F, Norata GD, Magnoni M, Camici PG. Markers of inflammation associated with plaque progression and instability in patients with carotid atherosclerosis. *Mediators Inflamm.* (2015) 2015:718329. doi: 10.1155/2015/718329

22. Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, et al. Antiinflammatory therapy with canakinumab for atherosclerotic disease. *N Engl J Med.* (2017) 377:1119–31. doi: 10.1056/NEJMoa1707914

23. Nording H, Baron L, Langer HF. Platelets as therapeutic targets to prevent atherosclerosis. *Atherosclerosis.* (2020) 307:97–108. doi: 10.1016/j.atherosclerosis.2020.05.018

24. Kunutsor SK, Apekey TA, Khan H. Liver enzymes and risk of cardiovascular disease in the general population: a meta-analysis of prospective cohort studies. *Atherosclerosis.* (2014) 236:7–17. doi: 10.1016/j.atherosclerosis.2014.06.006

25. Abdou AS, Magour GM, Mahmoud MM. Evaluation of some markers of subclinical atherosclerosis in Egyptian young adult males with abdominal obesity. *Br J Biomed Sci.* (2009) 66:143–7. doi: 10.1080/09674845.2009.11730261

26. Xing L, Li R, Zhang S, Li D, Dong B, Zhou H, et al. High burden of carotid atherosclerosis in rural northeast China: a population-based study. *Front Neurol.* (2021) 12:597992. doi: 10.3389/fneur.2021.597992

27. Rauramaa R, Halonen P, Väisänen SB, Lakka TA, Schmidt-Trucksäss A, Berg A, et al. Effects of aerobic physical exercise on inflammation and atherosclerosis in men: the DNASCO Study: a six-year randomized, controlled trial. *Ann Intern Med.* (2004) 140:1007–14. doi: 10.7326/0003-4819-140-12-200406150-00010

28. Marchio P, Guerra-Ojeda S, Vila JM, Aldasoro M, Valles SL, Soler C, et al. Chronic exercise impairs nitric oxide pathway in rabbit carotid and femoral arteries. *J Physiol.* (2018) 596:4361–74. doi: 10.1113/JP275611

29. Kim DK, Jee JH, Park WH. Effects of aerobic and resistance exercise training on carotid intima-media thickness in abdominal obese women. *Metab Syndr Relat Disord.* (2021) 19:200–4. doi: 10.1089/met.2020.0118

30. Xing H, Pang H, Du T, Yang X, Zhang J, Li M, et al. Establishing a risk prediction model for atherosclerosis in systemic lupus erythematosus. *Front Immunol.* (2021) 12:622216. doi: 10.3389/fimmu.2021.622216

31. Chen Z, Yang M, Wen Y, Jiang S, Liu W, Huang H. Prediction of atherosclerosis using machine learning based on operations research. *Math Biosci Eng.* (2022) 19:4892–910. doi: 10.3934/mbe.2022229

32. Wu W-T, Li Y-J, Feng A-Z, Li L, Huang T, Xu A-D, et al. Data mining in clinical big data: the frequently used databases, steps, and methodological models. *Mil Med Res.* (2021) 8:44. doi: 10.1186/s40779-021-00338-z

33. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol.* (2015) 16:e173–80. doi: 10.1016/S1470-2045(14)71116-7

34. Jiang F, Yu X, Wu C, Wang M, Wei K, Wang J, et al. A simple-to-use nomogram for predicting survival in children with acute myeloid leukemia. *Biomed Res Int.* (2021) 2021:7264623. doi: 10.1155/2021/7264623

35. Jiang F, Wei K, Lyu W, Wu C. Predicting risk of insulin resistance in a Chinese population with polycystic ovary syndrome: designing and testing a new predictive nomogram. *Biomed Res Int.* (2020) 2020:8031497. doi: 10.1155/2020/8031497

36. Wu Y, Hu H, Cai J, Chen R, Zuo X, Cheng H, et al. A prediction nomogram for the 3-year risk of incident diabetes among Chinese adults. *Sci Rep.* (2020) 10:21716. doi: 10.1038/s41598-020-78716-1

37. Wu J, Zhang H, Li L, Hu M, Chen L, Xu B, et al. A nomogram for predicting overall survival in patients with low-grade endometrial stromal sarcoma: a population-based analysis. *Cancer Commun (Lond).* (2020) 40:301–12. doi: 10.1002/cac2.12067

38. Shi R, Zhang T, Sun H, Hu F. Establishment of clinical prediction model based on the study of risk factors of stroke in patients with Type 2 diabetes mellitus. *Front Endocrinol (Lausanne).* (2020) 11:559. doi: 10.3389/fendo.2020.00559

39. Li W, Wang H, Dong S, Tang Z-R, Chen L, Cai X, et al. Establishment and validation of a nomogram and web calculator for the risk of new vertebral compression fractures and cement leakage after percutaneous vertebroplasty in patients with osteoporotic vertebral compression fractures. *Eur Spine J.* (2022) 31:1108–21. doi: 10.1007/s00586-021-07064-z

40. Dong S, Li W, Tang Z-R, Wang H, Pei H, Yuan B. Development and validation of a novel predictive model and web calculator for evaluating transfusion risk after spinal fusion for spinal tuberculosis: a retrospective cohort study. *BMC Musculoskelet Disord.* (2021) 22:825. doi: 10.1186/s12891-021-04715-6

41. Song P, Xia W, Zhu Y, Wang M, Chang X, Jin S, et al. Prevalence of carotid atherosclerosis and carotid plaque in Chinese adults: a systematic review and meta-regression analysis. *Atherosclerosis.* (2018) 276:67–73. doi: 10.1016/j.atherosclerosis.2018.07.020

42. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* (2010) 33:1–22.

43. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med.* (2007) 26:5512–28.

frontiers | Frontiers in Cardiovascular Medicine

TYPE Original Research
PUBLISHED 07 September 2022
DOI 10.3389/fcvm.2022.911987

Check for updates

OPEN ACCESS

EDITED BY
Andre Rodrigues Duraes,
Federal University of Bahia, Brazil

REVIEWED BY
Jun Lyu,
First Affiliated Hospital of Jinan University, China
Tsehay Admassu,
Injibara University, Ethiopia

*CORRESPONDENCE
Kun Ling Ma
klma@zju.edu.cn

SPECIALTY SECTION
This article was submitted to General Cardiovascular Medicine, a section of the journal Frontiers in Cardiovascular Medicine

RECEIVED 03 April 2022
ACCEPTED 15 August 2022
PUBLISHED 07 September 2022

CITATION
Liu WT, Liu XQ, Jiang TT, Wang MY, Huang Y, Huang YL, Jin FY, Zhao Q, Wu QY, Liu BC, Ruan XZ and Ma KL (2022) Using a machine learning model to predict the development of acute kidney injury in patients with heart failure.
Front. Cardiovasc. Med. 9:911987.
doi: 10.3389/fcvm.2022.911987

COPYRIGHT
© 2022 Liu, Liu, Jiang, Wang, Huang, Huang, Jin, Zhao, Wu, Liu, Ruan and Ma. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Using a machine learning model to predict the development of acute kidney injury in patients with heart failure

Wen Tao Liu[1], Xiao Qi Liu[1], Ting Ting Jiang[1], Meng Ying Wang[1], Yang Huang[1], Yu Lin Huang[1], Feng Yong Jin[1], Qing Zhao[1], Qin Yi Wu[1], Bi Cheng Liu[1], Xiong Zhong Ruan[2] and Kun Ling Ma[3]*

[1]School of Medicine, Institute of Nephrology, Zhongda Hospital, Southeast University, Nanjing, China, [2]John Moorhead Research Laboratory, Department of Renal Medicine, University College London (UCL) Medical School, London, United Kingdom, [3]Department of Nephrology, The Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China

Background: Heart failure (HF) is a life-threatening complication of cardiovascular disease. HF patients are more likely to progress to acute kidney injury (AKI) with a poor prognosis. However, it is difficult for doctors to distinguish which patients will develop AKI accurately. This study aimed to construct a machine learning (ML) model to predict AKI occurrence in HF patients.

Materials and methods: The data of HF patients from the Medical Information Mart for Intensive Care-IV (MIMIC-IV) database was retrospectively analyzed. A ML model was established to predict AKI development using decision tree, random forest (RF), support vector machine (SVM), K-nearest neighbor (KNN), and logistic regression (LR) algorithms. Thirty-nine demographic, clinical, and treatment features were used for model establishment. Accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUROC) were used to evaluate the performance of the ML algorithms.

Results: A total of 2,678 HF patients were engaged in this study, of whom 919 developed AKI. Among 5 ML algorithms, the RF algorithm exhibited the highest performance with the AUROC of 0.96. In addition, the Gini index showed that the sequential organ function assessment (SOFA) score, partial pressure of oxygen (PaO_2), and estimated glomerular filtration rate (eGFR) were highly relevant to AKI development. Finally, to facilitate clinical application, a simple

model was constructed using the 10 features screened by the Gini index. The RF algorithm also exhibited the highest performance with the AUROC of 0.95.

Conclusion: Using the ML model could accurately predict the development of AKI in HF patients.

KEYWORDS

heart failure, acute kidney injury, machine learning, prediction model, artificial intelligence

Introduction

Heart failure (HF) is the end stage of cardiovascular disease with a prevalence of around 1–2% in adults (1). HF patients are more likely to progress to acute kidney injury (AKI) with a poor prognosis (2). Studies have shown that more than 20% of HF inpatients would progress to AKI with a fatality rate of 4.1% (3, 4). Even mildly reversible AKI is associated with severe clinical outcomes, such as an increased risk of death (5, 6). Furthermore, HF has imposed a heavy financial burden on patients, with an annual cost ranging from $2,496 to $84,434 per patient (7).

Currently, it's difficult for doctors to distinguish which patients will develop AKI. The diagnosis of AKI mainly depends on serum creatinine (Scr) and urine output. However, the elevation of Scr is usually delayed relative to the kidney injury, and Scr can be affected by muscle mass and metabolism (8). In addition, urine output is easily affected by drugs such as diuretics, and thus cannot reflect the kidney injury accurately. Therefore, some researchers have analyzed the risk factors of AKI, hoping to identify patients at high risk of AKI in advance. Fan et al. (9) employed a multivariate logistic regression method to reveal that age, diabetes, New York Heart Association (NYHA) classification, estimated glomerular filtration rate (eGFR), highly sensitive C-reactive protein (hs-CRP), and urinary angiotensinogen (uAGT) were independently associated with AKI development in HF patients. A meta-analysis also revealed that baseline chronic kidney disease (CKD), history of hypertension and diabetes, age, and diuretic use were significant predictors for AKI occurrence (3). In recent years, some researchers have adopted new biomarkers to predict the occurrence of AKI. Schanz et al. (10) examined urinary tissue inhibitors of metalloprotease-2 (TIMP-2) and insulin-like growth factor-binding protein 7 (IGFBP7) in 400 patients with HF. They found that urinary [TIMP-2] × [IGFBP7] was a promising marker for AKI risk assessment with high sensitivity and specificity. Although the above studies analyzed the risk factors of AKI, these studies adopted traditional strategies of developing prediction models and were not supported by big data. Therefore, more studies are needed to be carried out to verify the correctness of the above viewpoints.

Using machine learning (ML) algorithms is another strategy for establishing prediction models. ML is a subset of artificial intelligence (AI) in computer science. It is a discipline that focuses on how computers simulate human behaviors to acquire new knowledge (11). ML algorithms include decision tree, random forest (RF), support vector machine (SVM), logistic regression (LR), and K-nearest neighbor (KNN) (12). Compared with classical statistical methods, ML algorithms can explore the relationship between data and solve classification problems better (13, 14). Currently, the connection between ML and medicine is getting closer and ML has been adopted in the scope of diagnosis, risk stratification, and treatment (11, 15). Kimura et al. (16) utilized ML algorithms to analyze peripheral blood smears and developed an automated diagnostic model for myelodysplastic syndrome and aplastic anemia. In a multicenter study, Tomašev et al. (17) successfully developed a ML model to predict the occurrence of AKI, and stratified the risk of AKI to provide the possibility for the prevention of AKI. In addition, due to the support of ML algorithm in the treatment of anemia in hemodialysis patients, it not only reduced the use of erythropoietic-stimulating agent but also optimized anemia management (18). Overall, ML algorithms have made great contributions to improving the quality of healthcare.

Clinical studies often need the support of a large amount of data, and public databases can provide the required data.

Abbreviations: HF, heart failure; AKI, acute kidney injury; ML, machine learning; MIMIC-IV, Medical Information Mart for Intensive Care-IV; RF, random forest; SVM, support vector machine; KNN, K-nearest neighbor; LR, logistic regression; AUROC, area under the receiver operating characteristic curve; SOFA, sequential organ function assessment; PaO_2, partial pressure of oxygen; eGFR, estimated glomerular filtration rate; Scr, serum creatinine; NYHA, New York Heart Association; hs-CRP, highly sensitive C-reactive protein; uAGT, urinary angiotensinogen; CKD, chronic kidney disease; TIMP-2, tissue inhibitor of metalloprotease-2; IGFBP7, insulin-like growth factor-binding protein 7; AI, artificial intelligence; ICU, intensive care unit; CKD-EPI, Chronic Kidney Disease Epidemiology Collaboration; CHD, coronary heart disease; COPD, chronic obstructive pulmonary disease; WBC, white blood cell; PT, prothrombin time; INR, international normalized ratio; $PaCO_2$, partial pressure of carbon dioxide; RAS, renin-angiotensin system.

By analyzing the data in the database, researchers may draw valuable conclusions and help doctors make clinical decisions (19, 20). Medical Information Mart for Intensive Care-IV (MIMIC-IV) database contains clinical data on over 60,000 Intensive Care Unit (ICU) stays at the Beth Israel Deaconess Medical Center. Individuals who completed the test in PhysioNet have access to the database (certification number = 33449415) (21).

Machine learning algorithms have many advantages in the field of data processing. However, they are rarely used in AKI prediction in HF patients. Therefore, this study examined whether a ML-derived model for predicting AKI development would achieve high accuracy and guide AKI prevention.

Materials and methods

Study design and population

The data of HF patients hospitalized in the Cardiac Vascular Intensive Care Unit and Coronary Care Unit (CCU) in the MIMIC-IV database were retrospectively analyzed. With PostgreSQL 13, we installed the database on the computer. Next, demographics, clinical features, etc., of HF patients were extracted according to the corresponding codes. The inclusion criteria were as follows: 1) patients were older than 18 years old; 2) patients were diagnosed with HF according to the ICD code; 3) patients should have at least two Scr tests within the first 48 h of ICU admission; 4) For patients who were admitted to the hospital multiple times, the clinical data of the first hospitalization were selected. The exclusion criteria were as follows: 1) eGFR $<$ 15 ml/min/1.73 m^2 at the time of ICU admission; 2) patients received renal replacement therapy, including hemodialysis and peritoneal dialysis; 3) Patients whose Scr had risen $\geq$ 0.3 mg/dl before ICU admission during the hospitalization; 4) patients were diagnosed with heart transplantation, kidney transplantation, malignant tumor, and pregnancy; 5) patients who stayed in the ICU for less than 48 h. The primary endpoint was AKI, defined as the increase in Scr by $\geq$ 0.3 mg/dl within the first 48 h of ICU admittance (22). Because of inadequate data and probable changes in the urine output caused by medical therapy, urine output criterion was not employed to diagnose AKI.

After selecting patients, the demographic, clinical, and treatment data were extracted. A total of thirty-nine features were considered as AKI predictors. The eGFR was calculated by the CKD-EPI (Chronic Kidney Disease Epidemiology Collaboration) formula (23). In the analysis, missing data were replaced by mean or median according to data distribution. In addition, features with more than 30% missing data were not included.

Establishment of the prediction model

Five ML algorithms: decision tree, RF, KNN, SVM, and LR were utilized to establish the model to predict the development of AKI. All the above features were incorporated into the ML model. A total of 70% of the dataset was randomly selected as the training set and the remaining 30% as the test set. The data in the training set was used to train the model, and the test set was used to examine the performance of the optimal model. The AKI status was classified as "Yes" or "No" (**Figure 1**). Accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUROC) were used to evaluate the predictive performance of the model. The Gini index in the RF algorithm was calculated to rank the predictive value of features. To make the prediction model more concise and easier to use in clinical practice, a simple model with ten features selected by the Gini index was established. Python 3.7 was used to establish the model.

Statistical analysis

For continuous variables, a *t*-test was used to compare the differences between two groups if they conformed to the normal distribution, otherwise rank-sum test was used. For categorical variables, the chi-square test was used for comparison. All tests of significance were 2-tailed, and the $P < 0.05$ was considered

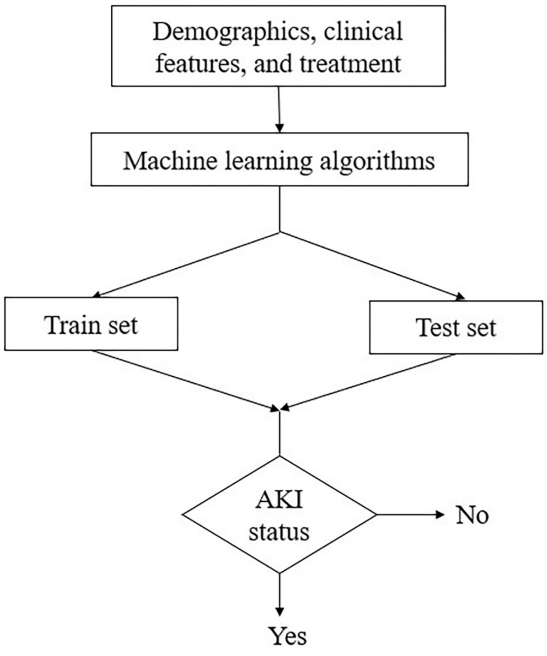


FIGURE 1
Process of establishing the prediction model. AKI, acute kidney injury.

statistically significant. StataMP software (Version 14) was used for statistical analysis.

Results

Comparisons between acute kidney injury and non-acute kidney injury groups

A total of 2,678 HF patients were engaged in the study (Figure 2). In our cohort, 919 HF patients progressed to AKI within the first 48 h of ICU admission. Males comprised 59.7 and 59.1% of AKI and non-AKI groups. Patients in the AKI group were significantly older than those in non-AKI group. Sequential organ function assessment (SOFA) score, Scr, and urea nitrogen were also higher in the AKI group. All comorbidities, including hypertension and diabetes, were highly related to AKI development. In addition, all treatments demonstrated significance between the two groups (Table 1).

Performance of the machine learning model

Five ML algorithms were applied to predict the AKI status. Table 2 shows that the RF algorithm achieved the highest accuracy with 88.36%. In addition, the algorithms with the highest sensitivity and specificity were RF (96.04%) and LR (77.02%) (Table 2). The RF algorithm performed the best for the ML model with the AUROC of 0.96, which was better than 0.92 of SVM, 0.83 of KNN, 0.82 of the decision tree, and 0.92 of LR (Figure 3).

Predictors of acute kidney injury status

By calculating the Gini index in the RF algorithm, the predictive value of features was ranked. The top ten predictors were: SOFA score, partial pressure of oxygen (PaO_2), eGFR, serum bicarbonate, hemoglobin, platelet count, blood lactic acid, Scr, serum magnesium, and blood glucose (Figure 4).

Establishment of a simple model

According to the ten features selected by the Gini index, a simple model was established. Same as the prediction model using all 39 features, in the simple model, the RF algorithm achieved the highest accuracy with 87.07% (Table 3). In addition, the RF algorithm also achieved the highest sensitivity (92.52%), specificity (79.68%), and AUROC (0.95). Interestingly, the algorithms of KNN and decision tree outperformed the initial model with an improved AUROC (Figure 5).

Discussion

Different from previous studies, the strength of our study was the implementation of ML algorithms to predict AKI development (24, 25). Traditional approaches to constructing

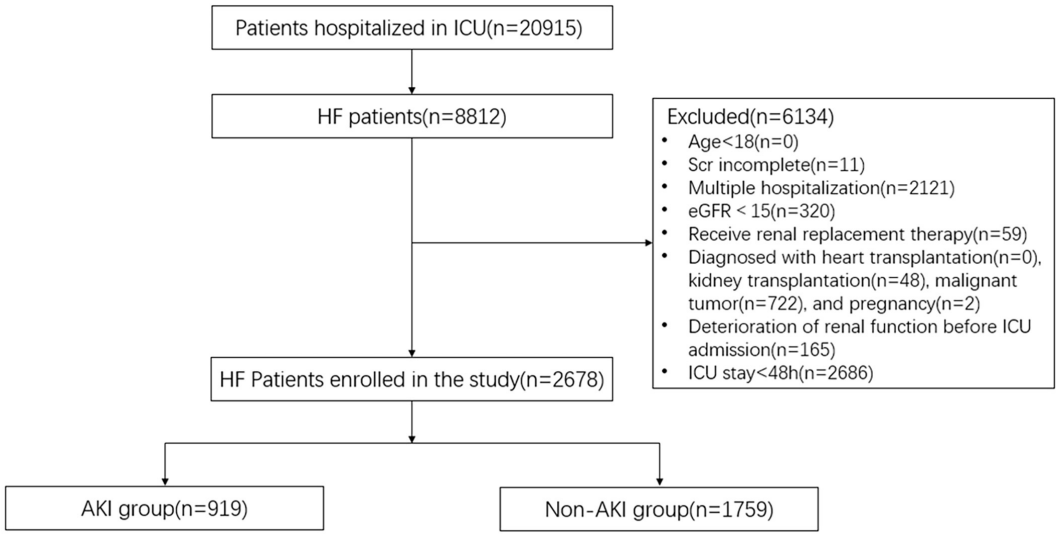


FIGURE 2
Consort flow chart. A total of 2,678 patients were selected from the database with 20,915 patients. ICU, intensive care unit; HF, heart failure; Scr, serum creatinine; eGFR, estimated glomerular filtration rate; AKI, acute kidney injury.

TABLE 1 Clinical characteristics of HF patients.

Features	AKI ($n = 919$)	Non-AKI ($n = 1759$)	*P*-value
Demographic/clinical characteristics			
Age (year)	73 (63,81)	70 (59,80)	< 0.001
Male (%)	549 (59.7%)	1040 (59.1%)	0.759
Height (cm)	169 (163,175)	169 (165,178)	0.002
Weight (kg)	80 (67.8,95)	80.8 (67.9,96.3)	0.587
Respiratory rate (bpm)	16 (14,20)	18 (15,23)	< 0.001
Body temperature (°C)	36.6 (36.3,36.8)	36.6 (36.4,36.9)	< 0.001
Heart rate (bpm)	81 (74,91)	85 (74,97)	0.001
Systolic blood pressure (mmHg)	115 (99,132)	116 (102,131)	0.389
Diastolic blood pressure (mmHg)	59 (51,70)	63 (54,75)	< 0.001
SOFA score	8 (5,10)	5 (3,7)	< 0.001
Ventilation (%)	536 (58.3%)	678 (38.5%)	< 0.001
Diabetes (%)	435 (47.3%)	652 (37.1%)	< 0.001
CHD (%)	431 (46.9%)	748 (42.5%)	0.030
Hypertension (%)	246 (26.8%)	648 (36.8%)	< 0.001
Atrial flutter or atrial fibrillation (%)	535 (58.2%)	898 (51.1%)	< 0.001
COPD (%)	54 (5.9%)	162 (9.2%)	0.003
Laboratory data			
Scr (mg/dL)	1.2 (0.9,1.7)	1.1 (0.8,1.4)	< 0.001
eGFR (mL/min/1.73 m^2)	54.9 (37.3,74.8)	65.3 (43.3,86.4)	< 0.001
Urea nitrogen (mg/dL)	25 (17,37)	22 (16,33)	< 0.001
WBC (K/μL)	12.3 (8.8,16.5)	11.2 (8.1,15)	< 0.001
Hemoglobin (g/dL)	9.4 (8,11.4)	10.7 (9,12.5)	< 0.001
Platelet (K/μL)	164 (120,227)	194 (145,250)	< 0.001
PT (s)	15.6 (13.9,18.5)	15 (13.1,17.5)	< 0.001
INR	1.4 (1.2,1.7)	1.4 (1.2,1.6)	< 0.001
PH	7.38 (7.33,7.44)	7.38 (7.36,7.43)	0.509
PaO_2 (mmHg)	232 (116,347)	190 (86,272)	< 0.001
$PaCO_2$ (mmHg)	41 (36,44)	42 (38,45)	< 0.001
Blood lactic acid (mmol/L)	1.9 (1.5,3)	1.9 (1.4,2.2)	< 0.001
Serum bicarbonate (mEq/L)	22 (20,24)	24 (22,27)	< 0.001
Serum potassium (mEq/L)	4.3 (3.9,4.8)	4.2 (3.8,4.6)	< 0.001
Serum sodium (mEq/L)	139 (136,141)	138 (136,141)	0.027
Serum calcium (mg/dL)	8.4 (8,8.8)	8.5 (8.1,8.9)	0.003
Serum magnesium (mg/dL)	2.2 (1.9,2.7)	2.1 (1.9,2.3)	< 0.001
Serum phosphate (mg/dL)	3.9 (3.3,4.8)	3.6 (3.1,4.2)	< 0.001
Blood glucose (mg/dL)	135 (108,178)	129 (108,168)	0.054
Treatments			
RAS inhibitor (%)	117 (12.7%)	442 (25.1%)	< 0.001
Diuretics (%)	806 (87.7%)	1329 (75.6%)	< 0.001
Digoxin (%)	30 (3.3%)	102 (5.8%)	0.004
β-receptor blocker (%)	459 (49.9%)	997 (56.7%)	0.001

Values are shown as median (interquartile range), absolute values, and percentages. SOFA, sequential organ function assessment; CHD, coronary heart disease; COPD, chronic obstructive pulmonary disease; Scr, serum creatinine; eGFR, estimated glomerular filtration rate; WBC, white blood cell; PT, prothrombin time; INR, international normalized ratio; PaO_2, partial pressure of oxygen; $PaCO_2$, partial pressure of carbon dioxide; RAS, renin-angiotensin system.

TABLE 2 Performance of the prediction model.

Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)
RF	88.36	96.04	73.91
SVM	86.85	92.41	76.40
Decision tree	79.53	86.14	67.08
KNN	80.39	93.73	55.28
LR	86.42	91.42	77.02

RF, random forest; SVM, support vector machine; KNN, K-nearest neighbor; LR, logistic regression.

prediction models have made great contributions to assisting doctors in medical decision-making (26, 27). However, they have inherent drawbacks that may result in the omission of crucial predictors and correlations. Compared with traditional approaches, ML algorithms have great advantages in constructing prediction models, such as high accuracy in predicting heart disease (28, 29). According to our findings, the RF algorithm exhibited the highest performance among the five algorithms in predicting AKI. This is not surprising since the RF algorithm has advantages in processing high-dimensional data (30). In addition, the Gini index in the RF algorithm can reflect the predictive value of features, which facilitates the application of the prediction model in clinical practice. Therefore, from our point of view, RF algorithm should be preferentially adopted in clinical research, especially when analyzing high-dimensional data.

At present, HF patients are more likely to progress to AKI. Therefore, it is of great significance to analyze the risk factors of AKI and take corresponding treatment. The reported risk factors of AKI include age, baseline eGFR, NYHA classification, Kidney injury molecular-1, neutrophil gelatinase-associated lipocalin, urinary C-C motif chemokine ligand 14, etc. (9, 31, 32). However, some of these features are not routinely examined in clinical practice, which is not conducive to promotion. In this study, the features used in the prediction model are common and easy to obtain, which is also a major strength.

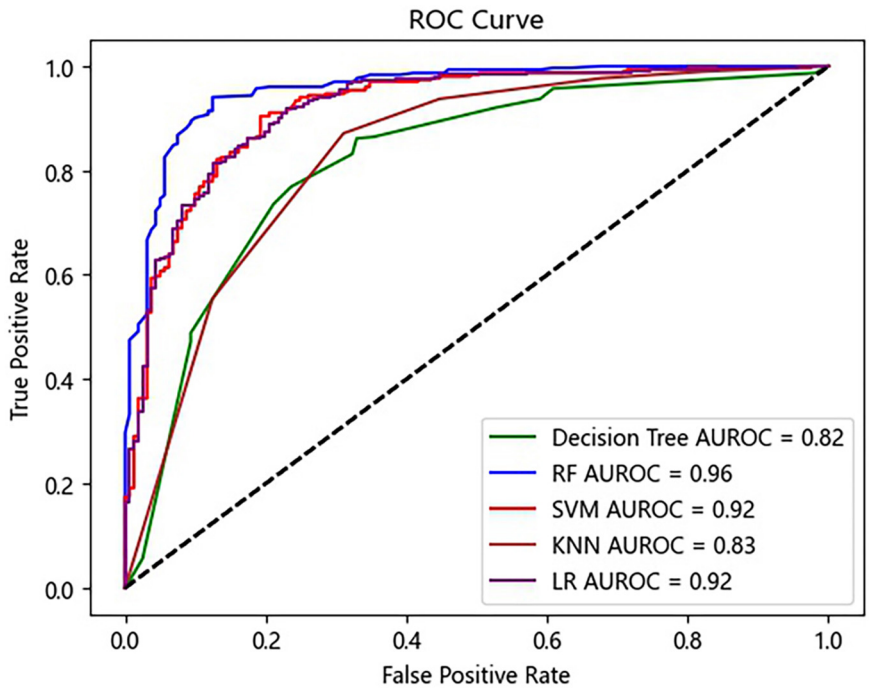


FIGURE 3
Receiver operating characteristic (ROC) curves of the prediction model. RF, random forest; SVM, support vector machine; KNN, K-nearest neighbor; LR, logistic regression.

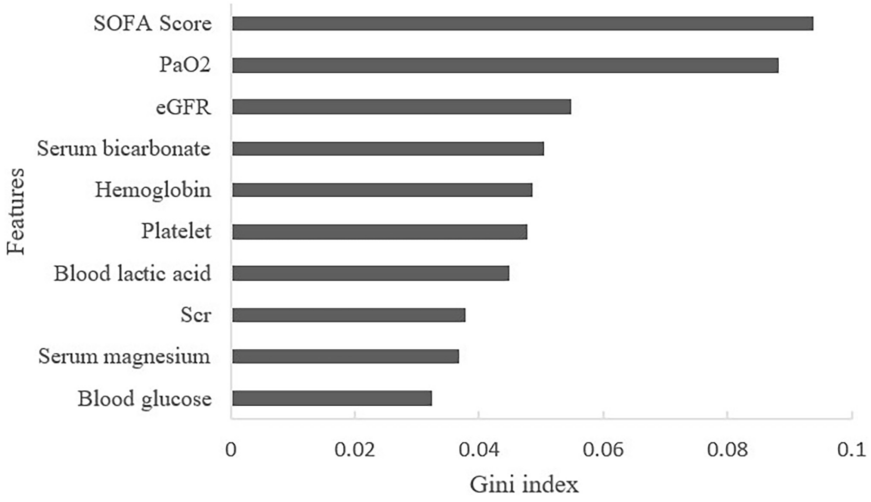


FIGURE 4
Contribution of features of AKI in HF patients (Top 10 displayed). SOFA, sequential organ function assessment score; PaO_2, partial pressure of oxygen; eGFR, estimated glomerular filtration rate; Scr, serum creatinine.

A total of 39 features were used to predict AKI in this study. Therefore, it is important to screen out the features related to the occurrence of AKI. Through the Gini index, we found that SOFA score, PaO_2, and eGFR exhibited the highest predictive value. SOFA score can reflect the function of the nervous system, respiratory system, circulatory system, etc., and is used to monitor organ dysfunction. Currently, the SOFA score has been considered an excellent score to predict short-term mortality in life-threatening conditions (33). In critically ill patients, the SOFA score was thought to be an important predictor of AKI with the AUROC of 0.957 (34). In our study, patients with higher SOFA score were apt to progress to AKI,

TABLE 3 Performance of the simple model.

Algorithm	Accuracy (%)	Sensitivity (%)	Specificity (%)
RF	87.07	92.52	79.68
SVM	80.73	86.61	72.73
Decision tree	83.45	90.16	74.33
KNN	84.13	92.13	73.26
LR	81.63	88.19	72.73

RF, random forest; SVM, support vector machine; KNN, K-nearest neighbor; LR, logistic regression.

showing that the degree of organ dysfunction was associated with AKI. These findings suggest that SOFA scores should be routinely calculated in HF patients. Patients with high SOFA scores should be monitored and treated more aggressively.

In our study, patients who progressed to AKI had higher PaO_2. This may be related to the high proportion of mechanical ventilation treatment in the AKI group. High PaO_2 is related to oxidative stress, which is thought to be a pathogenesis of AKI (35, 36). Furthermore, Chen et al. (37) found that LPS-induced AKI in mice could be alleviated by inhibiting oxidative stress. Therefore, PaO_2 may be a predictor of AKI. Currently, the relationship between hyperoxia and AKI is still inconclusive. Shen et al. (38) found that AKI was more common in patients with persistent hyperoxia than those with transient hyperoxia. In addition, according to an observational study by Bae, intraoperative hyperoxia was found to be strongly linked with the risk of AKI following cardiac surgery (39). Therefore, in clinical practice, we need to pay attention to the relationship between PaO_2 and AKI, and further investigate the mechanism behind it.

Currently, eGFR is used to assess glomerular filtration function. In our study, we found that the eGFR of the AKI group was lower than that of the non-AKI group. It suggested that patients with kidney injury were more likely to progress to AKI. This finding was consistent with Tuukka's study: lower baseline eGFR is an independent predictor of AKI (40). In a multicenter study, Patel et al. (41) performed a statistical analysis of more than 360,000 HF patients and found that 64% of them had eGFR $<$ 60 mL/min/1.73 m^2. In addition, they found that lower admission eGFR was associated with in-hospital mortality. Therefore, we recommend that eGFR should be calculated in each HF patient.

Finally, a simple model with ten selected features was established for the convenience of doctors. Compared with the prediction model using 39 features, the simple model was more usable and could also accurately predict AKI development. Interestingly, although there is less data in the simple model, the AUROC of KNN and decision tree algorithms were even higher. This phenomenon may be related to the removal of confounding factors. Therefore, screening predictors is also a key step when establishing the prediction model.

Our study had several limitations. First, in the MIMIC database, acute HF and chronic HF were not well distinguished. However, the severity of the two diseases was different, and the

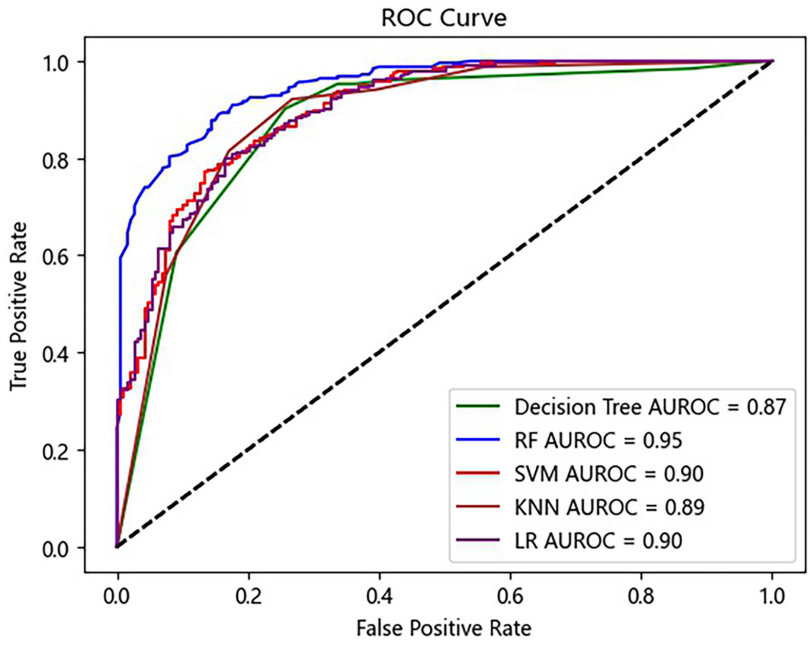


FIGURE 5
Receiver operating characteristic (ROC) curves of the prediction model using ten selected features. RF, random forest; SVM, support vector machine; KNN, K-nearest neighbor; LR, logistic regression.

mechanisms that led to AKI were also different. We did not distinguish between the two diseases affected the correctness of the results to some extent. Second, this study was a single-center retrospective study without validation from other centers. Hence, high-quality randomized controlled trials are needed to confirm our findings.

Conclusion

We successfully established a ML model to predict the development of AKI in HF patients. Among five ML algorithms, the RF algorithm exhibited the highest predictive performance. Our results provided the possibility for ML algorithms to guide AKI prevention in HF patients. Further studies are needed to verify whether our model can be applied to populations in other countries.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

KM and WL: study design and writing – original draft. WL: data collection. WL, XL, TJ, and MW: statistical analysis. WL, YH, YLH, FJ, QZ, and QW: software. KM, XR, and BL: manuscript revision. All authors read and approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (82170736 and 81970629), Project for Jiangsu Provincial Medical Talent (ZDRCA2016077), Fundamental Research Funds for the Central Universities (3224002110D), and Jiangsu Province Ordinary University Graduate Research Innovation Project (SJCX20-0055).

Acknowledgments

We thank Xue Liang Wang for his assistance with Python.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Groenewegen A, Rutten FH, Mosterd A, Hoes AW. Epidemiology of heart failure. *Eur J Heart Fail.* (2020) 22:1342–56. doi: 10.1002/ejhf.1858

2. Schefold JC, Filippatos G, Hasenfuss G, Anker SD, von Haehling S. Heart failure, and kidney dysfunction: epidemiology, mechanisms, and management. *Nat Rev Nephrol.* (2016) 12:610–23. doi: 10.1038/nrneph.2016.113

3. Damman K, Valente MA, Voors AA, O'Connor CM, van Veldhuisen DJ, Hillege HL. Renal impairment, worsening renal function, and outcome in patients with heart failure: an updated meta-analysis. *Eur Heart J.* (2014) 35:455–69. doi: 10.1093/eurheartj/eht386

4. Zhang Y, Zhang J, Butler J, Yang X, Xie P, Guo D, et al. Contemporary epidemiology, management, and outcomes of patients hospitalized for heart failure in China: results from the China heart failure (China-HF) registry. *J Card Fail.* (2017) 23:868–75. doi: 10.1016/j.cardfail.2017.09.014

5. Hoste EA, Clermont G, Kersten A, Venkataraman R, Angus DC, De Bacquer D, et al. RIFLE criteria for acute kidney injury are associated with hospital mortality in critically ill patients: a cohort analysis. *Crit Care.* (2006) 10:R73. doi: 10.1186/cc4915

6. Uchino S, Bellomo R, Goldsmith D, Bates S, Ronco C. An assessment of the RIFLE criteria for acute renal failure in hospitalized patients. *Crit Care Med.* (2006) 34:1913–7. doi: 10.1097/01.CCM.0000224227.70642.4F

7. Shafie AA, Tan YP, Ng CH. Systematic review of economic burden of heart failure. *Heart Fail Rev.* (2018) 23:131–45. doi: 10.1007/s10741-017-9661-0

8. Damman K, Voors AA, Navis G, van Veldhuisen DJ, Hillege HL. Current and novel renal biomarkers in heart failure. *Heart Fail Rev.* (2012) 17:241–50. doi: 10.1007/s10741-011-9254-2

9. Fan Z, Li Y, Ji H, Jian X. Nomogram model to predict cardiorenal syndrome type 1 in patients with acute heart failure. *Kidney Blood Press Res.* (2018) 43:1832–41. doi: 10.1159/000495815

10. Schanz M, Shi J, Wasser C, Alscher MD, Kimmel M. Urinary [TIMP-2] × [IGFBP7] for risk prediction of acute kidney injury in decompensated heart failure. *Clin Cardiol.* (2017) 40:485–91. doi: 10.1002/clc.22683

11. Deo RC. Machine learning in medicine. *Circulation.* (2015) 132:1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593

12. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. *J Intern Med.* (2018) 284:603–19. doi: 10.1111/joim.12822

13. Fleuren LM, Klausch TLT, Zwager CL, Schoonmade LJ, Guo T, Roggeveen LF, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive Care Med.* (2020) 46:383–400. doi: 10.1007/s00134-019-05872-y

14. Han X, Zheng X, Wang Y, Sun X, Xiao Y, Tang Y, et al. Random forest can accurately predict the development of end-stage renal disease in immunoglobulin a nephropathy patient. *Ann Transl Med.* (2019) 7:234. doi: 10.21037/atm.2018.12.11

15. Radakovich N, Nagy M, Nazha A. Machine learning in haematological malignancies. *Lancet Haematol.* (2020) 7:e541–50. doi: 10.1016/S2352-3026(20)30121-6

16. Kimura K, Tabe Y, Ai T, Takehara I, Fukuda H, Takahashi H, et al. A novel automated image analysis system using deep convolutional neural networks can assist to differentiate MDS and AA. *Sci Rep.* (2019) 9:13385. doi: 10.1038/s41598-019-49942-z

17. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature.* (2019) 572:116–9. doi: 10.1038/s41586-019-1390-1

18. Barbieri C, Molina M, Ponce P, Tothova M, Cattinelli I, Ion Titapiccolo J, et al. An international observational study suggests that artificial intelligence for clinical decision support optimizes anemia management in hemodialysis patients. *Kidney Int.* (2016) 90:422–9. doi: 10.1016/j.kint.2016.03.036

19. Yang J, Li Y, Liu Q, Li L, Feng A, Wang T, et al. Brief introduction of medical database and data mining technology in big data era. *J Evid Based Med.* (2020) 13:57–69. doi: 10.1111/jebm.12373

20. Wu WT, Li YJ, Feng AZ, Li L, Huang T, Xu AD, et al. Data mining in clinical big data: the frequently used databases, steps, and methodological models. *Mil Med Res.* (2021) 8:44. doi: 10.1186/s40779-021-00338-z

21. Johnson A, Bulgarelli L, Pollard T, Horng S, Celi LA, Mark R. *Physionet. MIMIC-IV (Version 1.0).* (2021). Available online at: https://Doi.Org/10.13026/S6n6-Xd98 (accessed Oct. 3, 2019).

22. Lameire NH, Levin A, Kellum JA, Cheung M, Jadoul M, Winkelmayer WC, et al. Harmonizing acute and chronic kidney disease definition and classification: report of a Kidney Disease: Improving Global Outcomes (KDIGO) consensus conference. *Kidney Int.* (2021) 100:516–26. doi: 10.1016/j.kint.2021.06.028

23. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF III, Feldman HI, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med.* (2009) 150:604–12. doi: 10.7326/0003-4819-150-9-200905050-00006

24. Nakada Y, Kawakami R, Matsui M, Ueda T, Nakano T, Takitsume A, et al. Prognostic value of urinary neutrophil gelatinase-associated lipocalin on the first day of admission for adverse events in patients with acute decompensated heart failure. *J Am Heart Assoc.* (2017) 6:e004582. doi: 10.1161/JAHA.116.004582

25. Yang X, Chen C, Tian J, Zha Y, Xiong Y, Sun Z, et al. Urinary angiotensinogen level predicts AKI in acute decompensated heart failure: a prospective, two-stage study. *J Am Soc Nephrol.* (2015) 26:2032–41. doi: 10.1681/ASN.2014040408

26. Zhou LZ, Yang XB, Guan Y, Xu X, Tan MT, Hou FF, et al. Development and validation of a risk score for prediction of acute kidney injury in patients with acute decompensated heart failure: a prospective cohort study in China. *J Am Heart Assoc.* (2016) 5:e004035. doi: 10.1161/JAHA.116.004035

27. Naruse H, Ishii J, Takahashi H, Kitagawa F, Nishimura H, Kawai H, et al. Predicting acute kidney injury using urinary liver-type fatty-acid binding protein and serum N-terminal pro-B-type natriuretic peptide levels in patients treated at medical cardiac intensive care units. *Crit Care.* (2018) 22:197. doi: 10.1186/s13054-018-2120-z

28. Assegie TA, Sushma SJ, Bhavya BG, Padmashree S. Correlation analysis for determining effective data in machine learning: detection of heart failure. *SN Comput.* (2021) 2:213. doi: 10.1007/s42979-021-00617-5

29. Suresh T, Assegie TA, Rajkumar S, Kumar NK. A hybrid approach to medical decision-making: diagnosis of heart disease with machine-learning model. *Int J Elec Comp Eng.* (2022) 12:1831–8. doi: 10.11591/ijece.v12i2

30. Speiser JL, Miller ME, Tooze J, Ip E. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst Appl.* (2019) 134:93–101. doi: 10.1016/j.eswa.2019.05.028

31. Hoste E, Bihorac A, Al-Khafaji A, Ortega LM, Ostermann M, Haase M, et al. Identification and validation of biomarkers of persistent acute kidney injury: the RUBY study. *Intensive Care Med.* (2020) 46:943–53. doi: 10.1007/s00134-019-05919-0

32. Xue W, Xie Y, Wang Q, Xu W, Mou S, Ni Z. Diagnostic performance of urinary kidney injury molecule-1 and neutrophil gelatinase-associated lipocalin for acute kidney injury in an obstructive nephropathy patient. *Nephrology.* (2014) 19:186–94. doi: 10.1111/nep.12173

33. Raith EP, Udy AA, Bailey M, McGloughlin S, MacIsaac C, Bellomo R, et al. Prognostic accuracy of the SOFA score, SIRS criteria, and qSOFA score for in-hospital mortality among adults with suspected infection admitted to the intensive care unit. *JAMA.* (2017) 317:290–300. doi: 10.1001/jama.2016.20328

34. Lee CW, Kou HW, Chou HS, Chou HH, Huang SF, Chang CH, et al. A combination of SOFA score and biomarkers gives a better prediction of septic AKI and in-hospital mortality in critically ill surgical patients: a pilot study. *World J Emerg Surg.* (2018) 13:41. doi: 10.1186/s13017-018-0202-5

35. Xu S, Chen YH, Tan ZX, Xie DD, Zhang C, Xia MZ, et al. Vitamin D3 pretreatment alleviates renal oxidative stress in lipopolysaccharide-induced acute kidney injury. *J Steroid Biochem Mol Biol.* (2015) 152:133–41. doi: 10.1016/j.jsbmb.2015.05.009

36. Quoilin C, Mouithys-Mickalad A, Lécart S, Fontaine-Aupart MP, Hoebeke M. Evidence of oxidative stress and mitochondrial respiratory chain dysfunction in an *in vitro* model of sepsis-induced kidney injury. *Biochim Biophys Acta.* (2014) 1837:1790–800. doi: 10.1016/j.bbabio.2014.07.005

37. Chen Y, Jin S, Teng X, Hu Z, Zhang Z, Qiu X, et al. Hydrogen sulfide attenuates lps-induced acute kidney injury by inhibiting inflammation and oxidative stress. *Oxid Med Cell Longev.* (2018) 2018:6717212. doi: 10.1155/2018/6717212

38. Shen Y, Ru W, Cao L, Jiang R, Xu X. Impact of partial pressure of oxygen trajectories on the incidence of acute kidney injury in patients undergoing cardiopulmonary bypass. *J Cardiol.* (2022) 79:545–50. doi: 10.1016/j.jjcc.2021

39. Bae J, Kim J, Lee S, Ju JW, Cho YJ, Kim TK, et al. Association between intraoperative hyperoxia and acute kidney injury after cardiac surgery: a retrospective observational study. *J Cardiothorac Vasc Anesth.* (2021) 35:2405–14. doi: 10.1053/j.jvca.2020.11.054

40. Tarvasmäki T, Haapio M, Mebazaa A, Sionis A, Silva-Cardoso J, Tolppanen H, et al. Acute kidney injury in cardiogenic shock: definitions, incidence, haemodynamic alterations, and mortality. *Eur J Heart Fail.* (2018) 20:572–81. doi: 10.1002/ejhf.958

41. Patel RB, Fonarow GC, Greene SJ, Zhang S, Alhanti B, DeVore AD, et al. Kidney function and outcomes in patients hospitalized with heart failure. *J Am Coll Cardiol.* (2021) 78:330–43. doi: 10.1016/j.jacc.2021.05.002

frontiers | Frontiers in Cardiovascular Medicine

TYPE Original Research
PUBLISHED 15 September 2022
DOI 10.3389/fcvm.2022.939103

Check for updates

OPEN ACCESS

EDITED BY
Gaurang Vaidya,
University of Kentucky, United States

REVIEWED BY
Jiandong Zhou,
University of Oxford, United Kingdom
Emma Louise Robinson,
University of Colorado, United States

*CORRESPONDENCE
Junzhang Tian
jz.tian@163.com
Zhongzhi Xu
zhongzxu@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Heart Failure and Transplantation, a section of the journal Frontiers in Cardiovascular Medicine

RECEIVED 08 May 2022
ACCEPTED 01 August 2022
PUBLISHED 15 September 2022

CITATION
Cheng W, Du Y, Zhang Q, Wang X, He C, He J, Jing F, Ren H, Guo M, Tian J and Xu Z (2022) Age-related changes in the risk of high blood pressure.
Front. Cardiovasc. Med. 9:939103.
doi: 10.3389/fcvm.2022.939103

COPYRIGHT
© 2022 Cheng, Du, Zhang, Wang, He, He, Jing, Ren, Guo, Tian and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Age-related changes in the risk of high blood pressure

Weibin Cheng[1,2,3†], Yumeng Du[2†], Qingpeng Zhang[3], Xin Wang[4], Chaocheng He[5], Jingjun He[1], Fengshi Jing[2], Hao Ren[2], Mengzhuo Guo[3], Junzhang Tian[1*] and Zhongzhi Xu[2,3*]

[1]Department of Health Management, Guangdong Second Provincial General Hospital, Guangzhou, China, [2]Institute for Healthcare Artificial Intelligence Application, Guangdong Second Provincial General Hospital, Guangzhou, China, [3]School of Data Science, City University of Hong Kong, Hong Kong, Hong Kong SAR, China, [4]School of Artificial Intelligence, Tianjin University, Tianjin, China, [5]School of Information Management, Wuhan University, Wuhan, China

Background and aims: Understanding the age-related trend of risk in high blood pressure (BP) is important for preventing heart failure and cardiovascular diseases. But such a trend is still underexplored. This study aims to (a) depict the relationship of BP patterns with age, and (b) understand the trend of high BP prevalence over time in different age groups.

Materials and methods: Health check-up data with an observational period of 8 years (January 1, 2011, to December 31, 2018) was used as the data source. A total of 71,468 participants aged over 18 years old with complete information on weight, height, age, gender, glucose, triglyceride, total cholesterol, systolic (SBP), and diastolic blood pressure (DBP) were included for analysis. Generalized additive models were adopted to explore the relationship between the risk of high BP and age. Variance analysis was conducted by testing the trend of high BP prevalence in age groups over time.

Results: Risk of high SBP showed a continuous rise from age 35 to 79 years and a concurrent early increase in the risk of high DBP; after age 50–65 years, high DBP risk declined. The risk of SBP rises linearly with age for men, whereas increases non-linearly for women. In addition, a significant increasing trend of high SBP risk among middle-aged people was found during the past decade, men experienced a later but longer period of increase in high SBP than women.

Conclusion: The high SBP risk progresses more rapidly in the early lifetime in women, compared to the lifetime thereafter. Thresholds of increasing trend of SBP suggest a possible need for hypertension screening in China after the age of 40.

KEYWORDS

age-related trend, high blood pressure, generalized additive models, South China, heart failure

Background

High blood pressure (BP), including high systolic blood pressure (SBP) and high diastolic blood pressure (DBP), has been recognized as triggers of the leading cause of mortality in the 21st century (1). It is estimated that an annual death of 7.7–10.4 million were associated with elevated blood pressure world-wide (2). High BP is also associated with the strongest evidence for causation of cardiovascular diseases (CVD) (3). The most important one was based on experience in 61 cohort studies that provided 12.7 million person-years of observation. The risk of CVD increased steadily with progressively higher levels of baseline SBP and DBP (4). For example, at ages 40–69 years, each difference of 20 mm Hg usual SBP is associated with more than a twofold difference in the stroke death rate (4).

Blood pressure is highly age-dependent. Existing literature has documented the age-related trend of absolute BP, which shows a linear rise of SBP with age after 30–40 years old (5–8) and reaching a plateau in late life (9). For DBP, an inverse U-shaped age-related trend was identified, with its peak at 40–60 years old (5–8). However, literature on this line manually split continuous age into discrete age groups, bringing in artificial influence. Moreover, age-BP trend is formulated by counting the number of subjects that fall into handcrafted age intervals. Therefore, there is a lack of adjustment for confounding factors. We found only one recent study (10) that enables an arbitrary complex age-BP relationship when investigating the life course BP trajectory, by using restricted cubic splines-based regression models. The authors focused on exploring sex differences in blood pressure trajectories over a lifetime, but the investigation of the protective or risk effect of aging and other covariates on high BP risk is lacking.

Previous literature also reported mixed findings about the high BP trend in developing countries (11–13): Studies conducted in Mexico and China reported that hypertension is shifting toward younger ages over time (11), whereas researchers from Iran found that younger generations are at lower risk of developing hypertension (13).

Against this backdrop, we present an analysis of 8 years' routine health check-up data with the aim of (a) articulating the relationship of BP patterns with age using advanced modeling techniques, and (b) understanding the trend of blood pressure over time in different age groups. For task (a), we employed generalized additive models (GAM) with interaction items (GA^2M). Compared to linear models and mixed linear models which put strict constraints on the potentially complex relationships between a dependent variable and independent variables, GA^2M permits arbitrary complex relationships between the two. Thus, it is potentially capable of sophisticatedly depicting the age-related BP elevation risk. For task (b), the large-scale dataset used in this study creates a unique opportunity to further the current understanding of the trend of high BP prevalence over time in the general population in south China. The findings of this article could provide insights that can aid high BP prevention.

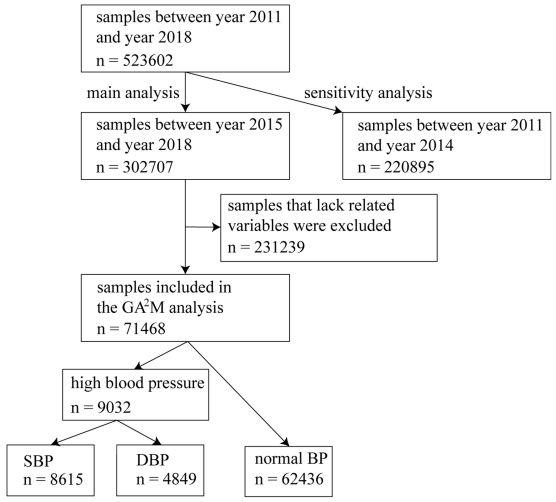


FIGURE 1
Sample selection process.

Participants and methods

Data source

This study followed a cross-sectional design. Health check-up data with an 8-year (January 1, 2011, to December 31, 2018) observational period from Guangdong Second Provincial General Hospital, Guangdong, China was used in this study. Records from January 1, 2015, to December 31, 2018 are used in the main analysis. Records in the earlier period (January 1, 2011, to December 31, 2014) were used for model validation. A health check-up was conducted in the physical examination department in Guangdong Second Provincial General Hospital. Check-up records would be eligible for analysis if the participants were aged 18 years or older, and had complete information on the participants' weight, height, age, gender, glucose (GLU), triglyceride (TG), total cholesterol (TC), and systolic blood pressure and diastolic blood pressure. **Figure 1** demonstrates the sample selection process.

Variables and measurements

Outcome variable

The outcome was defined as a binary variable (i.e., whether the participants have high SBP/DBP) according to their BP measurements. Participants were told not to drink a caffeinated beverage or smoke during the 30 min before the BP examination. They were also required to sit quietly for 5 min

TABLE 1 Characteristics of research samples.

	High blood pressure (n = 9032)	Normal blood pressure (n = 62436)	Significance test
	Mean (SD)/n (%)	Mean (SD)/n (%)	
Age	54.1 (16.5)	37.5 (12.8)	Statistic = 4.4×10^8, $p = 0.000$
BMI (kg/m^2)	25.3 (3.59)	22.9 (3.34)	Statistic = 3.8×10^8, $p = 0.000$
GLU (mmol/L)	5.60 (1.76)	4.95 (0.96)	Statistic = 3.7×10^8, $p = 0.000$
TG (mmol/L)	1.95 (1.60)	1.44 (1.13)	Statistic = 3.7×10^8, $p = 0.000$
Men	6494 (71.9%)	37524 (60.1%)	$\chi^2 = 5.7$, $p = 0.001$

before the test begins. To measure BP, participants shall sit in a chair with feet on the floor and right arm supported such that the elbow is at about heart level. A second measurement would be taken if the first reading is above the normal limit (i.e., >90 mmHg for DBP or >140 mmHg for SBP). An automatic electronic blood pressure monitor was applied to BP test. Only one reading for SBP/DBP measurement was recorded for each person, the second reading would be put down if the blood pressure was measured twice. Participants with high SBP/DBP were annotated as positive cases, with a dependent variable (y 1), those with SBP/DBP below the normal limit are classified as negative counterparts (y 0).

Independent variable

The age of the participants being recorded while taking the physical examination was regarded as the main independent variable. Other independent covariates included body mass index (i.e., person's weight in kilograms divided by the square of height in meters), glucose (GLU), triglyceride (TG), and total cholesterol (TC). All the independent variables were used as continuous variables.

TABLE 2 Prevalence of high SBP/DBP in different age groups among men.

Age groups	No. of records	No. of high SBP (%)	No. of high DBP (%)
18–39	26934	1766 (6.56%)	1182 (4.39%)
40–49	7026	944 (13.44%)	971 (13.82%)
50–59	4858	1129 (23.24%)	922 (18.98%)
60–69	3083	1139 (36.94%)	584 (18.94%)
70–79	1430	754 (52.73%)	199 (13.92%)
>=80	679	388 (57.14%)	51 (7.51%)

TABLE 3 Prevalence of high SBP/DBP in different age groups among women.

Age groups	No. of records	No. of high SBP (%)	No. of high DBP (%)
18–39	16505	150 (0.91%)	201 (1.22%)
40–49	4508	282 (6.26%)	194 (4.30%)
50–59	2997	548 (18.28%)	281 (9.38%)
60–69	2159	792 (36.68%)	195 (9.03%)
70–79	1068	593 (55.52%)	61 (5.71%)
>=80	221	130 (58.82%)	8 (3.62%)

Analysis

Empirical analysis

We conducted the empirical analysis by calculating the high SBP/DPB prevalence (without any adjustment for confounding influences), which is simply dividing the number of high SBP/DBP individuals by the total research samples for each age group. We also summarized the characteristics of the research samples. Age, BMI, GLU, and TG are continuous variables, for which we reported the mean and standard deviation (SD). Gender is a binary feature, so we calculated and reported the number and percentage of men. Mann–Whitney U test and χ^2 test with Yates' correction were performed to explore the statistical significance of the continuous variables and the binary variable, respectively.

Main analysis

We formulated the first task of this study, which aims to explore the relationship between BP elevation risk and age, as a binary classification problem. The relationship was adjusted for covariates including Body Mass Index (BMI), glucose (GLU), triglyceride (TG), and total cholesterol (TC), to control the confounders. A generalized additive model (GAM) was adopted to explore the relationship between dependent and independent variables. A standard GAM has the form:

$$g(y) = \beta_0 + \sum f_i(x_i)$$

Where $g(\cdot)$ is the link function and $f_i(\cdot)$ is the boosted trees-based shape function for term i. Compared to the standard logistic regression model, which has a form of $g(y)$ $1/1 + e^{-w_i x_i}$, GAM permits arbitrary complex relationships between individual features and the target to be captured while avoiding overfitting. Therefore, it is a powerful tool to help us understand the age-related BP elevation risk. Pairwise interactions can be added to standard GAMs, leading to

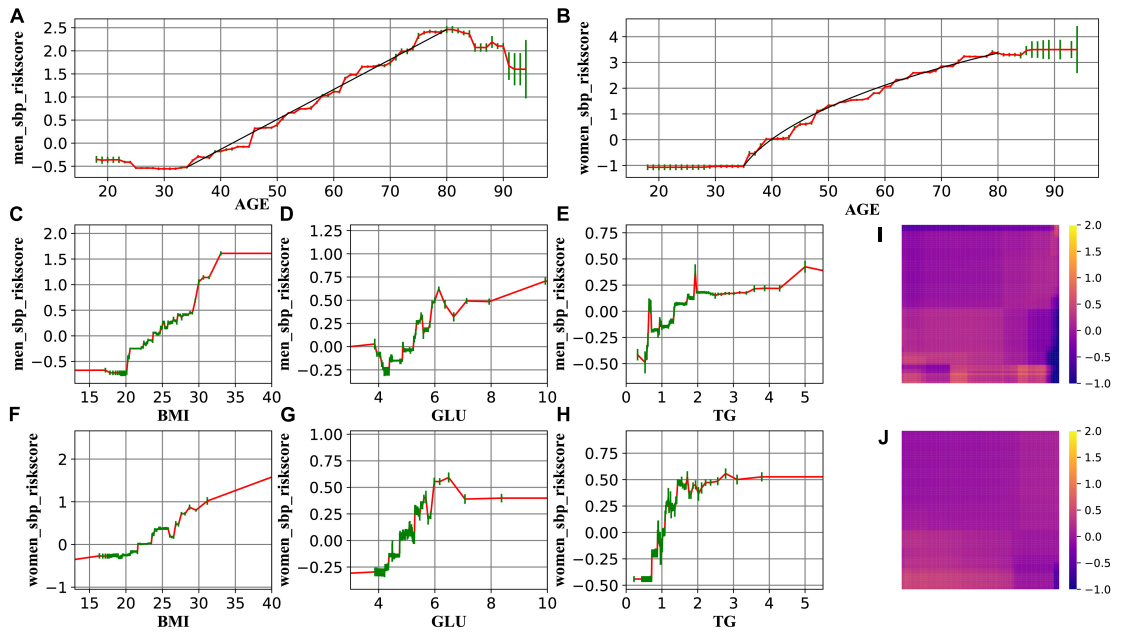


FIGURE 2
The change in the risk of high SBP related to age (**A** for men and **B** for women), BMI (**C** for men and **F** for women), GLU (**D** for men and **G** for women), and TG (**E** for men and **H** for women). **(I,J)** Demonstrate that the interaction effect is marginal. Red lines represent the spline-based GAM curves. Green lines represent the confidence intervals. Black lines in **A** and **B** are drawn to guide the eyes.

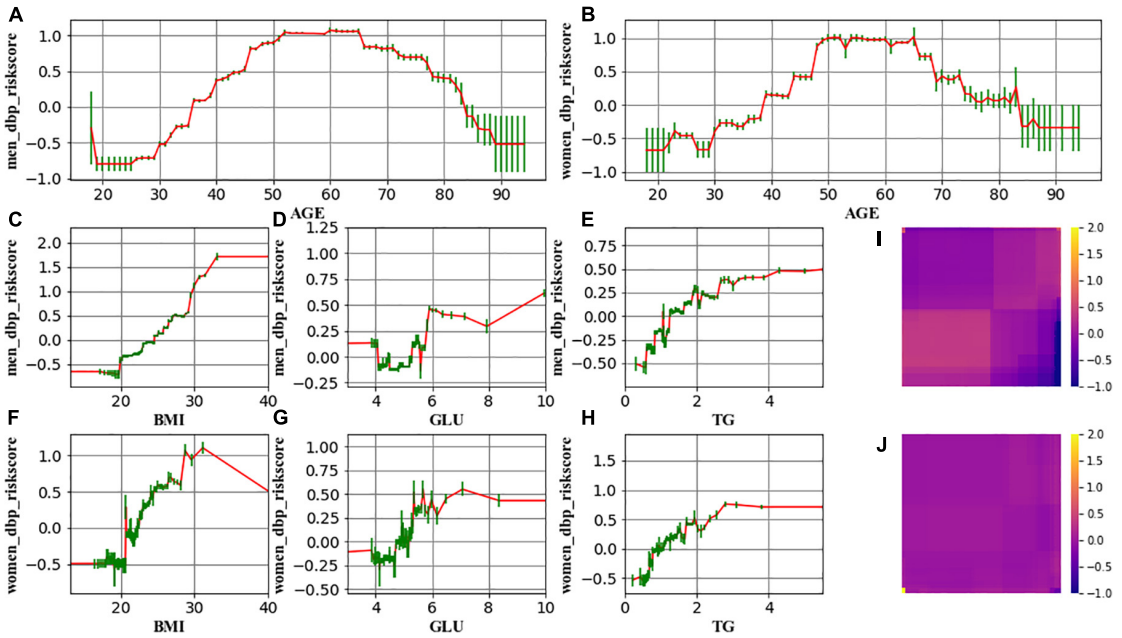


FIGURE 3
The change in the risk of high DBP related to age (**A** for men and **B** for women), BMI (**C** for men and **F** for women), GLU (**D** for men and **G** for women), and TG (**E** for men and **H** for women). **(I,J)** Demonstrate that the interaction effect is marginal. Red lines represent the spline-based GAM curves. Green lines represent the confidence intervals.

a model called GA^{2}M. We also performed the standard logistic regression analysis using the same samples and covariables of the GA^{2}M analysis. We adopted the Receiver Operating Characteristic of the Area Under the Curve (ROCAUC, also known as c-statistic) to evaluate the performances of GA^{2}M and the logistic regression model.

TABLE 4 Transition points for age, BMI, GLU, and TG in terms of the risk of high SBP.

	Men	Women
Age group	40–50	40–50
BMI interval	20–25	20–25
GLU interval (mmol/L)	4–6	4–6
TG (mmol/L)	1.4	1

TABLE 5 Transition points for age, BMI, GLU, and TG in terms of the risk of high DBP.

	Men	Women
Age group	40–50; 80–85	40–50; 80–85
BMI interval	20–25	20–25
GLU interval (mmol/L)	4–6	4–6
TG (mmol/L)	1.3	0.9

To cope with the second task that focuses on understanding the overall trend of high BP prevalence over the 8-year observational period within the different age groups, variance analysis was conducted by testing in five age groups if the variance of BP over time is statistically significant. A p value of <0.05 was considered statistically significant. Python (version 3.2.5) was used for analysis, visualization, and statistical tests. Mann–Whitney U test and χ^2 test were conducted using *scipy.stats* package, Mann–Kendall test was conducted using *PyMannKendall*.

Model sensitivity analysis

For task 1, we conducted model sensitivity analyses using an earlier dataset (January 1, 2011, to December 31, 2014) to see if our findings are robust.

Ethical approval

This study was approved by the Ethical Review Board of Guangdong Second Provincial General Hospital.

Results

From January 1, 2015, to December 31, 2018, we observed 71,468 eligible health check-up records. Of those, 12.05% (8615/71468) were with high SBP and 6.78% (4849/71468) were with high DBP. Characteristics of the study samples are shown

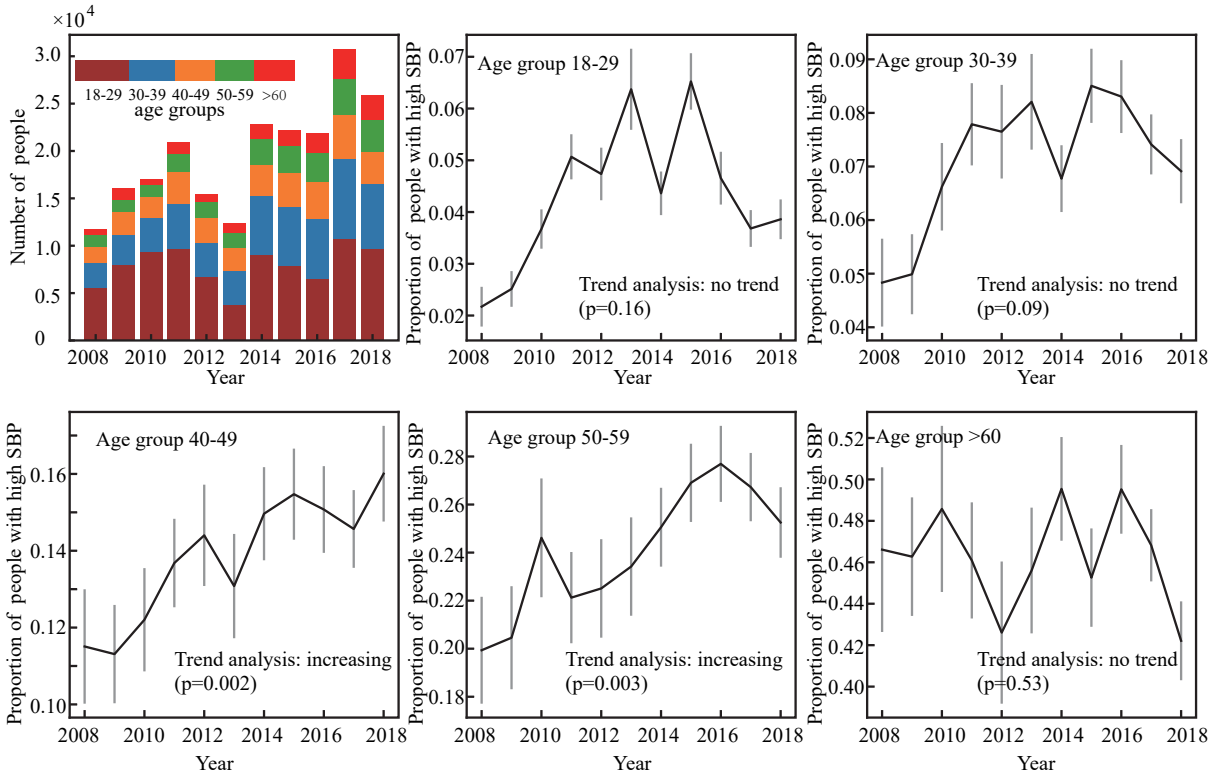


FIGURE 4
Variance of the number of high SBP men with time.

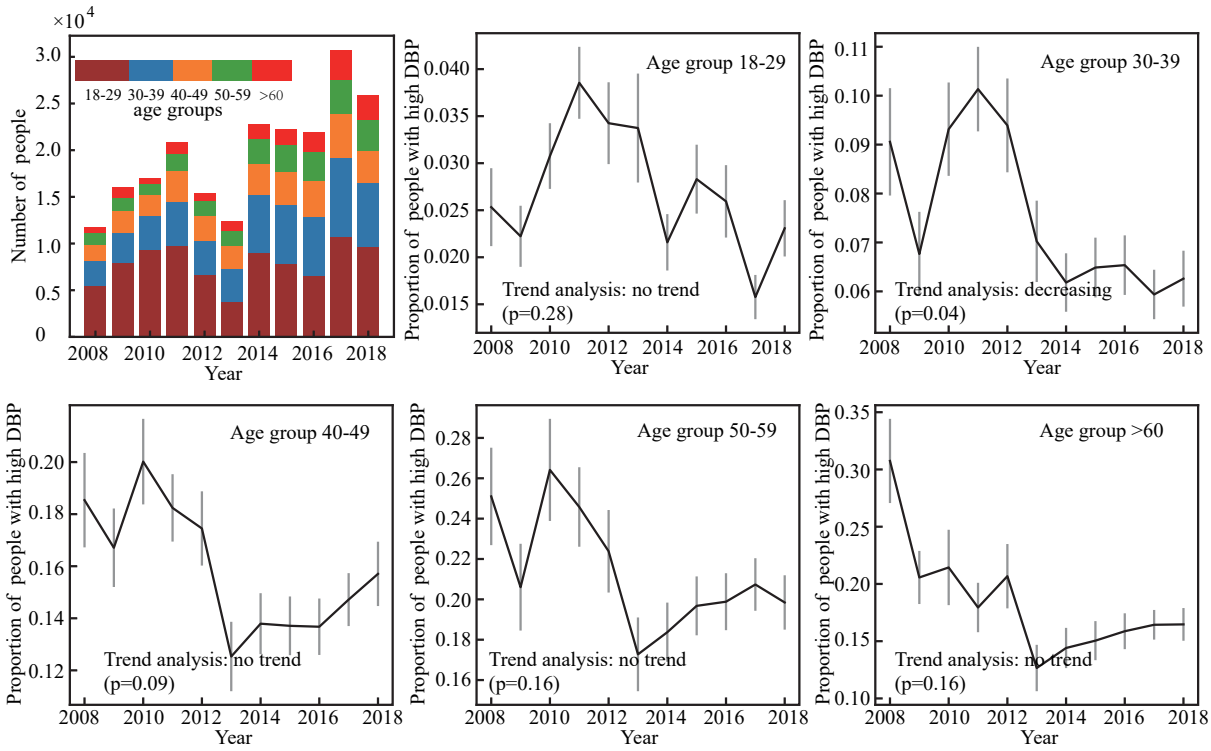


FIGURE 5
Varience of the number of high DBP male participants with time in Guangdong, China (*N* = 44010).

in **Table 1**. All variables demonstrate significant differences between the high blood pressure group and the normal group. The prevalence of high SBP/DPB in different age groups for men and women is shown in **Tables 2**, **3**.

Figures 2A,B demonstrates the relationship between age and the risk of high SBP for men and women, respectively. In general, the risk of high SBP increases with age, reaching a plateau in late life. For high DBP risk, an inverse U-shaped age-related risk trend was identified. Women compared with men exhibited a first-steeper-then-slower increase pattern in BP that began at age 42 years. The ROCAUC was calculated as 0.837 for the main GA^{2}M model and 0.803 for the standard logistic regression model. Results of the model validation analysis can be found in **Supplementary Figures 1, 2**. They exhibit very similar patterns to the main results, indicating that the patterns discovered in this study are robust and generalizable.

The resulting curves (**Figures 2**, **3**) also enable us to explore the transition points of the role of Age, BMI, GLU, and TG, by identifying the curve where the risk score (*y*-axis) rises above 0. Results of the transition point are reported in **Tables 4**, **5**.

Results of the high BP prevalence for each age group during the study period (2008–2018) were reported in **Figures 4–7**. For men, an increasing trend of high SBP risk was observed among those aged 40–49 years (p 0.002) and 50–59 years (p 0.002), and high DBP risk decreased among those aged 30–39 years (p 0.004). For women, an increasing trend of high SBP risk was witnessed among those aged 30–39 years (p 0.008), and high DBP risk decreased (p 0.005) among those aged 40–49 years. No significant trend in other age groups was observed.

Discussion

Our study depicted the age-related trend of risk in high SBP/DBP and identified the specific age group which showed significant changes in risk in high SBP/DBP prevalence over time among male and female adults. The subjects in this study showed a rise in SBP from age 35 to 79 years and a concurrent early increase in DBP; after age 50–65 years, DBP declined. GA^{2}M analysis of high SBP/DBP risk as a function of age, adjusting for BMI, GLU, and TG, showed that slope patterns of age-risk curves differed for men and women, whereas it was curvature that differed for SBP and DBP. A general picture of the BP trend in south China was also reported. For men aged 40–59 years and women aged 30–39 years, an increasing trend of high SBP was observed. For men aged 30–39 years and women aged 40–49 years, the trend of high DBP was deflated.

The present study indicated that the risk of SBP rises linearly with age for men, whereas it was a non-linear

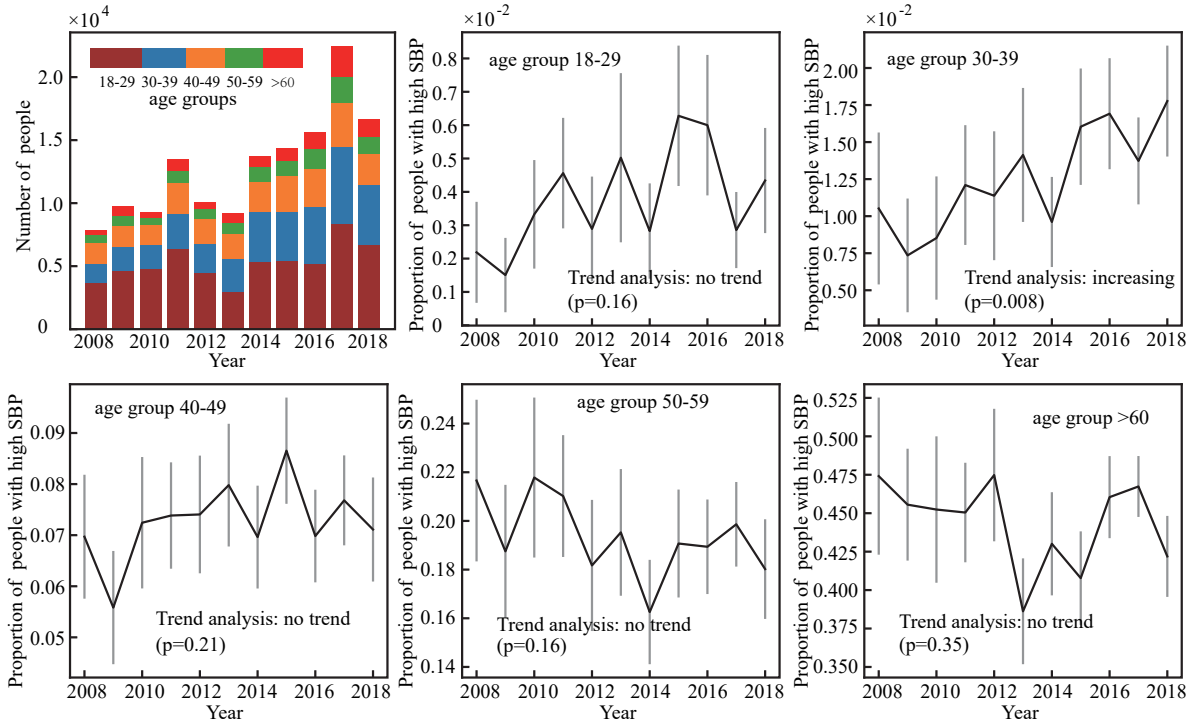


FIGURE 6
Variance of the number of high SBP women with time.

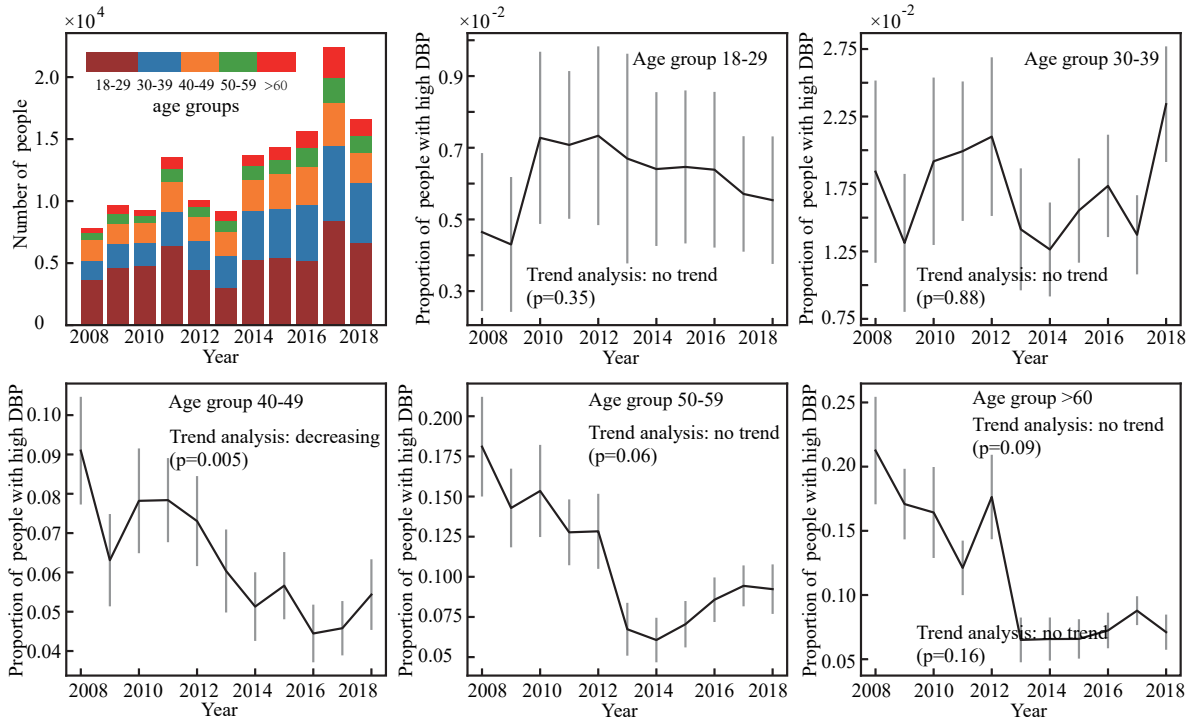


FIGURE 7
Variance of the number of high DBP women with time.

relationship between age and SBP risk for women. More concretely, SBP risk progresses more rapidly in the age period 35–55 (first phase) than lifetime thereafter (55–80, second phase). This finding differs from previous literature where a linear rise for both genders was reported (5, 7). From a methodological perspective, such discrepancy appears probably because (a) literature of this line manually splits continuous age into discrete age groups, bringing in human influence, and (b) through counting the number of subjects that fall into handcrafted age intervals, there lacks adjustment for confounding factors. This study took a step forward by leveraging the strength of GA^2M that permits arbitrary complex relationships between dependent and independent variables. The resulting c-statistics of GA^2M and a standard logistic regression model echoes our hypothesis that GA^2M enables a more precise way of depicting the feature–target relationship. Plausible explanations from the physiological perspective would be that (a) menopause might be attributable to the disparity between the two phases; (b) taking men as a reference, previous literature reports that women's blood pressure starts lower than men's, catches up around age 45, and frequently becomes slightly higher thereafter (5, 10). Therefore, SBP risk elevates more rapidly in the first phase to catch up to BP levels in men.

GA^2M analysis also enables us to quantify the transition point of the role of covariates. Such information can signal us when we should pay special attention to the measurement. For example, the transition point of the role of BMI in SBP is located in 20–25, indicating a protective role of BMI $\leq$ 20 in preventing SBP, and a harmful role of BMI $\geq$ 25 in inducing SBP. This happened to echo the stratification of BMI index alone. It is noteworthy that the transition point (80–85) in DBP seen in the elderly is probably the result rather than the cause of the disease process. Age-related stiffening of the aorta is associated with a decreased capacity of the elastic reservoir and hence a greater peripheral runoff of stroke volume during systole. Thus, with less blood remaining in the aorta at the beginning of diastole, and with diminished elastic recoil, diastolic pressure decreases with increased steepness of diastolic decay (5). The exaggerated fall in DBP risk seen in the elderly suggests a process of transmural pressure-induced arterial wall damage resulting in large artery stiffness, enforcing age $\geq$85 to be a phantom "protective role." Apart from this observation, there is a marginal difference for SBD and DBP, and for men and women with respect to the role of covariates in high BP risk. Generalized additive models have recently been used to explore the association between physical activity and mental health burden (14), to investigate the association of depression with dietary inflammatory index (15), and to predict postoperative acute kidney injury and acute respiratory failure (16). This study adds the age–blood pressure relationship to a growing list of fields of research that are actively capitalizing on GAM.

With respect to the second goal of this article, we observed an increasing trend of high SBP risk among middle-aged people during the very recent decade. This finding echoes previous evidence that the incidence of hypertension has been shifting toward younger generations (11, 12). Results also indicate that there exists gender discrepancy in terms of those experiencing an increasing high SBP risk: Age groups of men are older and wider compared to women (40–59 vs. 30–39). Plausible factors leading to such discrepancy would be sex hormones, lifestyles (e.g., alcohol intake) (17–19), social pressures (12), etc.

This study has implications for hypertension management and prevention from both disease and population perspectives. From the disease perspective, our results support the evidence that aging is an important risk factor for hypertension, for constant increase of risk in high SBP with age. According to the trends of risk in high SBP and high DBP, diagnosis of hypertension could become more dependent on SBP measurement among the elderly. From the population perspective, our study supports the threshold for significant changes in blood pressure at 40–45 years of age. The current national guideline for hypertension management in China (2019) defined people aged over 45 years as a high-risk groups for hypertensions. However, no recommendations for regular blood pressure measurement and hypertension screening have been provided for Chinese adults. Another implication informs the importance of controlling elevated blood pressure among middle-aged Chinese adults (30–59 years old) who might be the main contributors of trends of people suffering from hypertension at a younger age. About 30% of middle-aged people in China were estimated to suffer from hypertension (20), while 87% of them with ISH remain untreated (21). Future research on improving awareness, prevention, and treatment of hypertension among middle-aged adults in China could be taken.

Although analyses of this study were based on a well-constructed model and a large volume dataset, several inherent deficits of routine health check-up data should be noted when interpreting the results. First, a check-up BP record relies on a one-off BP test, but tests conducted more than one time on different days are recommended to acquire more solid BP results. Secondly, this study reports findings from a single center, whether it is generalizable to populations elsewhere needs further investigation.

Conclusion

In contrast with the notion that systolic blood pressure linearly increases with age in individuals, our results suggested that SBP risk progresses more rapidly in early lifetime in women, compared to lifetime thereafter. Thresholds of the increasing trend of SBP suggest a possible need for hypertension screening in China after the age of 40. The changing risk of high BP in

different age groups in the past decade (from 2008 to 2018) suggested the necessity of controlling elevated blood pressure of middle-aged Chinese adults to avoid the rejuvenation of hypertension among younger adults.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethical Review Board of Guangdong Second Provincial General Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

WC, ZX, and JT: conceptualization. YD, WC, and ZX: writing – original draft. ZX, YD, WC, QZ, XW, CH, JH, FJ, MG, and HR: review and editing. WC, QZ, and JT: supervision. WC and JT: project administration and funding acquisition. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by the National Key-Area Research and Development Program of China (2021YFC2009400).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcvm.2022.939103/full#supplementary-material

References

1. WHO. *The Top 10 Causes of Death – Factsheet. WHO Reports.* Geneva: WHO (2020).

2. Zhou B, Perel P, Mensah GA. Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension. *Nat Rev Cardiol.* (2021) 18:785–802. doi: 10.1038/s41569-021-00559-8

3. Fuchs FD, Whelton PK. High blood pressure and cardiovascular disease. *Hypertension.* (2020) 75:285–92. doi: 10.1161/HYPERTENSIONAHA.119.14240

4. Lewington S, Clark R, Qizilbash N, Peto R, Collins R. Mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet.* (2002) 360:1903–13.

5. Franklin SS, Gustin W, Wong ND, Larson MG, Weber MA, Kannel WB, et al. Hemodynamic patterns of age-related changes in blood pressure: the Framingham heart study. *Circulation.* (1997) 96:308–15. doi: 10.1161/01.CIR.96.1.308

6. Lee D-T, Lee Y-S. The age-related trend in blood pressure and the prevalence of hypertension in Korean adults. *J Life Sci.* (2012) 22: 148–55.

7. Satoh M, Metoki H, Asayama K, Murakami T, Inoue R, Tsubota-Utsugi M, et al. Age-related trends in home blood pressure, home pulse rate, and day-to-day blood pressure and pulse rate variability based on longitudinal cohort data: the Ohasama study. *J Am Heart Assoc.* (2019) 8:e012121. doi: 10.1161/JAHA.119.012121

8. Ntineri A, Stergiou GS, Thijs L, Asayama K, Boggia J, Boubouchairopoulou N, et al. Relationship between office and home blood pressure with increasing age: the International Database of HOme blood pressure in relation to Cardiovascular Outcome (IDHOCO). *Hypertens Res.* (2016) 39:612–7. doi: 10.1038/hr.2016.32

9. Zhou B, Danaei G, Stevens GA, Bixby H, Taddei C, Carrillo-Larco RM, et al. Long-term and recent trends in hypertension awareness, treatment, and control in 12 high-income countries: an analysis of 123 nationally representative surveys. *Lancet.* (2019) 394:639–51. doi: 10.1016/S0140-6736(19)31145-6

10. Ji H, Kim A, Ebinger JE, Niiranen TJ, Claggett BL, Bairey Merz CN, et al. Sex differences in blood pressure trajectories over the life course. *JAMA Cardiol.* (2020) 5:255–62. doi: 10.1001/jamacardio.2019.5306

11. Castro-Porras LV, Rojas-Martínez R, Aguilar-Salinas CA, Bello-Chavolla OY, Becerril-Gutierrez C, Escamilla-Nuñez C. Trends and age-period-cohort effects on hypertension mortality rates from 1998 to 2018 in Mexico. *Sci Rep.* (2021) 11:1–9. doi: 10.1038/s41598-021-96175-0

12. Tu Y-P, Jennings R, Hart B, Cangelosi GA, Wood RC, Wehber K, et al. Swabs collected by patients or health care workers for SARS-CoV-2 testing. *N Engl J Med.* (2020) 383:494–6. doi: 10.1056/nejmc2016321

13. Hosseini M, Yousefifard M, Baikpour M, Rafei A, Fayaz M, Heshmat R, et al. Twenty-year dynamics of hypertension in Iranian adults: age, period, and cohort analysis. *J Am Soc Hypertens.* (2015) 9:925–34. doi: 10.1016/j.jash.2015.09.005

14. Xu P, Huang Y, Hou Q, Cheng J, Ren Z, Ye R, et al. Relationship between physical activity and mental health in a national representative cross-section study: its variations according to obesity and comorbidity. *J Affect Disord.* (2022) 308:484–93. doi: 10.1016/j.jad.2022.04.037

15. Jiang C, Yin H, Liu A, Liu Q, Ma H, Geng Q. Dietary inflammatory index and depression risk in patients with chronic diseases and comorbidity. *J Affect Disord.* (2022) 301:307–14. doi: 10.1016/j.jad.2022.01.008

16. Cui Z, Fritz BA, King CR, Avidan MS, Chen Y. A factored generalized additive model for clinical decision support in the operating room. *AMIA Annu Symp Proc.* (2019) 2019:343–52.

17. Reckelhoff JF. Gender differences in the regulation of blood pressure. *Hypertension.* (2001) 37:1199–208. doi: 10.1161/01.HYP.37.5.1199

18. Millwood IY, Li L, Smith M, Guo Y, Yang L, Bian Z, et al. Alcohol consumption in 0.5 million people from 10 diverse regions of China: prevalence, patterns and socio-demographic and health-related correlates. *Int J Epidemiol.* (2013) 42:816–27. doi: 10.1093/ije/dyt078

19. Mi YJ, Zhang B, Wang HJ, Yan J, Han W, Zhao J, et al. Prevalence and secular trends in obesity among Chinese adults, 1991-2011. *Am J Prev Med.* (2015) 49:661–9.

20. Xia C. Prevalence of hypertension in older and middle-aged adults in China: estimates from two national longitudinal surveys. *Glob Clin Transl Res.* (2020) 2:78–84. doi: 10.36316/gcatr.02.0034

21. Mahajan S, Feng F, Hu S, Lu Y, Gupta A, Murugiah K, et al. Assessment of prevalence, awareness, and characteristics of isolated systolic hypertension among younger and middle-aged adults in China. *JAMA Netw Open.* (2020) 3:e209743. doi: 10.1001/jamanetworkopen.2020.9743

frontiers | Frontiers in Cardiovascular Medicine

TYPE Original Research
PUBLISHED 23 September 2022
DOI 10.3389/fcvm.2022.980625

Check for updates

OPEN ACCESS

EDITED BY
Qingpeng Zhang,
City University of Hong Kong,
Hong Kong SAR, China

REVIEWED BY
Hamayak Sisakian,
Yerevan State Medical
University, Armenia
Anne Pizard,
INSERM U955 Institut Mondor de
Recherche Biomédicale (IMRB), France

*CORRESPONDENCE
William E. Sanders
wsanders@corvista.com

SPECIALTY SECTION
This article was submitted to
Heart Failure and Transplantation,
a section of the journal
Frontiers in Cardiovascular Medicine

RECEIVED 28 June 2022
ACCEPTED 22 August 2022
PUBLISHED 23 September 2022

CITATION
Burton T, Ramchandani S, Bhavnani SP, Khedraki R, Cohoon TJ, Stuckey TD, Steuter JA, Meine FJ, Bennett BA, Carroll WS, Lange E, Fathieh F, Khosousi A, Rabbat M and Sanders WE (2022) Identifying novel phenotypes of elevated left ventricular end diastolic pressure using hierarchical clustering of features derived from electromechanical waveform data. *Front. Cardiovasc. Med.* 9:980625. doi: 10.3389/fcvm.2022.980625

COPYRIGHT
© 2022 Burton, Ramchandani, Bhavnani, Khedraki, Cohoon, Stuckey, Steuter, Meine, Bennett, Carroll, Lange, Fathieh, Khosousi, Rabbat and Sanders. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Identifying novel phenotypes of elevated left ventricular end diastolic pressure using hierarchical clustering of features derived from electromechanical waveform data

Timothy Burton[1], Shyam Ramchandani[1], Sanjeev P. Bhavnani[2], Rola Khedraki[2], Travis J. Cohoon[2], Thomas D. Stuckey[3], John A. Steuter[4], Frederick J. Meine[5], Brett A. Bennett[6], William S. Carroll[7], Emmanuel Lange[1], Farhad Fathieh[1], Ali Khosousi[1], Mark Rabbat[8] and William E. Sanders[9,10]*

[1]CorVista Health (Analytics For Life Inc., d.b.a CorVista Health) Toronto, Toronto, ON, Canada, [2]Scripps Clinic Division of Cardiology, San Diego, CA, United States, [3]Cone Health Heart and Vascular Center, Greensboro, NC, United States, [4]Bryan Heart, Lincoln, NE, United States, [5]Novant Health New Hanover Regional Medical Center, Wilmington, NC, United States, [6]Jackson Heart Clinic, Jackson, MS, United States, [7]Cardiology Associates of North Mississippi, Tupelo, MS, United States, [8]Division of Cardiology, Loyola University Medical Center, Maywood, IL, United States, [9]CorVista Health, Inc., Washington, DC, United States, [10]University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

Introduction: Elevated left ventricular end diastolic pressure (LVEDP) is a consequence of compromised left ventricular compliance and an important measure of myocardial dysfunction. An algorithm was developed to predict elevated LVEDP utilizing electro-mechanical (EM) waveform features. We examined the hierarchical clustering of selected features developed from these EM waveforms in order to identify important patient subgroups and assess their possible prognostic significance.

Materials and methods: Patients presenting with cardiovascular symptoms ($N = 396$) underwent EM data collection and direct LVEDP measurement by left heart catheterization. LVEDP was classified as non-elevated ($\leq$12 mmHg) or elevated ($\geq$25 mmHg). The 30 most contributive features to the algorithm output were extracted from EM data and input to an unsupervised hierarchical clustering algorithm. The resultant dendrogram was divided into five clusters, and patient metadata overlaid.

Results: The cluster with highest LVEDP (cluster 1) was most dissimilar from the lowest LVEDP cluster (cluster 5) in both clustering and with respect to clinical characteristics. In contrast to the cluster demonstrating the highest percentage of elevated LVEDP patients, the lowest was predominantly non-elevated LVEDP, younger, lower BMI, and males with a higher rate of significant coronary artery disease (CAD). The next adjacent cluster (cluster 2) to that of

the highest LVEDP (cluster 1) had the second lowest LVEDP of all clusters. Cluster 2 differed from Cluster 1 primarily based on features extracted from the electrical data, and those that quantified predictability and variability of the signal. There was a low predictability and high variability in the highest LVEDP cluster 1, and the opposite in adjacent cluster 2.

Conclusion: This analysis identified subgroups of patients with varying degrees of LVEDP elevation based on waveform features. An approach to stratify movement between clusters and possible progression of myocardial dysfunction may include changes in features that differentiate clusters; specifically, reductions in electrical signal predictability and increases in variability. Identification of phenotypes of myocardial dysfunction evidenced by elevated LVEDP and knowledge of factors promoting transition to clusters with higher levels of left ventricular filling pressures could permit early risk stratification and improve patient selection for novel therapeutic interventions.

KEYWORDS

machine learning, risk stratification, left ventricular filling pressures, artificial intelligence, digital health

Introduction

An elevated left ventricular end diastolic pressure (LVEDP), indicative of increased left-sided filling pressures of the heart, represents a critical and sensitive measurement used to aid in the identification of impending as well as decompensated heart failure (HF) (1–3). Cardiac performance, with regard to ventricular contractility, is one critical determinant of elevated LVEDP; however, ejection fraction (EF) alone rarely elucidates the accurate clinical status of a heart failure patient (4). Patients with diminished ventricular function are the cohort most frequently encountered with elevated LVEDP, but an increased filling pressure may be the manifestation of multiple myocardial disease states including cardiomyopathies of ischemic, constrictive, restrictive, or valvular origin (5). Thus, higher filling pressures typically precede clinical deterioration in heart failure with reduced ejection fraction (HFrEF) as well as heart failure with preserved ejection fractions (HFpEF) (6). An elevated LVEDP does not indicate a specific diagnosis but provides important information that serves as a guide regarding the need for further evaluations or testing and affords data necessary for the development of an appropriate patient care plan. In addition, elevated LVEDP is predictive of morbidity and mortality (7–11).

Awareness of possible impending or new onset heart failure (HF) as indicated by an elevation in LVEDP may aid in diagnosis and risk stratification of the patient and help inform optimal clinical care pathways. The available therapeutic modalities for HF have doubled in the last decade and new pharmacologic agents for the treatment of hypertrophic cardiomyopathy (HCM) and infiltrative diseases, such as amyloid, conditions associated with elevated LVEDP, are rapidly emerging (12, 13). However, there exists limited information regarding the risks of transition from early asymptomatic or mildly symptomatic stages (Stages A to B) to later overtly symptomatic stages (Stages C-D), thus complicating initiation of drug interventions at the times when such therapies might prove most beneficial (14).

Risk stratification in heart failure remains a significant challenge despite the development of multiple scoring models, many of which are designed to predict survival with a few attempting to predict future morbidity (15–19). These scoring models were primarily derived from highly selected cohorts of patients, with an established diagnosis of heart failure, recruited for randomized clinical trials occurring over the last several decades (20). Most have required input data that includes echocardiographic determination of EF, multiple blood analyses (recently biomarkers), and a significant number of clinical parameters (15–21). Predicting outcomes in those with known moderate to severe disease allows modification and testing of current therapies but provides little information that might permit interventions at early stages of heart failure to prevent progression. Machine learning affords the opportunity with a single rapid test to detected features of similarities between groups with early stage LVEDP elevation and those not yet manifesting hemodynamic changes (22). ML clustering techniques that identify such features facilitate the development and future validation of new risk models.

The clinical diagnosis of HF encompasses a broad and heterogeneous population of patients with complex pathophysiology, various etiologic mechanisms, and diverse genetic triggers. Unsupervised clustering analysis has permitted phenotyping in cardiomyopathies identifying subgroups which

may have similar mechanisms and outcomes (23, 24). Clustering is a useful approach to detect patterns in variables that aid in decrypting the heterogeneity present in datasets (24, 25). To date, the evaluation of cardiomyopathies based on clustering has focused on patient with diagnosed HF and those recently or in the past hospitalized for the HF, clearly in later stages of the disease (23, 24). In this trial, we recruited a cohort with new onset symptoms and no previously known HF. The goal was to detect changes in the features that differentiate between clusters, which might indicate the risk of transition from one cluster to an adjacent cluster with a higher prevalence of elevated LVEDP. Identification of distinct subtypes of myocardial dysfunction evidenced by elevated LVEDP and knowledge of factors that promote transition to clusters with higher levels of left ventricular filling pressures could permit early risk stratification and improve patient selection for novel therapeutic interventions, thus facilitating precision medicine.

FIGURE 1
Signal acquisition using the CorVista Capture, with the electrodes placed on the torso (electrode on the back not visible), and the PPG clip placed on the finger.

Methods

Features for clustering

An algorithm was previously developed to predict LVEDP elevation status based on manually engineered features calculated from CorVista Capture signals (22). The signal acquisition modality was previously described (26), but briefly, is the simultaneous acquisition of orthogonal voltage gradient (OVG) data *via* electrodes placed on the torso, and photoplethysmogram using transmission of red and infrared light *via* a clip placed on the finger. The acquisition configuration is shown in Figure 1. The OVG signal is related to the electrocardiogram (ECG) signal in that both measure the voltage changes within the myocardium that occur during the cardiac cycle, but OVG differs from ECG in its three-dimensional perspective of the heart, as well as its high frequency bandwidth (Figure 2).

The contribution of these features to the LVEDP algorithm were assessed using a permutation analysis, enabling the ranking of the features from most to least contributive. The permutation analysis was a generalization of the methodology first proposed by Breiman for use in Random Forests (27). The top 30 most contributive features were selected for use in this clustering exploration.

Population

The clinical population analyzed in the present work were subjects with symptoms suggestive of cardiovascular disease who were referred to cardiac catheterization for the assessment of coronary artery disease using angiography. Specifically, the cohort was patients in whom the treating physician chose to measure the LVEDP during the catheterization. LVEDP measurement was performed using standard catheterization laboratory procedures. Subjects with mid-range LVEDPs were removed, preserving subjects with definitive LVEDP non-elevation of $\leq$12 mmHg and subjects with definitive LVEDP elevation of $\geq$25 mmHg. The subjects were enrolled within the CADLAD and IDENTIFY (Group 2) studies, the records for which are available on clinicaltrials.gov (NCT02784197 and NCT03864081, respectively). Both studies have identical inclusion and exclusion criteria and are multi-center in nature. CADLAD is closed, and IDENTIFY enrollment is ongoing. Finally, subjects without a signal passing a series of automated signal quality assessment tests (11%) were excluded. Signal quality assessment was previously described (26), but included quantification of powerline interference and excessive high-frequency content in the OVG signal, and the presence of jumps, dropouts, and railing in the PPG signal.

Clustering

The features were pre-processed prior to clustering by applying the box-cox transformation to each feature individually, which is a monotonic power transform intended to approximate a normal distribution across the data (28). While deviations from normality occur frequently in realistic data, many statistical approaches rely on an assumption of normality, and therefore benefit from distribution transformations such as box-cox. For instance, nearest neighbor classification has

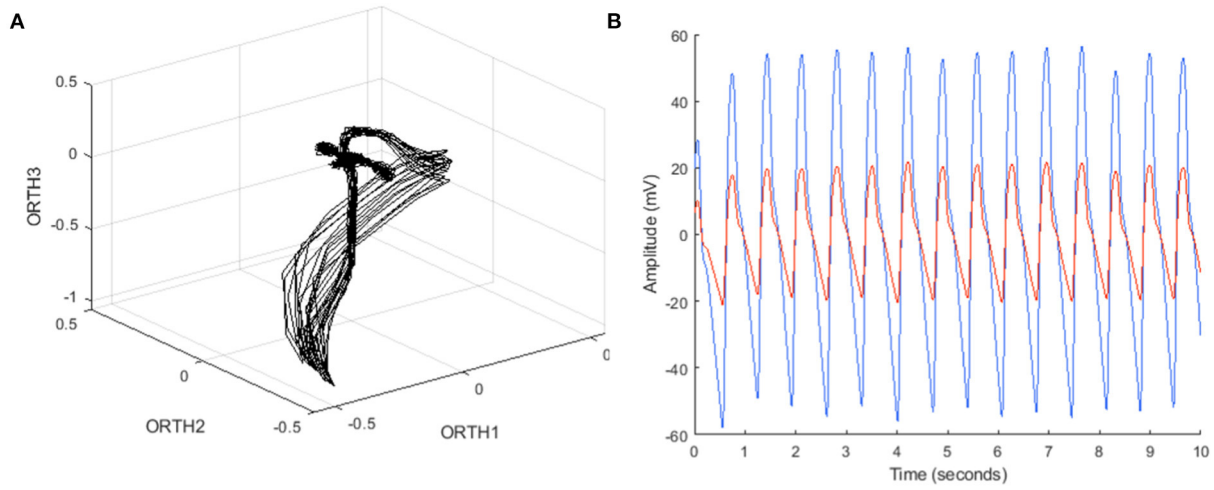


FIGURE 2
(A) Example OVG data in phase space, with coordinates from each bipolar channel (ORTH1, ORTH2, ORTH3) represented as a three-dimensional coordinate in that space, and **(B)** example PPG data in the time domain, containing both red and infrared time series.

been shown to improve considerably when the features are pre-processed using box-cox, as compared to classification using the raw feature values (29). While the intent of the present work is not nearest neighbor classification (or classification by any methodology), clustering similarly necessitates pairwise distance measurements across the dataset, and can therefore realize a similar benefit from box-cox. After box-cox transformation, the features were normalized using the z-score transformation to set the mean and standard deviation of each feature to zero and one, respectively.

Next, hierarchical clustering was applied to the processed data. While there exist a variety of clustering algorithms, hierarchical clustering was chosen to facilitate visual discovery of the appropriate numbers of clusters from the clustering result, rather than requiring pre-specification of the number of clusters, or quantitative evaluation of the optimal number of clusters. Hierarchical cluster is agglomerative, meaning that initially each subject is initially its own cluster, then clusters are merged in an agglomerative manner, until the end state is reached where all the subjects belong to the same cluster. Therefore, a cluster merger strategy, also known as linkage, must be selected. Possibilities include "single" (smallest pairwise distance between the subjects in one cluster and the subjects in another), "complete" (the largest pairwise difference between the subjects in one cluster and the subjects in another), and "average" (the average pairwise difference between the subjects in one cluster and the subjects in another). Note that the linkages are equivalent when considering clusters containing only one subject. Given that average considers all subjects in both clusters, it was chosen for the present work. However, as described for all linkage methods, average linkage requires the definition of a distance metric between subjects, which are represented herein by the values of 30 features. A distance metric increases in magnitude as subjects become more dissimilar, and must meet specific mathematical criteria to quality as a proper metric (30). Correlation distance (more precisely, 1-Pearson correlation) was shown to generally perform well across a variety of datasets and clustering algorithms, and so was chosen herein (31).

The output of hierarchical clustering is a dendrogram, which is a visual representation of subjects beginning in their own clusters, and iterative merging of clusters until all the subjects belong to a single cluster. The dendrogram was inspected, and an appropriate number of clusters chosen based on both the structure of the dendrogram and the appropriate amount of data in each cluster.

Results

A total of 396 subjects were included in the clustering, composed of 92 (23%) subjects with LVEDP≥25 mmHg and 304 (77%) subjects with LVEDP≤12 mmHg. Of the top 30 features from the LVEDP algorithm included in the clustering, 22 were derived from the OVG signal and had an average rank in the contribution of 17, and eight were derived from the PPG signal with an average rank of 10.

Figure 3 shows the dendrogram, with five clusters (labeled 1–5) chosen to segment the dataset, which for simplicity and recognition are colored (Purple=cluster 1, Green=cluster 2, Red=cluster 3, Yellow=cluster 4, Blue=cluster 5). In Figure 3A, the dendrogram is associated to a heatmap visualizing the magnitude of the feature values for each subject, with vertical lines delineating the boundary between adjacent clusters, and subjects aligning between the dendrogram and the heatmap.

10.3389/fcvm.2022.980625

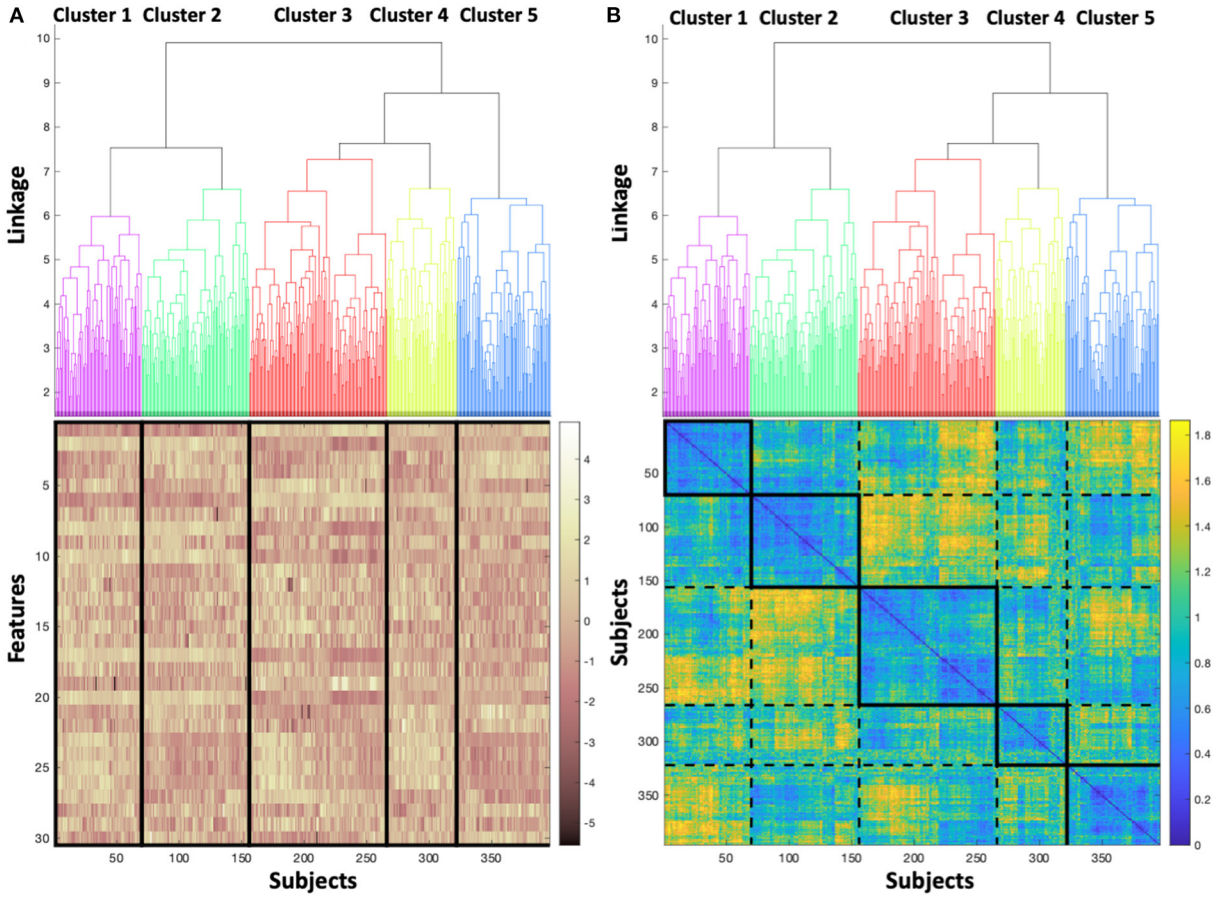


FIGURE 3
(A) Dendrogram colored to identify each cluster, associated to a heatmap visualizing the magnitude of the feature values for each subject, and **(B)** colored dendrogram associated with the pairwise distance matrix across the dataset, with bold boxes defining each cluster, and dotted lines delineating between adjacent clusters.

TABLE 1 Clusters demographics and measured parameters.

Property	Purple	Green	Red	Yellow	Blue
N	69	86	110	56	75
LVEDP≥25	36.2%	22.1%	22.7%	26.8%	10.7%
LVEDP (mmHg)	17.8 ± 10.8	13.5 ± 8.5	13.9 ± 8.4	15.0 ± 9.9	14.2 ± 9.0
Age(years)	65.3 ± 9.6	62.9 ± 10.4	63.4 ± 9.2	63.7 ± 10.4	61.6 ± 11.6
BMI(kg/m^2)	34.5 ± 7.6	31.0 ± 6.4	30.8 ± 7.2	32.0 ± 7.5	30.0 ± 6.0
Female	49.3%	40.7%	40.0%	42.9%	25.3%
Significant CAD*	40.6%	33.7%	38.9%	35.7%	44.0%
Diabetes	36.2%	32.6%	39.3%	34.9%	21.3%
Hypertension	78.3%	69.8%	76.8%	70.6%	68.0%
Hyperlipidemia	78.3%	75.6%	73.2%	72.5%	62.7%

*Significant CAD was defined as the presence of a ≥70% lesion or an FFR < 0.80, assessed during the same left heart catheterization procedure in which the LVEDP measurement was acquired.

Banding in the heatmap that differs between clusters is a visual manifestation of the feature values differing across clusters. In Figure 3B, the dendrogram is associated to the pairwise distance matrix across the dataset, with the bold boxes along the diagonal defining each cluster, and the dotted lines delineating between adjacent clusters. Similar to Figure 3A, the subjects are aligned

between the dendrogram and the distance matrix. As expected, the pairwise distances within clusters (i.e., within bolded boxes, occurring along the diagonal of the distance matrix) are low, showing that the distance between subjects within the same cluster is generally low. Therefore, subjects are cohesive within the clusters. The off-diagonal boxes defined by dashed lines in the distance matrix represent the distances between subjects not belonging to the same cluster, and as expected, exhibit larger distances on average than subjects belonging to the same cluster. Further, the distance between the cluster 1 (Purple) and cluster 2 (Green) is relatively low, as is supported by the dendrogram, which indicates that the next level of linkage would join these two clusters. Similarly, large distances in the heatmap are concentrated in the region comparing cluster 2 (Green) to cluster 3 (Red), which would not be joined together until the top linkage level, where all subjects are joined in a single cluster.

After the clustering was complete, clinical metadata was overlaid on the resultant clusters, as shown in Table 1. The overlaid clinical metadata can then be considered in conjunction with the observations from the feature clustering, as shown in Figure 3.

Cluster 1, with the highest LVEDP of all clusters, is the most dissimilar from the cluster 5 in both the clustering and with respect to the clinical properties of the subjects in each cluster. In contrast to cluster 1, cluster 5 is predominantly non-elevated LVEDP, younger, lower BMI, male and has a higher rate of significant coronary artery disease. The most substantial feature differences between cluster 1 and cluster 5 are those calculated from the PPG signal. Given the gender and age differences between these clusters, a plausible explanation of the PPG differentiation may be the vasculature changes that occur in post-menopausal women.

Cluster 1 is closest in the clustering to cluster 2 (i.e., the next cluster that would be joined to cluster 1 in the dendrogram is cluster 2), with the second lowest LVEDP of all clusters. To explore the mechanism underlying the difference between these clusters, the feature values were normalized across the full dataset using z-score (i.e., mean set to zero, and standard deviation set to one), and the normalized values mapped on to the clusters. The feature values were averaged within cluster 1 and cluster 2, then the averages differenced across these two clusters. The largest differences between cluster 1 and cluster 2 were found in features extracted from the electrical data, and specifically those that quantify predictability and variability of the signal. There is low ability to predict the signal (occurring as high feature values) and high signal variability (occurring as high feature values) in the cluster 1, and the opposite in the cluster 2 (occurring with low feature values). The predictability functions by fitting a model to a portion of the OVG data and evaluating the performance of that model on the remaining data. Should the ability of the model to predict on the withheld data be poor, then the resultant error between the signal and the model will be high, resulting in a high feature value. The variability feature

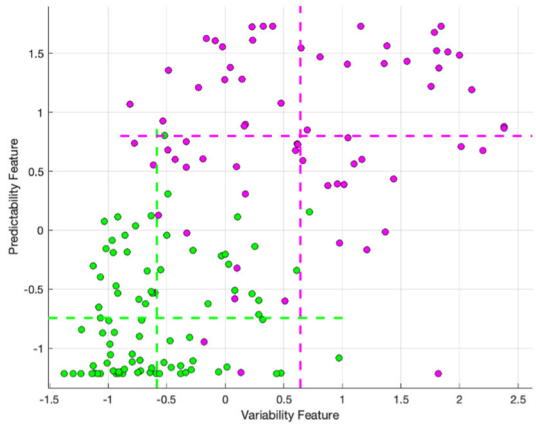


FIGURE 4
The variability and lack of predictability feature values for cluster 2 (Green) and cluster 1 (Purple), with the mean of each feature for each cluster marked with dashed lines.

derives a representative cardiac cycle for the subject based on the acquired data and compares that representative cardiac cycle to each acquired cycle to calculate the disparity – should it be high, then the representative cycle cannot sufficiently capture the variability in the signal, which is then represented as a high feature value.

The relationship between these features and clusters is shown in Figure 4, which visualizes the feature values for clusters 1 and 2, with the mean of each feature for each cluster marked with dashed lines. The mean predictability feature was −0.74 and 0.80, for cluster 2 and cluster 1, respectively. The mean variability feature was −0.58 and 0.64, for cluster 2 and cluster 1 respectively.

Discussion

Machine-learned algorithms now permeate many facets of medical practice and the magnitude of these techniques' utility and their diverse applications are rapidly being adapted to challenges in cardiology (32, 33). Cluster analysis of datasets identifies patterns and trends by using the relationships between variables to uncover hidden structure (34). In the past, the heterogeneous nature of HF populations and the complexity of the pathophysiologic mechanisms operative in specific HF disease states have made identification of similar phenotypes difficult. The utilization of clustering analysis allows detection of small unique phenogroups by examining the similarities and differences among quantitative variables (34). This type of unsupervised ML has be applied to populations of HF including HFpEF (23, 35), HFrEF (36), and acute HF (37). These studies have elucidated phenotype clusters and demonstrated subsequent clinical outcomes of patients previously diagnosed with HF as well as those acutely admitted with that diagnosis.

This approach is useful for risk stratification and prognostic prediction in later stages of the disease but does little for targeting early transition into pathophysiologic states at risk for progression to significant disease or to facilitate precision interventions at times when remodeling might be prevented.

In our trial, we identify novel phenotypes with varying prevalences of elevated LVEDP within a population presenting with new onset cardiovascular symptoms. Elevated LVEDP has served as a direct measurement of myocardial dysfunction as well as a marker indicative of pathophysiologic changes ultimately resulting in the remodeling observed in later stages of HF (38). The goal of this investigation was to analyze the phenotypes and determine if features within the clusters varied in a fashion that could be followed to predict transition to adjacent clusters with increased risk of LVEDP elevation.

This trial demonstrates five clusters resulting from the analysis of 396 subjects (Figure 3), composed of 92 (23%) subjects with LVEDP≥25 mmHg and 304 (77%) subjects with LVEDP≤12 mmHg. The heatmap of the magnitude of feature values (Figure 3A) illustrates the similarity of cluster 1 and cluster 2 (purple and green) phenogroups with regard to each of the 30 features. The degree of similarity can be appreciated visually, with the density of yellow and brown being closest between cluster 1 and 2 (purple and green). When moving from cluster 1 (shown in column 1) to cluster 5 (column 5), the coloration representing feature values becomes progressively more discrepant to cluster 1.

Correspondingly, the pairwise distance matrix (Figure 3B) showing the distance between pairs of subjects yields a diagonal line that elucidates the basis behind the formation of each cluster. Each solid box on the diagonal line starting with cluster 1 (top left, column 1) and progressing down the diagonal to bottom right to arrive at cluster 5, indicates very similar distancing between points in each cluster (high intensity of blue coloring). The dotted rectangle boxes permit comparison between the clusters with reference to the distance between subjects. In addition, this provides confirmation of the appropriateness and accuracy of the resultant clusters.

Cluster 1 and the adjacent cluster 2 demonstrate the closest feature values, but interestingly, have the widest divergence of prevalence of elevated LVEDP. Their proximity suggests that modest changes in feature values could result in transition of a subject in cluster 2 to cluster 1, and consequently may result in increased risk of elevation in LVEDP.

Of the top 30 features from the LVEDP algorithm included in the clustering, 22 were derived from the OVG signal, and eight were derived from the PPG signal. We found that features quantifying predictability and variability exhibited the most substantial differences between the cluster with the highest rate of LVEDP elevation (cluster 1) and the adjacent cluster 2, which presented with the second lowest rate of LVEDP elevation across the clusters. This feature difference suggests that it may be possible to follow subjects initially belonging to the cluster 2 phenotype to determine whether the signal properties shift over time to increased variability and decreased predictability, indicating that they may be transitioning to the adjacent cluster 1, with the associated higher prevalence, and therefore risk of, LVEDP elevation. Validation of the described phenotypes can be achieved through clinical follow-up to detect outcomes. In addition, opportunities exist to further characterize the phenotypes by inclusion of supplementary clinical data, including biomarkers such as BNP.

Conclusion

An approach to stratify the likelihood of movement between clusters and the possible progression of myocardial dysfunction for an individual patient could include changes in the features that differentiate these clusters; specifically, reductions in electrical signal predictability and increases in variability. Identification of distinct subtypes of myocardial dysfunction evidenced by elevated LVEDP and knowledge of factors that promote transition to clusters with higher levels of left ventricular filling pressures may permit early risk stratification and improve patient selection for novel therapeutic interventions thereby facilitating precision medicine.

Data availability statement

The datasets presented in this article are not readily available because signals obtained by proprietary device, not yet FDA approved. Requests to access the datasets should be directed to wsanders@corvista.com.

Ethics statement

The studies involving human participants were reviewed and approved by the Western Institutional Review Board, Inc. The patients provided their written informed consent to participate in this study. Written informed consent was obtained from the individual for the publication of any potentially identifiable images or data included in this article.

Author contributions

Concept and design: SB, WS, TB, RK, and SR. Drafting of the manuscript: WS, TB, SB, MR, FF, AK, and EL. Statistical analysis: HG, EL, TB, AK, and FF. Literature search: WS, TB, FF, AK, SR, and MR. Obtained funding and final approval of the version to be published: WS. Administrative, technical, and material support: WS and TB. Supervision: WS. Acquisition, analysis, interpretation of data, and critical revision of the manuscript for important intellectual content: All authors. All authors contributed to the article and approved the submitted version.

Funding

This study was supported by CorVista Health.

Acknowledgments

The authors would like to thank the contributing clinical centers and their research teams, as well as Horace R. Gillins and Charles Bridges MD ScD for their support, leadership, and insight.

Conflict of interest

CorVista Health funded the collection of subject data. Authors WS, HG, TB, FF, AK, EL, and SR are employees of CorVista Health. MR is a member of the Medical Advisory Board for CorVista Health.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcvm.2022.980625/full#supplementary-material

References

1. Reddy YN V, El-Sabbagh A, Nishimura RA. Comparing pulmonary arterial wedge pressure and left ventricular end diastolic pressure for assessment of left-sided filling pressures. *JAMA Cardiol.* (2018) 3:453–4. doi: 10.1001/jamacardio.2018.0318

2. Kern MJ, Christopher T. Hemodynamic rounds series II: the LVEDP. *Cathet Cardiovasc Diagn.* (1998) 44:70–4. - doi: 10.1002/(SICI)1097-0304(199805)44:1%3C70::AID-CCD17%3E3.0.CO;2-T

3. Bouchard JR, Gault HJ, Ross J. Evaluation of pulmonary arterial end-diastolic pressure as an estimate of left ventricular end-diastolic pressure in patients with normal and abnormal left ventricular performance. *Circulation.* (1971) 44:1072–9 doi: 10.1161/01.CIR.44.6.1072

4. Iskandrian AS, Segal BL, Hamid HAKKI A. Left ventricular end-diastolic pressure in evaluating left ventricular function. *Clin Cardiol.* (1981) 4:28–33. doi: 10.1002/clc.4960040107

5. Paulus WJ, Tschöpe C, Sanderson JE, Rusconi C, Flachskampf FA, Rademakers FE, et al. How to diagnose diastolic heart failure: a consensus statement on the diagnosis of heart failure with normal left ventricular ejection fraction by the Heart Failure and Echocardiography Associations of the European Society of Cardiology. *Eur Heart J.* (2007) 28:2539–50. doi: 10.1093/eurheartj/ehm037

6. Pahlevan NM, Matthews R V. Cardiac triangle mapping: a new systems approach for noninvasive evaluation of left ventricular end diastolic pressure. *Fluids.* (2019) 4:16. doi: 10.3390/fluids4010016

7. Uriel N, Sayer G, Annamalai S, Kapur NK, Burkhoff D. Mechanical unloading in heart failure. *J Am Coll Cardiol.* (2018) 72:569–80. doi: 10.1016/j.jacc.2018.05.038

8. Basir MB, Schreiber TL, Grines CL, Dixon SR, Moses JW, Maini BS, et al. Effect of early initiation of mechanical circulatory support on survival in cardiogenic shock. *Am J Cardiol.* (2017) 119:845–51. doi: 10.1016/j.amjcard.2016.11.037

9. Salem R, Denault AY, Couture P, Bélisle S, Fortier A, Guertin MC, et al. Left ventricular end-diastolic pressure is a predictor of mortality in cardiac surgery independently of left ventricular ejection fraction. *Br J Anaesth.* (2006) 97:292–7. doi: 10.1093/bja/ael140

10. Mielniczuk LM, Lamas GA, Flaker GC, Mitchell G, Smith SC, Gersh BJ, et al. Left ventricular end-diastolic pressure and risk of subsequent heart failure in patients following an acute myocardial infarction. *Congest Hear Fail.* (2007) 13:209–14. doi: 10.1111/j.1527-5299.2007.06624.x

11. Brienesse SC, Davies AJ, Khan A, Boyle AJ. Prognostic value of LVEDP in acute myocardial infarction: a systematic review and meta-analysis. *J Cardiovasc Transl Res.* (2018) 11:33–5. doi: 10.1007/s12265-017-9776-7

12. Masri A, Olivotto I. Cardiac myosin inhibitors as a novel treatment option for obstructive hypertrophic cardiomyopathy: addressing the core of the matter. *J Am Heart Assoc.* (2022) 11:e024656. doi: 10.1161/JAHA.121.024656

13. Kittleson MM, Maurer MS, Ambardekar AV, Bullock-Palmer RP, Chang PP, Eisen HJ, et al. American Heart Association Heart Failure and Transplantation Committee of the Council on Clinical Cardiology. Cardiac amyloidosis: evolving diagnosis and management: a scientific statement from the American Heart Association. *Circulation.* (2020) 142:e7–22. doi: 10.1161/CIR.0000000000000792

14. Hershberger RE, Zareba KM. Dilated cardiomyopathy: new distinct phenotypes or temporal phases of disease? *J Am Coll Cardiol.* (2022) 79:2233–5. doi: 10.1016/j.jacc.2022.04.008

15. Aaronson KD, Schwartz JS, Chen T-M, Wong K-L, Goin JE, Mancini DM. Development and prospective validation of a clinical index to predict survival in ambulatory patients referred for cardiac transplant evaluation. *Circulation.* (1997) 95:2660–7. doi: 10.1161/01.CIR.95.12.2660

16. Pocock SJ, Wang D, Pfeffer MA, Yusuf S, McMurray JJ V, Swedberg KB, et al. Predictors of mortality and morbidity in patients with chronic heart failure. *Eur Heart J.* (2006) 27:65–75. doi: 10.1093/eurheartj/ehi555

17. Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, et al. The Seattle heart failure model: prediction of survival in heart failure. *Circulation.* (2006) 113:1424–33. doi: 10.1161/CIRCULATIONAHA.105.584102

18. Pocock SJ, Ariti CA, McMurray JJ V, Maggioni A, Køber L, Squire IB, et al. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. *Eur Heart J.* (2013) 34:1404–13. doi: 10.1093/eurheartj/ehs337

19. Rich JD, Burns J, Freed BH, Maurer MS, Burkhoff D, Shah SJ. Meta-Analysis Global Group in Chronic (MAGGIC) heart failure risk score: validation of a simple tool for the prediction of morbidity and mortality in heart failure with preserved ejection fraction. *J Am Heart Assoc.* (2018) 7:e009594. doi: 10.1161/JAHA.118.009594

20. Lupón J, De Antonio M, Vila J, Peñafiel J, Galán A, Zamora E, et al. Development of a novel heart failure risk tool: the Barcelona bio-heart failure risk calculator (BCN bio-HF calculator). *PLoS ONE.* (2014) 9:e85466. doi: 10.1371/journal.pone.0085466

21. Vazquez R, Bayes-Genis A, Cygankiewicz I, Pascual-Figal D, Grigorian-Shamagian L, Pavon R, et al. The MUSIC Risk score: a simple method for predicting mortality in ambulatory patients with chronic heart failure. *Eur Heart J.* (2009) 30:1088–96. doi: 10.1093/eurheartj/ehp032

22. Bhavnani S, Khedraki R, Cohoon T, et al. Multicenter validation of a machine learning phase space electro-mechanical pulse wave analysis to predict elevated left ventricular end diastolic pressure at the point-of-care. *PloS ONE*.

23. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghiade M, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation.* (2015) 131:269–79. doi: 10.1161/CIRCULATIONAHA.114.010637

24. Verdonschot JAJ, Merlo M, Dominguez F, Wang P, Henkens MTHM, Adriaens ME, et al. Phenotypic clustering of dilated cardiomyopathy patients highlights important pathophysiological differences. *Eur Heart J.* (2021) 42:162–74. doi: 10.1093/eurheartj/ehaa841

25. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science.* (2006) 313:504–7. doi: 10.1126/science.1127647

26. Fathieh F, Paak M, Khosousi A, Burton T, Sanders WE, Doomra A, et al. Predicting cardiac disease from interactions of simultaneously-acquired hemodynamic and cardiac signals. *Comput Methods Programs Biomed.* (2021) 202:105970. doi: 10.1016/j.cmpb.2021.105970

27. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/A:1010933404324

28. Box GEP, Cox DR. An analysis of transformations. *J R Stat Soc Ser B.* (1964) 26:211–43. doi: 10.1111/j.2517-6161.1964.tb00553.x

29. Van Der Heiden R, Groen FCA. The Box-Cox metric for nearest neighbor classification improvement. *Pattern Recognit.* (1997) 30:273–9. doi: 10.1016/S0031-3203(96)00077-5

30. Chen S, Ma B, Zhang K. On the similarity metric and the distance metric. *Theor Comput Sci.* (2009) 410:2365–76. doi: 10.1016/j.tcs.2009.02.023

31. Kumar V, Chhabra JK, Kumar D. Performance evaluation of distance metrics in the clustering algorithms. *INFOCOMP J Comput Sci.* (2014) 13:38–52.

32. Quer G, Arnaout R, Henne M, Arnaout R. Machine learning and the future of cardiovascular care: JACC state-of-the-art review. *J Am Coll Cardiol.* (2021) 77:300–13. doi: 10.1016/j.jacc.2020.11.030

33. Krittanawong C, Johnson KW, Rosenson RS, Wang Z, Aydar M, Baber U, et al. Deep learning for cardiovascular medicine: a practical primer. *Eur Heart J.* (2019) 40:2058–73. doi: 10.1093/eurheartj/ehz056

34. Shalev-Shwartz S, Ben-David S. Understanding machine learning: From theory to algorithms. *Cambridge university press.* (2014). doi: 10.1017/CBO9781107298019

35. Kao DP, Lewsey JD, Anand IS, Massie BM, Zile MR, Carson PE, et al. Characterization of subgroups of heart failure patients with preserved ejection fraction with possible implications for prognosis and treatment response. *Eur J Heart Fail.* (2015) 17:925–35. doi: 10.1002/ejhf.327

36. Ahmad T, Lund LH, Rao P, Ghosh R, Warier P, Vaccaro B, et al. Machine learning methods improve prognostication, identify clinically distinct phenotypes, and detect heterogeneity in response to therapy in a large cohort of heart failure patients. *J Am Heart Assoc.* (2018) 7:e008081. doi: 10.1161/JAHA.117.008081

37. Horiuchi Y, Tanimoto S, Latif AHMM, Urayama KY, Aoki J, Yahagi K, et al. Identifying novel phenotypes of acute heart failure using cluster analysis of clinical variables. *Int J Cardiol.* (2018) 262:57–63. doi: 10.1016/j.ijcard.2018.03.098

38. Burchfield JS, Xie M, Hill JA. Pathological ventricular remodeling: mechanisms: part 1 of 2. *Circulation.* (2013) 128:388–400. doi: 10.1161/CIRCULATIONAHA.113.001878

frontiers | Frontiers in Cardiovascular Medicine

TYPE Systematic Review
PUBLISHED 06 October 2022
DOI 10.3389/fcvm.2022.918051

Check for updates

OPEN ACCESS

EDITED BY
Marco Giuseppe Del Buono,
Agostino Gemelli University Polyclinic (IRCCS), Italy

REVIEWED BY
Hirotsugu Yamada,
Tokushima University, Japan
Mohanad Alkhodari,
Khalifa University, United Arab Emirates

*CORRESPONDENCE
Fan Xu
xufan@cmc.edu.cn
Xiechuan Weng
wengxc2000@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Heart Failure and Transplantation, a section of the journal Frontiers in Cardiovascular Medicine

RECEIVED 12 April 2022
ACCEPTED 12 September 2022
PUBLISHED 06 October 2022

CITATION
Dao L, Huang M, Lin X, Li L, Feng X, Wei C, Guo M, Yang Y, Xu F and Weng X (2022) A systemic review and meta-analysis comparing the ability of diagnostic of the third heart sound and left ventricular ejection fraction in heart failure. *Front. Cardiovasc. Med.* 9:918051. doi: 10.3389/fcvm.2022.918051

COPYRIGHT
© 2022 Dao, Huang, Lin, Li, Feng, Wei, Guo, Yang, Xu and Weng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

A systemic review and meta-analysis comparing the ability of diagnostic of the third heart sound and left ventricular ejection fraction in heart failure

Lin Dao[1†], Min Huang[2†], Xinghong Lin[1†], Liuyin Li[3], Xixi Feng[4], Changyou Wei[4], Mingjin Guo[5], Yifan Yang[6], Fan Xu[4*] and Xiechuan Weng[7*]

[1]Department of Clinic Medicine, Chengdu Medical College, Chengdu, Sichuan, China, [2]Department of Physiology, Chengdu Medical College, Chengdu, Sichuan, China, [3]Department of Chinese Medicine Heart Disease, Zigong City First People's Hospital, Zigong, Sichuan, China, [4]Department of Public Health, Chengdu Medical College, Chengdu, Sichuan, China, [5]Department of Vascular Surgery, The Affiliated Hospital of Qingdao University, Qingdao, China, [6]Department of Pain Medicine, Peking University Peoples Hospital, Beijing, China, [7]Department of Neuroscience, Beijing Institute of Basic Medical Sciences, Beijing, China

Objective: This study aimed to compare the sensitivity and specificity of diagnosis between the third heart sound (S3) and left ventricular ejection fraction (LVEF) in heart failure (HF).

Methods: Relevant studies were searched in PubMed, SinoMed, China National Knowledge Infrastructure, and the Cochrane Trial Register until February 20, 2022. The sensitivity, specificity, likelihood ratio (LR), and diagnostic odds ratio (DOR) were pooled. The symmetric receiver operator characteristic curve (SROC) and Fagan's nomogram were drawn. The source of heterogeneity was explored by meta-regression and subgroup analysis.

Results: A total of 19 studies, involving 5,614 participants, were included. The combined sensitivity of S3 was 0.23 [95% confidence interval (CI) (0.15–0.33), specificity was 0.94 [95% CI (0.82–0.98)], area under the SROC curve was 0.49, and the DOR was 4.55; while the sensitivity of LVEF was 0.70 [95% CI (0.53–0.83)], specificity was 0.79 [95% CI (0.75–0.82)], area under the SROC curve was 0.79, and the DOR was 8.64. No publication bias was detected in Deeks' funnel plot. The prospective design, partial verification bias, and blind contributed to the heterogeneity in specificity, while adequate description of study participants contributed to the heterogeneity in sensitivity. In Fagan's nomogram, the post-test probability was 48% when the pre-test probability

was set as 20%, while in LVEF, the post-test probability was 45% when the pre-test probability was set as 20%.

Conclusion: The use of S3 alone presented lower sensitivity in diagnosing HF compared with LVEF, whereas it was useful in early pathological assessment.

KEYWORDS

acute heart failure, LVEF, meta-analysis, third heart sound, heart failure

Introduction

Heart failure (HF) is a pathological process during the pumping of blood in the heart. The cardiac output becomes insufficient to fully meet the needs of body metabolism (1–3). Currently, the 5 year mortality rate for HF has remained around 50% (4). The traditional diagnosis of HF relies mainly on the history and physical examination; the clinical diagnostic methods for HF include bio-standard object examination, electrocardiogram, echocardiogram, cardiac magnetic resonance imaging, and invasive hemodynamic monitoring (5). The ratio of stroke volume to ventricular end-diastolic volume is called the ejection fraction. The ejection fraction accurately reflects the pumping function of the heart, which is important for the early detection of cardiac pumping dysfunction. Left ventricular ejection fraction (LVEF) is an important diagnostic index of HF and an important basis for its classification (5).

The non-invasive detection method has been used for the effective diagnosis of early HF without organic heart disease or clinical symptoms (6). As routine cardiac physical examination, heart sound auscultation helps in cardiac function evaluation and initial screening of cardiac structure abnormalities, and has important value for the early diagnosis of cardiovascular diseases (7). Heart sound signals, especially the third heart sound (S3) signals, are associated with increased left ventricular end-diastolic pressure, and considered ideal confirmatory markers (8). Recently, the application of heart sound analysis in the diagnosis and classification of HF has emerged gradually. However, the value of diagnosis using heart sounds in HF remains controversial. Therefore, this study aimed to compare the sensitivity and specificity of diagnosis between S3 and LVEF in HF.

Methods

Search strategy

Two reviewers (DL and LXJ) searched the PubMed, Embase, Cochrane Library, China National Knowledge Infrastructure, and Wan Fang databases up to February 2022 independently. The search terms were as follows: #1 TS = ("HF, Diastolic" OR "HF, Systolic" OR "Ventricular Dysfunction, Right" OR "Ventricular Dysfunction, Left"); #2 TS = (the S3 OR Heart Auscultation OR Heart Sounds OR Sounds, Heart OR Cardiac Sounds OR Cardiac Sound OR Sound, and Cardiac OR Sounds, Cardiac); #3 DT = (Clinical Trial OR Article); #4 DOP = (1971-01-01/2022-2-22); #5 #1 AND #2 AND #3 AND #4.

Inclusion and exclusion criteria

The inclusion criteria were as follows: (1) randomized controlled experiments using patients with HF as the experimental group and healthy people or patients with benign disease as the control group; (2) well-defined patients with HF included as study participants; (3) diagnostic tests including S3 or/and LVEF; (4) number of true-positive (TP) cases, false-negative (FN) cases, false-positive (FP) cases, and true-negative (TN) cases obtained directly or calculated through the literature; (5) age, sex, and race not considered; and (6) studies published in any language. The exclusion criteria were as follows: (1) animal studies; (2) non-case-control trials; (3) studies with incomplete or no experimental data, duplicate published literature, reviews, and abstracts; (4) poor equilibrium between groups and different baselines, and the two groups not compared with the literature; and (5) no described diagnostic tests.

Data extraction

Two authors (LD and XL) independently extracted the demographic data and treatment information; the third author (MH) was consulted when disagreement occurred. The baseline information extracted from 23 studies contained the first author's name, year of publication, title, design type, study participants (number, age, and male/female ratio), disease degree, and length of the disease. The primary outcomes included FN, TN, TP, and FP with S3 and LVEF.

Statistical analysis

A meta-analysis was performed with Stata 15.0 software (Stata Corp., College Station, TX, USA). The combined

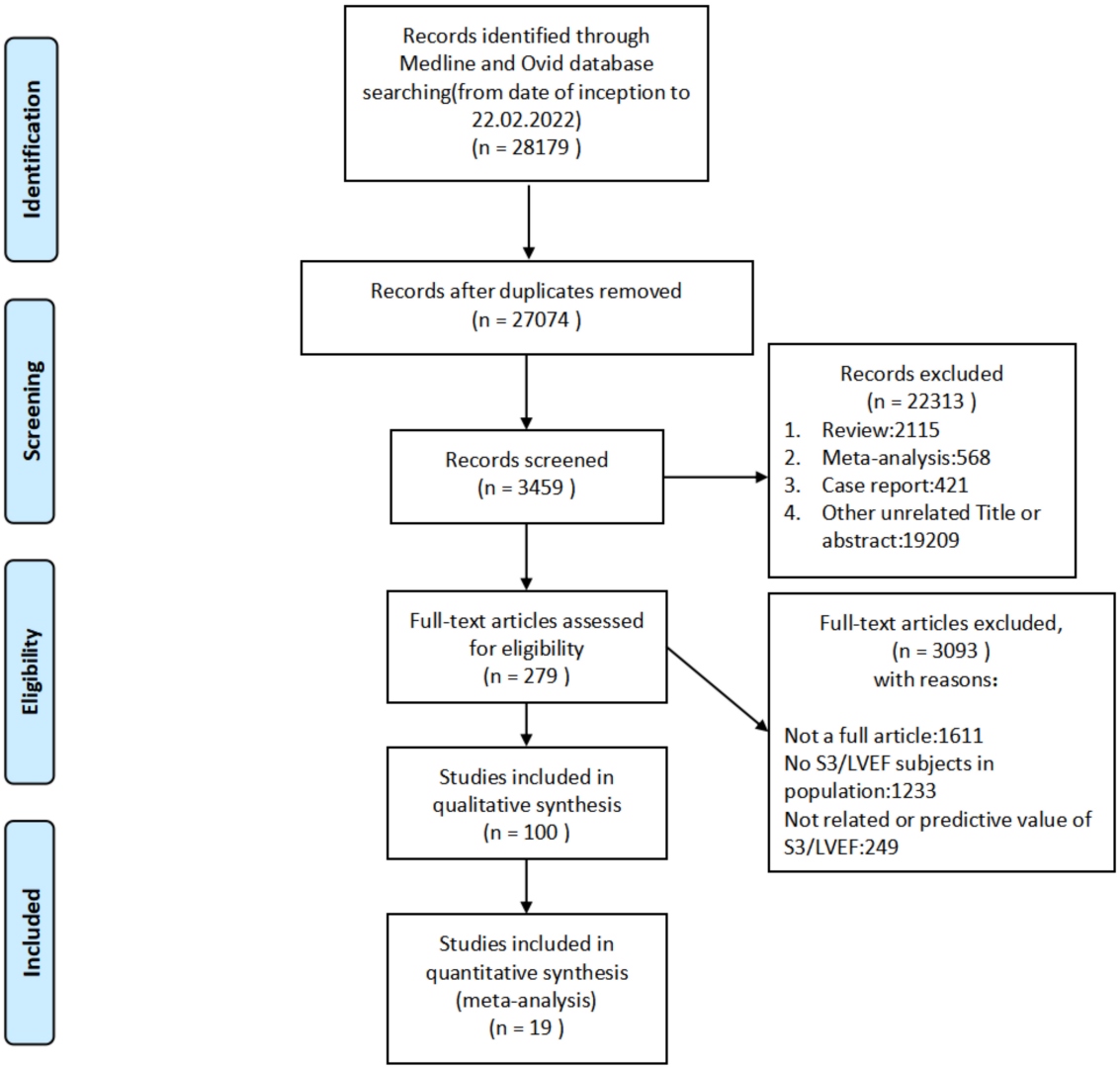


FIGURE 1
Literature screening process of the meta-analysis.

sensitivity, specificity, positive/negative likelihood ratio (PLR/NLR), and diagnostic odds ratio (DOR) were calculated using the bivariate model. The total diagnostic accuracy was estimated by drawing the symmetric receiver operator characteristic curve (SROC). Post-test probability was used to determine whether the probability of diagnosis increased or reduced compared with pre-test probability, which was estimated from routine data, practice data, or clinical judgment. Heterogeneity was assessed using Cochrane's Q statistics (chi-square) or inverse variance (I^2). $I^2 < 50\%$ and $P > 0.1$ indicated that these studies could be considered homogeneous using a fixed-effects model. If $I^2 \geq 50\%$ and $P < 0.10$, the random-effects model was used for meta-analysis. A P-value < 0.05 indicated a significant difference.

Results

Flow chart and study quality

A total of 28,179 studies (including documents, reviews, animal experiments, case reports, and repeated studies) were retrieved from each database. After removing 27,074 duplicate records, 279 relevant studies were included. Among these studies, 2,115 were excluded for being reviews, meta-analyses, or case reports, while 19,209 studies did not have related titles and abstracts. The full text of the remaining 279 studies was read and 3,093 studies were removed after reading the full text due to incomplete data. The remaining 19 studies were extracted from the corresponding data according to the data extraction requirements. Twelve studies used S3, and seven used

LVEF. The literature screening process is shown in Figure 1. The basic characteristics and inclusion and exclusion criteria of each study included are shown in Table 1 and Supplementary Table 1.

Third heart sound against heart failure

The combined sensitivity of S3 in HF was 0.23 [95% CI (0.15–0.33)], specificity was 0.94 [95% CI (0.82–0.98)], PLR was 3.74 95% CI (1.33–10.50)], NLR was 0.82 [95% CI (0.74–0.92)], and DOR was 4.55, indicating that S3 had a medium value in the screening of HF. The random-effects model was used when the heterogeneity was $I^2 > 50\%$. The details of the combined sensitivity and specificity forest are shown in Figure 2A; the combined likelihood ratio (LR) forest is shown in Figure 2B; and the combined diagnosis ratio forest is shown in Figure 2C.

Publication bias and heterogeneity

The Deeks' funnel plots were used to assess potential publication bias in detecting HF with S3. As shown in Figure 3, no publication bias existed, with a *P*-value of 0.35. The bivariate boxplot showed that three studies were out of the circles, indicating heterogeneity between included studies, as shown in Figure 4.

Threshold effect

The symmetric receiver operator characteristic curve curve plane test was used for the threshold effect. No typical "shoulder arm" was found, indicating no threshold effect. A moderate predictive value could be concluded by the value of the area under the SROC curve (AUC), which was 0.49 [95% CI (0.45–0.54)], as shown in Figure 5.

TABLE 1 Basic characteristics of enrolled studies.

References	Study	Region	N	Age (mean ± SD)	Sex (male%)
Dao et al. (9)	Retrospective	America	250	CHF (n = 250): 63 ± 0.86;	CHF: male/female = 94:6
Knudsen et al. (10)	Retrospective	America	880	HF (n = 447): 64 ± 16; No-HF (n = 433): 64 ± 16	HF: 482 (55)
Knudsen et al. (11)	Retrospective	Norway	155	CHF (n = 155): men (n = 69): 74 (66–79); women (n = 86): 78 (71–84)	Men: 69 (44.5)
Zhang (12)	Prospective	China	78	CHF (n = 42): 63 ± 12; non-CHF (n = 36): 54 ± 12	CHF: 29 (69.0); non-CHF: 21 (58.3)
Collins et al. (13)	Prospective	America	343	Primary HF (n = 133): 69 (30–97); secondary HF (n = 60): 68 (39–93); non-HF (n = 150): 55 (20–95)	Primary HF: 70 (52.6); secondary HF: 26 (43.3); non-HF: 63 (42.0)
Collins et al. (14)	Prospective	America	1,076	ADHF (n = 413): 68 (40–95); no-ADHF (n = 506): 59.5 (40–95)	ADHF: 246 (59.6) No-ADHF: 255 (50.4)
Wang et al. (15)	Retrospective	China	292	HBP (n = 94): 54 ± 10; HFREF (n = 89): 73 ± 13; HFNEF (n = 109): 77 ± 10	HBP: 46 (49); HFREF: 66 (74); HFNEF: 46 (42)
Dieplinger et al. (16)	Prospective	Austria	251	CHF (n = 137): 76 (69–82); non-CHF (n = 114): 69 (58–78)	CHF: 128 (93); non-CHF: 106 (93)
Miller et al. (17)	Prospective	USA	89	AHF (n = 35): 72 ± 10; non-AHF (n = 54): 65 ± 10	AHF: 24 (69); non-AHF; 28 (52)
Wang et al. (18)	Retrospective	China	127	HF: 72 ± 13 (36–97)	HF: 91 (71.7)
Logeart et al. (19)	Prospective	America	163	CHF (n = 115): 68.3 ± 14.7; non-CHF (n = 48): 65.1 ± 15.1	Male/female CHF (n = 115): 80/35; non-CHF (n = 48): 29/19
Steg et al. (20)	Prospective	America	709	CHF (n = 492): 68.5 ± 14.1; non-CHF (n = 217): 61.6 ± 14.8	Male/female CHF (n = 492): 217/275 non-CHF (n = 217): 90/127
Nazerian et al. (21)	Prospective	Italy	145	aLVHF (n = 64): 8 ± 8; Others (n = 81): 75 ± 12	aLVHF (female): 33 (54); others (female): 41 (51)
Anderson et al. (22)	Prospective	America	101	ADHF (n = 44): 63 (53–91); non-ADHF (n = 57): 62 (52–88)	ADHF: 25 (56); non-ADHF: 27 (47)
Kajimoto et al. (23)	Prospective	Japan	90	AHFS group (n = 53): 77.7 ± 10.3; pulmonary group (n = 37):78.6 ± 9.2	AHFS group (female): 29 (54.7); pulmonary group (female): 16 (43.2)
Hu (24)	Prospective	China	100	CHF (n = 50): 63.55 ± 2.4; non-CHF (n = 50): 63.5 ± 2.35	CHF: 33 (66); Non-CHF: 35 (70)
Jiang (25)	Prospective	China	60	CHF (n = 48): 59.14 ± 6.82; non-CHF (n = 12): 59.14 ± 6.82	CHF: 37 (61.7) non-CHF: NA
Logeart et al. (19)	Prospective	America	163	CHF (n = 115): 68.3 ± 14.7; non-CHF (n = 48): 65.1 ± 15.1	CHF: 80 (52.7); non-CHF: 29 (17.8)
Steg et al. (20)	Prospective	America	709	CHF (n = 492): 68.5 ± 14.1; No CHF (n = 217): 61.6 ± 14.8	CHF: 217 (44.1); No CHF: 90 (41.5)

NA, not available from original study paper or supplementary or registration information; ED, emergency department.

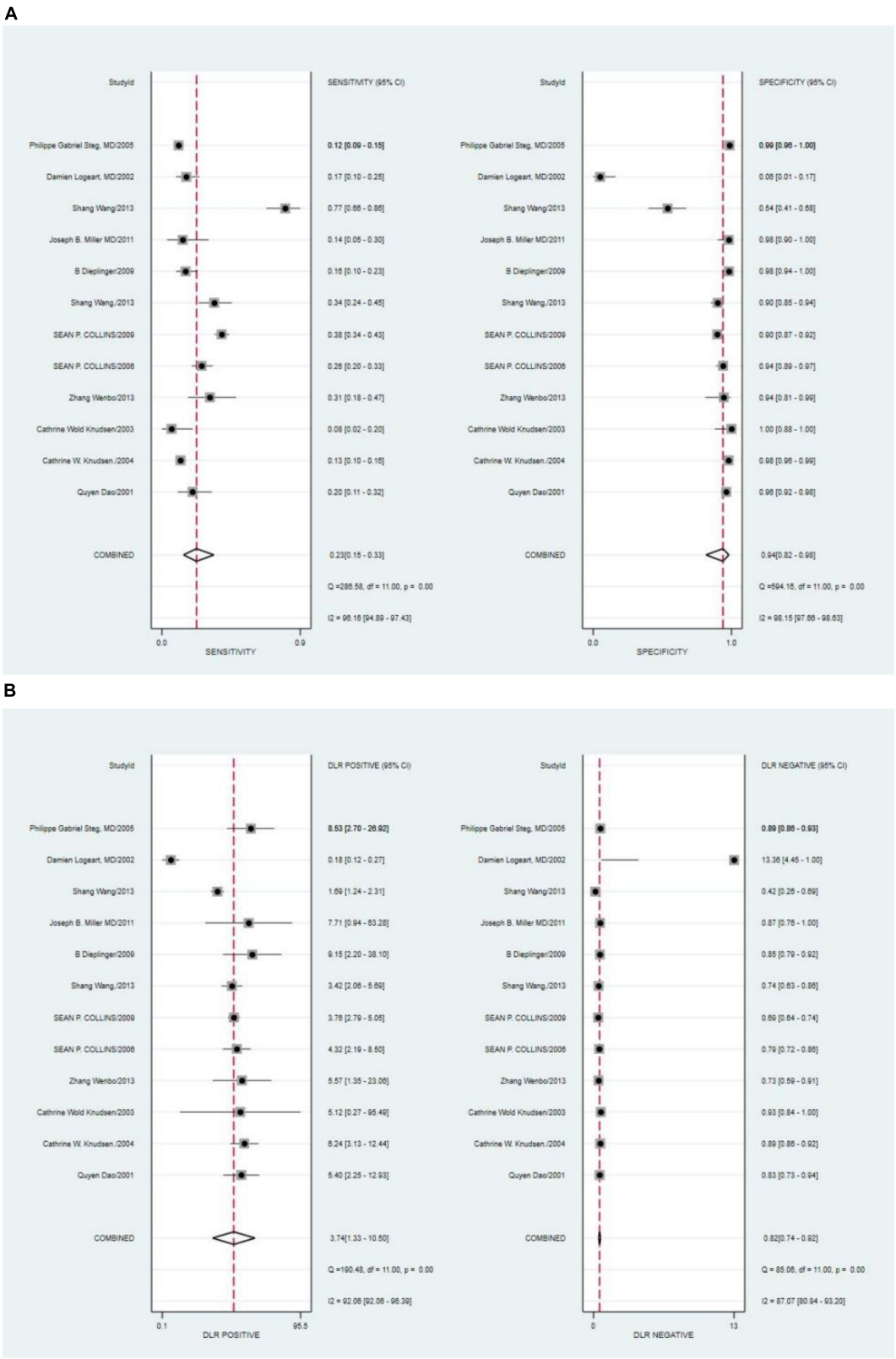


FIGURE 2
(Continued)

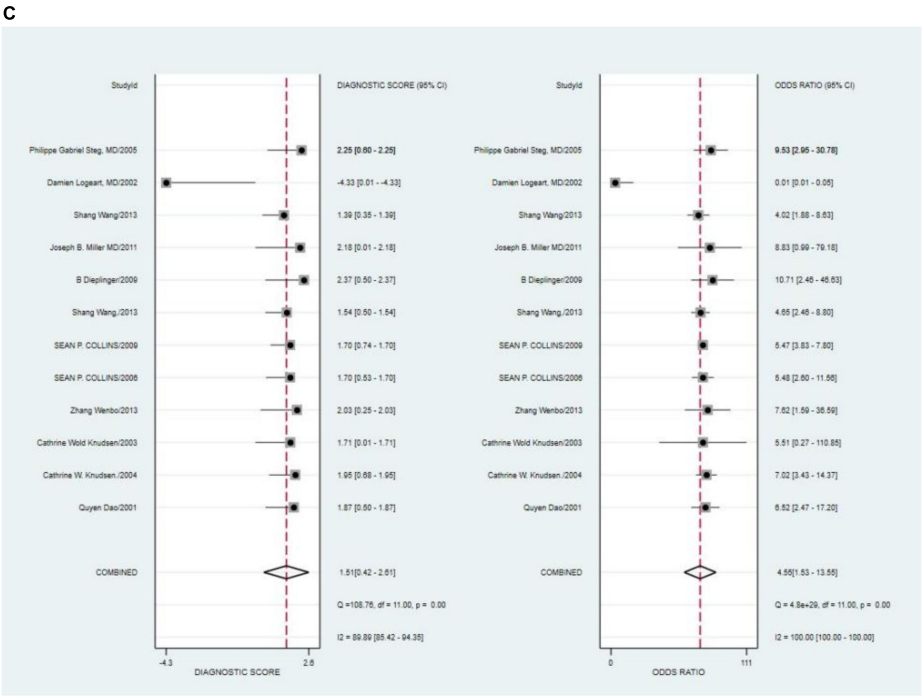


FIGURE 2
(A) Forest plot of sensitivity and specificity of third heart sound (S3) in the diagnosis of heart failure (HF). **(B)** Forest plot of DLR positives and negatives of HF. **(C)** Forest plot of the diagnostic odds ratio (DOR) of S3 in the diagnosis of HF.

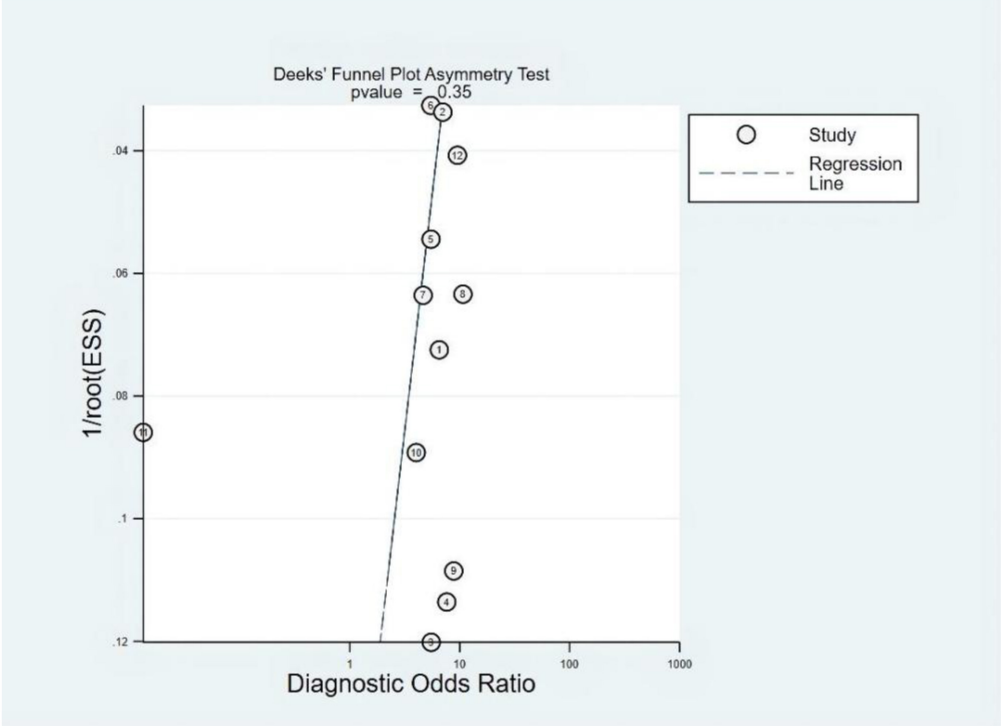


FIGURE 3
Deeks' funnel plot.

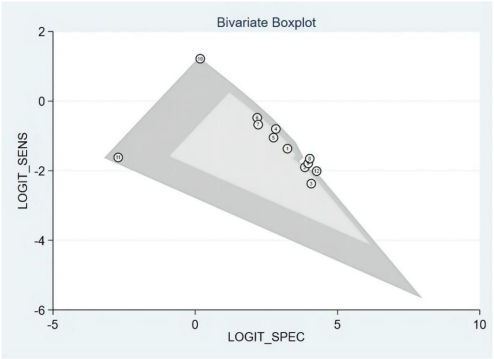


FIGURE 4
Bivariate boxplot.

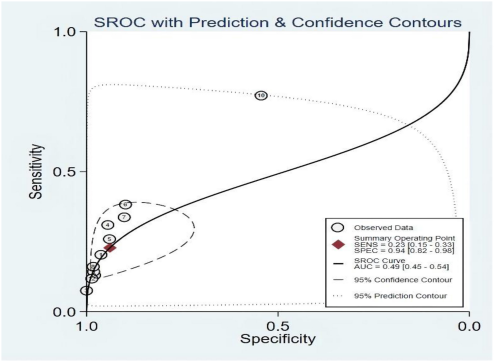


FIGURE 5
Summary receiver operating characteristic of third heart sound (S3).

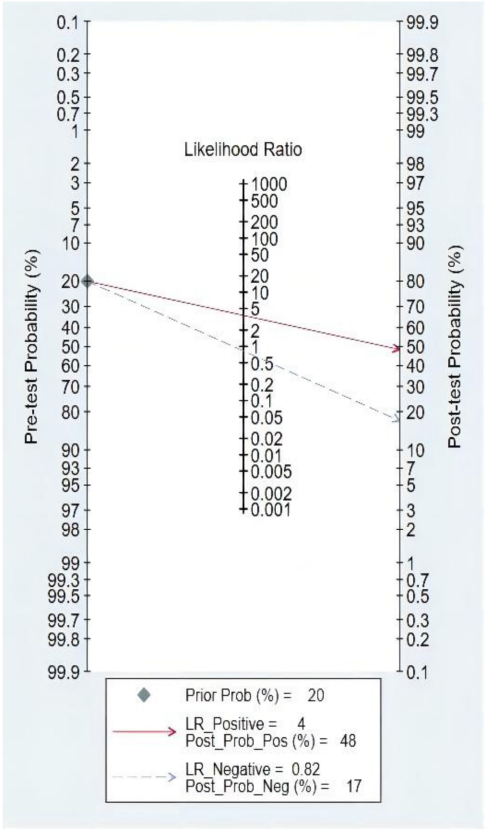


FIGURE 6
Fagan diagram of third heart sound (S3) in the diagnosis of heart failure (HF).

Pre-test probability, likelihood ratio, and post-test probability

The Fagan graph was plotted to show the relationship among the prior probability, the LR, and the posterior probability. The pre-test probability was 20% and the post-test probability of HF was 48%. In addition, the positive likelihood ratio (LRP) was <10 (LRP = 4) and the negative likelihood ratio (LRN) was >0.1 (LRN = 0.82), indicating that the diagnosis could neither be confirmed nor excluded. The predictive value of S3 in HF was limited, as shown in **Figure 6**.

Meta-regression and subgroup analysis

Among the S3 studies, the factors that might affect the heterogeneity, including prospective design (prodesign), partial verification bias (fulverif), an adequate description of study participants (subjdescr), report, a broad spectrum of diseases (brdspect), and whether the test results were evaluated by a blind method, were evaluated. The meta-regression analysis of the aforementioned factors revealed that the sources of heterogeneity of sensitivity were statistically related to subjdescr and the sources of heterogeneity of specificity were related to prodesign, as shown in **Figure 7**.

Left ventricular ejection fraction against heart failure

The combined sensitivity was 0.70 (95% CI, 0.53–0.83), specificity was 0.79 (95% CI, 0.75–0.82), PLR was 3.31 (95% CI, 2.46–4.44), NLR was 0.38 (95% CI, 0.23–0.64), and DOR was 8.64, indicating that the LVEF had a medium value in the screening of HF. The heterogeneity was $I^2 > 50\%$; therefore, the random model was used, as shown in **Figure 8**.

Publication bias and heterogeneity

The *P*-value of Deeks' funnel plots asymmetry test was 0.90 ($P > 0.05$). As shown in **Figure 9**, no evidence of publication

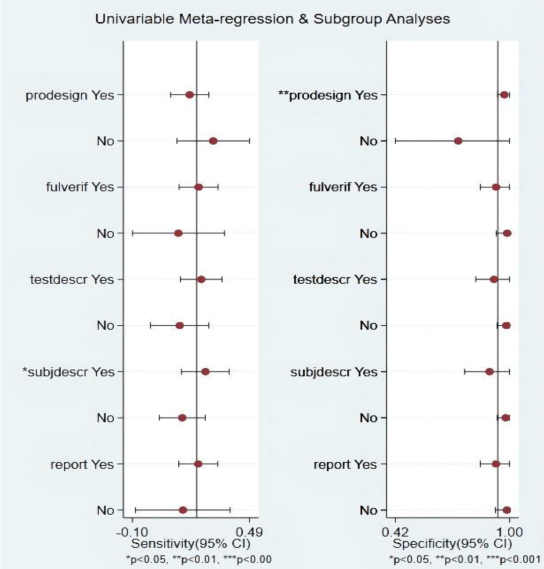


FIGURE 7
Multiple univariate meta-regression and subgroup analysis. Prospective design: prodesign; fulverif: partial verification bias; subjdescr: adequate description of study participants; brdspect: broad spectrum of disease.

bias was found. It demonstrated that three sets of data were out of the circles, indicating heterogeneity between included studies. The details are shown in **Figure 10**.

Threshold effect

The threshold effect was assessed using the SROC curve plane test. As no typical "shoulder arm" was found, no threshold effect was observed. A moderate predictive value could be concluded using the value of the AUC, which was 0.79 (95% CI, 0.75–0.83). The details are shown in **Figure 11**.

Pre-test probability, likelihood ratio, and post-test probability

The pre-test probability was 20%, and the probability of HF was 45%. In addition, the LRP was <10 (LRP = 3) and the LRN was >0.1 (LRN = 0.38), indicating that the diagnosis could be neither confirmed nor excluded. Their predictive value of LVEF in HF was also limited. The details are also shown in **Figure 12**.

Meta-regression and subgroup analysis

Among the LVEF studies, the factors that might affect the heterogeneity, including prospective design (prodesign), partial verification bias (fulverif), an adequate description of study participants (subjdescr), report, a broad spectrum of disease (brdspect), and whether the test results were evaluated by a blind method, were evaluated. The meta-regression analysis of the aforementioned factors revealed that although the sources of heterogeneity of specificity were statistically related to the prodesign, fuverif, and blind, the sources of heterogeneity of sensitivity were not related to these factors, as shown in **Figure 13**.

Comparison of third heart sound and left ventricular ejection fraction

Third heart sound and LVEF were compared using SROC, sensitivity, and specificity analysis. Among them, the predictive value of LVEF was better. The details are shown in **Table 2**.

Discussion

Heart failure is a global public health issue of epidemic proportions and represents a tremendous burden to the overall healthcare costs (26). Meanwhile, it affects the quality of life of patients and their families seriously. Therefore, early recognition and accurate diagnosis are essential, and meaningful for a positive outcome.

This systematic review and meta-analysis was novel in comparing the ability to diagnose HF between S3 and LVEF. In this meta-analysis, 19 studies, including 5,614 participants, were analyzed. The combined sensitivity and DOR of S3 was less than that of LVEF. On the contrary, S3 had a higher specificity than LVEF, and the AUC of S3 was less than that of LVEF. Moreover, after using the S3 or LVEF, the post-test probability was equally improved. This suggested that LVEF had the highest diagnostic value compared with S3 and S3 alone was not of high diagnostic value for HF.

The 2021 ESC guidelines (5) pointed out that the identification of the etiology of the underlying cardiac dysfunction was imperative in the diagnosis of HF, making it convenient for subsequent treatment decision-making. In general, HF is due to systolic, diastolic, or both dysfunction. However, the pathology of the valves, pericardium, or endocardium and the abnormalities of heart rhythm, and conduction can contribute to HF (27). Heart sound intensity and frequency and their relationship or the occurrence of heart murmur are closely related to the condition of a cardiac valve, myocardial contraction, and blood flow in the heart (28). The aforementioned arguments laid the solid foundation for the diagnosis of HF *via* heart sound detection and analysis.

Kosmicki et al. (29) found that the diagnostic efficacy of the discriminative model constructed based on the characteristics of heart sound-electrocardiogram fusion, which included the

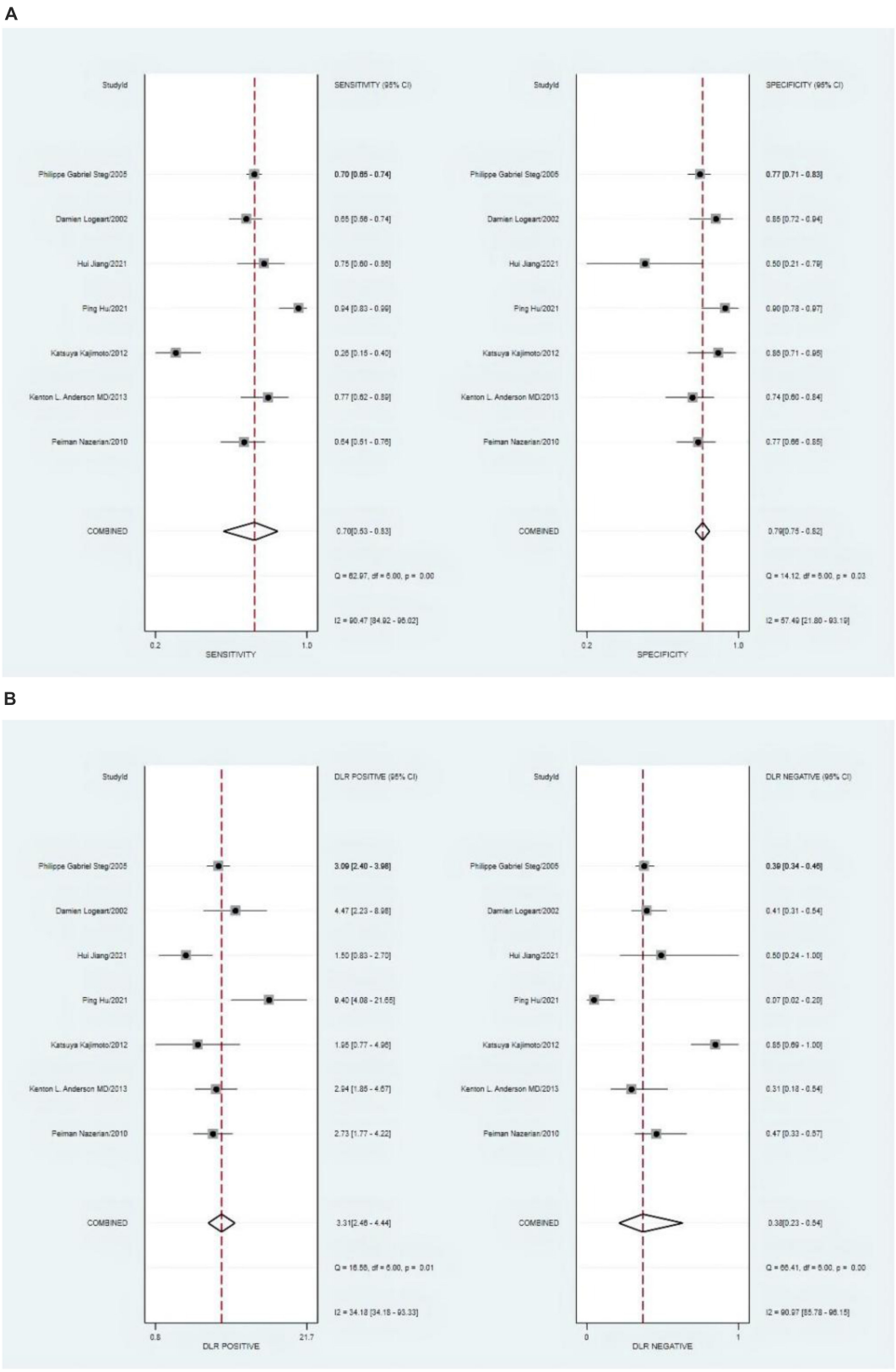


FIGURE 8
(Continued)

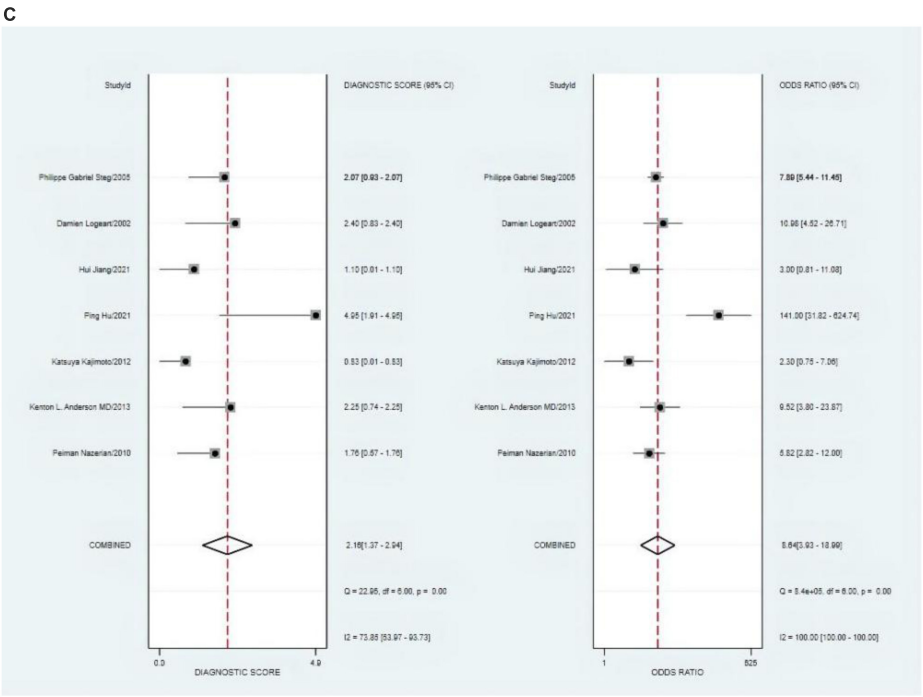


FIGURE 8
(A) Forest plot of sensitivity and specificity of left ventricular ejection fraction (LVEF) in diagnosing heart failure (HF). **(B)** Forest plot of DLR positives and negatives of HF. **(C)** Forest plot of the diagnostic odds ratio (DOR) of LVEF in diagnosing HF.

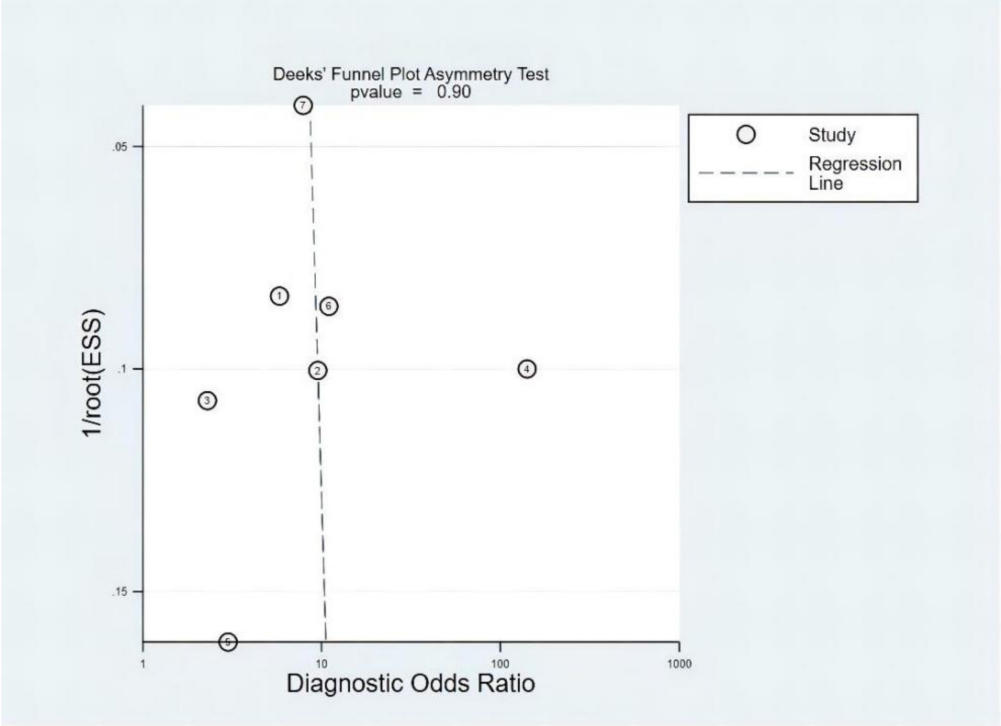


FIGURE 9
Deeks' funnel plot.

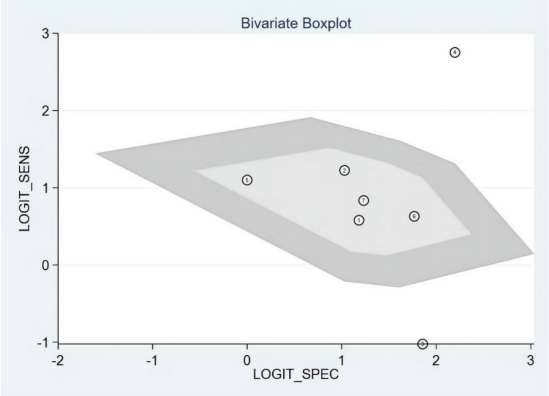


FIGURE 10
Bivariate boxplot.

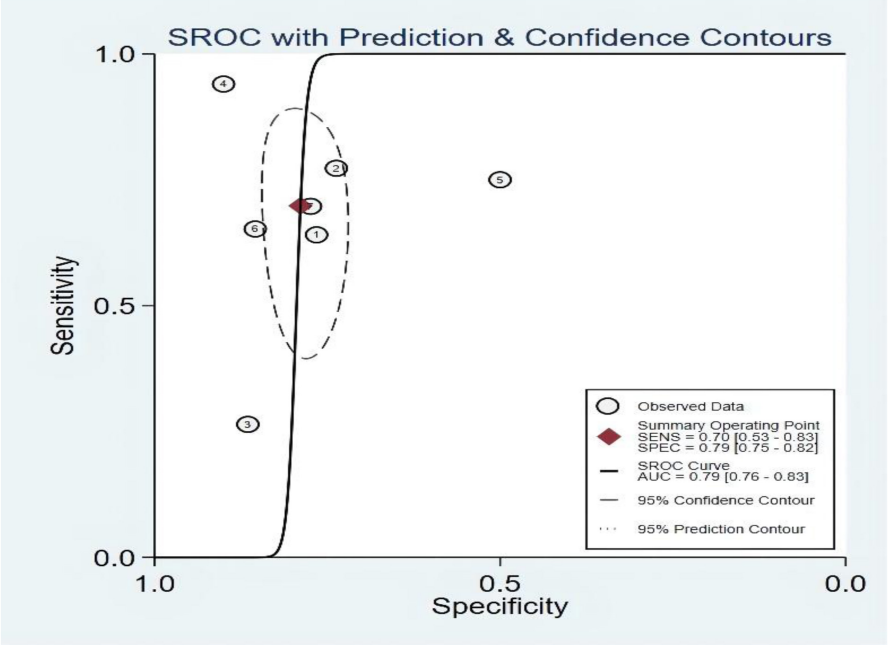


FIGURE 11
Summary receiver operating characteristic of left ventricular ejection fraction (LVEF).

S3 strength, left ventricular systolic time, electromechanical activation time, QR interval, and QRS interval, was significantly better than that of B-type natriuretic peptide (BNP). Moreover, the identification ability of patients with CHF in the "gray zone" of BNP significantly improved. Moreover, Maisel et al. (30) found that the strength of the S3 provided rapid results that assisted with the identification of acute HF (AHF) in selected populations. This evidence disclosed that S3 had high diagnostic value as an auxiliary diagnostic indicator. However, S3 was used as the single diagnostic indicator, reducing the sensitivity and specificity of diagnosis.

Meanwhile, the deficit in auscultation technology and the techniques of sound deciphering pulled down the sensitivity and specificity of diagnosis in HF *via* heart sound. The detection method for heart sound has been continuously improving, with new feature extraction algorithm and computer-aided diagnosis system based on machine learning or deep learning (31, 32), which further improves its diagnostic value. Liu et al. (31). reported that using extreme learning machine and heart sound characteristics to assist in diagnosing HF with preserved ejection fraction (HFpEF) showed an accuracy of 96.32%, a sensitivity of 95.48%, and a specificity of 97.10, which demonstrated

10.3389/fcvm.2022.918051

TABLE 2 Diagnostic performance of third heart sound (S3) and left ventricular ejection fraction (LVEF).

Indicator	Sensitivity	Specificity	AUC	Sensitivity combined	Specificity combined	Prior *P*	PLR (%)	NLR (%)
LVEF	0.70	0.79	0.79	0.70	0.79	20	45	9
S3	0.23	0.94	0.49	0.23	0.94	20	48	17

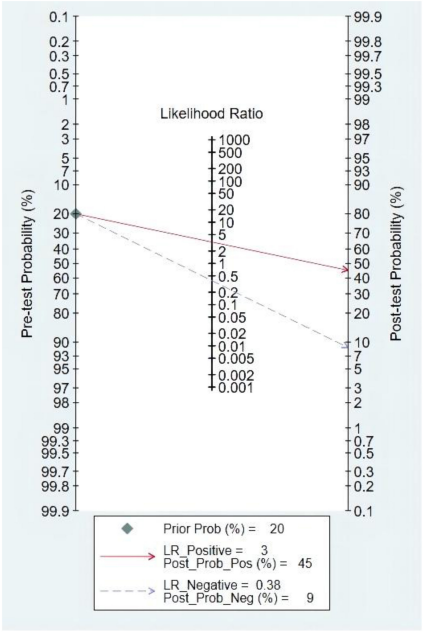


FIGURE 12
Fagan diagram of left ventricular ejection fraction (LVEF) in the diagnosis of heart sound.

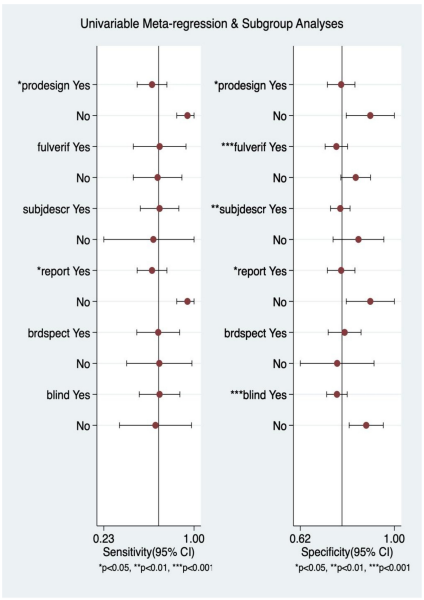


FIGURE 13
Multiple univariate meta-regression and subgroup analysis. Prospective design: prodesign; fulverif: partial verification bias; subjdescr: adequate description of study participants; brdspect: broad spectrum of disease.

the effectiveness of HS for HFpEF diagnosis.Alkhodari et al. (33) showed that the potential of implementing deep learning-based models clinically to ensure faster, yet accurate, automatic

prediction of HF based on the ASE/EACVI LVEF guidelines with only clinical profiles, and corresponding information as input to the models. In addition, Alkhodari et al. (34) also found that applied support vector regression (SVR) models to estimate LVEF from ECG derived heart rate variability (HRV) data which ensure the best possible estimations of LVEF levels. Although the diagnostic value of S3 as a single indicator for HF is not high, heart sound shows promise as a diagnostic and prognostic tool in HF with the update of heart sound feature extraction methods.

Conclusion

The use of S3 alone presented lower sensitivity in the diagnosis of HF compared with LVEF, whereas it was useful in early pathological assessment. Future prospective studies are needed to explore the diagnostic value of heart sound analysis based on new feature extraction algorithm and computer-aided diagnosis system based on machine learning or deep learning, so as to improve the early recognition rate of HF.

Study limitations

First, less and reduplicative studies made the original data incomplete. Second, moderate heterogeneity existed across studies, and meta-regression and subgroup analyses failed output due to limited S3 data. Third, few included studies did not explicitly exclude participants. These shortcomings should be further investigated and addressed in future studies.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding authors.

Author contributions

LD, MH, and XL performed the literature screening, the data extraction and analysis, result representation, and drafting of the manuscript for intellectual content. FX performed statistical analysis, interpreted the data, and revised the manuscript for intellectual content. XW initiated the study and revised the manuscript for intellectual content. XF and CW revised the manuscript. All authors contributed to the study and approved the submitted version of the manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (Nos. 8216050478 and 82073833), the Chengdu Science and Technology Bureau Focuses on Research and Development Support Plan (No. 2019-YF09-00097-SN), the Popular Scientific Research Project of Sichuan Health Commission (No. 20PJ171), the Yunnan Education Program (No. SYSX202036), and the Key Discipline Construction Project of Sichuan Administration of Traditional Chinese Medicine (No. 202072).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcvm.2022.918051/full#supplementary-material

References

1. Sacks D, Baxter B, Campbell BCV, Carpenter JS, Cognard C, Dippel D, et al. Multisociety consensus quality improvement revised consensus statement for endovascular therapy of acute ischemic stroke. *Int J Stroke.* (2018) 13: 612–32.

2. Heidenreich PA, Albert NM, Allen LA, Bluemke DA, Butler J, Fonarow GC, et al. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. *Circ Heart Fail.* (2013) 6:606–19. doi: 10.1161/HHF.0b013e318291329a

3. Ambrosy AP, Fonarow GC, Butler J, Chioncel O, Greene SJ, Vaduganathan M, et al. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. *J Am Coll Cardiol.* (2014) 63:1123–33. doi: 10.1016/j.jacc.2013.11.053

4. Taylor CJ, Ryan R, Nichols L, Gale N, Hobbs FR, Marshall T. Survival following a diagnosis of heart failure in primary care. *Fam Pract.* (2017) 34:161–8. doi: 10.1093/fampra/cmx040

5. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. Corrigendum to: 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure: developed by the task force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur Heart J.* (2021) 42:4901. doi: 10.1093/eurheartj/ehab670

6. Long B, Koyfman A, Gottlieb M. Diagnosis of acute heart failure in the emergency department: an evidence-based review. *West J Emerg Med.* (2019) 20:875–84. doi: 10.5811/westjem.2019.9.43732

7. Voin V, Oskouian RJ, Loukas M, Tubbs RS. Auscultation of the heart: the basics with anatomical correlation. *Clin Anat.* (2017) 30:58–60. doi: 10.1002/ca.22780

8. Stevenson LW, Perloff JK. The limited reliability of physical signs for estimating hemodynamics in chronic heart failure. *JAMA.* (1989) 261:884–8. doi: 10.1001/jama.261.6.884

9. Dao Q, Krishnaswamy P, Kazanegra R, Harrison A, Amirnovin R, Lenert L, et al. Utility of B-type natriuretic peptide in the diagnosis of congestive heart failure in an urgent-care setting. *J Am Coll Cardiol.* (2001) 37:379–85. doi: 10.1016/S0735-1097(00)01156-6

10. Knudsen CW, Omland T, Clopton P, Westheim A, Abraham WT, Storrow AB, et al. Diagnostic value of B-type natriuretic peptide and chest radiographic findings in patients with acute dyspnea. *Am J Med.* (2004) 116:363–8. doi: 10.1016/j.amjmed.2003.10.028

11. Knudsen CW, Riis JS, Finsen AV, Eikvar L, Müller C, Westheim A, et al. Diagnostic value of a rapid test for B-type natriuretic peptide in patients presenting with acute dyspnoe: effect of age and gender. *Eur J Heart Fail.* (2004) 6:55–62. doi: 10.1016/j.ejheart.2003.10.006

12. Zhang W. *Usefulness of Acoustic Cardiography in the Diagnosis Of Chronic Systolic Heart Failure.* Zhejiang: ©1994–2021 China Academic Journal Electronic Publishing House (2013).

13. Collins SP, Lindsell CJ, Peacock WF, Hedger VD, Askew J, Eckert DC, et al. The combined utility of an S3 heart sound and B-type natriuretic peptide levels in emergency department patients with dyspnea. *J Card Fail.* (2006) 12:286–92. doi: 10.1016/j.cardfail.2006.01.012

14. Collins SP, Peacock WF, Lindsell CJ, Clopton P, Diercks DB, Hiestand B, et al. S3 detection as a diagnostic and prognostic aid in emergency department patients with acute dyspnea. *Ann Emerg Med.* (2009) 53:748–57. doi: 10.1016/j.annemergmed.2008.12.029

15. Wang S, Lam YY, Liu M, Fang F, Wang J, Shang Q, et al. Acoustic cardiography helps to identify heart failure and its phenotypes. *Int J Cardiol.* (2013) 167:681–6. doi: 10.1016/j.ijcard.2012.03.067

16. Dieplinger B, Gegenhuber A, Haltmayer M, Mueller T. Evaluation of novel biomarkers for the diagnosis of acute destabilised heart failure in patients with shortness of breath. *Heart.* (2009) 95:1508–13. doi: 10.1136/hrt.2009.170696

17. Miller JB, Sen A, Strote SR, Hegg AJ, Farris S, Brackney A, et al. Inferior vena cava assessment in the bedside diagnosis of acute heart failure. *Am J Emerg Med.* (2012) 30:778–83. doi: 10.1016/j.ajem.2011.04.008

18. Wang S, Fang F, Liu M, Lam YY, Wang J, Shang Q, et al. Rapid bedside identification of high-risk population in heart failure with reduced ejection fraction by acoustic cardiography. *Int J Cardiol.* (2013) 168:1881–6. doi: 10.1016/j.ijcard.2012.12.064

19. Logeart D, Saudubray C, Beyne P, Thabut G, Ennezat PV, Chavelas C, et al. Comparative value of Doppler echocardiography and B-type natriuretic peptide assay in the etiologic diagnosis of acute dyspnea. *J Am Coll Cardiol.* (2002) 40:1794–800. doi: 10.1016/s0735-1097(02)02482-8

20. Steg PG, Joubin L, McCord J, Abraham WT, Hollander JE, Omland T, et al. B-type natriuretic peptide and echocardiographic determination of ejection fraction in the diagnosis of congestive heart failure in patients with acute dyspnea. *Chest.* (2005) 128:21–9. doi: 10.1378/chest.128.1.21

21. Nazerian P, Vanni S, Zanobetti M, Polidori G, Pepe G, Federico R, et al. Diagnostic accuracy of emergency Doppler echocardiography for identification of acute left ventricular heart failure in patients with acute dyspnea: comparison with Boston criteria and N-terminal prohormone brain natriuretic peptide. *Acad Emerg Med.* (2010) 17:18–26. doi: 10.1111/j.1553-2712.2009.00630.x

22. Anderson KL, Jenq KY, Fields JM, Panebianco NL, Dean AJ. Diagnosing heart failure among acutely dyspneic patients with cardiac, inferior vena cava, and lung ultrasonography. *Am J Emerg Med.* (2013) 31:1208–14. doi: 10.1016/j.ajem.2013.05.007

23. Kajimoto K, Madeen K, Nakayama T, Tsudo H, Kuroda T, Abe T, et al. Rapid evaluation by lung-cardiac-inferior vena cava (LCI) integrated ultrasound for differentiating heart failure from pulmonary disease as the cause of acute dyspnea in the emergency setting. *Cardiovasc Ultrasound.* (2012) 10:49. doi: 10.1186/1476-7120-10-49

24. Hu P. Accuracy and clinical application value of cardiac color ultrasound in the diagnosis of multiple causes of chronic heart failure. *Imaging Res Med Appl.* (2021) 5:144–5.

25. Jiang H. Diagnosis value analysis of echocardiography combined with BNP detection in chronic heart failure. *Forum Essent Med.* (2021) 25:978–9.

26. Tsao L, Gibson CM. Heart failure: an epidemic of the 21st century. *Crit Pathw Cardiol.* (2004) 3:194–204. doi: 10.1097/01.hpc.0000146867.90558.ca

27. Bozkurt, B, Coats AJ, Tsutsui H, Abdelhamid M, Adamopoulos S, Albert N, et al. Universal definition and classification of heart failure: a report of the Heart Failure Society of America, Heart Failure Association of the European Society of Cardiology, Japanese Heart Failure Society and Writing Committee of the Universal Definition of Heart Failure. *J Card Fail.* (2021) 27:387–413. doi: 10.1002/ejhf.2115

28. Tang H, Gao J, Ruan C, Qiu T, Park Y. Modeling of heart sound morphology and analysis of the morphological variations induced by respiration. *Comput Biol Med.* (2013) 43:1637–44. doi: 10.1016/j.compbiomed.2013.08.005

29. Kosmicki DL, Collins SP, Kontos MC, Zuber M, Kipfer P, Attenhofer Jost C, et al. Noninvasive prediction of left ventricular systolic dysfunction in patients with clinically suspected heart failure using acoustic cardiography. *Congest Heart Fail.* (2010) 16:249–53. doi: 10.1111/j.1751-7133.2010.00191.x

30. Maisel AS, Peacock WF, Shah KS, Clopton P, Diercks D, Hiestand B, et al. Acoustic cardiography S3 detection use in problematic subgroups and B-type natriuretic peptide "gray zone": secondary results from the heart failure and audicor technology for rapid diagnosis and initial treatment multinational investigation. *Am J Emerg Med.* (2011) 29:924–31. doi: 10.1016/j.ajem.2010.03.032

31. Liu Y, Guo X, Zheng Y. An automatic approach using ELM classifier for HFpEF identification based on heart sound characteristics. *J Med Syst.* (2019) 43:285. doi: 10.1007/s10916-019-1415-1

32. Zheng Y, Guo X, Qin J, Xiao S. Computer-assisted diagnosis for chronic heart failure by the analysis of their cardiac reserve and heart sound characteristics. *Comput Methods Programs Biomed.* (2015) 122:372–83. doi: 10.1016/j.cmpb.2015.09.001

33. Alkhodari M, Jelinek HF, Karlas A, Soulaidopoulos S, Arsenos P, Doundoulakis I, et al. Deep learning predicts heart failure with preserved, mid-range, and reduced left ventricular ejection fraction from patient clinical profiles. *Front Cardiovasc Med.* (2021) 8:755968. doi: 10.3389/fcvm.2021.755968

34. Alkhodari M, Jelinek HF, Werghi N, Hadjileontiadis LJ, Khandoker AH. Estimating left ventricle ejection fraction levels using circadian heart rate variability features and support vector regression models. *IEEE J Biomed Health Inform.* (2021) 25:746–54. doi: 10.1109/JBHI.2020.3002336

frontiers | Frontiers in Cardiovascular Medicine

TYPE Brief Research Report
PUBLISHED 18 October 2022
DOI 10.3389/fcvm.2022.1011071

Check for updates

OPEN ACCESS

EDITED BY
Qingpeng Zhang,
City University of Hong Kong,
Hong Kong SAR, China

REVIEWED BY
Gen-Min Lin,
Hualien Armed Forces General Hospital, Taiwan
Suneet Kr Gupta,
Bennett University, India

*CORRESPONDENCE
Tatiana Kuznetsova
tatiana.kouznetsova@kuleuven.be

SPECIALTY SECTION
This article was submitted to Heart Failure and Transplantation, a section of the journal Frontiers in Cardiovascular Medicine

RECEIVED 03 August 2022
ACCEPTED 03 October 2022
PUBLISHED 18 October 2022

CITATION
Sabovčik F, Ntalianis E, Cauwenberghs N and Kuznetsova T (2022) Improving predictive performance in incident heart failure using machine learning and multi-center data.
Front. Cardiovasc. Med. 9:1011071.
doi: 10.3389/fcvm.2022.1011071

COPYRIGHT
© 2022 Sabovčik, Ntalianis, Cauwenberghs and Kuznetsova. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Improving predictive performance in incident heart failure using machine learning and multi-center data

František Sabovčik, Evangelos Ntalianis, Nicholas Cauwenberghs and Tatiana Kuznetsova*

Research Unit of Hypertension and Cardiovascular Epidemiology, KU Leuven Department of Cardiovascular Sciences, University of Leuven, Leuven, Belgium

Objective: To mitigate the burden associated with heart failure (HF), primary prevention is of the utmost importance. To improve early risk stratification, advanced computational methods such as machine learning (ML) capturing complex individual patterns in large data might be necessary. Therefore, we compared the predictive performance of incident HF risk models in terms of (a) flexible ML models and linear models and (b) models trained on a single cohort (single-center) and on multiple heterogeneous cohorts (multi-center).

Design and methods: In our analysis, we used the meta-data consisting of 30,354 individuals from 6 cohorts. During a median follow-up of 5.40 years, 1,068 individuals experienced a non-fatal HF event. We evaluated the predictive performance of survival gradient boosting (SGB), CoxNet, the PCP-HF risk score, and a stacking method. Predictions were obtained iteratively, in each iteration one cohort serving as an external test set and either one or all remaining cohorts as a training set (single- or multi-center, respectively).

Results: Overall, multi-center models systematically outperformed single-center models. Further, c-index in the pooled population was higher in SGB (0.735) than in CoxNet (0.694). In the precision-recall (PR) analysis for predicting 10-year HF risk, the stacking method, combining the SGB, CoxNet, Gaussian mixture and PCP-HF models, outperformed other models with PR/AUC 0.804, while PCP-HF achieved only 0.551.

Conclusion: With a greater number and variety of training cohorts, the model learns a wider range of specific individual health characteristics. Flexible ML algorithms can be used to capture these diverse distributions and produce more precise prediction models.

KEYWORDS

heart failure, incidence, machine learning, prediction model, multi-center data

Introduction

A major burden of modern society is progressive increase in age-associated disorders such as cardiovascular (CV) diseases. Due to population aging and unhealthy lifestyle, the prevalence of heart failure (HF) in low- to middle-income countries will rise by 50% in the next 5–10 years (1). According to the World Health Organization, proper early risk stratification and management could help reduce the burden of this chronic disorder.

There are currently a number of clinical risk scores indented for specific populations and risk groups (2–6). For example, the Pooled Cohort Equations to Prevent HF (PCP-HF) score is recommended to use for 10-year incident HF prediction in a general population (7). Most of the recommended scores are, however, based on a linear model and might therefore lack the specificity and sensitivity in certain subgroups. The wide variety of scores may also cause slow adoption in clinical practice.

The use of advanced analytic techniques such as machine learning (ML) might improve the predictive performance of models by employing a higher number of interrelated and non-linear features. In addition, training ML models on a wide range of patient groups might create tools generally applicable in different settings. Recently, the number of publications applying ML for both prognosis and diagnosis of CV disease sharply increased, with 85% of these decided in favor of ML as opposed to traditional linear methods (8). The most popular choices include tree-based boosting and bagging methods, such as survival gradient boosting (SGB) and random survival forests (RSF) (9). However, a substantial obstacle in the adoption of ML in CV risk prediction, including HF, is the lack of adequate external validation in a large number of individuals with varied characteristics, as well as the ability to exploit flexible models when trained in various populations using a large number of relevant features (8).

Therefore, we proposed to test predictive model performance in a spectrum of train and test populations systematically instead of selecting a single derivation and validation cohort. Similarly, choosing an appropriate evaluation strategy (internal, external, etc.) is crucial (10). Thus, we additionally evaluated the influence of obtaining more diverse training data (multi-center) on the predictive performance of incident HF prediction models.

Objectives

The main objective of our analysis was to evaluate the predictive performance of the incident HF risk prediction models in the general population. Specifically, we compared the predictive performance of a linear model (CoxNet), non-linear model (SGB), currently used HF risk score (PCP-HF), and a stacking method (combining prediction of tested models). We also evaluated the predictive performance of models when trained on multiple heterogeneous cohorts (multi-center) rather than on single cohort. In addition, we reported features selected by the models and assessed their achieved predictive performance for the given number of features.

Methods and materials

Study design is outlined in Figure 1.

Cohorts

In our analysis, we included 6 cohorts from the Heart "Omics" in Aging (HOMAGE) meta-data—the Anglo-Scandinavian Cardiac Outcomes Trial (ASCOT), the Flemish Study on Environment, Genes, and Health Outcomes (FLEMENGHO), the Health Aging and Body Composition (Health ABC), HVC database, Valutazione della PREvalenza di DIsfunzione Cardiaca asinTOmatica e di scompenso cardiaco (PREDICTOR), and the Prospective Study of Pravastatin in the Elderly at Risk (PROSPER). The final dataset included 33 features consisting of clinical (e.g., medical history, HR, SBP), biochemical (e.g., blood glucose, creatinine), and ECG (e.g., duration QRS) variables. For a complete list of included features, see Supplemental Table 1. The unfiltered input data of the individuals consisted of 43,817 individuals. We removed participants younger than 30 and older than 80 years old (n = 4,709), with HF diagnosis at baseline (n = 3,462), with missing outcome (n = 5,058), with missing blood pressure measurements (n = 107), and with invalid follow-up time (n = 127). The final study population included 6 cohorts consisting of 30,354 subjects.

By sending anonymized data, the contributing partners confirmed that their study complies with good clinical practice (Helsinki Declaration), that all participants provided written informed consent, and that at the time of its conduct the study conformed to national regulations on clinical research in humans and on the protection of privacy. The HOMAGE database was described in depth elsewhere (11, 12).

Outcome of interest

The primary outcome of interest in this study is incident non-fatal HF, defined as HF hospitalization. The specificities of

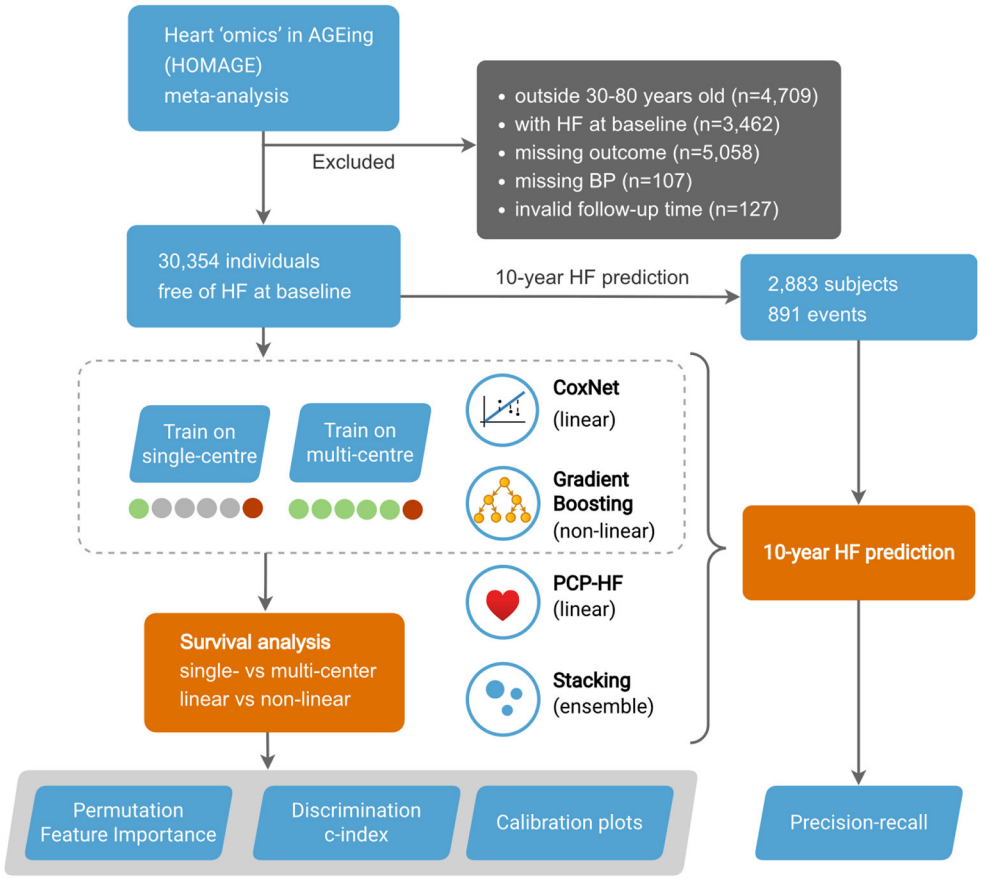


FIGURE 1
Study design.

non-fatal HF in each of the cohorts were described elsewhere (11, 12).

ML algorithms

We evaluated the predictive performance of the following two models: a flexible survival ML model (Survival Gradient Boosting, SGB) and a linear Cox proportional hazard model-CoxNet. SGB is a non-linear machine learning method based on training regression trees with the objective of optimizing Cox partial likelihood. CoxNet is a standard linear Cox proportional hazard model, regularized by both L1 and L2 norms. As an alternative, we also employed the stacking method, consisting of the above-mentioned CoxNet and SGB, together with PCP-HF score and unsupervised Gaussian mixture model. The stacking method works in two layers, the output of the first layer of base learners is the input of the second "meta" layer, consisting of another model. For CoxNet, the features were standardized. We used the Optuna library with the tree-structured parzen estimator to optimize the model hyperparameters on an internal validation set. For ML pipelines, we used the scikit-learn and scikit-survival Python libraries. The code of the analysis is available online[1].

Model evaluation and statistical analysis

The discrimination of the models was evaluated using the c-index. Predictive performance was evaluated iteratively, with each cohort serving as a test set (external cohort validation). In each iteration, the multi-center models were trained on the remaining 5 cohorts (1 cohort test), while the single-center models were trained on a single cohort. The final predictive performance was obtained by evaluating the merged predictions and, in the case of single-center models, averaged. The goal was to compare the predictive performance of models trained

1 https://github.com/hcve/incidence-hf

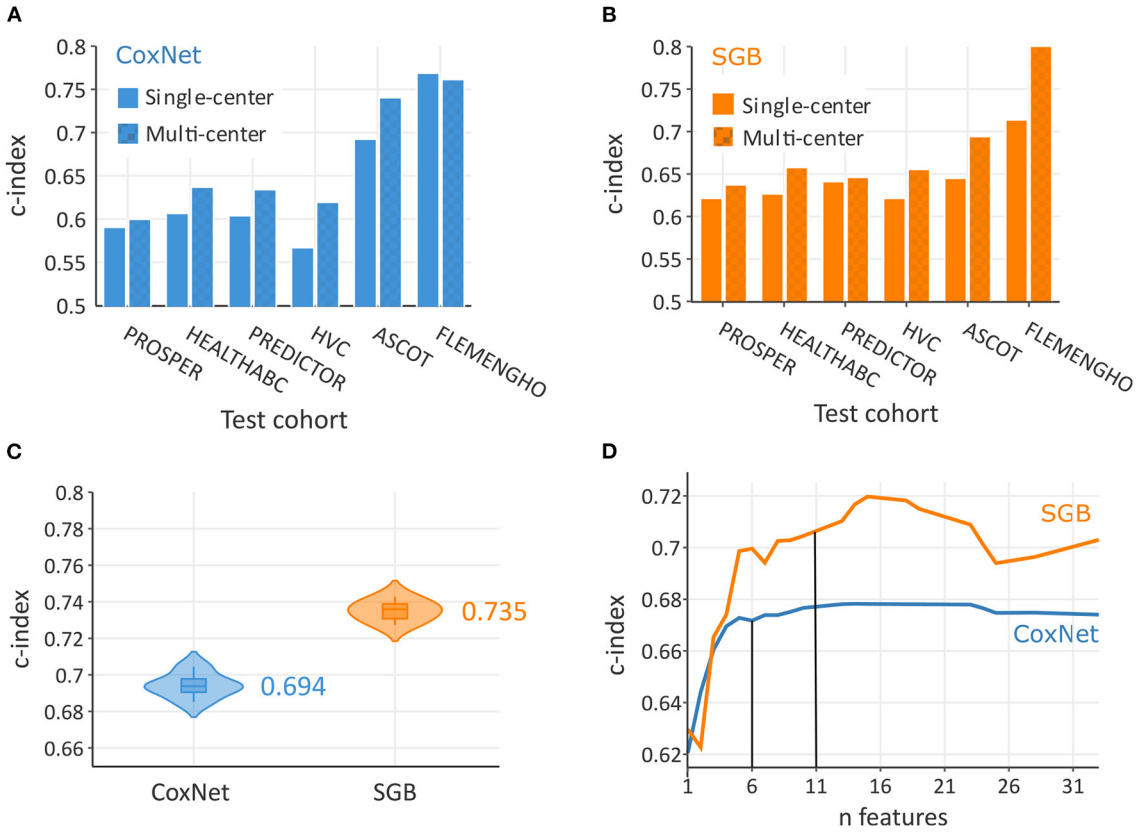


FIGURE 2
Predictive performance increased with multi-rather than single-cohort training data both in CoxNet **(A)** and survival gradient boosting (SGB) **(B)**. On the pooled dataset, SBG achieved a greater overall c-index than CoxNet **(C)**. In terms of number required features to achieve maximum predictive performance, SBG required 11 features and CoxNet 6 features **(D)**.

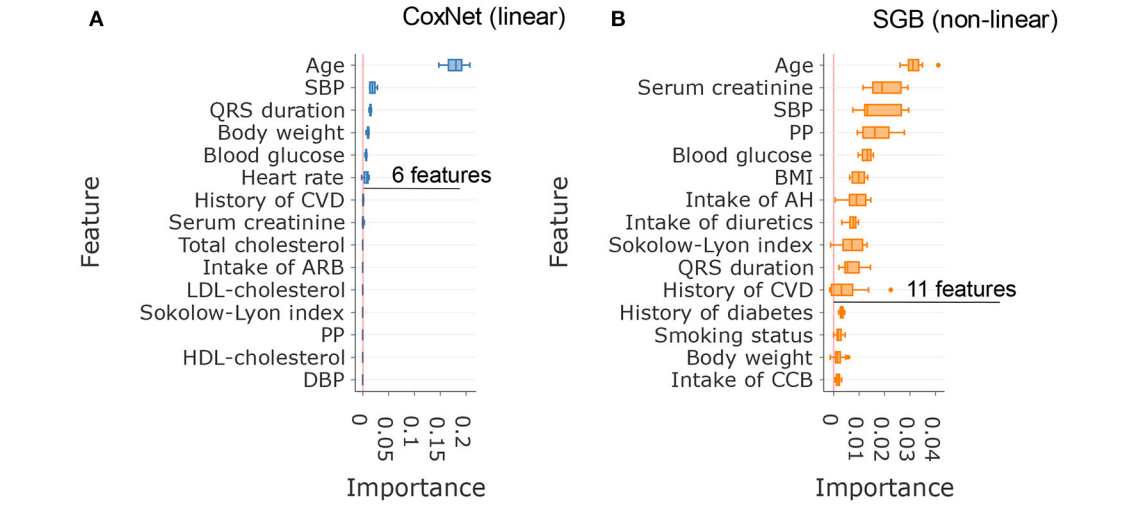


FIGURE 3
Comparison of the most important features in CoxNet **(A)** and survival gradient boosting **(B)**.

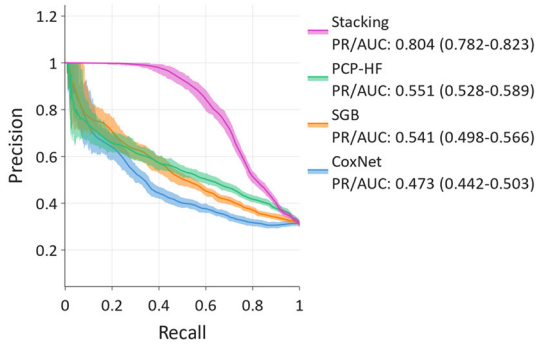


FIGURE 4
Precision-recall curves for 10-year prediction showing stacking, Survival Gradient Boosting (SGB), CoxNet, and PCP-HF. Stacking exhibits the best discrimination.

in limited data (single-center) and those trained in a more comprehensive sample (multi-center) (Figure 1).

All optimization/training pipelines were run 10 times, each time with a different random seed, producing 10 slightly different models. To provide a more stable prediction, each prediction was a mean of predictions of these 10 models. To obtain a more balanced estimate, samples were weighted inversely to the proportions of sizes of their corresponding cohorts. Therefore, in effect, each cohort had the same influence on predictive performance. We performed the two-sample Kolmogorov-Smirnov test to evaluate the statistical significance of the difference in c-index of the models. We also performed a feature importance analysis to illustrate which features were important for specific models and, therefore, might carry clinically useful information.

In addition, we used the 10-year binary endpoint to evaluate the models in clinically relevant precision-recall (PR) analysis (Figure 1). PR analysis is more sensitive to differences in false positives and thus better captures the practical aspect of clinical decision making. In the process of binarizing the outcome, subjects censored before the endpoint were removed. The predictions in the calibration plot were divided into bins on the x axis according to the predicted probability of an event, and this probability was put into comparison with empirical event incidence in the given bin on the y axis. 95% confidence intervals were calculated from 100 bootstrapped test scores, thus quantifying uncertainty on the unseen data (assuming the training data were given).

Results

Study population

The dataset consisted of 30,354 individuals (mean age 66 ± 9 years, 66.57% male), free of HF at the baseline. For detailed cohort characteristics and missing values, see Supplemental Tables 1, 2, respectively. During a median follow-up of 5.40 years (IQR 4.28–6.52), 1,068 individuals experienced at least one non-fatal HF event (6.46 events per 1,000 person-years). For a detailed overview of the outcome statistics, see Supplemental Table 3. In this study, we also evaluated the HF prediction at the endpoint of 10 years. After removing censored subjects, we obtained 2,883 subjects (891 events) for the 10-year prediction (Figure 1).

Comparing predictive performance of single-/multi-center and linear/non-linear models

As illustrated in Figures 2A,B, multi-center models systematically outperformed single-center models. For example, when testing on the FLEMENGHO cohort, multi-center SGB achieved c-index of 0.812, while only 0.714 in a single-center setting. Overall, the c-index in the pooled population (Figure 2C) was significantly higher ($P < 0.0001$) in non-linear SGB (0.735; 95% CI: 0.728–0.742) than in linear CoxNet (0.694; 95% CI: 0.686–0.704). These c-index values corresponded to risk discrimination between all individuals, and, therefore, represent the situation of a heterogeneous population. The calibration plot in Supplemental Figure 1 showed good calibration of both methods, but SGB showed a certain overestimation of risk in some individuals with lower risk.

Important features for incident HF prediction

In this analysis, SGB achieved maximal predictive performance with about 11 features, whereas CoxNet exploited fewer features and achieved its maximal predictive performance with only 6 features (Figure 2D). Figure 3 showed the importance of the permutation obtained from changes in predictive performance when supplied with shuffled input features. The CoxNet predictive model relied heavily on age, with a smaller influence of several other predictors, including SBP, QRS duration, and body weight. On the contrary, SGB employed a wider range of features beyond age, including serum creatinine, blood pressure, blood glucose, BMI, ECG features, antihypertensive treatment, and CV disease history.

Precision-recall of 10-year HF prediction using novel and traditional scores

For predicting the 10-year risk of incident HF, the stacking method, combining the SGB, CoxNet, Gaussian mixture, and

PCP-HF models, achieved the best discrimination in the PR analysis (Figure 4), with PR/AUC 0.804 (95% CI: 0.782–0.823). SGB and PCP-HF performed similarly with PR/AUC of 0.541 (0.498–0.566) and 0.551 (0.528–0.589), respectively. CoxNet achieved PR/AUC of 0.473 (0.442–0.503).

Discussion

Predictive performance of ML and linear methods in incident HF prediction

In the line with previous studies focusing on the comparison of ML methods with linear-based scores (9, 13–15), our study supported the findings that the ML model (SGB) detected more subtle patterns for incidence HF prediction than the linear CoxNet. Simultaneously, the predictive performance of ML algorithms varies dependent on a given task. Given the heterogeneity of predictive performance, stacking is a suitable tool capable of increasing predictive performance by combining the output of several estimators (models) using another ML meta-model. In our study, the stacking model, consisting of SGB, CoxNet, Gaussian Mixture, and the PCP-HF score, outperformed all tested models for 10-year incident HF prediction.

The full capacity of ML in the deep characterization of clinical phenotypes and therefore the delivery of personalized medicine is still unknown. The potential of using ML in HF prediction (and CV more broadly) depends strongly on the depth and volume of representative data points available. For example, Balabaeva et al. reported that temporal features increased the predictivity of ML in assessment of symptomatic HF prognosis (16). Therefore, integration of the temporal domain into prediction models might be the next logical step in the application of ML in CV/HF risk stratification. Using two or more data points could help extrapolate the rate of change in one's health status more precisely.

Multi-center data increased predictive performance in HF prediction

The collection of representative data is crucial for success in the development of reliable risk stratification tools. Our analysis showed that models trained on a combination of cohorts outperformed models trained on a single cohort (Figures 2A,B). These results advocated for use of large and representative training data for developing robust prediction tools. Indeed, flexible ML models, such as gradient boosting, require a large training sample to cover the full feature space. However, clinical data use must be in accordance with data privacy and might struggle with pitfalls in data sharing.

Dataset shift and multi-center data

One of the main reasons for maximizing the volume of training data is to overcome the so-called "dataset shift" problem. This problem relates to a shift in the distribution of derivation and validation (test, real-world) cohorts (17). It occurs for various reasons, such as differences in population characteristics, hospital procedures, selection bias, etc. The dataset shift manifests itself as a decrease in predictive performance when the model is tested on an unseen population, and therefore the external validation is crucial in estimating the generalization of the model. One straightforward and effective approach to combating the dataset shift is to train flexible ML models on a large volume of representative data (17). An example of difficulties originating from the lack of robust predictive models trained on such diverse data is the need to recalibrate current HF and/or CV risk scores for each tested population (5). With a greater number and variety of training samples, ML models can learn a wider range of health profiles. Flexible ML algorithms, such as gradient boosting, can thus be used to capture these diverse distributions and produce more precise, personalized models for the prediction of adverse events. For instance, the ML model that captures personalized characteristics (e.g., ethnic differences) achieves better predictive performance compared to the traditional one (9).

Data sharing issues and regulation

As outlined above, to effectively evaluate the benefit of ML in HF risk prediction, models need to be trained and evaluated on diverse populations with extensive deep phenotypes. On the other hand, there are some problems related to data sharing and the aggregation of sensitive healthcare information, as the protection of personal information is of the utmost importance. This problematic is materialized in General Data Protection Regulation (GDPR), which through various means, including data minimization, strives to limit sharing and storage of sensitive information. However, without employing the full potential of high-quality data collected globally, we will end up with numerous ML models struggling to generalize and therefore deliver clinical value. To overcome this issue, the use of federated learning (FL) could be an alternative (18).

Federated learning

The objective of FL is to train ML models without the requirement to collect data in a central place. FL algorithms extract only the generally applicable clinical, behavioral, and physiological characteristics, without the

requirement for sensitive personal information. However, there is a need for user-friendly FL tools to perform such analyses in a practical way. Such tools could allow not only federated learning (privacy-preserving collaborative model training) but also federated analysis which could supercharge the scientific progress in the CV field (and others). Similarly, the integration of deep phenotypes and the temporal domain will require specific tools for clinical practice.

AI-based HF risk stratification tool in clinical practice

To effectively deploy a validated and robust predictive model in clinical practice, integration into the clinical workflow must be as seamless as possible. This could require integrating information from the electronic health records (EHR). However, it might be notoriously difficult, e.g., in the European settings, as there is currently a limited interoperable infrastructure. The use of EHR should be human-centric with individuals having full control over the use of their data. FL is compatible with this approach by protecting personal data. One approach to overcoming the heterogeneity of the healthcare systems is to connect systems through a system-specific bridge (adapter). Another approach is to use deep learning, which learns feature representations directly from the EHR data. This system can then transform the unstructured data from different systems into a structure that can be pipelined into another ML algorithm. These systems should be however open to ensure trust and fairness and to facilitate integration into other systems. Additionally, these systems might raise questions regarding reliability and fairness. With many current scores, there are online calculators allowing easy-to-use estimation of individual risk. However, use in the clinical setting should be cautious, as data privacy policy is not always clear and could be a possible target of 3rd-party malware attacks (the browser is more vulnerable). Therefore, an AI-based app would serve as a more secure way to handle personal data and would be more robust (e.g., working under internet connectivity disruptions). Independently on the data input method, the user experience (UX) needs to be user-friendly and informative to aid in doctor-patient communication and to provide the benefit of improved accuracy of ML-based predictive models.

Conclusion

With a greater number and variety of training cohorts, the model learns a wider range of specific individual health characteristics. Flexible ML algorithms as well as the stacking methods can be used to capture these diverse distributions and produce more precise, personalized models for the prediction of adverse events.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: The use of database should be approved by the PI of the HOMAGE study. Requests to access these datasets should be directed to Prof. Faiez Zannad, f.zannad@chru-nancy.fr.

Ethics statement

The studies involving human participants were reviewed and approved by Ethics Committee Research UZ/KU Leuven. The patients/participants provided their written informed consent to participate in this study.

Author contributions

FS and TK: design the sub-analysis and drafted the manuscript. TK: HOMAGE database was maintained at the Research Unit Hypertension and Cardiovascular Epidemiology in Leuven, Belgium. FS: wrote the code and performed the statistical analysis. EN and NC: provided critical revisions, read, and approved the final version of the manuscript.

Funding

The Research Unit Hypertension and Cardiovascular Epidemiology currently received grants from Internal Funds KU Leuven (C24M/21/025) and the Research Foundation Flanders (FWO Grants 1225021N, 1S07421N, and G0C5319N).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those

of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcvm.2022.1011071/full#supplementary-material

References

1. Lippi G, Sanchis-Gomar F. Global epidemiology and future trends of heart failure. *AME Med J.* (2020) 5:15. doi: 10.21037/amj.2020.03.03

2. Butler J, Kalogeropoulos A, Georgiopoulou V, Belue R, Rodondi N, Garcia M, et al. Incident heart failure prediction in the elderly. *Circ Heart Fail.* (2008) 1:2. doi: 10.1161/CIRCHEARTFAILURE.108.768457

3. Agarwal SK, Chambless LE, Ballantyne CM, Astor B, Bertoni AG, Chang PP, et al. Prediction of incident heart failure in general practice. *Circ Heart Fail.* (2012) 5:4. doi: 10.1161/CIRCHEARTFAILURE.111.964841

4. Kannel WB, D'Agostino RB, Silbershatz H, Belanger AJ, Wilson PW, Levy D. Profile for estimating risk of heart failure. *Arch Intern Med.* (1999) 159:11. doi: 10.1001/archinte.159.11.1197

5. SCORE2 Working Group and ESC Cardiovascular Risk Collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J.* (2021) 42:25. doi: 10.1093/eurheartj/ehab369

6. SCORE2-OP Working Group and ESC Cardiovascular Risk Collaboration. SCORE2-OP risk prediction algorithms: estimating incident cardiovascular event risk in older persons in four geographical risk regions. *Eur Heart J.* (2021) 42:25. doi: 10.1093/eurheartj/ehab312

7. Khan SS, Ning H, Shah SJ, Yancy CW, Carnethon M, Berry JD, et al. 10-Year risk equations for incident heart failure in the general population. *J Am Coll Cardiol.* (2019) 73:19. doi: 10.1016/j.jacc.2019.02.057

8. Friedrich S, Groß S, König IR, Engelhardt S, Bahls M, Heinz J, et al. Applications of artificial intelligence/machine learning approaches in cardiovascular medicine: a systematic review with recommendations. *Eur Heart J Digit Health.* (2021) 2:3. doi: 10.1093/ehjdh/ztab054

9. Segar MW, Jaeger BC, Patel KV, Nambi V, Ndumele CE, Correa A, et al. Development and validation of machine learning–based race-specific models to predict 10-year risk of heart failure: a multicohort analysis. *Circulation.* (2021) 143:24. doi: 10.1161/CIRCULATIONAHA.120.053134

10. Miller RJH, Sabovčik F, Cauwenberghs N, Vens C, Khush KK, Heidenreich PA, et al. Temporal shift and predictive performance of machine learning for heart transplant outcomes. *J Heart Lung Transplant.* (2022) 41:7. doi: 10.1016/j.healun.2022.03.019

11. Jacobs L, Thijs L, Jin Y, Zannad F, Mebazaa A, Rouet P, et al. Heart 'omics' in AGEing (HOMAGE): design, research objectives and characteristics of the common database. *J Biomed Res.* (2014) 28:5. doi: 10.7555/JBR.28.20140045

12. Jacobs L, Efremov L, Ferreira JP, Thijs L, Yang W, Zhang Z, et al. Risk for incident heart failure: a subject-level meta-analysis from the heart 'OMics' in AGEing (HOMAGE) Study. *J Am Heart Assoc.* (2017) 6:5. doi: 10.1161/JAHA.116.005231

13. Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, et al. Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. *Circ Res.* (2017) 121:9. doi: 10.1161/CIRCRESAHA.117.311312

14. Segar MW, Vaduganathan M, Patel KV, McGuire DK, Butler J, Fonarow GC. et. al Machine learning to predict the risk of incident heart failure hospitalization among patients with diabetes: The WATCH-DM risk score. *Diabetes Care.* (2019) 42:12. doi: 10.2337/dc19-0587

15. Guo CY, Wu MY, Cheng HM. The Comprehensive machine learning analytics for heart failure. *Int J Environ Res Public Health.* (2021) 18:9. doi: 10.3390/ijerph18094943

16. Balabaeva K, Kovalchuk S. Comparison of temporal and non-temporal features effect on machine learning models quality and interpretability for chronic heart failure patients. *Procedia Comput Sci.* (2019) 156:1. doi: 10.1016/j.procs.2019.08.183

17. Dockès J, Varoquaux G, Poline JB. Preventing dataset shift from breaking machine-learning biomarkers. *Gigascience.* (2021) 10:9. doi: 10.1093/gigascience/giab055

18. Li Q, Wen Z, Wu Z, Hu S, Wang N, Li Y, et al. A survey on federated learning systems: vision, hype and reality for data privacy and protection. *IEEE Trans Knowl Data Eng.* (2021). doi: 10.1109/TKDE.2021.3124599

frontiers | Frontiers in Cardiovascular Medicine

TYPE Original Research
PUBLISHED 28 October 2022
DOI 10.3389/fcvm.2022.916429

Check for updates

OPEN ACCESS

EDITED BY
Ling Yang,
First People's Hospital of Changzhou, China

REVIEWED BY
Jiannan Yang,
City University of Hong Kong, Hong Kong SAR, China
Zhongzhi Xu,
City University of Hong Kong, Hong Kong SAR, China

*CORRESPONDENCE
Zhifu Guo
13918459432@163.com
Xiaowei Song
xiao_wei_song@163.com
Xianxian Zhao
zhaoxianxian2022@163.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Heart Failure and Transplantation, a section of the journal Frontiers in Cardiovascular Medicine

RECEIVED 09 April 2022
ACCEPTED 30 September 2022
PUBLISHED 28 October 2022

CITATION
Tu D, Ma C, Zeng Z, Xu Q, Guo Z, Song X and Zhao X (2022) Identification of hub genes and transcription factor regulatory network for heart failure using RNA-seq data and robust rank aggregation analysis. *Front. Cardiovasc. Med.* 9:916429. doi: 10.3389/fcvm.2022.916429

COPYRIGHT
© 2022 Tu, Ma, Zeng, Xu, Guo, Song and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Identification of hub genes and transcription factor regulatory network for heart failure using RNA-seq data and robust rank aggregation analysis

Dingyuan Tu[1†], Chaoqun Ma[1†], ZhenYu Zeng[1], Qiang Xu[2], Zhifu Guo[1*], Xiaowei Song[1*] and Xianxian Zhao[1*]

[1]Department of Cardiology, Changhai Hospital, Naval Medical University, Shanghai, China, [2]Department of Cardiology, Navy 905 Hospital, Naval Medical University, Shanghai, China

Background: Heart failure (HF) is the end stage of various cardiovascular diseases with a high mortality rate. Novel diagnostic and therapeutic biomarkers for HF are urgently required. Our research aims to identify HF-related hub genes and regulatory networks using bioinformatics and validation assays.

Methods: Using four RNA-seq datasets in the Gene Expression Omnibus (GEO) database, we screened differentially expressed genes (DEGs) of HF using Removal of Unwanted Variation from RNA-seq data (RUVSeq) and the robust rank aggregation (RRA) method. Then, hub genes were recognized using the STRING database and Cytoscape software with cytoHubba plug-in. Furthermore, reliable hub genes were validated by the GEO microarray datasets and quantitative reverse transcription polymerase chain reaction (qRT-PCR) using heart tissues from patients with HF and non-failing donors (NFDs). In addition, R packages "clusterProfiler" and "GSVA" were utilized for enrichment analysis. Moreover, the transcription factor (TF)–DEG regulatory network was constructed by Cytoscape and verified in a microarray dataset.

Results: A total of 201 robust DEGs were identified in patients with HF and NFDs. STRING and Cytoscape analysis recognized six hub genes, among which *ASPN*, *COL1A1*, and *FMOD* were confirmed as reliable hub genes through microarray datasets and qRT-PCR validation. Functional analysis showed that the DEGs and hub genes were enriched in T-cell-mediated immune response and myocardial glucose metabolism, which were closely associated with myocardial fibrosis. In addition, the TF–DEG regulatory network was constructed, and 13 significant TF–DEG pairs were finally identified.

Conclusion: Our study integrated different RNA-seq datasets using RUVSeq and the RRA method and identified *ASPN*, *COL1A1*, and *FMOD* as potential diagnostic biomarkers for HF. The results provide new insights into the underlying mechanisms and effective treatments of HF.

KEYWORDS

heart failure, RNA-seq dataset, RUVSeq, robust rank aggregation, hub gene, biomarker, functional enrichment analysis, transcription factor

Introduction

Heart failure (HF) is a complex clinical syndrome that results from dysfunction of ventricular filling or ejection, characterized by a variety of worsening symptoms and signs, including dyspnea, fatigue, and fluid retention (1). The occurrence of HF is predominantly caused by underlying myocardial diseases, while cardiac lesions from valves, vasculature, pericardium, heart rate/rhythm, or a combination of cardiac abnormalities may also result in cardiac malfunction (2). Despite the development of drug therapy and surgical interventional therapy, the morbidity and mortality of HF are increasing annually worldwide, which seriously threatens human health and quality of life (3, 4). Therefore, to improve the curative efficacy, it remains urgent to investigate the in-depth underlying molecular mechanisms of HF to facilitate its accurate diagnosis, early intervention, and precision therapy.

In recent years, the rapid progress of transcriptome sequencing technology provides new directions for the exploration of epigenetic changes and molecular mechanisms in different diseases, including neoplastic and non-neoplastic diseases (5, 6). Accordingly, an increasing volume of RNA sequencing (RNA-seq) and microarray datasets of HF has been uploaded in the Gene Expression Omnibus (GEO) database, providing opportunities for bioinformatics data mining of marker genes associated with HF (7). However, in comparison to cancer-related surgery, the number of heart transplantation surgeries is relatively small, which results in the small sample size and large batch effects of RNA sequencing or microarray datasets of HF. Therefore, to date, the bioinformatics data mining of HF still faces great challenges, especially regarding the integration and analysis of the RNA-seq data (RUVSeq) related to HF.

The robust rank aggregation (RRA) method, first proposed in 2012 by Kolde et al., is a rigorous approach using probabilistic models to analyze the significant probability of all elements in different sequencing or microarray datasets (8). Recently, the RRA algorithm has been extensively used to integrate data in different microarray platforms to screen the differentially expressed genes (DEGs) in multiple diseases, including thyroid carcinoma (9), prostate cancer (PCa) (10), and DCM (11). For example, Song et al. utilized the RRA method to integrate 10 eligible PCa microarray datasets from the GEO and identify four candidate biomarkers for prognosis of PCa (10). Ma et al. integrated four eligible dilated cardiomyopathy (DCM) microarray datasets from the GEO database and developed a 7-gene signature predictive of DCM by utilizing the RRA method (11). However, due to the greater difficulty in integrating sequencing data, the application of the previous RRA algorithm was limited to microarray data, and the RRA analysis of RUVSeq was still rarely reported. Removal of Unwanted Variation from RUVSeq, a Bioconductor package that generalizes a linear model to regress variance estimated from the expression of housekeeping genes, has been reported to be used to reduce batch effects due to different sequencing conditions (12), which provides a huge possibility for the combination of RUVSeq and the RRA method in integrating different RUVSeq sets and identifying HF-associated DEGs.

In the present study, RUVSeq and RRA analysis were performed for the first time based on four RNA-seq datasets in the GEO database to identify robust DEGs in HF samples and non-failing donor (NFD) samples, followed by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis for the DEGs. Moreover, three reliable HF-related hub genes with differential expression and excellent diagnostic efficiency, *ASPN*, *COL1A1*, and *FMOD*, were selected and validated using microarray datasets and human heart tissue assays. Gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA) were further utilized to investigate potential functions of the hub genes. In addition, the transcription factor (TF)–DEG regulatory network was constructed based on the HF datasets and websites.

Materials and methods

Datasets search and inclusion criteria

The GEO database[1] was searched to obtain the sequencing datasets based on the search terms of "heart failure" or/and "HF." The search results and relevant datasets were filtered according to the following inclusion criteria: (i) the organism was filtered by "homo sapiens"; (ii) the study type was

1 www.ncbi.nlm.nih.gov/geo/

set as "expression profiling by high throughput sequencing"; (iii) RUVSeq for both HF samples and NFDs should be included in the dataset; (iv) the total number of samples should not be < 5; and (v) the raw data of the RNA-seq should be provided for reanalysis. Datasets that did not meet the aforementioned criteria were excluded. The selected HF sequencing datasets from the NCBI Sequence Read Archive (SRA)[2] were downloaded as SRA files and converted to FASTQ files *via* the SRA toolkit.

Compilation of gene expression matrices

To obtain high-quality reads, raw data from the GEO dataset were pre-processed using the fastp tool (13), and sample quality was assessed by FastQC and MultiQC (14). The sequences were then aligned against the human reference genome hg38 using STAR (15). Furthermore, the expression values (count matrices) for either gene bodies or called peaks were generated by featureCounts (16).

Identification of robust differentially expressed genes by the RNA-seq data and robust rank aggregation method

For RNA-seq expression analysis, batch effects were adjusted using the R package RUVSeq, which applies a generalized linear model to regress out the variation estimated from the expression of the housekeeping gene. First, the initial DEGs were detected using the edgeR program package within a single RNA-seq dataset. Second, the RUVg function in RUVSeq was utilized to remove additional sources of unwanted variation (parameter $k = 1$) (17). The remaining non-DEGs were considered as negative control genes and used as housekeeping genes to correct for relative gene expression levels between different samples. Third, based on the corrected gene expression matrix, the corrected DEGs were further obtained by the edgeR package. Fourth, the RRA method-based R package "RobustRankAggreg" was used to integrate the results of RUVSeq analysis of each RNA-seq dataset to identify the final DEGs in patients with HF compared with NFDs. The threshold of DEGs was set as $|\log FC| > 1$, and the significance criterion was set as an adjusted p-value < 0.05.

Functional enrichment analysis

To further investigate the possible functions of DEGs identified by the RUVSeq and the RRA method, GO enrichment and KEGG pathway analyses were performed in the upregulated and downregulated DEGs separately, using the R package "clusterProfiler" (18). The GO term or KEGG pathway with adjusted $p < 0.05$ was considered with significant enrichment. The results were visualized by dot plots using the "dotplot" function of the R package.

Identification of hub genes

The robust DEG list was uploaded to the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database[3] (19), and the significant protein interaction was determined at the criterion of confidence (combined score) > 0.4. Next, we used Cytoscape software[4] and cytoHubba (20) plug-in to investigate node composition and pick out hub nodes with a high degree of connectivity in the network.

Validation of the hub genes using microarray datasets

RNA-seq datasets for HF samples are limited due to a small volume of heart transplant surgeries and the difficulty in obtaining human heart samples. Therefore, in our study, the four eligible HF sequencing datasets (GSE46224, GSE116250, GSE133054, and GSE135055), including 95 HF and NFD samples, were all used for the identification of DEGs, hub genes, and functional enrichment analysis. To further validate the analysis results, HF microarray datasets were acquired from the GEO database. The inclusion criterion was identical to the RUVSeq sources, except that the study type was set as "expression profiling by array." For the study, four microarray datasets were finally included for the validation: GSE16499 (21), GSE26887 (22), GSE57338 (23), and GSE79962 (24). The gene expression profiling was annotated using the annotation document of corresponding platforms, and the gene expression matrices were column-normalized by the R package "limma" (25) and log-transformed, if necessary. Next, the differential expression of the identified hub genes between patients with HF and NFDs in the microarray datasets was validated and visualized by column graphs.

Validation of the hub genes using quantitative reverse transcription polymerase chain reaction

For further validation, total RNAs of the heart tissues from patients with HF and NFDs were extracted for the

2 https://www.ncbi.nlm.nih.gov/sra/

3 https://string-db.org/

4 www.Cytoscape.org/

qRT-PCR validation assay. Heart tissues from six patients with HF and eight NFDs were obtained from the Specimen Bank of Cardiovascular Surgery Laboratory and Department of Pathology of Changhai Hospital, Shanghai, China. Written informed consents were obtained from all patients or their family members, and the study was approved by the Institute Ethics Committee of Changhai Hospital.

Total RNAs from the heart tissues were isolated using TRIzol reagent (TRIzol^{TM} Reagent, Invitrogen). RNAs were then reverse-transcribed into cDNAs using a TOYOBO

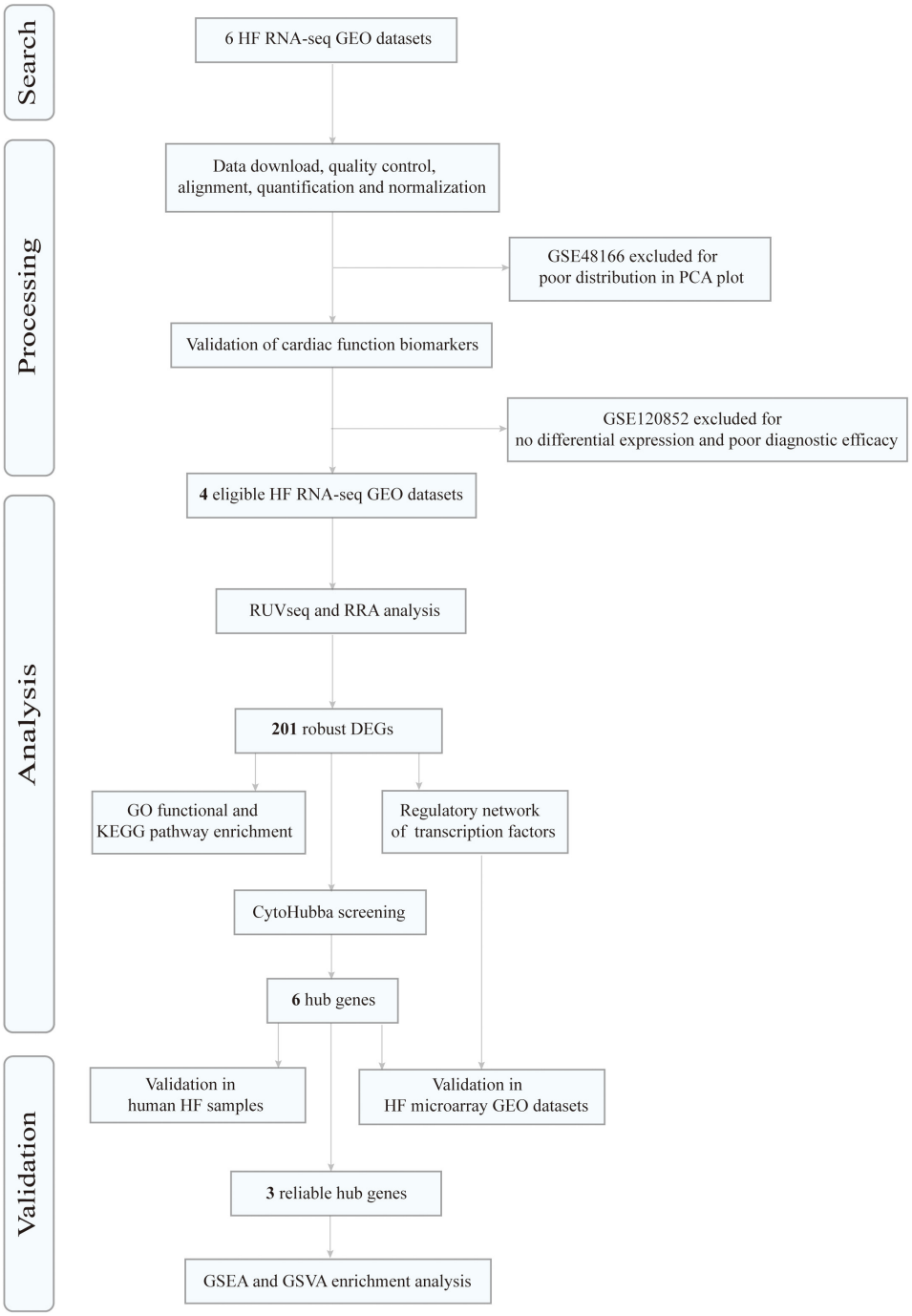


FIGURE 1
Flowchart of data search, processing, analysis, and validation. HF, heart failure; RNA-seq, RNA sequencing; GEO, Gene Expression Omnibus; PCA, principal component analysis; RRA, robust rank aggregation; DEGs, differentially expressed genes; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; GSEA, gene set enrichment analysis; GSVA, gene set variation analysis.

ReverTra Ace® qRT-PCR RT Kit (TOYOBO, Japan). SYBR® GREEN (TOYOBO, Japan) was used for qRT-PCR, and the primer sequences used are listed as follows: asporin (*ASPN*) forward, 5′-GGGTGACGGTGTTCCATATC-3′ and reverse, 5′-TTGGCACTGTTGGACAGAAG-3′; collagen type I alpha 1 chain (*COL1A1*) forward, 5′-TCG TGGAAATGATGGTGCTA-3′ and reverse, 5′-ACCAGGTT CACCGCTGTTAC-3′; collagen type IX alpha 2 chain (*COL9A2*) forward, 5′-AAGAGCAACTGGCAGAGGTC-3′ and reverse, 5′-GACCCTCGATCTCCATCCTT-3′; collagen type X alpha 1 chain (*COL10A1*) forward, 5′-TGGGACCCCTC TTGTTAGTG-3′ and reverse, 5′-GCCACACCTGGTCA TTTTCT-3′; cartilage oligomeric matrix protein (*COMP*) forward, 5′-CAGGACGACTTTGATGCAGA-3′ and reverse, 5′-AAGCTGGAGCTGTCCTGGTA-3′; and fibromodulin (*FMOD*) forward, 5′-AGAGAGCTCCAT CTCGACCA-3′ and reverse, 5′-GCAGCTGGTTGT AGGAGAGG-3′. The expression levels of mRNAs relative to glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) were detected using the $2^{-\Delta\Delta Ct}$ method.

Gene set enrichment analysis and gene set variation analysis of the validated hub genes

To further explore potential functions of the hub genes in HF, we performed GSEA and GSVA in the microarray dataset with the maximum HF sample size (GSE57338). The flow of GSEA is as follows: First, correlation analyses were conducted between hub genes and other genes in the gene expression matrix of 54 patients with HF, and genes with the absolute value of correlation coefficient > 0.5 and p-value < 0.05 were defined as hub genes-related genes. Then, KEGG pathway enrichment analysis was conducted on these hub genes-related genes using the ClusterProfiler package. For GSVA, 54 patients with HF in the GSE57338 dataset were divided into two groups based on the median expression level of each hub gene (high- and low-expression groups). Then, the "GSVA" package was used to explore the pathways associated with the hub genes. The annotated gene set "c2.cp.kegg.v7.4.entrez.gmt" in the Molecular Signatures Database (MsigDB)[5] was selected as the reference.

Construction of the transcription factor–differentially expressed gene regulatory network

It has been reported that binding of TFs to the regulatory regions of genes is a key transcriptional regulatory mechanism to control chromatin and transcription, forming a complex system that guides expression of the genome (26). The TF–DEG regulatory network is constructed by using the following methods: First, the NetworkAnalyst database (27)[6] and the TF–gene interactions module from the JASPAR database (28) were utilized to explore the possible TFs that could bind to the RRA-identified DEGs. Second, a novel significant TF–DEG regulatory pair was defined in our study according to the following criteria: (i) both the TF and DEG were present in the TF regulatory network constructed by the JASPAR database, and there was predicted interaction between them; (ii) the TF was differentially expressed in patients with HF and NFD samples in the validation set GSE57338 ($p < 0.05$); and (iii) there was a statistically significant relationship between the expression level of TF and its target gene in the validation dataset GSE57338 (the absolute value of correlation coefficient > 0.5 and $p < 0.05$). Third, the constructed TF–DEG regulatory network was visualized using Cytoscape.

Statistical analysis

Independent two-sample t-tests were used to analyze variables with homogeneous variance and normal distribution, whereas Mann–Whitney non-parametric tests were used to analyze variables without homogeneous variance and normal distribution. P-values were adjusted for multiple testing by using the Benjamini–Hochberg method. The DEG threshold was set as $|\log FC| > 1$, and the significance criterion was set as an adjusted p-value < 0.05. The hypergeometric test was used to calculate the statistical significance of enrichment analysis. An absolute value of the correlation coefficient $|r| > 0.3$ ($p < 0.05$) indicates a significant interaction relationship (29). All data analyses in the present study were performed by using R (version 3.5.3) and Rstudio (version 1.2.1335). Graphic representations were generated by using GraphPad Prism 9.0 (GraphPad, San Diego, CA, USA) and Cytoscape (Version 3.7.1).

Results

Characteristics of the screened heart failure RNA-seq datasets

Figure 1 depicts the flow diagram of our study. After screening and exclusion according to the aforementioned criteria, six datasets from the GEO database were finally included in this analysis: GSE46224 (30), GSE48166, GSE116250 (31), GSE120852 (32), GSE133054 (33), and GSE135055 (34). The characteristics of these six datasets are summarized in

5 http://www.gsea-msigdb.org/gsea/downloads.jsp

6 https://www.networkanalyst.ca/

Supplementary Table 1, including the GSE accession number, study country, number of patients with HF and NFDs, and sequencing platform.

Pre-processing of RNA-seq data

After the quality-filtering using the fastp tool, the reads with a base quality < 20 or the sequence length ≤ 36 nt were discarded. Then, FastQC was used to assess the sequence quality of the dataset. The final all-in-one quality control report of each dataset was generated using MultiQC. The per base sequence quality and per sequence GC content across all samples of 6 RNA-seq datasets are demonstrated in **Figure 2**.

Determination of the selected datasets

Reads were mapped to the human genome (UCSC, hg38) using STAR, and the unique alignments were filtered and presented in **Supplementary Table 2**. Samples from each dataset were characterized by principal component analysis (PCA) after normalization and adjustment for batch effects using the RUVSeq package. 2D plots of PCA distribution showed that complete separation between samples of patients with HF and NFD samples was observed in five datasets (GSE46224, GSE116250, GSE120852, GSE133054, and GSE135055), except GSE48166 (**Figure 3**). Hence, dataset GSE48166 was excluded from subsequent analysis. Next, the expression difference and diagnostic efficacy of the four cardiac function markers, namely, natriuretic peptide A (*NPPA*), natriuretic peptide B (*NPPB*), myosin heavy chain 6 (*MYH6*), and myosin heavy chain 7 (*MYH7*), were examined between samples of patients with HF and NFD samples in the five sequencing databases. As shown in **Figure 4**, the markers showed no differential expression and poor diagnostic performance between the two groups in dataset GSE120852, which was also eliminated from further analysis.

Identification of robust differentially expressed genes by RNA-seq data and robust rank aggregation method

Using the RUVSeq package, DEGs (patients with HF vs. NFDs) were screened for adjusted $p < 0.05$ and $|\log FC| > 1$ in the four identified datasets, respectively, which were visualized by volcano plots (**Figures 5A–D**). Furthermore, an integrated analysis was performed using the R package "RobustRankAggreg" to generate the differentially expressed ranked gene list. A total of 201 highly ranked DEGs were identified in patients with HF vs. NFD samples, and **Supplementary Table 3** exhibits the top 50 upregulated and the top 50 downregulated DEGs. The top 20 upregulated and the 20 most downregulated genes consistently expressed across all datasets were visualized by heatmap, as shown in **Figure 5E**.

Functional enrichment analysis of differentially expressed genes

To explore the potential biological functions of these DEGs, GO term enrichment and KEGG pathway analyses were performed. The upregulated genes were significantly enriched in extracellular structure organization, skeletal system development, extracellular matrix organization, T-cell activation, and connective tissue development in the biological process (BP) term; the extracellular matrix, collagen-containing extracellular matrix, endoplasmic reticulum lumen, basement membrane, and collagen trimer in the cellular component (CC) term; and extracellular matrix structural constituent, glycosaminoglycan binding, heparin binding, growth factor activity, and extracellular matrix structural constituent conferring tensile strength in the molecular function (MF) term (**Figure 6A**). For the downregulated genes, the enriched GO functions included purine ribonucleotide metabolic process, coenzyme metabolic process, energy derivation by oxidation of organic compounds, cellular respiration, and citrate metabolic process in the BP category; organelle inner membrane, mitochondrial inner membrane, mitochondrial matrix, mitochondrial protein complex, and mitochondrial membrane part in the CC category; and cofactor binding, coenzyme binding, and NAD binding in the MF category (**Figure 6B**).

Regarding KEGG pathway analysis, the MAPK signaling pathway, TGF-β signaling pathway, T-cell receptor signaling pathway, Th17 cell differentiation, and ECM–receptor interaction were mostly associated with the upregulated genes (**Figure 6C**), while the downregulated genes were most enriched in the calcium signaling pathway, carbon metabolism, valine, leucine and isoleucine degradation, citrate cycle, and propanoate metabolism (**Figure 6D**).

Hub gene determination

The PPI network of the 201 DEGs in patients with HF was constructed by using the STRING database (**Figure 7A**). Next, the PPI network was loaded into Cytoscape to screen hub genes by degree using the cytoHubba plug-in. As shown in **Figure 7B**, genes in the inner concentric circles have higher degrees, while genes in the outer concentric circles have relatively lower degrees. Therefore, hub genes were the six genes with the highest degree of connectivity (degree ≥ 10) in the innermost concentric circle: *COL1A1*, *COMP*, *ASPN*, *COL10A1*, *FMOD*, and *COL9A2*.

Furthermore, the relative expression of the identified hub genes in patients with HF and NFD samples was assessed in the

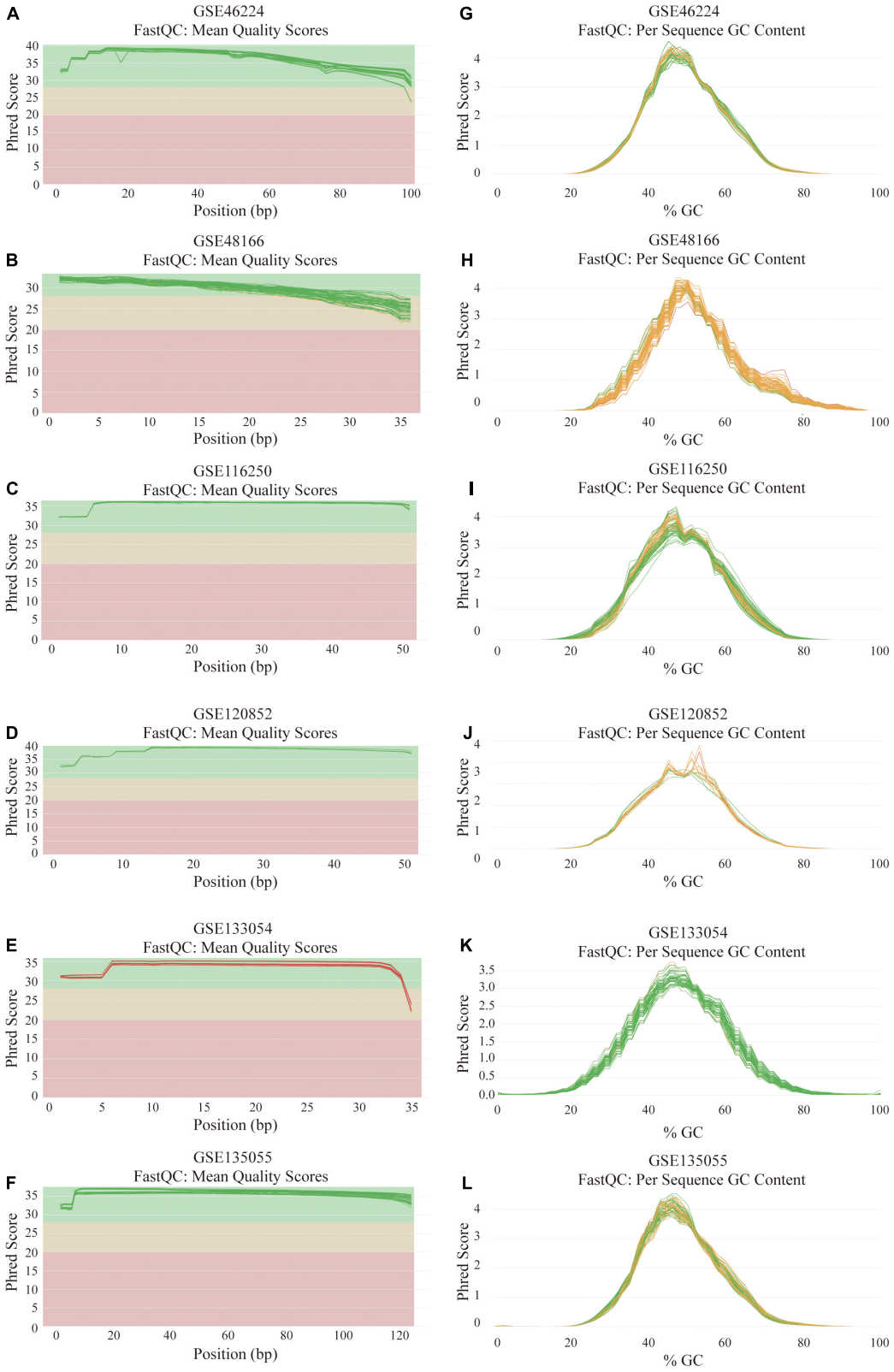


FIGURE 2
Quality assessment and GC count evaluation of the data from six RNA sequencing datasets. **(A–F)** Per base sequence quality across all samples of GSE46224 **(A)**, GSE48166 **(B)**, GSE116250 **(C)**, GSE120852 **(D)**, GSE133054 **(E)**, and GSE135055 **(F)**. **(G–L)** Per sequence GC content across all samples of GSE46224 **(G)**, GSE48166 **(H)**, GSE116250 **(I)**, GSE120852 **(J)**, GSE133054 **(K)**, and GSE135055 **(L)**.

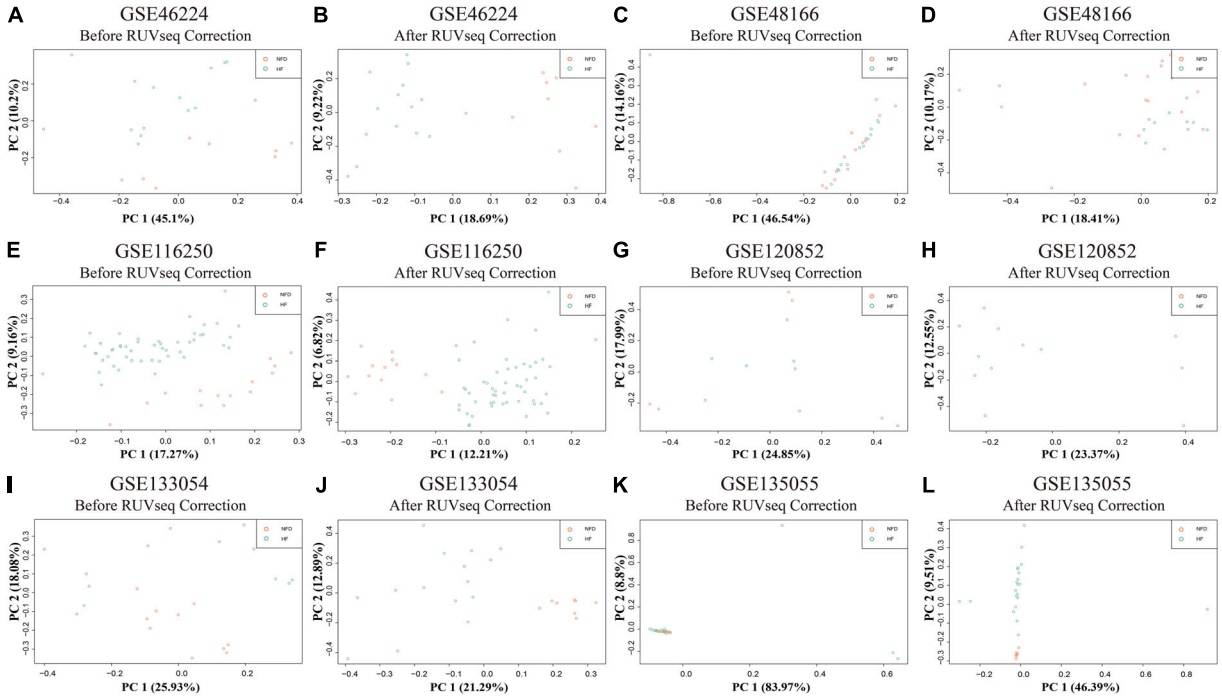


FIGURE 3
PCA plots of the six RNA-seq datasets in GEO database. PCA distribution plots showed that complete separation between patients with HF and NFD samples was observed in five datasets after RUVSeq correction, except GSE48166 **(C,D)**, namely, GSE46224 **(A,B)**, GSE116250 **(E,F)**, GSE120852 **(G,H)**, GSE133054 **(I,J)**, and GSE135055 **(K,L)**. PCA, principal component analysis; GEO, Gene Expression Omnibus; HF, heart failure; NFD, non-failing donor.

four RNA-seq datasets. The results showed that *COL1A1*, *ASPN*, and *FMOD* were consistently upregulated in the HF samples of the four datasets (Figures 7C–F). In addition, univariate ROC analysis was performed to determine the diagnostic accuracy of independent hub genes, suggesting that *COL1A1*, *ASPN*, and *FMOD* had a good diagnostic value in HF (Figures 7G–J).

Hub gene validation

After normalization, four microarray datasets (GSE16499, GSE26887, GSE57338, and GSE79962) containing human left ventricular samples of HF and NFDs were used to validate the expression of these hub genes (Supplementary Table 4 and Supplementary Figure 1). As shown in Figures 8A–D, the expression of *ASPN* or *FMOD* in HF samples was significantly higher than that in the NFD samples in all four datasets, and *COL1A1* or *COMP* showed the similar upregulation in three datasets. However, the expression of *COL9A2* or *COL10A1* was not statistically different in the HF and NFD samples in these datasets. Consistently, the diagnostic values of the hub genes suggested by the ROC curves revealed the same trend (Figures 8E–H).

In addition to the microarray datasets, the expression of hub genes was further validated by qRT-PCR experiments using 14 heart tissues from patients with HF or NFDs. As described in Figure 9, *ASPN*, *COL1A1*, and *FMOD* were significantly upregulated in the six heart tissues of patients with HF compared with NFDs. Taken together, these validation results confirmed the differential expression and diagnostic value of *ASPN*, *COL1A1*, and *FMOD* as reliable hub genes in HF development.

Gene set enrichment analysis and gene set variation analysis reveal a close relationship between hub genes and glucose metabolism-related pathways

To reveal the underlying mechanism of the three reliable hub genes (*ASPN*, *COL1A1*, and *FMOD*) involved in HF, GSEA was conducted to explore significantly enriched pathways associated with the hub genes in the validation dataset GSE57338. As shown in Figures 10A–C, the top three signaling pathways enriched by the DEGs between subgroups were identified, among which citrate cycle (TCA cycle) and propanoate metabolism pathways were significantly enriched in the subgroups of all the three hub genes. In addition, the enrichment in glucose metabolism-related pathways was further confirmed by GSVA (Figures 10D–F).

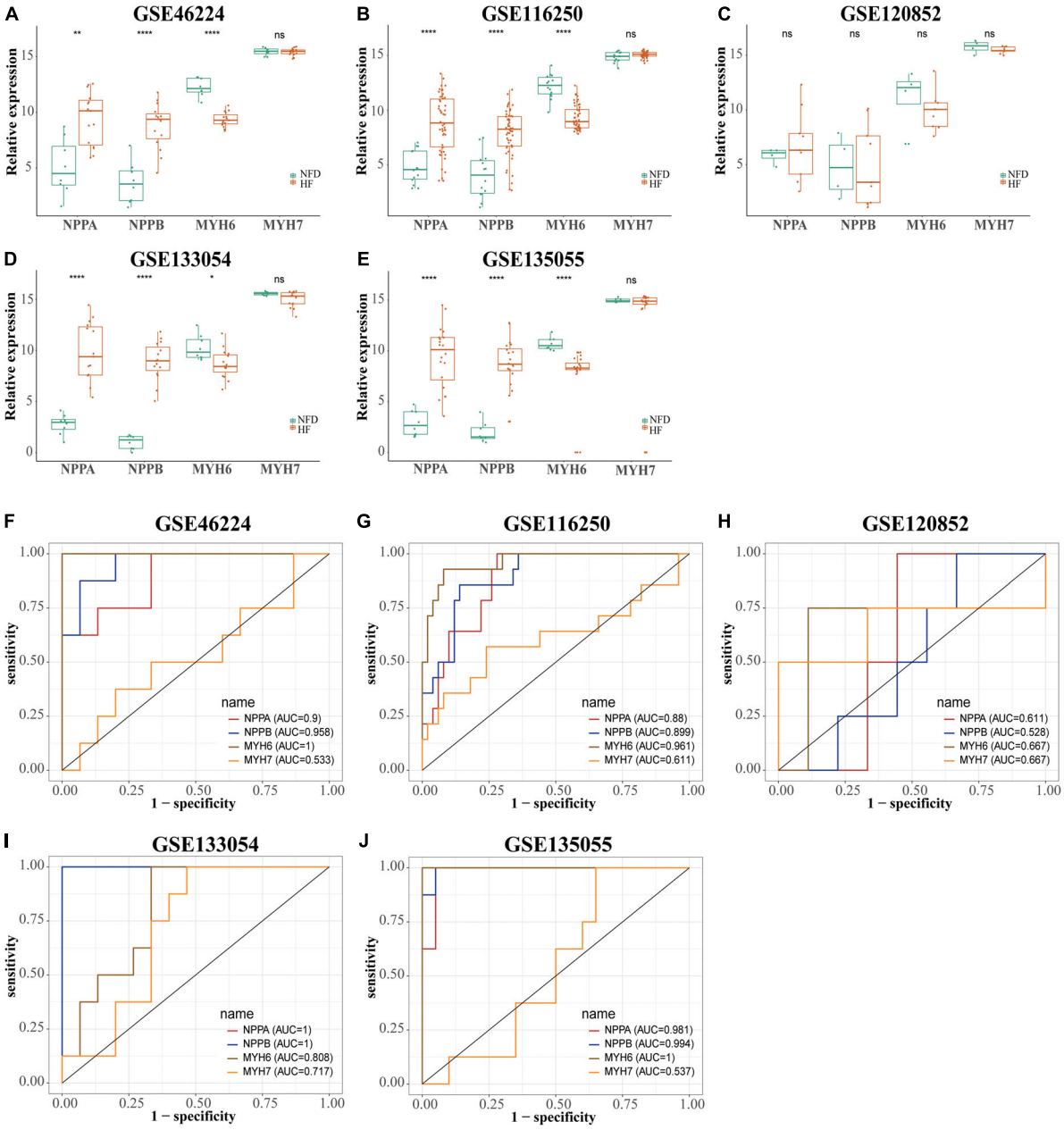


FIGURE 4
Expression level and diagnostic value of *NPPA*, *NPPB*, *MYH6*, and *MYH7* in the five HF-related RNA-seq datasets. The expression level of *NPPA*, *NPPB*, *MYH6*, and *MYH7* in GSE46224 **(A)**, GSE116250 **(B)**, GSE120852 **(C)**, GSE133054 **(D)**, and GSE135055 **(E)**, respectively. ns, not significant vs. the NFD group; *$p < 0.05$ vs. the NFD group; **$p < 0.01$ vs. the NFD group; ****$p < 0.0001$ vs. the NFD group. The diagnostic values of *NPPA*, *NPPB*, *MYH6*, and *MYH7* in GSE46224 **(F)**, GSE116250 **(G)**, GSE120852 **(H)**, GSE133054 **(I)**, and GSE135055 **(J)**, respectively, as determined by ROC curves. *NPPA*, natriuretic peptide A; *NPPB*, natriuretic peptide B; *MYH6*, myosin heavy chain 6; *MYH7*, myosin heavy chain 7; HF, heart failure; NFD, non-failing donor; ROC, receiver operating characteristic.

Identification of signatures of transcription factor–differentially expressed gene regulatory network

To determine the potential roles of TFs in regulating the transcriptional expression of DEGs, the specific TF–gene regulatory network was established based on the top 20 upregulated and the 20 most downregulated integrated DEGs (Figure 11A). As demonstrated in Figures 11B,C, several TFs, including CEBPB, MEF2A, PPARG, BRCA1, TEAD1, TFAP2A, TP63, SREBF1, and PDX1, showed significant correlation with multiple DEGs and were differentially expressed in patients

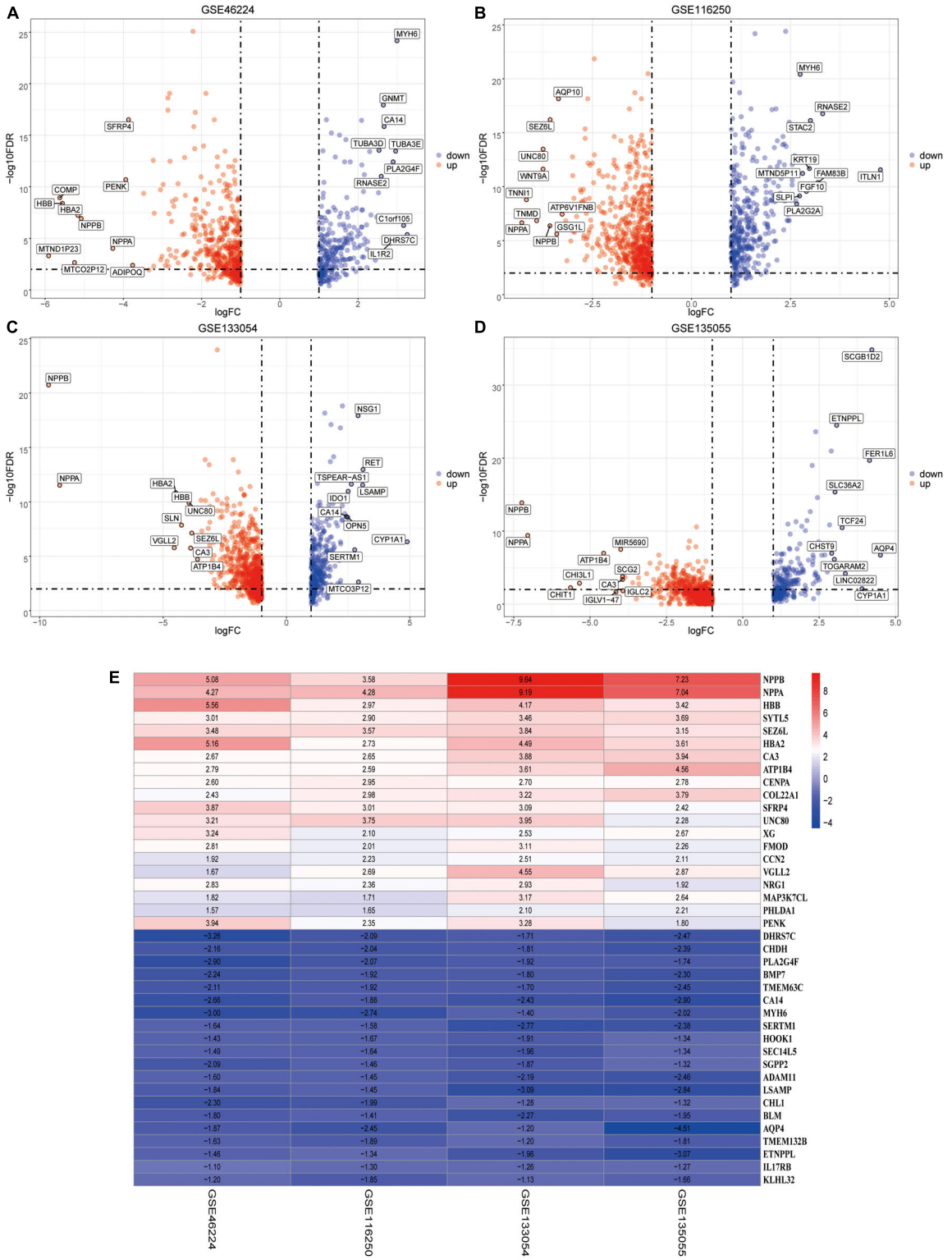


FIGURE 5
Identification of DEGs by RUVSeq and the RRA method. **(A–D)** Using RUVSeq package, DEGs (patients with HF vs. NFDs) were screened in the four selected RNA-seq datasets GSE46224 **(A)**, GSE116250 **(B)**, GSE133054 **(C)**, and GSE135055 **(D)**, as visualized by volcano plots. Adjusted $p < 0.05$ and $|\log FC| > 1$. **(E)** Heatmap of the top 20 upregulated and the 20 most downregulated DEGs screened from the four selected RNA-seq datasets using RRA analysis. DEGs, differentially expressed genes; RRA, robust rank aggregation; HF, heart failure; NFD, non-failing donor; FC, fold change.

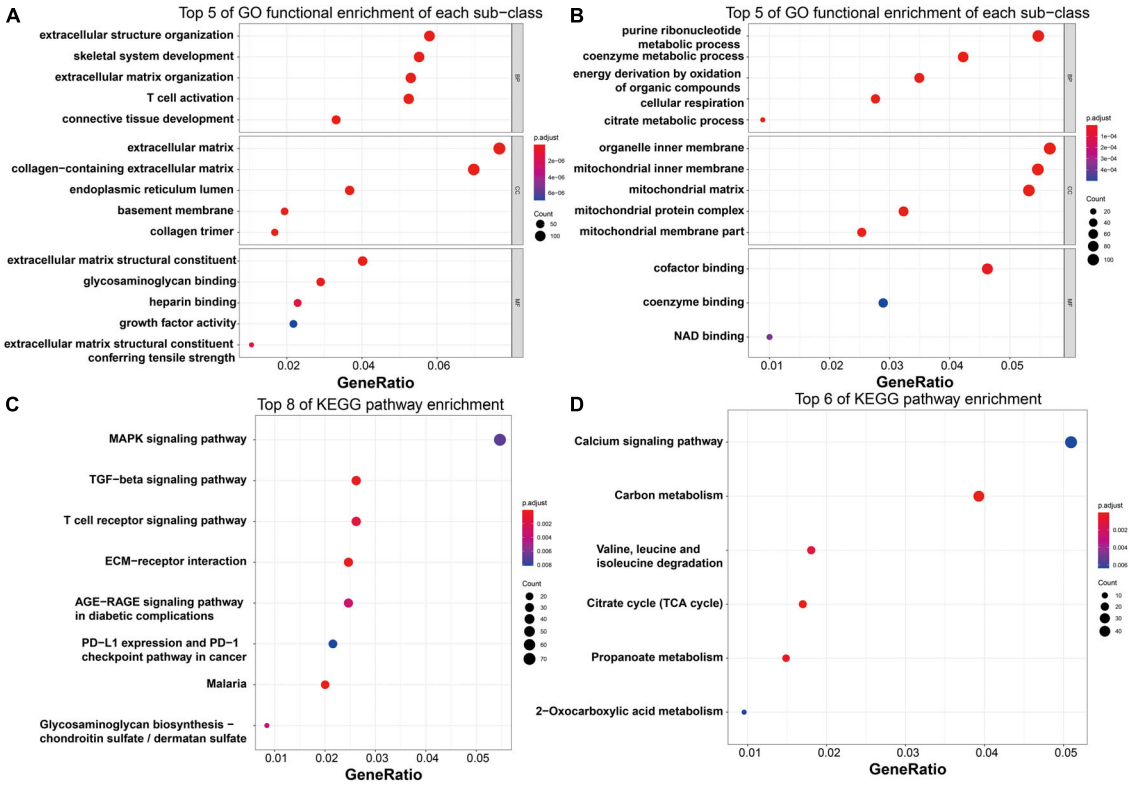


FIGURE 6
Functional enrichment analysis of the robust HF-related DEGs. **(A)** Top five enriched GO functions of the upregulated genes regarding BP, CC, and MF terms, as visualized by bubble plots. **(B)** Top five enriched GO functions of the downregulated genes regarding BP, CC, and MF terms, as visualized by bubble plots. **(C)** Top eight enriched KEGG pathways of the upregulated genes, as visualized by bubble plots. **(D)** Top eight enriched KEGG pathways of the downregulated genes, as visualized by bubble plots. HF, heart failure; DEGs, differentially expressed genes; GO, gene ontology; BP, biological process; CC, cellular component; MF, molecular function; KEGG, Kyoto Encyclopedia of Genes and Genomes.

with HF and NFDs in GSE57338 ($p < 0.05$). According to the defining criteria of the significant TF–DEG regulation pair, we identified TP63-*SERTM1*/*SYTL5*/*UNC80*, PPARG-*XG*, BRCA1-*NRG1*, MEF2A-*LSAMP*, SREBF1-*NPPA*/*HOOK1*/*CENPA*, TEAD1-*CA14*/*MYH6*/*PENK*, and PDX1-*SEC14L5* as significant TF–DEG pairs (**Figure 11D**).

Discussion

In the present study, four HF RNA-seq GEO datasets (GSE46224, GSE116250, GSE133054, and GSE135055) were finally included, involving a total of 100 patients with HF and 38 NFDs. In total, 201 robust HF-related DEGs were obtained utilizing RUVSeq and RRA method, and *ASPN*, *COL1A1*, *COL9A2*, *COL10A1*, *COMP*, and *FMOD* were identified as hub genes with the highest degree of connectivity using STRING database and cytoHubba plug-in. Among them, *ASPN*, *COL1A1*, and *FMOD* exhibited differential expressions and excellent diagnostic efficiency in all four RNA-seq datasets, which were further validated using data from the four screened HF microarray datasets (GSE16499, GSE26887, GSE57338, and GSE79962). Moreover, the significant upregulation of *ASPN*, *COL1A1*, and *FMOD* was experimentally confirmed by qRT-PCR using the heart tissues of patients with HF and NFD samples. In addition, functional enrichment analysis showed that myocardial fibrosis-related pathways resulted from T-cell-mediated immune response and myocardial glucose metabolism were closely associated with the onset and progression of HF. In addition to this, the TF–DEG regulatory network was established, and 13 significant TF–DEG pairs were identified.

Despite the great advancement in HF medical treatment, it remains the major and growing public health problem that leads to considerable morbidity and mortality (35). Robust biomarkers for early diagnosis of HF are the key for novel targeted pharmacological approaches and for improving the prognosis of patients (36). Consistently, serum type B natriuretic peptide (*BNP*) has been recognized as a well-established biomarker for the diagnosis of HF. However, a recent study reported that a subset (4.9%) of hospitalized patients with confirmed HF had unexpectedly low BNP levels (<50 pg/ml), and a small proportion (0.1–1.1%) had BNP levels even below

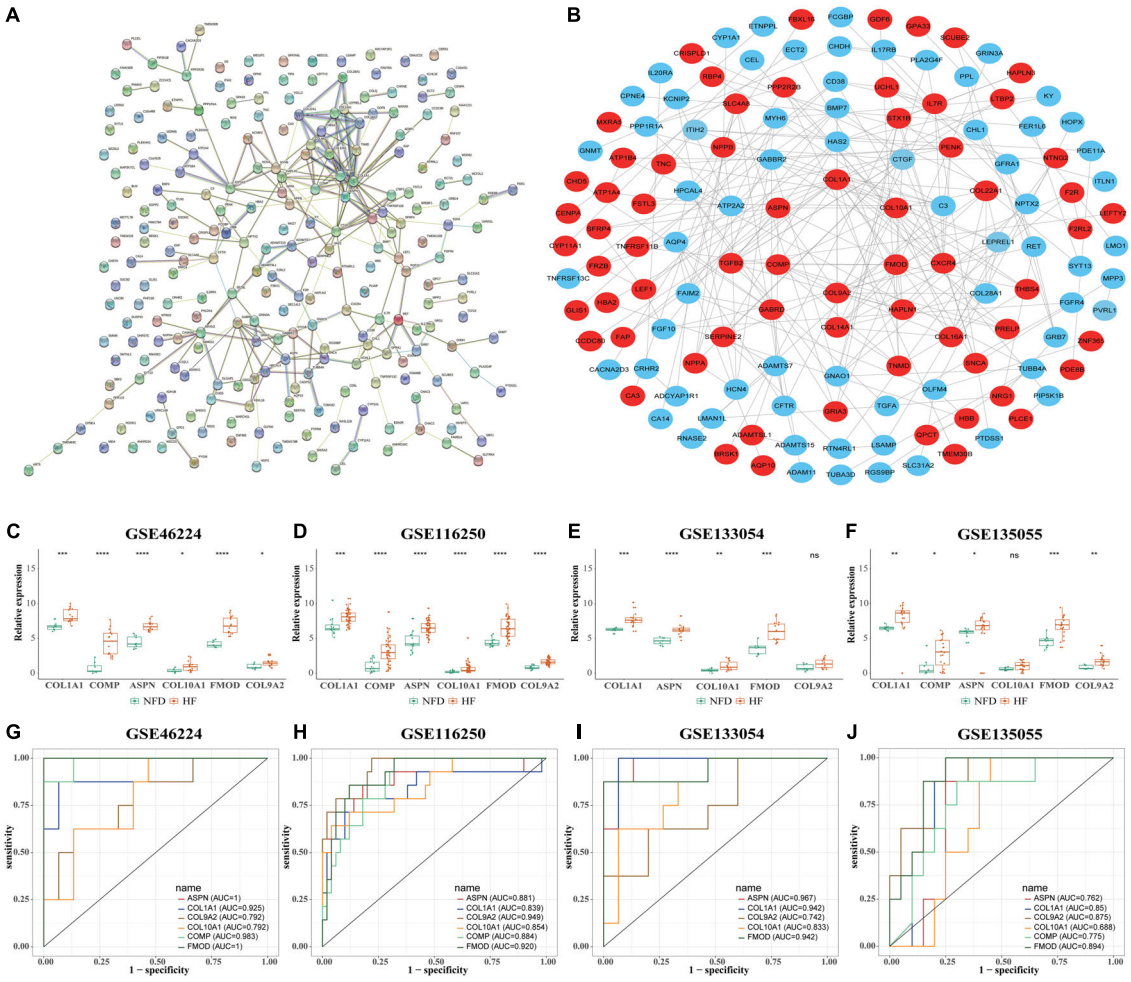


FIGURE 7
Identification of HF-related hub genes from the four selected RNA-seq datasets. **(A)** PPI network of the 201 DEGs in patients with HF was constructed by using the STRING database. The nodes represent proteins, and the edges represent the interactions. **(B)** Robust hub genes were screened by degree using the cytoHubba plug-in in Cytoscape. The inner the concentric circles, the larger the degree values of the genes. **(C–F)** Relative expression of the six identified hub genes in GSE46224 **(C)**, GSE116250 **(D)**, GSE133054 **(E)**, and GSE135055 **(F)**, respectively. ns, not significant vs. the NFD group; *$p < 0.05$ vs. the NFD group; **$p < 0.01$ vs. the NFD group; ***$p < 0.001$ vs. the NFD group; ****$p < 0.0001$ vs. the NFD group. **(G–J)** Diagnostic values of the six identified hub genes in GSE46224 **(G)**, GSE116250 **(H)**, GSE133054 **(I)**, and GSE135055 **(J)**, respectively, as determined by ROC curves. HF, heart failure; RNA-seq, RNA sequencing; PPI, protein–protein interaction; DEGs, differentially expressed genes; STRING, search tool for the retrieval of interacting genes; NFD, non-failing donor; ROC, receiver operating characteristic.

detection limits (37). Therefore, it remains urgent to explore novel molecules with potentially new mechanisms for the development of HF.

Recently, gene mining using microarrays or RNA-seq datasets has been widely used to generate the transcriptomic profiles of HF development. Zhang et al. identified six hub genes (*BCL3*, *HCK*, *PPIF*, *S100A9*, *SERPINA1*, and *TBC1D9B*) as potential biomarkers of HF by using the weighted gene co-expression network analysis (WGCNA) method through three HF datasets, namely, GSE59867, GSE1869, and GSE42955 (38). Tian et al. constructed a random forest algorithm and artificial neural network and detected six hub genes by mining of two HF datasets (GSE57345 and GSE141910) (39). However, the aforementioned studies were based on the integration of DEGs, rather than raw data from different datasets. To date, the inconsistency between different platforms and datasets remains the major hurdle blocking the bioinformatics mining of HF-related genes, especially for RNA-seq datasets.

The RRA method, a recently emerging analysis method, has been widely used to integrate different datasets and produce a ranked list of the DEGs (40). For example, Ma et al. utilized the RRA method to integrate four eligible DCM microarray datasets from the GEO and developed a 7-gene signature predictive model of DCM (11). While in the present study, using RUVSeq to substantially decrease batch effects, we integrated, for the first time, the different RNA-seq datasets of the GEO database to

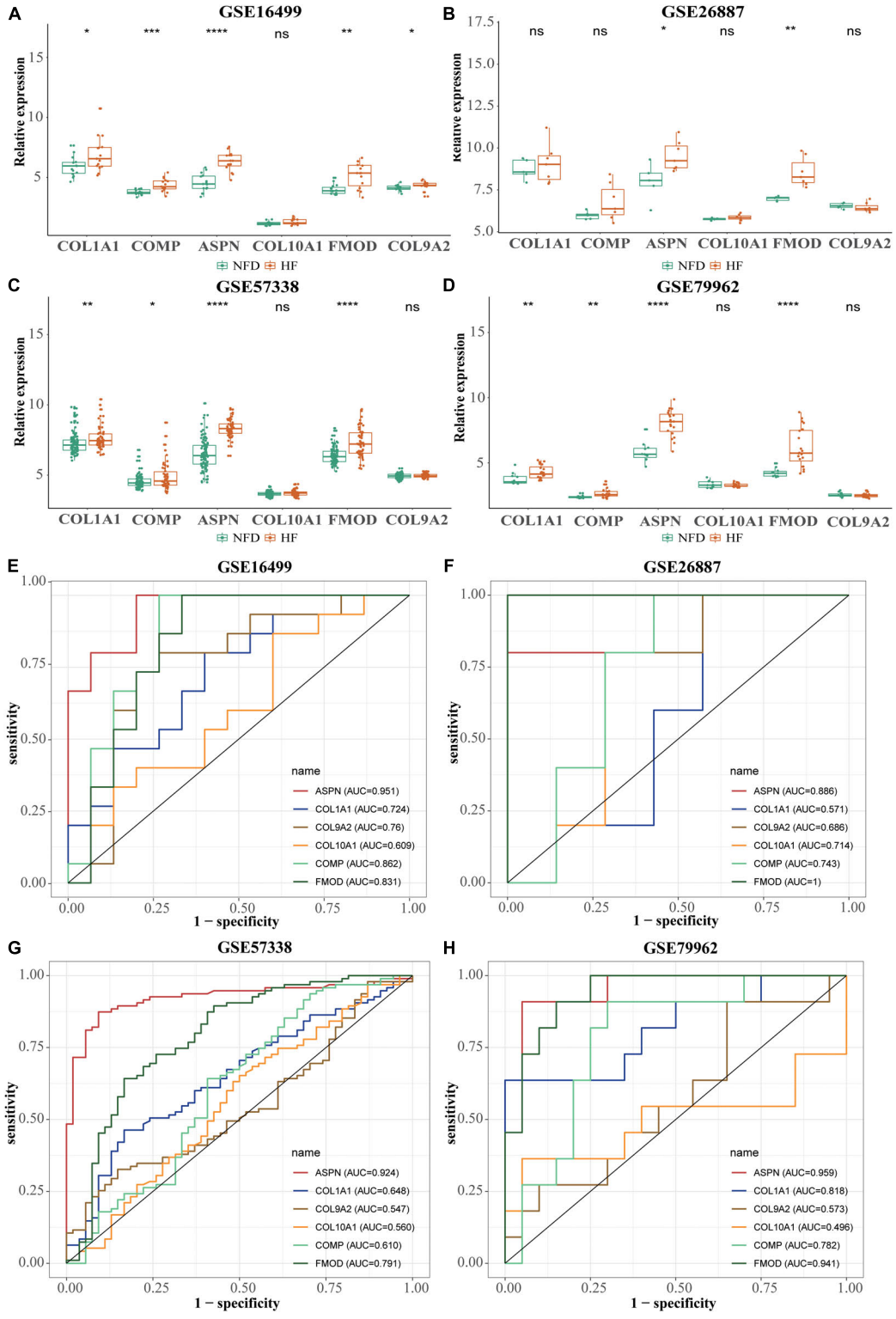


FIGURE 8
Validation of the six hub genes using four normalized HF-related microarray datasets from the GEO database. **(A–D)** Relative expression of the six hub genes in GSE16499 **(A)**, GSE26887 **(B)**, GSE57338 **(C)**, and GSE79962 **(D)**, respectively. ns, not significant vs. the NFD group; *$p < 0.05$ vs. the NFD group; **$p < 0.01$ vs. the NFD group; ***$p < 0.001$ vs. the NFD group; ****$p < 0.0001$ vs. the NFD group. **(E–H)** Diagnostic values of the six identified hub genes in GSE16499 **(E)**, GSE26887 **(F)**, GSE57338 **(G)**, and GSE79962 **(H)**, respectively, as determined by ROC curves. HF, heart failure; GEO, Gene Expression Omnibus; NFD, non-failing donor; ROC, receiver operating characteristic.

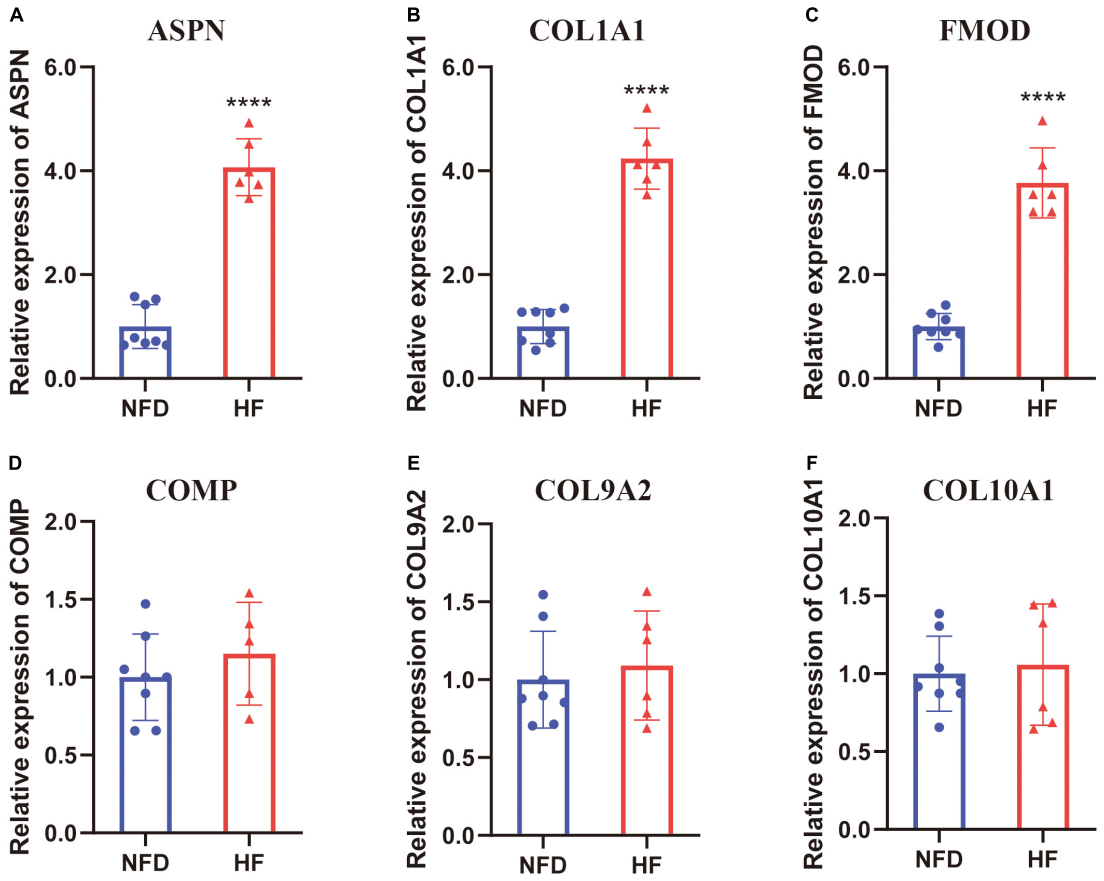


FIGURE 9
Validation of the six hub genes by qRT-PCR using human heart tissues from patients with HF and NFDs. The expression level of *ASPN* **(A)**, *COL1A1* **(B)**, *FMOD* **(C)**, *COMP* **(D)**, *COL9A2* **(E)**, and *COL10A1* **(F)** in heart tissues from HF patients and NFDs, as determined by qRT-PCR. Data are presented with mean ± SD. ****$p < 0.0001$ vs. the NFD group. QRT-PCR, quantitative real-time reverse transcription PCR; HF, heart failure; NFD, non-failing donor; SD, standard deviation.

explore DEGs and hub genes associated with HF by using the RRA method. Through internal RNA-seq dataset and external microarray dataset validation, *ASPN*, *COL1A1*, and *FMOD* were finally identified as real hub genes of HF, which were further confirmed by qRT-PCR using the heart tissues from patients with HF and NFDs.

Interestingly, the identified hub genes *ASPN* (41), *COL1A1* (42), and *FMOD* (43), all belong to the type I collagen members in the extracellular matrix (ECM) composition and have been reported to play important roles in the development and progression of various diseases, especially malignant tumors. For example, *ASPN* was reported to enhance tumor invasion and cancer-associated fibroblasts *via* activation of the CD44-Rac1 pathway in gastric cancer (41). Ma et al. highlighted the role of *COL1A1* as a potential diagnostic biomarker and therapeutic target in early development and metastasis of hepatocellular carcinoma (42). Ao et al. revealed that *FMOD* could promote tumor angiogenesis by upregulating the expression of angiogenic factors in human small-cell lung cancer (43). Regarding the function of hub genes in HF development, a multi-level transcriptomic study conducted by Hua et al. suggested that *COL1A1* might be a plasma biomarker of HF and associated with HF progression, especially to predict the 1-year survival from HF onset to transplantation. A *COL1A1* content ≥ 256.5ng/ml in plasma was found to be associated with poor survival within 1 year of heart transplantation from HF (34). In the study conducted by Andenæs et al., *FMOD* was found 3–10-fold upregulated in hearts of patients with HF and mice, and *FMOD*-KO mice showed a relatively mild hypertrophic phenotype (44). However, to the best of our knowledge, there are no experimental studies focusing on the role of *ASPN* in HF. Therefore, our multi-dataset RRA analysis, followed by microarray dataset and experimental validation, provides more robust and comprehensive evidence for the value of the three ECM-related genes, namely, *COL1A1*, *FMOD*, and *ASPN*, in HF development.

Recent advances have highlighted the crucial role of immune activation in the development and progression of HF. A study

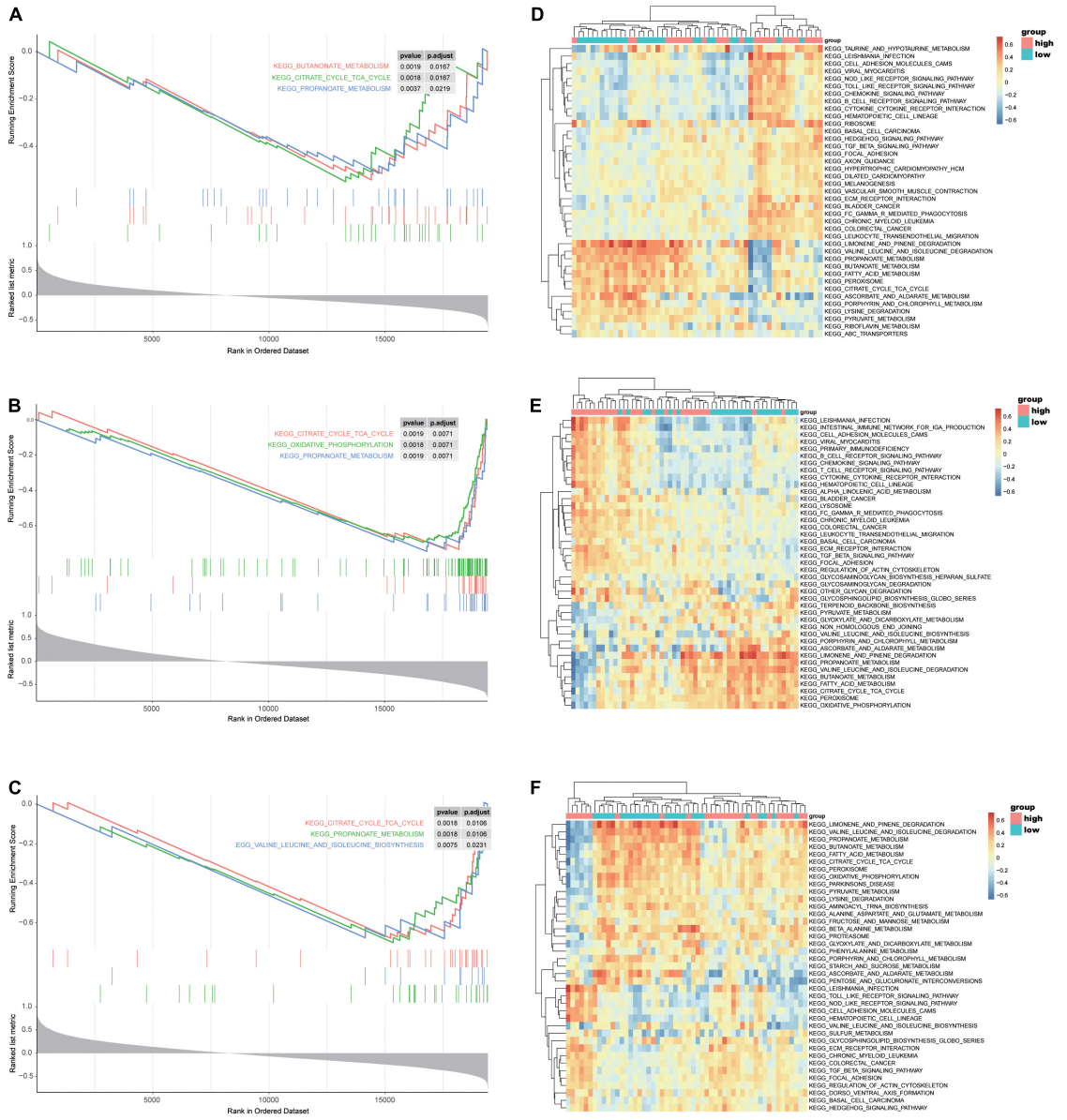


FIGURE 10
GSEA and GSVA of *ASPN*, *COL1A1*, and *FMOD* in the selected microarray dataset GSE57338. **(A–C)** GSEA-enriched pathways of DEGs related to *ASPN* **(A)**, *COL1A1* **(B)**, or *FMOD* **(C)** expression in the GSE57338 dataset. **(D–F)** GSVA-derived clustering heatmaps showing the enriched pathways of DEGs related to *ASPN* **(D)**, *COL1A1* **(E)**, or *FMOD* **(F)** expression in GSE57338 dataset. GSEA, gene set enrichment analysis; GSVA, gene set variation analysis.

by Aghajanian et al. demonstrated that adoptive transfer of T cells that express a chimeric antigen receptor against fibroblast activation protein can inhibit myocardial fibrosis and improve cardiac function in mice (45). Consistently, according to the GO term analysis in our study, the upregulated HF-related DEGs were enriched in T-cell activation of the "BP" term, the extracellular matrix of "CC" terms, and the extracellular matrix structural constituent of "MF" terms. Moreover, regarding the KEGG pathway analysis, the T-cell receptor signaling pathway and ECM–receptor interaction were identified as the significantly enriched pathways of the upregulated DEGs. Considering that all the three hub genes—*ASPN*, *COL1A1*, and *FMOD*—are closely associated with the ECM, we thus speculate a potentially key pathway in the development of HF, that is, T-cell-mediated immune responses lead to the imbalance in ECM anabolism and catabolism, ultimately resulting in myocardial fibrosis and HF.

To further explore the potential mechanism of *ASPN*, *COL1A1*, and *FMOD* in HF, we performed GSEA and GSVA on the validation dataset of GSE57338. Results showed that

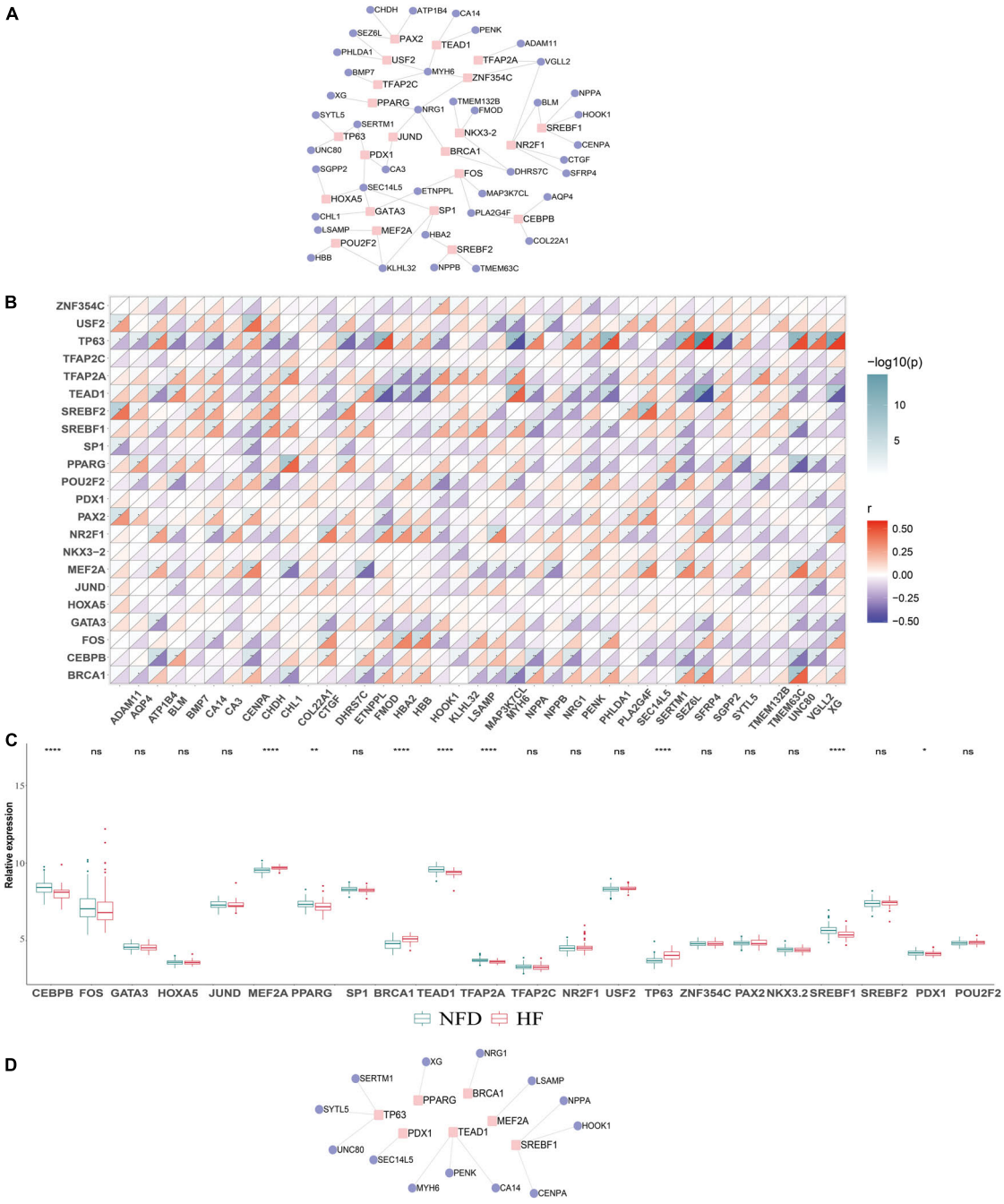


FIGURE 11
Construction of the TF–DEGs regulatory network in HF. **(A)** TF–DEG regulatory network was established based on the top 20 upregulated and the 20 most downregulated DEGs. **(B)** Correlation matrix between the identified TFs and DEGs. An absolute value of correlation coefficient $|r| > 0.3$ and $p < 0.05$ indicates a statistically significant relationship. **(C)** Relative expression of the identified TFs in the validation microarray dataset GSE57338. ns, not significant vs. the NFD group; *$p < 0.05$ vs. the NFD group; **$p < 0.01$ vs. the NFD group; ****$p < 0.0001$ vs. the NFD group. **(D)** Identified significant TF–DEG regulation pairs according to the criteria. TFs, transcription factors; DEGs, differentially expressed genes; HF, heart failure; NFD, non-failing donor.

ASPN-, *COL1A1*-, or *FMOD*-related DEGs were enriched in the "citric acid cycle (TCA cycle)" and "propionic acid metabolism" pathways, both of which are closely associated with glucose metabolism (46, 47). Notably, targeting cardiac glucose metabolism has been recognized as a promising therapeutic strategy for HF treatment. Liu et al. reported that dichloroacetate, a pyruvate dehydrogenase kinase inhibitor, could alter glucose metabolism in cardiomyocytes by

stimulating the activity of pyruvate dehydrogenase complex, thereby improving cardiac efficiency (48). In addition, inhibitors of fatty acid oxidation such as trimetazidine (49), perhexiline (50), and etomoxir (51) can improve cardiac function in patients with HF by increasing glucose oxidation.

Aberrant regulation of TFs is strongly associated with the onset and progression of HF (52). Therefore, in our research, we further investigated the TF–gene interactions to detect the transcriptional regulators of the robust DEGs. Among the seven identified significant TFs, MEF2A (53) and PPARG (54) have been reported to play a role in cardiac remodeling and water retention in HF, respectively. Liu et al. found that suppressing expression of TEAD1, the Hippo signaling effector, could activate the necroptotic pathway and induce massive cardiomyocyte necroptosis, ultimately leading to impaired cardiac function (55). Moreover, loss of BRCA1 in mouse cardiomyocytes resulted in adverse cardiac remodeling and poor ventricular function (56). Although the functions of these TFs in HF have been partially reported, the regulatory relationship of the TF–DEG pairs and the in-depth molecular mechanisms remain to be further validated through HF-related experimental studies.

Our study has several limitations. First, the sample size of patients with HF is relatively small, although we have included as many datasets that met the criteria as possible. Future studies with larger sample sizes are needed to confirm these findings. Second, this study is mainly based on bioinformatics analysis and qRT-PCR validation of hub gene expression. Further experimental research is needed to clarify the in-depth mechanism of the hub gene-related HF regulation. Third, information about disease grades, treatment methods, and prognosis of patients with HF is not available in the database, leading to the failure to analyze correlation between hub genes and clinical characteristics or prognosis of HF. Fourth, the etiology of HF is complex, involving multiple environmental factors in addition to genetic factors (57), such as behavioral factors, socioeconomic and psychosocial factors, air quality, and meteorological factors (58–60). Horton et al. reported that the influence of modifiable lifestyle factors cannot be ignored in the development of direct-to-consumer (DTC) genetic tests (61). In recent years, emerging evidence has shown that gene–environment interactions play an important role in complex disease progression. Bentley et al. revealed that the genetic associations with lipids could be modified by smoking (62). Therefore, future research needs to further focus on the role of environmental factors and gene–environment interactions in HF.

Conclusion

In conclusion, the present study integrated, for the first time, the different RNA-seq datasets of HF from the GEO database and identified robust HF-related DEGs utilizing RUVSeq and the RRA method. Furthermore, three reliable hub genes—*ASPN*, *COL1A1*, and *FMOD*—were screened and validated by bioinformatics and experimental assays. Functional enrichment analysis showed that DEGs and hub genes were associated with T-cell-mediated immune response and the glucose metabolism signaling pathway. In addition, significant TF–DEG regulatory network of HF was further established. However, high-quality basic or clinical research is required to deeply investigate the mechanisms by which these hub genes are involved in HF and to confirm their values as biomarker for HF diagnosis and treatment.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary material**.

Ethics statement

The studies involving human participants were reviewed and approved by the Institute Ethics Committee of Changhai Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

ZG, XS, and XZ contributed to the conception of the study. DT and CM performed the study execution and experiments. ZZ and QX contributed to part of bioinformatic analysis. DT, CM, and XS prepared the manuscript. XZ and ZG contributed to the funding of the study. All authors reviewed the manuscript, provided critical revision, and have approved the final version for publication.

Funding

This work was supported in part by the grants from the Chinese Natural Science Foundation (82170233 to XZ) and 234 Discipline Promotion Foundation of Changhai (2020YXK057 to ZG).

Acknowledgments

We thank Sirui Zhang and Xiaojuan Fan for their technical guidance on the analysis. In addition, we also thank the donors and their families for research organ donations.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcvm.2022.916429/full#supplementary-material

References

1. McMurray JJ, Pfeffer MA. Heart failure. *Lancet.* (2005) 365:1877–89. doi: 10.1016/s0140-6736(05)66621-4

2. Tanai E, Frantz S. Pathophysiology of heart failure. *Compre Physiol.* (2015) 6:187–214. doi: 10.1002/cphy.c140055

3. Snipelisky D, Chaudhry SP, Stewart GC. The many faces of heart failure. *Card Electrophysiol Clin.* (2019) 11:11–20. doi: 10.1016/j.ccep.2018.11.001

4. Mosterd A, Hoes AW. Clinical epidemiology of heart failure. *Heart.* (2007) 93:1137–46. doi: 10.1136/hrt.2003.025270

5. Saeidian AH, Youssefian L, Vahidnezhad H, Uitto J. Research techniques made simple: whole-transcriptome sequencing by rna-seq for diagnosis of monogenic disorders. *J Invest Dermatol.* (2020) 140:1117–26.e1. doi: 10.1016/j.jid.2020.02.032

6. Hanauer DA, Rhodes DR, Sinha-Kumar C, Chinnaiyan AM. Bioinformatics approaches in the study of cancer. *Curr Mol Med.* (2007) 7:133–41. doi: 10.2174/156652407779940431

7. Joshi A, Rienks M, Theofilatos K, Mayr M. Systems biology in cardiovascular disease: a multiomics approach. *Nat Rev Cardiol.* (2021) 18:313–30. doi: 10.1038/s41569-020-00477-1

8. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics.* (2012) 28:573–80. doi: 10.1093/bioinformatics/btr709

9. Liu Y, Chen TY, Yang ZY, Fang W, Wu Q, Zhang C. Identification of hub genes in papillary thyroid carcinoma: robust rank aggregation and weighted gene co-expression network analysis. *J Transl Med.* (2020) 18:170. doi: 10.1186/s12967-020-02327-7

10. Song ZY, Chao F, Zhuo Z, Ma Z, Li W, Chen G. Identification of hub genes in prostate cancer using robust rank aggregation and weighted gene co-expression network analysis. *Aging.* (2019) 11:4736–56. doi: 10.18632/aging.102087

11. Ma X, Mo C, Huang L, Cao P, Shen L, Gui C. An robust rank aggregation and least absolute shrinkage and selection operator analysis of novel gene signatures in dilated cardiomyopathy. *Front Cardiovasc Med.* (2021) 8:747803. doi: 10.3389/fcvm.2021.747803

12. Risso D, Ngai J, Speed TP, Dudoit S. Normalization of rna-seq data using factor analysis of control genes or samples. *Nat Biotechnol.* (2014) 32:896–902. doi: 10.1038/nbt.2931

13. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one fastq preprocessor. *Bioinformatics.* (2018) 34:i884–90. doi: 10.1093/bioinformatics/bty560

14. Ewels P, Magnusson M, Lundin S, Kaller M. Multiqc: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics.* (2016) 32:3047–8. doi: 10.1093/bioinformatics/btw354

15. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. Star: ultrafast universal rna-seq aligner. *Bioinformatics.* (2013) 29:15–21. doi: 10.1093/bioinformatics/bts635

16. Liao Y, Smyth GK, Shi W. Featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* (2014) 30:923–30. doi: 10.1093/bioinformatics/btt656

17. Pang B, Xu J, Hu J, Guo F, Wan L, Cheng M, et al. Single-cell rna-seq reveals the invasive trajectory and molecular cascades underlying glioblastoma progression. *Mol Oncol.* (2019) 13:2588–603. doi: 10.1002/1878-0261.12569

18. Yu G, Wang LG, Han Y, He QY. Clusterprofiler: an r package for comparing biological themes among gene clusters. *OMICS.* (2012) 16:284–7. doi: 10.1089/omi.2011.0118

19. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. String V11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* (2019) 47:D607–13. doi: 10.1093/nar/gky1131

20. Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY. Cytohubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol.* (2014) 8(Suppl 4.):S11. doi: 10.1186/1752-0509-8-s4-s11

21. Kong SW, Hu YW, Ho JW, Ikeda S, Polster S, John R, et al. Heart failure-associated changes in rna splicing of sarcomere genes. *Circ Cardiovasc Genet.* (2010) 3:138–46. doi: 10.1161/circgenetics.109.904698

22. Greco S, Fasanaro P, Castelvecchio S, D'Alessandra Y, Arcelli D, Di Donato M, et al. Microrna dysregulation in diabetic ischemic heart failure patients. *Diabetes.* (2012) 61:1633–41. doi: 10.2337/db11-0952

23. Liu Y, Morley M, Brandimarto J, Hannenhalli S, Hu Y, Ashley EA, et al. Rna-Seq identifies novel myocardial gene expression signatures of heart failure. *Genomics.* (2015) 105:83–9. doi: 10.1016/j.ygeno.2014.12.002

24. Matkovich SJ, Al Khiami B, Efimov IR, Evans S, Vader J, Jain A, et al. Widespread down-regulation of cardiac mitochondrial and sarcomeric genes in patients with sepsis. *Crit Care Med.* (2017) 45:407–14. doi: 10.1097/ccm.0000000000002207

25. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for rna-sequencing and microarray studies. *Nucleic Acids Res.* (2015) 43:e47. doi: 10.1093/nar/gkv007

26. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The human transcription factors. *Cell.* (2018) 172:650–65. doi: 10.1016/j.cell.2018.01.029

27. Zhou G, Soufan O, Ewald J, Hancock REW, Basu N, Xia J. Networkanalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res.* (2019) 47:W234–41. doi: 10.1093/nar/gkz240

28. Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, van der Lee R, et al. Jaspar 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* (2018) 46:D260–6. doi: 10.1093/nar/gkx1126

29. Mukaka MM. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. *Malawi Med J.* (2012) 24:69–71.

30. Yang KC, Yamada KA, Patel AY, Topkara VK, George I, Cheema FH, et al. Deep rna sequencing reveals dynamic regulation of myocardial noncoding rnas in failing human heart and remodeling with mechanical circulatory support. *Circulation.* (2014) 129:1009–21. doi: 10.1161/circulationaha.113.003863

31. Sweet ME, Cocciolo A, Slavov D, Jones KL, Sweet JR, Graw SL, et al. Transcriptome analysis of human heart failure reveals dysregulated cell adhesion in dilated cardiomyopathy and activated immune pathways in ischemic heart failure. *BMC Genom.* (2018) 19:812. doi: 10.1186/s12864-018-5213-9

32. Tzimas C, Rau CD, Buergisser PE, Jean-Louis G Jr, Lee K, Chukwuneke J, et al. Wipi1 is a conserved mediator of right ventricular failure. *JCI Insight.* (2019) 5:e122929. doi: 10.1172/jci.insight.122929

33. Ren Z, Yu P, Li D, Li Z, Liao Y, Wang Y, et al. Single-cell reconstruction of progression trajectory reveals intervention principles in pathological cardiac hypertrophy. *Circulation.* (2020) 141:1704–19. doi: 10.1161/circulationaha.119.043053

34. Hua X, Wang YY, Jia P, Xiong Q, Hu Y, Chang Y, et al. Multi-level transcriptome sequencing identifies col1a1 as a candidate marker in human heart failure progression. *BMC Med.* (2020) 18:2. doi: 10.1186/s12916-019-1469-4

35. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE Jr, Colvin MM, et al. 2017 Acc/Aha/Hfsa focused update of the 2013 Accf/Aha guideline for the management of heart failure: a report of the American college of cardiology/American heart association task force on clinical practice guidelines and the heart failure society of America. *J Card Fail.* (2017) 23:628–51. doi: 10.1016/j.cardfail.2017.04.014

36. Savic-Radojevic A, Pljesa-Ercegovac M, Matic M, Simic D, Radovanovic S, Simic T. Novel biomarkers of heart failure. *Adv Clin Chem.* (2017) 79:93–152. doi: 10.1016/bs.acc.2016.09.002

37. Bachmann KN, Gupta DK, Xu M, Brittain E, Farber-Eger E, Arora P, et al. Unexpectedly low natriuretic peptide levels in patients with heart failure. *JACC Heart Fail.* (2021) 9:192–200. doi: 10.1016/j.jchf.2020.10.008

38. Niu X, Zhang J, Zhang L, Hou Y, Pu S, Chu A, et al. Weighted gene co-expression network analysis identifies critical genes in the development of heart failure after acute myocardial infarction. *Front Genet.* (2019) 10:1214. doi: 10.3389/fgene.2019.01214

39. Tian Y, Yang J, Lan M, Zou T. Construction and analysis of a joint diagnosis model of random forest and artificial neural network for heart failure. *Aging.* (2020) 12:26221–35. doi: 10.18632/aging.202405

40. Song H, Ding N, Li S, Liao J, Xie A, Yu Y, et al. Identification of hub genes associated with hepatocellular carcinoma using robust rank aggregation combined with weighted gene co-expression network analysis. *Front Genet.* (2020) 11:895. doi: 10.3389/fgene.2020.00895

41. Satoyoshi R, Kuriyama S, Aiba N, Yashiro M, Tanaka M. Asporin activates coordinated invasion of scirrhous gastric cancer and cancer-associated fibroblasts. *Oncogene.* (2015) 34:650–60. doi: 10.1038/onc.2013.584

42. Ma HP, Chang HL, Bamodu OA, Yadav VK, Huang TY, Wu ATH, et al. Collagen 1a1 (Col1a1) is a reliable biomarker and putative therapeutic target for hepatocellular carcinogenesis and metastasis. *Cancers.* (2019) 11:786. doi: 10.3390/cancers11060786

43. Ao Z, Yu S, Qian P, Gao W, Guo R, Dong X, et al. Tumor angiogenesis of sclc inhibited by decreased expression of fmod via downregulating angiogenic factors of endothelial cells. *Biomed Pharmacother.* (2017) 87:539–47. doi: 10.1016/j.biopha.2016.12.110

44. Andenæs K, Lunde IG, Mohammadzadeh N, Dahl CP, Aronsen JM, Strand ME, et al. The extracellular matrix proteoglycan fibromodulin is upregulated in clinical and experimental heart failure and affects cardiac remodeling. *PLoS One.* (2018) 13:e0201422. doi: 10.1371/journal.pone.0201422

45. Aghajanian H, Kimura T, Rurik JG, Hancock AS, Leibowitz MS, Li L, et al. Targeting cardiac fibrosis with engineered T cells. *Nature.* (2019) 573:430–3. doi: 10.1038/s41586-019-1546-z

46. Judge A, Dodd MS. Metabolism. *Essays Biochem.* (2020) 64:607–47. doi: 10.1042/ebc20190041

47. Wang Z, Yang ST. Propionic acid production in glycerol/glucose co-fermentation by *Propionibacterium freudenreichii* subsp. shermanii. *Biores Technol.* (2013) 137:116–23. doi: 10.1016/j.biortech.2013.03.012

48. Liu B, Clanachan AS, Schulz R, Lopaschuk GD. Cardiac efficiency is improved after ischemia by altering both the source and fate of protons. *Circ Res.* (1996) 79:940–8. doi: 10.1161/01.res.79.5.940

49. Fragasso G, Palloshi A, Puccetti P, Silipigni C, Rossodivita A, Pala M, et al. A randomized clinical trial of trimetazidine, a partial free fatty acid oxidation inhibitor, in patients with heart failure. *J Am Coll Cardiol.* (2006) 48:992–8. doi: 10.1016/j.jacc.2006.03.060

50. Lee L, Campbell R, Scheuermann-Freestone M, Taylor R, Gunaruwan P, Williams L, et al. Metabolic modulation with perhexiline in chronic heart failure: a randomized, controlled trial of short-term use of a novel treatment. *Circulation.* (2005) 112:3280–8. doi: 10.1161/circulationaha.105.551457

51. Schmidt-Schweda S, Holubarsch C. First clinical trial with etomoxir in patients with chronic congestive heart failure. *Clin Sci.* (2000) 99:27–35.

52. Minerath RA, Hall DD, Grueter CE. Targeting transcriptional machinery to inhibit enhancer-driven gene expression in heart failure. *Heart Fail Rev.* (2019) 24:725–41. doi: 10.1007/s10741-019-09792-3

53. Tobin SW, Hashemi S, Dadson K, Turdi S, Ebrahimian K, Zhao J, et al. Heart failure and mef2 transcriptome dynamics in response to β-blockers. *Sci Rep.* (2017) 7:4476. doi: 10.1038/s41598-017-04762-x

54. Bao LZ, Shen M, Qudirat H, Shi JB, Su T, Song JW, et al. Obestatin ameliorates water retention in chronic heart failure by downregulating renal aquaporin 2 through Gpr39, V2r and Pparg signaling. *Life Sci.* (2019) 231:116493. doi: 10.1016/j.lfs.2019.05.049

55. Liu J, Wen T, Dong K, He X, Zhou H, Shen J, et al. Tead1 protects against necroptosis in postmitotic cardiomyocytes through regulation of nuclear DNA-encoded mitochondrial genes. *Cell Death Differ.* (2021) 28:2045–59. doi: 10.1038/s41418-020-00732-5

56. Shukla PC, Singh KK, Quan A, Al-Omran M, Teoh H, Lovren F, et al. Brca1 is an essential regulator of heart function and survival following myocardial infarction. *Nat Commun.* (2011) 2:593. doi: 10.1038/ncomms1601

57. Wald NJ, Old R. The illusion of polygenic disease risk prediction. *Genet Med.* (2019) 21:1705–7. doi: 10.1038/s41436-018-0418-5

58. Yang Q, Zhang Z, Gregg EW, Flanders WD, Merritt R, Hu FB. Added sugar intake and cardiovascular diseases mortality among us adults. *JAMA Internal Med.* (2014) 174:516–24. doi: 10.1001/jamainternmed.2013.13563

59. Zeng J, Zhang X, Yang J, Bao J, Xiang H, Dear K, et al. Humidity may modify the relationship between temperature and cardiovascular mortality in Zhejiang province, China. *Int J Environ Res Public Health.* (2017) 14:1383. doi: 10.3390/ijerph14111383

60. Yusuf S, Joseph P, Rangarajan S, Islam S, Mente A, Hystad P, et al. Modifiable risk factors, cardiovascular disease, and mortality in 155 722 individuals from 21 high-income, middle-income, and low-income countries (pure): a prospective cohort study. *Lancet.* (2020) 395:795–808. doi: 10.1016/s0140-6736(19)32008-2

61. Horton R, Crawford G, Freeman L, Fenwick A, Wright CF, Lucassen A. Direct-to-consumer genetic testing. *BMJ.* (2019) 367:l5688. doi: 10.1136/bmj.l5688

62. Bentley AR, Sung YJ, Brown MR, Winkler TW, Kraja AT, Ntalla I, et al. Multi-ancestry genome-wide gene-smoking interaction study of 387,272 individuals identifies new loci associated with serum lipids. *Nat Genet.* (2019) 51:636–48. doi: 10.1038/s41588-019-0378-y

frontiers | Frontiers in Cardiovascular Medicine

TYPE Original Research
PUBLISHED 07 November 2022
DOI 10.3389/fcvm.2022.1017673

Check for updates

OPEN ACCESS

EDITED BY
Leonardo Roever,
Federal University of Uberlândia, Brazil

REVIEWED BY
Roberto Latini,
Mario Negri Pharmacological Research Institute (IRCCS), Italy
Bart De Geest,
KU Leuven, Belgium

*CORRESPONDENCE
Imen Gtif
imengtif@gmail.com

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Heart Failure and Transplantation, a section of the journal Frontiers in Cardiovascular Medicine

RECEIVED 12 August 2022
ACCEPTED 17 October 2022
PUBLISHED 07 November 2022

CITATION
Gtif I, Abdelhedi R, Ouarda W, Bouzid F, Charfeddine S, Zouari F, Abid L, Rebai A and Kharrat N (2022) Oxidative stress markers-driven prognostic model to predict post-discharge mortality in heart failure with reduced ejection fraction. *Front. Cardiovasc. Med.* 9:1017673. doi: 10.3389/fcvm.2022.1017673

COPYRIGHT
© 2022 Gtif, Abdelhedi, Ouarda, Bouzid, Charfeddine, Zouari, Abid, Rebai and Kharrat. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Oxidative stress markers-driven prognostic model to predict post-discharge mortality in heart failure with reduced ejection fraction

Imen Gtif[1]*, Rania Abdelhedi[1†], Wael Ouarda[2†], Fériel Bouzid[1], Salma Charfeddine[3], Fatma Zouari[3], Leila Abid[3], Ahmed Rebai[1] and Najla Kharrat[1]

[1]Laboratory of Screening Cellular and Molecular Process, Centre of Biotechnology of Sfax, University of Sfax, Sfax, Tunisia, [2]Digital Research Center of Sfax, University of Sfax, Sfax, Tunisia, [3]Unit of Cardiology in Hospital of Hedi Chaker, Faculty of Medicine, University of Sfax, Sfax, Tunisia

Background: Current predictive models based on biomarkers reflective of different pathways of heart failure with reduced ejection fraction (HFrEF) pathogenesis constitute a useful tool for predicting death risk among HFrEF patients. The purpose of the study was to develop a new predictive model for post-discharge mortality risk among HFrEF patients, based on a combination of clinical patients' characteristics, N-terminal pro-B-type Natriuretic peptide (NT-proBNP) and oxidative stress markers as a potentially valuable tool for routine clinical practice.

Methods: 116 patients with stable HFrEF were recruited in a prospective single-center study. Plasma levels of NT-proBNP and oxidative stress markers [superoxide dismutase (SOD), glutathione peroxidase (GPX), uric acid (UA), total bilirubin (TB), gamma-glutamyl transferase (GGT) and total antioxidant capacity (TAC)] were measured in the stable predischarge condition. Generalized linear model (GLM), random forest and extreme gradient boosting models were developed to predict post-discharge mortality risk using clinical and laboratory data. Through comprehensive evaluation, the most performant model was selected.

Results: During a median follow-up of 525 days (7–930), 33 (28%) patients died. Among the three created models, the GLM presented the best performance for post-discharge death prediction in HFrEF. The predictors included in the GLM model were age, female sex, beta blockers, NT-proBNP, left ventricular ejection fraction (LVEF), TAC levels, admission systolic blood pressure (SBP), angiotensin-converting enzyme inhibitors /angiotensin receptor II blockers (ACEI/ARBs) and UA levels. Our model had a good discriminatory power for post-discharge mortality [The area under the curve (AUC) = 74.5%]. Based on the retained model, an online calculator was

developed to allow the identification of patients with heightened post-discharge death risk.

Conclusion: In conclusion, we created a new and simple tool that may allow the identification of patients at heightened post-discharge mortality risk and could assist the treatment decision-making.

KEYWORDS

heart failure, oxidative stress, mortality, prediction, models

Introduction

Heart failure (HF) is a complex clinical syndrome resulting from any functional or structural heart disorder, leading to a reduction of cardiac output or an increase in intracardiac pressures (1). HF is a major clinical and public health concern that affects around 63.4 million people worldwide, accounting for an economic burden of 346.17 billion US $ worldwide in 2017 (2–4). For African countries, HF is still health challenging and was associated with significant rates of hospitalizations and mortality (5–8).

Based on the measurement of the left ventricular ejection fraction (LVEF), HF with reduced ejection fraction (HFrEF) and HF with preserved ejection fraction (HFpEF) are the two major HF subtypes (9). HFrEF is a progressive and multifactorial disease, mainly associated with left ventricular systolic dysfunction and adverse cardiac remodeling (10). It develops as the final and serious stage of various cardiac diseases, including coronary artery disease, myocarditis, valve disease, arterial hypertension and arrhythmias (11). Although the significant advance in HF management, HFrEF remains a serious public health problem with high morbidity, hospitalizations and mortality rates (12–14). A study combining the Cardiovascular Health Study and Framingham Heart Study cohorts reported that 67% of HFrEF patients died within 5 years after diagnosis (15). Thus, prediction of mortality risk for HFrEF patients becomes essential to guide therapy decision-making. Indeed, several demographic characteristics, comorbidities, clinical variables and HF medications have been identified as relevant predictors of mortality among patients with HFrEF (16–18). The Natriuretic peptides, including B-type Natriuretic peptide (BNP) and N-terminal pro-B-type Natriuretic peptide (NT-proBNP), are the gold standard biomarkers used in diagnosis, risk stratification and prediction of future cardiac events in HFrEF patients (19–21). Furthermore, the measurement of specific biomarkers, associated with the different pathways of HFrEF pathogenesis has emerged as the most appropriate approach to facilitate the prediction of mortality risk in patients with HFrEF (19, 22–26).

An ever-growing body of evidence supports that increased oxidative stress, resulting from an imbalance between the production of reactive oxygen species (ROS) and antioxidant defense mechanisms, is involved in the pathogenesis of HFrEF (27, 28). Indeed, increased production of ROS causes cellular dysfunction, protein oxidation, lipid peroxidation, and nucleic acid damage. These alterations contribute to myocyte apoptosis, cardiomyocyte hypertrophy, collagen deposition and matrix remodeling eventually leading to progressive left ventricular remodeling and dysfunction driving HFrEF (29). The components of the antioxidant defense systems, responsible for the inactivation of ROS, consist of antioxidant enzymes such as superoxide dismutase (SOD), catalase, glutathione peroxidase (GPx), peroxiredoxins; non-enzymatic antioxidants, including glutathione (GSH), vitamins, uric acid (UA), total bilirubin (TB) and albumin (30). The assessment of markers relevant to antioxidant defense systems had indicated an association with the progression and severity of HFrEF (31–33). Furthermore, there is growing evidence that antioxidant parameters may provide valuable new insight into the prognosis of HFrEF. Indeed, a large number of studies have been conducted to prove the potential role of UA as a prognostic marker in HFrEF (34–36). The gamma-glutamyl transférase (GGT), the first enzyme of the gamma glutamyl cycle that regulates the antioxidant GSH, has emerged as a promising biomarker for predicting mortality among patients with HFrEF (37, 38). In 2019, Romuk et al. demonstrated that SOD activity was associated with long-term outcomes in HFrEF (39). Other studies showed that total antioxidant capacity (TAC) and bilirubin levels were associated with an increased risk of death in patients with HFrEF (36, 40).

Based on these findings, we hypothesized that oxidative stress markers in combination with NT-proBNP and relevant clinical factors may provide a good predictive potential for mortality risk in HFrEF. Accordingly, the present study aimed to develop a new predictive model for post-discharge mortality risk among HFrEF patients, based on a combination of clinical patient characteristics, NT-proBNP and oxidative stress markers, as a potentially valuable tool for routine clinical practice.

Materials and methods

Patients and study design

This study is a prospective single-center study. A total of 116 consecutive patients, admitted for newly diagnosed or exacerbated HFrEF to the Cardiology Department of CHU Hedi Chaker from November 2017 to December 2019, were recruited. This study was approved by the local ethics committee of CHU Hédi Chaker of Sfax (Tunisia), in accordance with the principles expressed in the Declaration of Helsinki (CPP Sud 0276/2017). Written informed consent was obtained from all enrolled patients.

The diagnosis of HFrEF was based on the Framingham criteria and the presence of left ventricular systolic reduction (LVEF $<$ 50%) (41, 42). Only patients discharged alive were evaluated in the present study. The exclusion criteria were: Age $<$ 20 years, HFpEF (LVEF $\geq$ 50%), acute myocardial infarction, a severe valvular disease requiring surgery, renal failure requiring dialysis, presence of inflammatory disease, autoimmune diseases and malignant diseases.

Data collection

Patients' demographic and clinical characteristics, comorbidities and treatments are known to influence the prognosis of HFrEF were documented from medical records and through patient interviews. For each patient, the following characteristics were collected: age sex, comorbidities [hypertension, diabetes mellitus, hyperlipidemia, chronic kidney disease (CKD), chronic obstructive pulmonary disease (COPD) and anemia], HF characteristics [previous history of HF, New York Heart Association (NYHA) class and ischemic etiology], clinical measures [body mass index (BMI), systolic blood pressure (SBP), LVEF, electrocardiogram indicators (atrial fibrillation (AF), left Bundle Branch Block (LBBB) and QRS duration) and creatinine clearance (CC)] and discharge medications [beta blockers, angiotensin-converting enzyme inhibitors /angiotensin receptor II blockers (ACEI/ARBs), loop diuretics, aldosterone antagonist and statins]. The etiology of HFrEF was classified as ischemic or non-ischemic, based on a history of myocardial infarction and/or coronary angiography. LVEF was determined by two-dimensional echocardiography, using the biplane Simpson's method (43). CC was estimated using the Cockcroft-Gault Equation (44). The prognostic outcome of the present study was post-discharge all-cause mortality. Information regarding outcomes was obtained through hospital records and telephone contact with patients or their close family members. The follow-up time was calculated from discharge to all-cause mortality (time to death) or termination of the study.

Biochemical measurements

Fasting blood samples were collected under stable conditions before discharge. Samples were centrifuged upon permanent cooling at 3,500 rpm for 5 min. Obtained plasma was stored immediately at −20°C temperature until assay. UA, TB, and GGT were measured using the Hitachi 912 analyzer (Roche).

SOD activity was measured by the method of Beyer and Fridovich (45), based on the ability of SOD to inhibit the oxidation of nitro blue tetrazolium (NBT) in the presence of oxygen. The reduction of NBT was measured by a spectrophotometer at 560 nm. SOD activity was calculated by determining the percentage inhibition per min under standard conditions. A 50% of inhibition corresponds to one unit of SOD activity.

GPx activity was determined according to the method of Flohé and Günzler (46), based on glutathione oxidation by GPx in the presence of Ellman's Reagent (DTNB). The absorbance was measured at 412 nm. GPx activity was expressed as nmoles of disappeared GSH/min/mg of proteins.

TAC was measured by colorimetric method using the Colorimetric Assay Kit (Catalog #K274-100; BioVisionIncorporated; CA 95,035 USA). According to the manufacturer's instructions, the antioxidant equivalent concentrations were measured at 570 nm as a function of Trolox concentration. TAC was expressed as mM Trolox equivalent.

NT-proBNP levels were assessed by the Human NT Pro-BNP DuoSet ELISA kit (DY3604-05, R&D, Minneapolis, MN, USA). According to the manufacturer's protocol, the double-antibody sandwich method was applied in this assay. The measurement range of the NT-proBNP assay was 312–10000 pg/ml.

Statistical analysis

For descriptive statistics, the Shapiro-Wilk test was used to assess the normality of continuous variables. Continuous variables were presented as mean and standard deviations (SD) or median and interquartile range [(IQR): Q3–Q1] according to their distribution. Categorical variables were expressed as numbers and percentages. To examine the differences in biomarker levels and clinical characteristics between survivors and non-survivors, *T*-tests were used for parametric variables, U Mann–Whitney tests for non-parametric variables and Chi-square tests for categorical variables. The level of statistical significance was set at a two-tailed p-values < 0.05.

The association between oxidative stress markers and post-discharge mortality risk was evaluated by Kaplan-Meier (KM) survival analysis, log-rank test and Cox proportional hazards regression. Receiver operating characteristic (ROC) curves were used to determine the relevant cut-off of biomarkers statistically

associated with post-discharge mortality for the identification of low-risk and high-risk subjects. KM survival curves were then generated to illustrate survival of patients, according to cut-off values of these biomarkers and Log rank tests were used to compare between the curves. Univariate Cox proportional hazards regression analysis was performed to determine the predictive value of each biomarker and each baseline patient characteristic. Variables with statistical significance in the univariate Cox analysis (p-values < 0.05) were then adjusted for age, sex and BMI in a multivariable model. Multivariate

TABLE 1 Baseline characteristics of study patients stratified according to prognosis outcome.

		Post-discharge mortality during follow-up		
	Total (116)	**Yes (33)**	**No (83)**	***P*-value**
Demographics				
Age, years	62.5 ± 11.6	65.7 ± 12.3	61.3 ± 11.1	0.067
Sex *n* (%)				
Male	81 (72.4)	20 (60.6)	64 (77.1)	0.061
Female	32 (27.6)	13 (39.4)	19 (22.9)	
Comorbidities				
Hypertension (Yes) *n* (%)	44 (37.9)	13 (39.4)	31 (37.3)	0.5
Diabetes mellitus (Yes) *n* (%)	33 (28.4)	10 (30.3)	23 (27.3)	0.474
Hyperlipidemia (Yes) *n* (%)	30 (25.9)	10 (30.3)	20 (24.3)	0.321
CKD (Yes) *n* (%)	22 (18.9)	10 (30.3)	12 (14.5)	**0.047**
COPD (Yes) *n* (%)	8 (6.9)	4 (12.12)	4 (4.8)	0.159
Anemia (Yes) *n* (%)	53 (45.7)	18 (54.5)	35 (42.16)	0.158
Heart failure characteristics				
Previous history of HF *n* (%)	55 (47.4)	21 (63.6)	34 (40.9)	**0.023**
NYHA class III/IV *n* (%)	83 (71.5)	29 (87.9)	54 (65.1)	**0.010**
LVEF (%)	30 (15–48)	25 (15–45)	30 (15–48)	**0.034**
Ischemic etiology *n* (%)	47 (40.5)	12 (36.4)	35 (42.2)	0.360
Clinical measures				
BMI (Kg/m^2)	25.5 (17.4–36.3)	24.2 (18.8–33.9)	25.9 (17.3–36.3)	0.664
Admission SBP (mm Hg)	120 (77–180)	110 (77–170)	120 (88–180)	**0.003**
QRS duration (ms)	118 (74–196)	108 (80–196)	100 (74–196)	**0.044**
AF *n* (%)	46 (39.6)	13 (39.4)	33 (39.7)	0.571
LBBB *n* (%)	49 (42.2)	18 (54.5)	31 (37.3)	0.069
CC (ml/min)	73.5 (33–174)	60 (33–141)	78 (33–174)	**0.025**
Discharge medications				
Beta blockers (yes) *n* (%)	90 (77.6)	21 (63.6)	69 (83.1)	**0.024**
ACEI/ARBs (yes) *n* (%)	62 (53.4)	11 (33.3)	51 (61.4)	**0.006**
Loop diuretics (yes) *n* (%)	100 (86.2)	31 (93.9)	69 (83.1)	0.106
Aldosterone antagonist (yes) *n* (%)	56 (48.3)	14 (42.4)	42 (50.6)	0.278
Statins (yes) *n* (%)	58 (50)	17 (51.5)	41 (49.4)	0.5
Biochemical variables				
UA (μmol/l)	429.5 (71–1000)	530 (71–970)	401 (224–1000)	**0.005**
TB (g/l)	15 (4–76)	17 (5–73)	15 (4–76)	0.210
GGT (UI/l)	37 (8.2–197)	36 (14–127)	37 (8.2–197)	0.788
SOD (UI/l)	117.4 (74–174)	120.2 (82.9–174.8)	109.7 (74–174.1)	0.650
GPx (nmol/min/mg protein)	2.6 (1–6.32)	2.5 (1.3–6.3)	2.6 (1–5.7)	0.753
TAC (mM Trolox equivalent)	10.9 ± 1.7	11.4 ± 1.5	10.6 ± 1.7	**0.023**
NT-proBNP (pg/ml)	3550 (354–7,000)	4393.33 (1,140–7,000)	3380 (354–6733.33)	**0.001**

ACEI/ARBs, angiotensin-converting enzyme inhibitors/angiotensin receptor II; AF, atrial fibrillation; BMI, body mass index; CC, creatinine clearance; COPD, chronic obstructive pulmonary disease; CKD, chronic kidney disease; GPx, glutathione peroxidase; HF, heart failure; LBBB, left Bundle Branch Block; LVEF, left ventricular ejection fraction; NT-proBNP, N-terminal pro-B-type Natriuretic peptide; NYHA, New York Heart Association; SBP, systolic blood pressure; SOD, superoxide dismutase; UA, uric acid; TAC, total antioxidant capacity; TB, total bilirubin. Bold values indicate the p-values < 0.05.

analyses were eventually conducted using the backward stepwise selection process. Variables with p-values ≤ 0.1 were selected in the multivariate proportional hazards regression analysis. Results are presented as hazard ratios (HR) with a 95% confidence interval (CI).

Development of post-discharge mortality risk prediction models

We developed three predictive models, including the generalized linear model (GLM), random forest (RF, based on bootstrap model aggregation of classification trees) model and extreme gradient boosting (XGBoost) model. Thirty baseline variables were put into the prognostic models, including demographics, comorbidities, clinical factors, medications and biochemical variables (**Table 1**). A bi-directional stepwise procedure (backward and forward), method that minimize the akaike information criterion (AIC), was used for GLM variable selection with a significance level at 0.10 as criteria to retain significant variables in the model. The AIC was used to avoid model overfitting. Results are reported as odds ratios (OR) with 95% CI. Shapley additive explanation (SHAP) values were used to evaluate the variables' importance in the RF and XGBoost models. Leave-One-Out Cross-Validation (LOOCV), a special case of k-fold cross validation with k equal to n (the number of observations in the data set) (47), was applied in order to evaluate models' performance. The evaluation metrics used in this study were area under the receiver operating characteristic curve (AUC), accuracy, recall, precision, F1-score and Matthews correlation coefficient (MCC). The best performing model was then selected to predict post-discharge mortality risk in this study. Subsequently, we calculated the probability of death using the predictors of the selected model. Finally, patients were classified into high and low risk groups according to this probability and the KM curve was performed for survival analysis. Statistical analysis was performed using SPSS (Statistical Package for the Social Sciences) version 23 and R statistical software version 3.3.3 (R Project for Statistical Computing). Machine learning algorithms were performed using Python version 3.9 (Python Software Foundation) (**Supplementary material**).

Results

Study population characteristics

Clinical patients' characteristics

A total of 116 patients with HFrEF were followed for 525 days (7–930). Baseline patients' characteristics and the difference between died patients and those surviving during the follow-up period are summarized in **Table 1**. Overall, the study patients present a mean age of 62.5 ± 11.6 years and were predominantly male (72%). Indeed, anemia and hypertension were the most prevalent comorbidities among study patients. A total of 55 (47.4%) patients had a previous history of HF and the majorities (71.5%) were in NYHA class III/IV. More than 50% of patients had a non-ischemic etiology for HF. The median LVEF was 30% (15–48). Loop diuretics (86.2%), beta blockers (77.6%) and angiotensin-converting enzyme inhibitors /angiotensin receptor II blockers (ACEI/ARBs) (53.4%) were the most common medications prescribed to patients at hospital discharge (**Table 1**).

During follow-up period, 33 (28%) patients died. The most frequent cause of death was HF in 60% of cases. Non-survivors were more likely to have CKD (30.3 vs. 14.5%, $p = 0.047$), previous history of HF (63.6 vs. 40.6%, $p = 0.023$) and NYHA class III/IV symptoms (78.9 vs. 65.1%, $p = 0.010$). They also had a lower SBP [110 mm Hg (77–170) vs. 120 mm Hg (88–180), $p = 0.003$], lower LVEF [25% (15–45) vs. 30% (15–48), $p = 0.034$] and lower CC rate [60 ml/min (33–141) vs. 78 ml/min (33–174), $p = 0.025$] compared with survivors. The death group had a higher QRS duration [108 ms (80–196) vs. 100 ms (74–196), $p = 0.044$]. The dead patients were less likely to be treated with ACEI/ARBs (33.3 vs. 61.4%, $p = 0.006$) and beta blockers (63.6 vs. 83.1%, $p = 0.024$).

Biochemical parameters

Among study patients, the median plasma concentrations of UA and TB were 429.5 (71–1000) μmol/l and 15 (4–76) g/l, respectively. Median plasma GGT, GPx and SOD activities were 37 (8.2–197) UI/l, 2.6 (1–6.32) to nmol/min/mg protein and 117.4 (74–174) UI/l, respectively. The mean plasma TAC levels were 10.9 ± 1.7 mM Trolox equivalents. The median plasma NT-proBNP levels were 3550 (354–7000) pg/ml. When analyzing oxidative stress marker levels, non-survivors had higher values of UA [530 μmol/l (71–970) vs. 401 μmol/l (224–1000), $p = 0.005$] and TAC (11.4 ± 1.5 mM trolox equivalent vs. 10.6 ± 1.7 mM trolox equivalent, $p = 0.023$). However, GPx, SOD, and GGT activities and TB concentration were not statistically different between the two groups. The death group had also elevated levels of NT-proBNP [4393.33 pg/ml (1140–7000) vs. 3380 pg/ml (354–6733.33), $p = 0.001$].

Association between oxidative stress markers and the risk of all-cause mortality

Kaplan-Meier survival analysis

We performed the KM analysis to estimate survival probabilities for all-cause mortality, according to cut-off values of UA, TAC, and NT-proBNP. The ROC curve analysis showed that the best cut-off value for UA, TAC, and NT-proBNP to predict all-cause mortality risk were 460 μmol/l

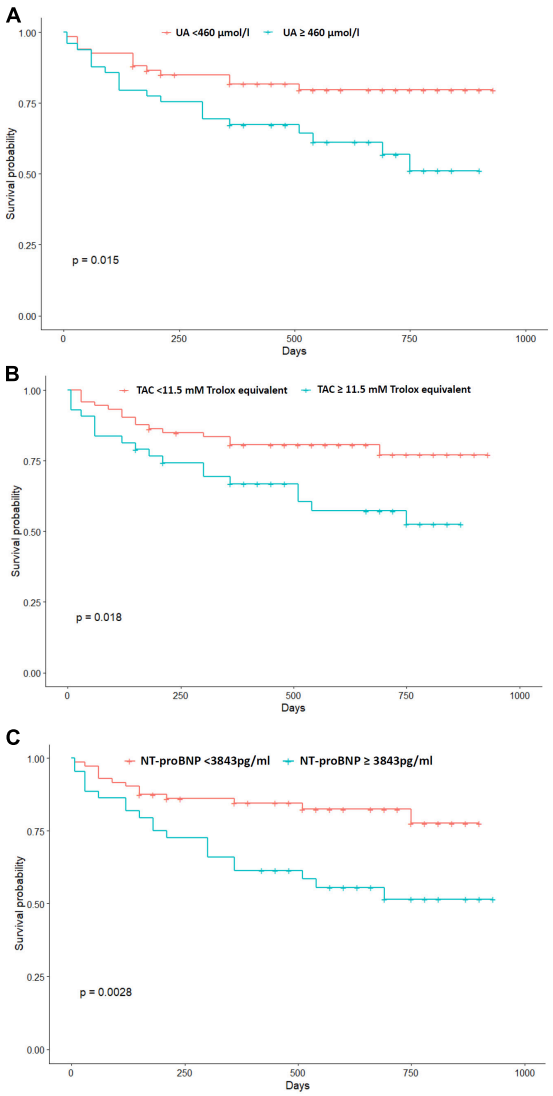


FIGURE 1
Kaplan-Meier event-free survival curves for post-discharge mortality relative to plasma levels of UA **(A)**, TAC **(B)** and NT-proBNP **(C)** above or below cut-off values.

(60% sensitivity, 65% specificity, 67% AUC), 11.5 mM trolox equivalent (55%sensitivity, 70% specificity, 65% AUC) and 3843 pg/ml (60% sensitivity, 71% specificity, 70% AUC), respectively. KM survival curves illustrate an increasing risk of mortality rate among patients with UA levels above 460 µmol/l (log-rank test $p = 0.015$). Furthermore, subjects with TAC levels above 11.5 mM trolox equivalent were more likely to die during follow-up period (log-rank test $p = 0.018$) (Figure 1). The predictive utility of NT-proBNP levels for death risk among HFrEF patients was also evaluated by KM survival curves. Log-rank test showed that patients with NT-proBNP levels above 3843 pg/ml are more likely to experience death (log-rank test $p = 0.0028$) (Figure 1).

TABLE 2 Univariate Cox proportional hazards regression analysis for predictors of post-discharge mortality.

Variable	HR	95% CI	*P*-value
Age	1.028	0.997–1.016	0.076
Female sex	1.798	0.894–3.616	0.1
Hypertension	1.011	0.503–2.033	0.976
Diabetes mellitus	1.055	0.502–2.219	0.887
Hyperlipidemia	1.283	0.609–2.701	0.520
CKD	1.966	0.948–4.199	0.069
COPD	2.179	0.762–6.235	0.146
Anemia	1.547	0.779–3.072	0.212
Previous history of HF	2.198	1.080–4.471	**0.030**
NYHA class III-IV	3.316	1.165–9.438	**0.025**
LVEF	0.951	0.912–0.991	**0.017**
BMI	0.976	0.890–1.076	0.612
Admission SBP	0.972	0.954–0.990	**0.003**
QRS duration	1.016	1.003–1.029	**0.021**
AF	0.946	0.471–1.903	0.877
LBBB	1.843	0.928–3.658	0.080
CC	0.986	0.972–1.000	**0.045**
ACEI/ARBs	0.359	0.173–743	**0.006**
Beta blockers	0.426	0.209–0.867	**0.019**
Diuretics	2.609	0.624–10.91	0.189
Aldosterone antagonist	0.784	0.393–1.565	0.491
Statins	1.058	0.543–2.096	0.872
UA	1.002	1.000–1.004	**0.042**
TB	1.018	0.997–1.040	0.098
GGT	0.998	0.990–1.007	0.715
GPx	1.046	0.713–1.534	0.819
SOD	1.006	0.991–1.020	0.457
TAC	1.226	1.016–1.478	**0.033**
NT-proBNP	1.001	1.000–1.001	$< \mathbf{10^{-3}}$

ACEI/ARBs, angiotensin-converting enzyme inhibitors/angiotensin receptor II; AF, atrial fibrillation; BMI, body mass index; CC, creatinine clearance; COPD, chronic obstructive pulmonary disease; CKD, chronic kidney disease; GPx, glutathione peroxidase; HF, heart failure; LBBB, left Bundle Branch Block; LVEF, left ventricular ejection fraction; NT-proBNP, N-terminal pro-B-type Natriuretic peptide; NYHA, New York Heart Association; SBP, systolic blood pressure; SOD, superoxide dismutase; UA, uric acid; TAC, total antioxidant capacity; TB, total bilirubin. Bold values indicate the *p*-values < 0.05.

Uni and multivariate cox regression analysis

In univariate Cox-regression analysis, highest UA levels (HR 1.002, 95% CI 1.000–1.004, $p = 0.042$) and elevated TAC levels (HR 1.126, 95% CI 1.016–1.478, $p = 0.033$) were significant predictors of post-discharge mortality. Additional significant determinants of mortality risk were revealed in univariate analysis, including: previous history of HF, NYHA class III/IV, NT-proBNP levels, LVEF, admission SBP, QRS duration, CC rate, beta blockers and ACEI/ARBs (Table 2). In order to evaluate the independent association of UA and TAC in the context of other common clinically available data, a multivariate model was performed, including sex, age, BMI

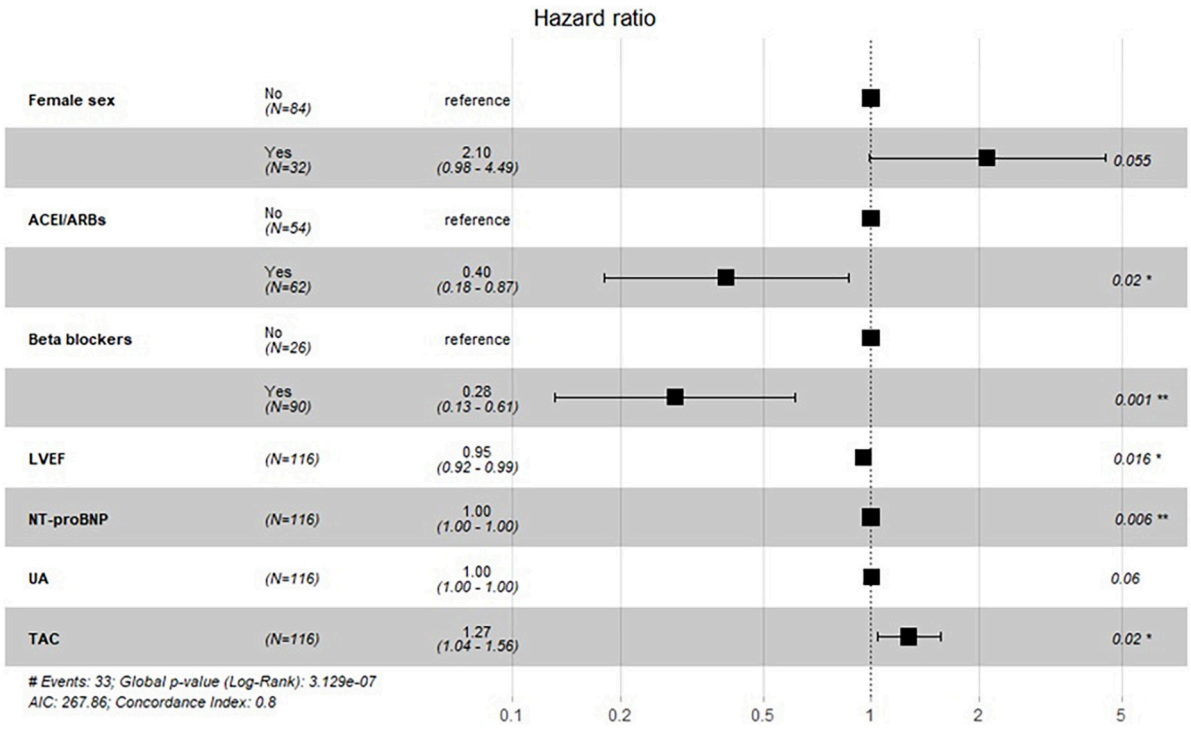


FIGURE 2
Multivariable Cox regression for post-discharge mortality prediction.

and all significant clinical predictors. Stepwise multivariate Cox-regression analysis revealed that elevated UA levels (HR 1.001, 95% CI 1.000–1.003, $p = 0.06$) of high TAC levels (HR 1.272, 95% CI 1.040–1.560, $p = 0.02$) remained independent predictors of death. In the multivariate Cox model, female sex, lower LVEF, and high NT-proBNP levels were also independent predictors for post-discharge mortality. Furthermore, multivariate analysis showed that patients taking of beta blockers or ACEI/ARBs after hospital discharge faced a lower risk of death during the follow-up period (Figure 2).

TABLE 3 Predictive model for post-discharge mortality based on stepwise generalized linear model.

Variable	β-coefficient	OR (95% CI)	*P*-value
Age	0.054	1.055 (1.002–1.117)	0.048
Female sex	1.480	4.392 (1.304–16.411)	0.020
LVEF	0.778	0.925 (0.859–0.988)	0.027
Admission SBP	−0.022	0.978 (0.947–1.006)	0.142
NT-proBNP	0.001	1.001 (1.0002–1.001)	0.009
UA	0.003	1.003 (1.0001–1.006)	0.044
TAC	0.347	1.414 (1.027–2.009)	0.039
ACEI/ARBs	−1.180	0.307 (0.093–0.949)	0.044
Beta blockers	−1.825	0.161 (0.038–0.596)	0.008

ACEI/ARBs, angiotensin-converting enzyme inhibitors/angiotensin receptor II; LVEF, left ventricular ejection fraction; NT-proBNP, N-terminal pro-B-type Natriuretic peptide; SBP, systolic blood pressure; UA, uric acid; TAC, total antioxidant capacity.

Risk prediction model for post-discharge mortality

In order to predict the post-discharge mortality risk among HFrEF patients, GLM, RF, and XGBoost models were performed using clinical and laboratory data. Regarding GLM, 9 independent predictors of mortality were retained in the final model. They included beta blocker, NT-proBNP levels, female sex, LVEF, TAC levels, admission SBP, ACEI/ARBs, UA levels and advanced age (Table 3). In the RF and XGBoost models, SHAP values were used to explain how the selected features affect the mortality prediction. In each variable importance row, all patients' attribution to post-discharge death risk were plotted with dots of different colors where the blue dots represent the lowest risk value and the red dots represent the highest risk value (Figure 3). In RF model, the top 5 related variables in mortality prediction were CC rate, UA levels, TAC levels, BMI and ACEI/ARBs. In XGBoost model, UA levels were the most important identified feature, followed by admission SBP, NT-proBNP levels, beta blockers and QRS duration. Among the three predictive models created, the GLM presented the best performance. This model achieved the highest AUC (74.5%),

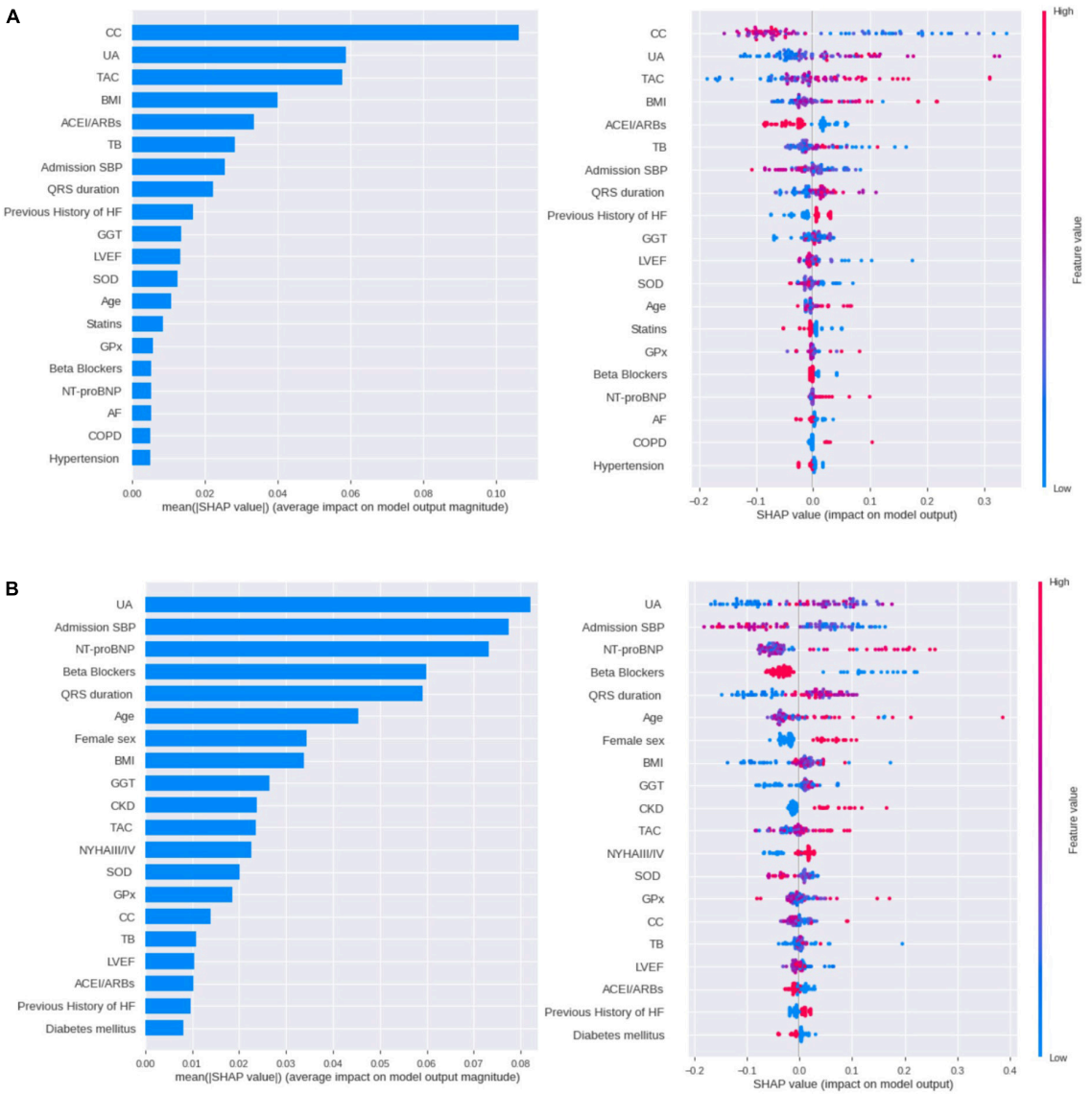


FIGURE 3
SHAP plots for the ML models in predicting post-discharge mortality using **(A)** RF and **(B)** XGBoost. In each variable importance row, all patients' attribution to mortality risk was plotted using different color dots. The red dots represent the highest risk of death.

accuracy (81.9%), recall (58%), precision (65%), F1-score (64%) and MCC (53%) compared to the respective values in the RF and XGboost models (Table 4). Therefore, the GLM model was selected to predict the risk of post-discharge death in the present study. The estimated β-coefficients of Glm-selected variables were used to estimate the logit for a patient using the standard GLM equation. The estimated individual probability (P) of dying was then calculated using the following formula:

$$P = \frac{e^{Logit}}{1 + e^{Logit}}$$

According to the estimated probability, study patients were divided into high ($P > 0.5$) and low ($P \leq 0.5$) risk groups and KM curve survival analysis was applied. The result of the Log-rank test showed a gradual decline in survival among high-risk subjects during follow-up period, indicating that patients with

TABLE 4 Performance comparison between the three models.

	AUC	Accuracy	Precision	Recall	F1-score	MCC
GLM	74.5%	81.89%	73%	58%	64%	53.1%
RF	63%	75.86%	52%	36%	43%	33.3%
XGBoost	61.5%	72.4%	65%	33%	44%	26.1%

AUC, area under curve; GLM, generalized linear model; MCC, matthews correlation coefficient; RF, random forest; XGBoost, extreme gradient boosting.

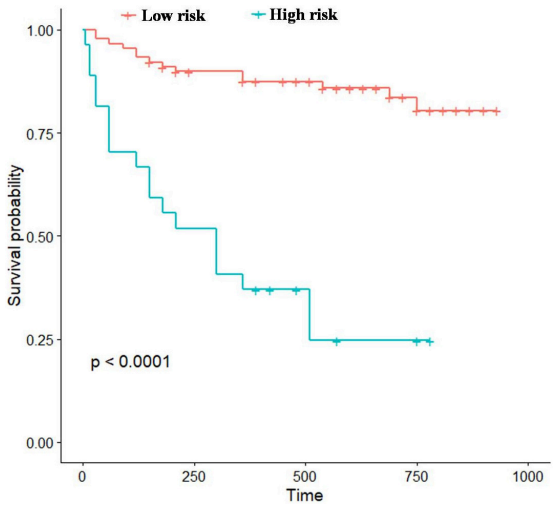


FIGURE 4
Kaplan–Meier survival analysis between the high-and low-risk groups.

higher mortality risk probability are more likely to die (log-rank test $p < 10^{-3}$) (**Figure 4**).

Based on the GLM model, an online post-discharge mortality risk calculator was created.[1] This tool calculated the estimated individual probability of dying using clinical characteristics and laboratory tests (**Figure 5**).

Discussion

In the present study, we developed a new predictive model for post-discharge mortality among patients admitted for newly diagnosed or exacerbated HFrEF. In addition to clinical patient characteristics and NT-proBNP, 6 oxidative stress markers were considered as candidate variables for risk prediction. Among the three developed models, the GLM presented the best performance for death prediction in HFrEF. Hence, this model was selected to predict mortality risk among our study patients. The predictors included in the GLM model were age, female sex, beta blockers, NT-proBNP levels, LVEF, TAC levels, admission SBP, ACEI/ARBs and UA levels.

To our knowledge, our study presents the first predictive model for post-discharge mortality risk among HFrEF patients, based on a combination of clinical patients' characteristics, NT-proBNP and oxidative stress markers. Previous mortality risk prediction models among patients with HFrEF incorporated clinical characteristics and NT-proBNP levels (24, 48, 49). The integration of few readily obtainable variables is also a great advantage of our model highlighting the potential use and implementation of this artificial intelligence tool in clinical practice. A number of the predictors identified in our model were also included in the PREDICT-HF models, including age, beta blockers, NT-proBNP levels, LVEF, admission SBP and UA levels; though the former has fewer variables (48).

On the statistical significance level, our model had a good discriminative power of post-discharge death with an AUC of 74.5%, which is comparable to the Seattle HF Model, one of the most extensively used models, achieving an AUC of 72.9% for 1-year survival (16). Likewise, the AUCs from recently developed mortality risk prediction models in HFrEF ranked from 67 to 78% (24, 48–50). Furthermore, our model was developed based on the original statistical approach. The traditional statistical method based on logistic regression, commonly applied in previous predictive models for HFrEF, was explored (51). In addition, two novel machine learning approaches (random forest and extreme gradient boosting) were also applied to predict post-discharge mortality in patients with HFrEF.

Regarding oxidative stress markers, UA and TAC were retained as significant predictors of post-discharge mortality risk in our model. Our analysis showed that high level of plasma UA before discharge was significantly associated with all increased post-discharge death risk among HFrEF patients using univariate and multivariate analysis. Patients with UA levels > 460 μmol/l were at high risk to die during follow up periods. In addition, UA was one of the top five features selected in RF and XGBoost models. Several clinical studies showed that elevated level of UA was an important risk factor of mortality in HFrEF (52, 53). Further, UA has been incorporated in a clinically validated model to predict mortality in HFrEF, displaying independent predictive ability in the Seattle Heart Failure Model (16). UA, the final product of purine degradation, is one of the major endogenous antioxidants in the human plasma (54, 55). As a putative protective mechanism, increased levels of plasma UA may be a compensatory mechanism to limit the damage of inappropriate ROS production (56). Xanthine oxidoreductase (XO) is the enzyme that catalyzes the conversion of hypothanthine to UA in the final steps of purine catabolism (57). Elevated UA levels reflect the amplified activity of XO, a key enzyme in the production of ROS (58). Elevated UA levels reflect the amplified activity of XO, a key enzyme in the production of ROS (28). Accordingly, previous studies reported that an increased level of UA was associated with disease severity and correlated positively with left ventricular remodeling indices in patients with HFrEF (31, 59).

TAC has also emerged as an important prognostic marker of mortality in the GLM model. Our study elucidated that an elevated level of plasma TAC was an independent predictor for post-discharge mortality even after complete adjustment, including sex, age, BMI, NT-proBNP and significant risk factors. We also introduced the 11.5 mM trolox equivalent cut-off of plasma TAC before discharge as a novel tool for risk stratification. Previous studies reported an association

1 https://heartcheckapp.herokuapp.com/

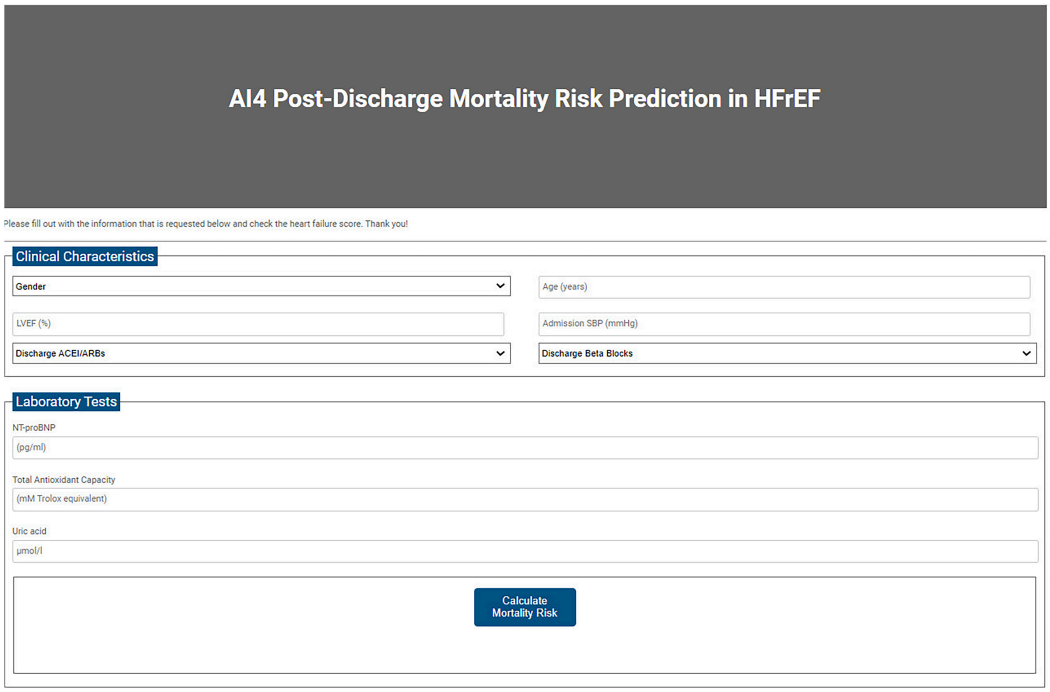


FIGURE 5
Post-discharge mortality risk calculator.

between high levels of TAC and death risk in HFrEF (60) and urgent heart transplantation among patients with non-ischemic cardiomyopathy (36). Similar predictive significance of the TAC level was also found in other cardiovascular diseases such as coronary artery disease (61). Indeed, Tomandlova et al. showed that TAC level was significantly higher in patients with more severe coronary artery disease and worse prognosis. In addition, previous studies showed an association between high TAC level and mortality among patients with ischemic stroke and severe septic (62, 63). TAC is an integrated parameter rather than a simple sum of measurable antioxidants, representing the cumulative action of all plasma antioxidants (64). A higher level of TAC may reflect a greater antioxidant response due to intensified production of ROS. It has also been suggested that elevated TAC level in non-survivors patients represents a compensating mechanism of an organism for depleted antioxidative components (65). Overall, the association between increased mortality risk and higher levels of UA and TAC, observed in this study and earlier studies, suggested that these easy and accessible markers of oxidative stress could be valuable biomarkers and prognostic factors in patients with HFrEF.

Our multivariable model confirmed the strong predictive value of NT-proBNP. Results obtained in this study showed that an elevated level of plasma NT-proBNP was a strong predictor of post-discharge mortality risk in both univariate and multivariate analysis. Previous clinical studies have shown that NT-proBNP level was significantly associated with increased mortality risk in HFrEF (24, 66). Moreover, the American Heart Association/American College of Cardiology HF guidelines have recommended measuring Natriuretic peptide biomarkers for prognostication among patients with HFrEF (20). Indeed, the Natriuretic peptide tests are still underutilized in different Tunisian centers as in many African countries which could implement this test in routine clinical practice (5). In the present study, we present the first prospective evaluation of NT-proBNP levels among Tunisian patients with HFrEF. In this context, our study confirms the predictive value of this test and encourages its implementation in routine clinical practice in Africa.

In addition, the clinical variables identified in our model include age, female sex, lower admission SBP and lower LVEF. These variables were established as prognostic markers in HFrEF (17, 18). Beta blockers and ACEI/ARBs were also retained in the model. Indeed, our findings reported a lower risk of post-discharge mortality in patients taking these medications before discharge, which are similar to those observed in several reports (67, 68). The protective role of these drugs impacting mortality risk among HFrEF patients may be partially due to their potent antioxidant properties (69, 70), indicating the important prognostic role of oxidative stress in patients with HFrEF.

Using the GLM model, we created a simple calculator allowing the identification of patients with heightened post-discharge death risk. Regarding its practical application, this calculator is promising to be applied in future clinical

practice. Indeed, this easy-to-use calculator can be easily implemented in clinical practice. It is anticipated to aid physicians to calculate the estimated mortality risk. Indeed, the identification of patients at heightened post-discharge death risk can be used to alter care, with closer follow-up and potential earlier consideration of advanced therapies. Accurate estimation of mortality risk in patients with HFrEF may allow clinicians and patients to make important decisions regarding the appropriateness and timing of disease-modifying treatments and advanced therapies (48). In addition, identifying factors common to patients at high risk of post-discharge mortality may reveal potential targets for interventions to improve prognosis. The implementation of these risk prediction tools is relevant to healthcare, particularly in the clinical decision-making (71). Shared decision making can improve motivation for therapy adherence and lifestyle change. Furthermore, the application of these prognosis tools can guide the allocation of healthcare resources and reduce costs (71).

Limitations

We present a single center study conducted in Hedi Chaker Hospital of Sfax presenting a relatively small number of patients and regional limitations. However, this sample size had sufficient statistical power to detect mean differences. At the clinical level, our study was restricted to HFrEF patients due to its high prevalence in Tunisia (8). Then, this model may not be generalizable to HFpEF patients and HFrEF patients with major life-altering comorbidities including, acute myocardial infarction, severe valvular disease requiring surgery and renal failure requiring dialysis. We also precise that data regarding the use of devices, such as implantable cardiac defibrillator and cardiac resynchronization therapy, were not available. However, the inclusion of such variables may add prognostic power to our model. At statistical level, we consider that the absence of external validation represents an acknowledged limitation, which was circumvented by good discrimination in internal validation of our model.

Future directions

Our pilot study should be expanded to different medical centers in order to include patients from different parts of Tunisia. We also encourage further validation of our risk model in other populations of HFrEF patients. Larger future prospective multicenter studies with larger numbers of patients are needed to confirm the predictive value of oxidative stress markers, assess for cost effectiveness and to define the implications for earlier interventions to improve prognosis.

Conclusion

In conclusion, we developed a new predictive model for post-discharge all-cause death in patients with HFrEF based on a combination of clinical patient characteristics, NT-proBNP and oxidative stress markers. This new model assisted by a simple-to-use calculator may allow the identification of patients at heightened post-discharge mortality risk and could assist the treatment decision-making.

Data availability statement

The original contributions presented in this study are included in the article/**Supplementary material**, further inquiries can be directed to the corresponding author.

Ethics statement

This study was approved by the Local Ethics Committee of CHU Hédi Chaker of Sfax (Tunisia), in accordance with the principles expressed in the Declaration of Helsinki (CPP Sud 0276/2017). The patients/participants provided their written informed consent to participate in this study.

Author contributions

NK contributed to the conceptualization and methodology. IG, SC, FZ, and LA made substantial contributions to samples and data acquisition. RA and WO contributed to data analysis. IG contributed to data interpretation and drafting of the manuscript. IG, FB, WO, NK, and AR reviewed and edited the manuscript. All authors had full access to all the data in the study and agreed to the published version of the manuscript.

Funding

This study was financially supported by the Tunisian Government Program "Programme d'encouragement des jeunes chercheurs" (n°18JEC06-05) funded by the ministry of higher education and scientific research.

Acknowledgments

We acknowledge the ministry of higher education and scientific research of the republic of Tunisia for the financial support. We thank the Cardiology Department Staff of CHU Hédi Chaker of Sfax (Tunisia).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fcvm.2022.1017673/full#supplementary-material

References

1. Yancy, CW, Jessup M, Bozkurt B, Butler J, Casey DE, Drazner MH, et al. 2013 ACCF/AHA guideline for the management of heart failure. *Circulation.* (2013) 128:e240–327. doi: 10.1161/CIR.0b013e31829e8776

2. James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 Diseases and Injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet.* (2018) 392:1789–858. doi: 10.1016/S0140-6736(18)32279-7

3. Groenewegen A, Rutten FH, Mosterd A, Hoes AW. Epidemiology of heart failure. *Eur J Heart Fail.* (2020) 22:1342–56. doi: 10.1002/ejhf.1858

4. Lippi G, Sanchis-Gomar F. Global epidemiology and future trends of heart failure. *AME Med J.* (2020) 5:15–15. doi: 10.21037/amj.2020.03.03

5. Gtif I, Bouzid F, Charfeddine S, Abid L, Kharrat N. Heart failure disease: an African perspective. *Arch Cardiovasc Dis.* (2021) 114:680–90. doi: 10.1016/j.acvd.2021.07.001

6. Agbor VN, Essouma M, Ntusi NAB, Nyaga UF, Bigna JJ, Noubiap JJ. Heart failure in sub-Saharan Africa: a contemporaneous systematic review and meta-analysis. *Int J Cardiol.* (2018) 257:207–15. doi: 10.1016/j.ijcard.2017.12.048

7. Ntusi NBA, Mayosi BM. Epidemiology of heart failure in sub-Saharan Africa. *Expert Rev Cardiovasc Ther.* (2009) 7:169–80. doi: 10.1586/14779072.7.2.169

8. Abid L, Charfeddine S, Kammoun I, Ben Halima M, Ben Slima H, Drissa M, et al. Epidemiology of heart failure and long-term follow-up outcomes in a north-African population: results from the NAtional TUnisian REgistry of Heart Failure (NATURE-HF). *PLoS One.* (2021) 16:e0251658. doi: 10.1371/journal.pone.0251658

9. McMurray JJV, Adamopoulos S, Anker SD, Auricchio A, Böhm M, Dickstein K, et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: the task force for the diagnosis and treatment of acute and chronic heart failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. *Eur Heart J.* (2012) 33:1787–847. doi: 10.1093/eurheartj/ehs104

10. Murphy SP, Ibrahim NE, Januzzi JL. Heart failure with reduced ejection fraction: a review. *JAMA.* (2020) 324:488–504. doi: 10.1001/jama.2020.10262

11. Schwinger RHG. Pathophysiology of heart failure. *Cardiovasc Diagn Ther.* (2021) 11:263–76. doi: 10.21037/cdt-20-302

12. Cheng RK, Cox M, Neely ML, Heidenreich PA, Bhatt DL, Eapen ZJ, et al. Outcomes in patients with heart failure with preserved, borderline, and reduced ejection fraction in the Medicare population. *Am Heart J.* (2014) 168:721–30. doi: 10.1016/j.ahj.2014.07.008

13. Shah KS, Xu H, Matsouaka RA, Bhatt DL, Heidenreich PA, Hernandez AF, et al. Heart failure with preserved, borderline, and reduced ejection fraction: 5-year outcomes. *J Am Coll Cardiol.* (2017) 70:2476–86. doi: 10.1016/j.jacc.2017.08.074

14. Butler J, Yang M, Manzi MA, Hess GP, Patel MJ, Rhodes T, et al. Clinical course of patients with worsening heart failure with reduced ejection fraction. *J Am Coll Cardiol.* (2019) 73:935–44. doi: 10.1016/j.jacc.2018.11.049

15. Tsao CW, Lyass A, Enserro D, Larson MG, Ho JE, Kizer JR, et al. Temporal trends in the incidence of and mortality associated with heart failure with preserved and reduced ejection fraction. *JACC Heart Fail.* (2018) 6:678–85. doi: 10.1016/j.jchf.2018.03.006

16. Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, et al. The seattle heart failure model. *Circulation.* (2006) 113:1424–33. doi: 10.1161/CIRCULATIONAHA.105.584102

17. Ho JE, Enserro D, Brouwers FP, Kizer JR, Shah SJ, Psaty BM, et al. Predicting heart failure with preserved and reduced ejection fraction: the international collaboration on heart failure subtypes. *Circ Heart Fail.* (2016) 9:10.1161/CIRCHEARTFAILURE.115.003116e003116. doi: 10.1161/CIRCHEARTFAILURE.115.003116

18. O'Connor CM, Whellan DJ, Wojdyla D, Leifer E, Clare RM, Ellis SJ, et al. Factors related to morbidity and mortality in patients with chronic heart failure with systolic dysfunction. *Circulation.* (2012) 5:63–71. doi: 10.1161/CIRCHEARTFAILURE.111.963462

19. Feng Z, Akinrimisi OP, Gornbein JA, Truong QA, Das S, Singh JP, et al. Combination biomarkers for risk stratification in patients with chronic heart failure biomarkers prognostication in HF. *J Cardiac Fail.* (2021) 27:1321–7. doi: 10.1016/j.cardfail.2021.05.028

20. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Colvin MM, et al. 2017 ACC/AHA/HFSA Focused Update of the 2013 ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. *Circulation.* (2017) 136:e137–61. doi: 10.1161/CIR.0000000000000509

21. Berezin AE, Berezin AA. *ÁåðåçíÎC Biomarker-Based Guideline-Directed Medical Therapy of Heart Failure: the Gap between Guidelines and Clinical Practice.* (2021). Available online at: http://dspace.zsmu.edu.ua/handle/123456789/15477 (accessed June 26, 2022).

22. Ky B, French B, Levy WC, Sweitzer NK, Fang JC, Wu AHB, et al. Multiple biomarkers for risk prediction in chronic heart failure. *Circ Heart Fail.* (2012) 5:183–90. doi: 10.1161/CIRCHEARTFAILURE.111.965020

23. Richter B, Koller L, Hohensinner PJ, Zorn G, Brekalo M, Berger R, et al. A multi-biomarker risk score improves prediction of long-term mortality in patients with advanced heart failure. *Int J Cardiol.* (2013) 168:1251–7. doi: 10.1016/j.ijcard.2012.11.052

24. Pocock SJ, Ferreira JP, Gregson J, Anker SD, Butler J, Filippatos G, et al. Novel biomarker-driven prognostic models to predict morbidity and mortality in chronic heart failure: the EMPEROR-Reduced trial. *Eur Heart J.* (2021) 42:4455–64. doi: 10.1093/eurheartj/ehab579

25. Subramanian D, Subramanian V, Deswal A, Mann D. New predictive models of heart failure mortality using time-series measurements and ensemble models. *Circ Heart Fail.* (2011) 4:456–62. doi: 10.1161/CIRCHEARTFAILURE.110.958496

26. Bouwens E, Schuurman A-S, Akkerhuis KM, Baart SJ, Caliskan K, Brugts JJ, et al. Serially measured cytokines and cytokine receptors in relation to clinical outcome in patients with stable heart failure. *Can J Cardiol.* (2020) 36:1587–91. doi: 10.1016/j.cjca.2020.08.010

27. Münzel T, Gori T, Keaney JF Jr., Maack C, Daiber A. Pathophysiological role of oxidative stress in systolic and diastolic heart failure and its therapeutic implications. *Eur Heart J.* (2015) 36:2555–64. doi: 10.1093/eurheartj/ehv305

28. Aimo A, Castiglione V, Borrelli C, Saccaro LF, Franzini M, Masi S, et al. Oxidative stress and inflammation in the evolution of heart failure: from pathophysiology to therapeutic strategies. *Eur J Prev Cardiol.* (2020) 27:494–510. doi: 10.1177/2047487319870344

29. Tsutsui H, Kinugawa S, Matsushima S. Oxidative stress and heart failure. *Am J Physiol Heart Circ Physiol.* (2011) 301:H2181–90. doi: 10.1152/ajpheart.00554.2011

30. He L, He T, Farrar S, Ji L, Liu T, Ma X. Antioxidants maintain cellular redox homeostasis by elimination of reactive oxygen species. *CPB.* (2017) 44:532–53. doi: 10.1159/000485089

31. Wojciechowska C, Romuk E, Tomasik A, Skrzep-Poloczek B, Nowalany-Kozielska E, Birkner E, et al. Oxidative stress markers and C-reactive protein are related to severity of heart failure in patients with dilated cardiomyopathy. *Mediators Inflamm.* (2014) 2014:e147040. doi: 10.1155/2014/147040

32. Wojciechowska C, Romuk E, Nowalany-Kozielska E, Jacheæ W. Total antioxidant capacity, uric acid, and bilirubin in patients with heart failure due to non-ischemic cardiomyopathy. *Reactive Oxygen Spec.* (2017) 3:66–80.

33. Romuk E, Wojciechowska C, Jacheæ W, Nowak J, Niedziela J, Malinowska-Borowska J, et al. Comparison of oxidative stress parameters in heart failure patients depending on ischaemic or nonischaemic aetiology. *Oxid Med Cell Longev.* (2019) 2019:7156038.

34. Anker SD, Doehner W, Rauchhaus M, Sharma R, Francis D, Knosalla C, et al. Uric acid and survival in chronic heart failure. *Circulation.* (2003) 107:1991–7. doi: 10.1161/01.CIR.0000065637.10517.A0

35. Tamariz L, Harzand A, Palacio A, Verma S, Jones J, Hare J. Uric acid as a predictor of all-cause mortality in heart failure: a meta-analysis. *Congest Heart Fail.* (2011) 17:25–30. doi: 10.1111/j.1751-7133.2011.00200.x

36. Wojciechowska C, Jacheæ W, Romuk E, Ciszek A, Bodnar P, Chwalba T, et al. Serum sulfhydryl groups, malondialdehyde, uric acid, and bilirubin as predictors of adverse outcome in heart failure patients due to ischemic or nonischemic cardiomyopathy. *Oxid Med Cell Longev.* (2021) 2021:6693405. doi: 10.1155/2021/6693405

37. Poelzl G, Eberl C, Achrainer H, Doerler J, Pachinger O, Frick M, et al. Prevalence and prognostic significance of elevated gamma-glutamyltransferase in chronic heart failure. *Circ Heart Fail.* (2009) 2:294–302. doi: 10.1161/CIRCHEARTFAILURE.108.826735

38. Ess M, Mussner-Seeber C, Mariacher S, Lorsbach-Koehler A, Pachinger O, Frick M, et al. Gamma-glutamyltransferase rather than total bilirubin predicts outcome in chronic heart failure. *J Cardiac Fail.* (2011) 17:577–84. doi: 10.1016/j.cardfail.2011.02.012

39. Romuk E, Jacheæ W, Kozielska-Nowalany E, Birkner E, Zemła-Woszek A, Wojciechowska C. Superoxide dismutase activity as a predictor of adverse outcomes in patients with nonischemic dilated cardiomyopathy. *Cell Stress Chaperones.* (2019) 24:661–73. doi: 10.1007/s12192-019-00991-3

40. Ewa R, Celina W, Wojciech J, Aleksandra Z-W, Alina M, Marta B, et al. Malondialdehyde and uric acid as predictors of adverse outcome in patients with chronic heart failure. *Oxid Med Cell Longev.* (2019) 2019:9246138. doi: 10.1155/2019/9246138

41. Ho KK, Anderson KM, Kannel WB, Grossman W, Levy D. Survival after the onset of congestive heart failure in Framingham Heart Study subjects. *Circulation.* (1993) 88:107–15. doi: 10.1161/01.CIR.88.1.107

42. Paulus WJ, Tschöpe C, Sanderson JE, Rusconi C, Flachskampf FA, Rademakers FE, et al. How to diagnose diastolic heart failure: a consensus statement on the diagnosis of heart failure with normal left ventricular ejection fraction by the Heart Failure and Echocardiography Associations of the European Society of Cardiology. *Eur Heart J.* (2007) 28:2539–50. doi: 10.1093/eurheartj/ehm037

43. Folland ED, Parisi AF, Moynihan PF, Jones DR, Feldman CL, Tow DE. Assessment of left ventricular ejection fraction and volumes by real-time, two-dimensional echocardiography. A comparison of cineangiographic and radionuclide techniques. *Circulation.* (1979) 60:760–6. doi: 10.1161/01.CIR.60.4.760

44. Cockcroft D, Gault M, Cockcroft DW, Gault MH. Prediction of creatine clearance from serum creatinine. *Nephron.* (1976) 16:31–41. doi: 10.1159/000180580

45. Beyer WF, Fridovich I. Assaying for superoxide dismutase activity: some large consequences of minor changes in conditions. *Anal Biochem.* (1987) 161:559–66. doi: 10.1016/0003-2697(87)90489-1

46. Flohé L, Günzler WA. Assays of glutathione peroxidase. *Methods Enzymol.* (1984) 105:114–21. doi: 10.1016/s0076-6879(84)05015-1

47. Cheng H, Garrick DJ, Fernando RL. Efficient strategies for leave-one-out cross validation for genomic best linear unbiased prediction. *J Anim Sci Biotechnol.* (2017) 8:38. doi: 10.1186/s40104-017-0164-6

48. Simpson J, Jhund PS, Lund LH, Padmanabhan S, Claggett BL, Shen L, et al. Prognostic models derived in PARADIGM-HF and validated in ATMOSPHERE and the Swedish heart failure registry to predict mortality and morbidity in chronic heart failure. *JAMA Cardiol.* (2020) 5:432–41. doi: 10.1001/jamacardio.2019.5850

49. Muñoz MA, Calero E, Duran J, Navas E, Alonso S, Argemí N, et al. Short-term mortality in patients with heart failure at the end-of-life stages: hades study. *J Clin Med.* (2022) 11:2280. doi: 10.3390/jcm11092280

50. Hung W-K, Liu H-T, Wang C-C, Chou C-C, Wen M-S, Chang P-C. One-year mortality risk stratification in patients hospitalized for acute decompensated heart failure: construction of TSOC-HFrEF risk scoring model. *Acta Cardiol Sin.* (2020) 36:240–50. doi: 10.6515/ACS.202005_36(3).20190826B

51. Adderley NJ, Taverner T, Price MJ, Sainsbury C, Greenwood D, Chandan JS, et al. Development and external validation of prognostic models for COVID-19 to support risk stratification in secondary care. *BMJ Open.* (2022) 12:e049506. doi: 10.1136/bmjopen-2021-049506

52. Ambrosio G, Leiro MGC, Lund LH, Coiro S, Cardona A, Filippatos G, et al. Serum uric acid and outcomes in patients with chronic heart failure through the whole spectrum of ejection fraction phenotypes: analysis of the ESC-EORP Heart Failure Long-Term (HF LT) Registry. *Eur J Intern Med.* (2021) 89:65–75. doi: 10.1016/j.ejim.2021.04.001

53. Palazzuoli A, Ruocco G, De Vivo O, Nuti R, McCullough PA. Prevalence of hyperuricemia in patients with acute heart failure with either reduced or preserved ejection fraction. *Am J Cardiol.* (2017) 120:1146–50. doi: 10.1016/j.amjcard.2017.06.057

54. Ames BN, Cathcart R, Schwiers E, Hochstein P. Uric acid provides an antioxidant defense in humans against oxidant- and radical-caused aging and cancer: a hypothesis. *Proc Natl Acad Sci.* (1981) 78:6858–62. doi: 10.1073/pnas.78.11.6858

55. Davies KJ, Sevanian A, Muakkassah-Kelly SF, Hochstein P. Uric acid-iron ion complexes. A new aspect of the antioxidant functions of uric acid. *Biochem J.* (1986) 235:747–54. doi: 10.1042/bj2350747

56. Nieto FJ, Iribarren C, Gross MD, Comstock GW, Cutler RG. Uric acid and serum antioxidant capacity: a reaction to atherosclerosis? *Atherosclerosis.* (2000) 148:131–9. doi: 10.1016/s0021-9150(99)00214-2

57. Battelli MG, Polito L, Bortolotti M, Bolognesi A. Xanthine oxidoreductase in drug metabolism: beyond a role as a Detoxifying Enzyme. *Curr Med Chem.* (2016) 23:4027–36. doi: 10.2174/0929867323666160725091915

58. Yu W, Cheng J-D. Uric acid and cardiovascular disease: an update from molecular mechanism to clinical perspective. *Front Pharmacol.* (2020) 11:582680. doi: 10.3389/fphar.2020.582680

59. Radovanovic S, Savic-Radojevic A, Pekmezovic T, Markovic O, Memon L, Jelic S, et al. Uric acid and gamma-glutamyl transferase activity are associated with left ventricular remodeling indices in patients with chronic heart failure. *Rev EspCardiol.* (2014) 67:632–42. doi: 10.1016/j.rec.2013.11.017

60. Romuk E, Wojciechowska C, Jacheæ W, Zemła-Woszek A, Momot A, Buczkowska M, et al. Malondialdehyde and uric acid as predictors of adverse outcome in patients with chronic heart failure. *Oxid Med Cell Longev.* (2019) 2019:e9246138.

61. Tomandlova M, Parenica J, Lokaj P, Ondrus T, Kala P, Miklikova M, et al. Prognostic value of oxidative stress in patients with acute myocardial infarction complicated by cardiogenic shock: a prospective cohort study. *Free Radic Biol Med.* (2021) 174:66–72. doi: 10.1016/j.freeradbiomed.2021.07.040

62. Lorente L, Martín MM, Pérez-Cejas A, Abreu-González P, Ramos L, Argueso M, et al. Association between total antioxidant capacity and mortality in ischemic stroke patients. *Ann Intensive Care.* (2016) 6:39. doi: 10.1186/s13613-016-0143-7

63. Lorente L, Martín MM, Almeida T, Abreu-González P, Ferreres J, Solé-Violán J, et al. Association between serum total antioxidant capacity and mortality in severe septic patients. *J Crit Care.* (2015) 30:217.e7–e12. doi: 10.1016/j.jcrc.2014.09.012

64. Ghiselli A, Serafini M, Natella F, Scaccini C. Total antioxidant capacity as a tool to assess redox status: critical view and experimental data. *Free Radic Biol Med.* (2000) 29:1106–14. doi: 10.1016/s0891-5849(00)00394-4

65. MacKinnon KL, Molnar Z, Lowe D, Watson ID, Shearer E. Measures of total free radical activity in critically ill patients. *Clin Biochem.* (1999) 32:263–8. doi: 10.1016/S0009-9120(98)00109-X

66. Salah K, Stienen S, Pinto YM, Eurlings LW, Metra M, Bayes-Genis A, et al. Prognosis and NT-proBNP in heart failure patients with preserved versus reduced ejection fraction. *Heart.* (2019) 105:1182–9. doi: 10.1136/heartjnl-2018-314173

67. Khan MS, Fonarow GC, Ahmed A, Greene SJ, Vaduganathan M, Khan H, et al. Dose of angiotensin-converting enzyme inhibitors and angiotensin receptor blockers and outcomes in heart failure. *Circulation.* (2017) 10:e003956. doi: 10.1161/CIRCHEARTFAILURE.117.003956

68. Park JJ, Park H-A, Cho H-J, Lee H-Y, Kim KH, Yoo B-S, et al. β-blockers and 1-year postdischarge mortality for heart failure and reduced ejection fraction and slow discharge heart rate. *J Am Heart Assoc.* (2019) 8:e011121. doi: 10.1161/JAHA.118.011121

69. Münzel T, Keaney JF. Are ACE inhibitors a "magic bullet" against oxidative stress? *Circulation.* (2001) 104:1571–4. doi: 10.1161/hc3801.095585

70. Book WM. Carvedilol: a nonselective β blocking agent with antioxidant properties. *Congestive Heart Fail.* (2002) 8:173–90. doi: 10.1111/j.1527-5299.2002.00718.x

71. Vogenberg FR. Predictive and prognostic models: implications for healthcare decision-making in a modern recession. *Am Health Drug Benefits.* (2009) 2:218–22.

frontiers | Frontiers in Cardiovascular Medicine

TYPE Original Research
PUBLISHED 15 November 2022
DOI 10.3389/fcvm.2022.1009821

Check for updates

OPEN ACCESS

EDITED BY
Qingpeng Zhang,
City University of Hong Kong,
Hong Kong SAR, China

REVIEWED BY
Ming Huang,
Nara Institute of Science
and Technology (NAIST), Japan
Emmanuel Andrès,
Hôpitaux Universitaires de Strasbourg,
France

*CORRESPONDENCE
David Susič
david.susic@ijs.si
Gregor Poglajen
gregor.poglajen@kclj.si
Anton Gradišek
anton.gradisek@ijs.si

SPECIALTY SECTION
This article was submitted to
Heart Failure and Transplantation,
a section of the journal
Frontiers in Cardiovascular Medicine

RECEIVED 02 August 2022
ACCEPTED 26 October 2022
PUBLISHED 15 November 2022

CITATION
Susič D, Poglajen G and Gradišek A
(2022) Identification
of decompensation episodes
in chronic heart failure patients based
solely on heart sounds.
Front. Cardiovasc. Med. 9:1009821.
doi: 10.3389/fcvm.2022.1009821

COPYRIGHT
© 2022 Susič, Poglajen and Gradišek.
This is an open-access article
distributed under the terms of the
Creative Commons Attribution License
(CC BY). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Identification of decompensation episodes in chronic heart failure patients based solely on heart sounds

David Susič[1,2]*, Gregor Poglajen[3,4]* and Anton Gradišek[1]*

[1]Department of Intelligent Systems, Jožef Stefan Institute, Ljubljana, Slovenia, [2]Jožef Stefan Postgraduate School, Ljubljana, Slovenia, [3]Advanced Heart Failure and Transplantation Program, Department of Cardiology, University Medical Centre Ljubljana, Ljubljana, Slovenia, [4]Medical Faculty, University of Ljubljana, Ljubljana, Slovenia

Decompensation episodes in chronic heart failure patients frequently result in unplanned outpatient or emergency room visits or even hospitalizations. Early detection of these episodes in their pre-symptomatic phase would likely enable the clinicians to manage this patient cohort with the appropriate modification of medical therapy which would in turn prevent the development of more severe heart failure decompensation thus avoiding the need for heart failure-related hospitalizations. Currently, heart failure worsening is recognized by the clinicians through characteristic changes of heart failure-related symptoms and signs, including the changes in heart sounds. The latter has proven to be largely unreliable as its interpretation is highly subjective and dependent on the clinicians' skills and preferences. Previous studies have indicated that the algorithms of artificial intelligence are promising in distinguishing the heart sounds of heart failure patients from those of healthy individuals. In this manuscript, we focus on the analysis of heart sounds of chronic heart failure patients in their decompensated and recompensated phase. The data was recorded on 37 patients using two types of electronic stethoscopes. Using a combination of machine learning approaches, we obtained up to 72% classification accuracy between the two phases, which is better than the accuracy of the interpretation by cardiologists, which reached 50%. Our results demonstrate that machine learning algorithms are promising in improving early detection of heart failure decompensation episodes.

KEYWORDS

heart failiure, cardiac decompensation, heart sound, machine learing, phonocardiogram (PCG), artificial intelligence-AI, decompensation detection, classification

Introduction

Chronic heart failure (CHF) is a complex chronic condition, characterized by the inability of the heart muscle to provide sufficient perfusion to meet the metabolic demands of the body; alternatively the failing heart stabilizes the circulation by operating at the higher filling pressures, which generate the majority of the symptoms and signs, characteristic of CHF. Globally, CHF has reached epidemic proportions, affecting roughly 2% of the world's overall population, with the incidence increasing at 2% annually. The prevalence of CHF reaches around 10% in the overall population aged over 65 years and additionally carries a significant burden in terms of healthcare costs and personnel expenditure. Importantly, the prognosis of CHF patients remains dismal with 50% mortality at 5 years, which is largely related to heart failure decompensation episodes that require in-hospital management (1). Available literature suggests that identifying the CHF decompensation episodes in their pre-symptomatic phase (when the patient does not yet subjectively feel worse) may enable clinicians to make appropriate and timely changes to patient's medical therapy thus preventing overt CHF decompensation to occur or to occur in much milder forms that do not require hospitalization (2). This may significantly improve patients' outcomes (2). As microelectromechanical systems technology is invasive, of limited availability and, at least for now, prohibitively expensive, it has not yet exerted a wider impact on the management of heart failure. Importantly, studies using easily obtainable (but non-specific) clinical parameters (body weight, blood pressure, heart rate, etc.) displayed only limited success in accurately predicting CHF decompensation episodes. There is thus a significant unmet need in the heart failure community for effective, cost-efficient and robust protocols for early detection of CHF decompensation episodes. First automatic detections of CHF and other cardiovascular diseases were performed with electrocardiogram data (3, 4), heart rate variability data, photoplethysmogram, and clinical data such as respiratory rate, weight, pulse rate, age, and blood pressure (5). Recently, automated methods for analysing heart sounds and detecting cardiovascular disease from heart sounds have been increasingly developed as more and more datasets of heart sound recordings have become publicly available (6–8).

Heart sound classification algorithms found in the literature include classical ML models, statistical models, and artificial neural networks (NN) (9). There are a few papers that specifically address CHF. In the work of Gjoreski et al. (10), a stack of ML classifiers was used to classify normal sounds and heart failure. In the preliminary study, Gjoreski et al. (11) used a simple decision tree classifier to classify compensated and decompensated stages of CHF, using only the portion of our dataset recorded under the first experimental setup. Gao et al. (12) compared the fully convolutional NN, gated recurrent unit, long short-term memory, and support-vector machine (SVM) models to classify normal heart and two subtypes of CHF. Liu et al. (13) compared NN and SVM in classifying normal heart and subtype of heart failure. In the work of Zheng et al. (14), the SVM, NN, and a statistical hidden Markov model were compared in classifying normal and CHF sounds. In our first study, we tested ML algorithms for detecting decompensation in CHF using general audio features generated with a dedicated audio feature tool. Features were extracted from segments with a fixed length of 2 s from a subset of our current data set (15).

As previous studies (10, 11, 16) have demonstrated promising results in distinguishing between the heart sounds of healthy people from those of CHF patients, we now focus on a more specific task. In this manuscript, we explore how ML algorithms can be employed to identify decompensation episodes based on the heart sounds of CHF patients. The study was performed on the recordings of 37 patients in both decompensated and recompensated phases, using two types of electronic stethoscopes. We discuss the performances of several ML algorithms in view of feasibility of this approach for telemedicine application. We compare the classification accuracies of the ML models with that of cardiologists, who are domain experts.

Materials and methods

Data

Our dataset consists of phonocardiograms (PCG) of 37 CHF patients (average age of 51.3 ± 13.3 years). The dataset was obtained by two different setups. The first part (21 subjects) was obtained with a 3M^{TM} Littmann Electronic Stethoscope Model 3200 (17) digital stethoscope and consists of PCGs 30 s in length. The second part (16 subjects) was obtained with the Eko DUO ECG + Digital Stethoscope (18) and consists of PCGs 15 s in length. Both devices use built-in filters to reduce ambient noise and record single channel audio signals at a sampling rate of 4 kHz. According to the principal component analysis, the

TABLE 1 Pathophysiology of heart sounds (9).

Heart sound	Frequency range	Characteristics	Duration/Location
S1	10–200 Hz	Dull and prolonged	0.12–0.15 s
S2	20–250 Hz	Sharp and short	0.08–0.12 s
S3	25–70 Hz	Soft and thudding quality	0.04 s, early diastole
S4	15–70 Hz	Weak and rumbling	Slightly before S1
Gallop	15–50 Hz	Galloping rhythm	0.08–0.2 s, diastole
Murmurs	Up to 600 Hz	Whooshing, rumbling	Systole, diastole
Opening snaps	100–800 Hz	Snapping sound	Diastole
Rubs	100–800 Hz	Scratching sound	Systole, early/Late diastole
Clicks	100–800 Hz	Short and loud	Early systole

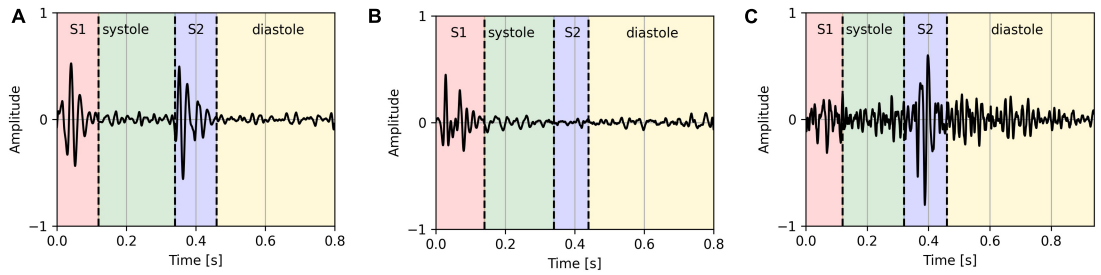


FIGURE 1
Clear PCG segment **(A)**, PCG segment with missing S2 sound **(B)**, and a noisy PCG segment **(C)**.

TABLE 2 List of features extracted from the phonocardiograms segments.

Feature type (N)	Per segment state	Domain	Description
BPM (1)	RR	Time	Inverse of segment duration in beats per minute
Dur_*state* (4)	S1, Sys, S2, Dia	Time	Duration in milliseconds
Dur_Ratio_*ratio* (8)		Time	Duration ratios
MeanEnv_Ratio_*ratio* (8)		Statistical	Mean envelope ratios
RMS_*state* (4)	S1, Sys, S2, Dia	Statistical	Root-mean-square of a signal
RMS_Ratio_*ratio* (8)		Statistical	Root-mean-square ratios
ZC_*state* (4)	S1, Sys, S2, Dia	Statistical	Zero crossings
SE_*state* (4)	S1, Sys, S2, Dia	Statistical	Sample entropy
Skewness_*state* (4)	S1, Sys, S2, Dia	Statistical	Skewness
Kurtosis_*state* (4)	S1, Sys, S2, Dia	Statistical	Kurtosis
PSD_*region_band* (24)	Sys, Dia	Frequency	Power spectral density for different frequency bands
mfcc1-13_*state* (52)	S1, Sys, S2, Dia	Frequency	13 Mel-frequency cepstral coefficients
SpecCentroid_*state* (4)	S1, Sys, S2, Dia	Frequency	Spectral centroid
SpecBandwidth_*state* (4)	S1, Sys, S2, Dia	Frequency	2nd order spectral bandwidth
SpecContrast2-5_*state* (16)	S1, Sys, S2, Dia	Frequency	Spectral contrast for different frequency bands
SpecFlatness_*state* (4)	S1, Sys, S2, Dia	Frequency	Spectral flatness
SpecRolloff_*state* (4)	S1, Sys, S2, Dia	Frequency	Frequency below which 85% of the total spectral energy lies
PolyFeatures_*state* (4)	S1, Sys, S2, Dia	Frequency	Coefficients of degree-1 polynomial fit to the spectrogram
dwt1-4_*state* (16)	S1, Sys, S2, Dia	Wavelet	Level 4 discrete wavelet transform coefficients

difference between the recordings from the two devices after preprocessing were small, thus it was reasonable to consider both as the same, device-independent dataset. The subjects were recorded in both the decompensated and the recompensated phase. The decompensated episode was recorded when the patient was admitted to the hospital for worsening heart failure episode, while the recompensated one was recorded upon discharge from the hospital when the patient was optimally recompensated and was deemed optivolemic. The PCGs were collected by medical professionals at University Medical Centre Ljubljana from left parasternal 3rd intercostal space body position. Overall, our dataset consists of 75 PCGs, 37 and 38 for compensated and decompensated phases, respectively, and adds up to 29 min and 15 s in length. The study protocol was reviewed and approved by the Republic of Slovenia National Medical Ethics Committee (decision number 0120-276/2016-5).

A phonocardiogram consists of regular S1 and S2 sounds, which are caused by the closing and opening of the heart valves, and several additional sounds that may be present. These include the S3 and S4 sounds, gallops, murmurs, opening snaps, rubs, and clicks. While the S3 sound may also be present in normal hearts of young children and athletes, other abnormal sounds are never present in a normal heart. The pathophysiology of the heart sounds can be found in **Table 1**.

Methods

The outcome of interest used to evaluate the ML models was a binary variable indicating whether a PCG represents a decompensated or a recompensated CHF phase. The steps of the methodology pipeline included preprocessing of

TABLE 3 Patient clinical characteristics.

Parameter	Study population (N = 37)
Age, y	56.6 ± 12
Gender (male), %	88
Heart failure etiology (ischemic), %	27
Cause of decompensation	
Volume overload, %	78
Infection, %	15
Arrhythmia, %	7
LVEF, %	26.7 ± 7.7
NT-proBNP, pg/ml	4,593 (953, 5,102)
Medical therapy	
ARNI/ACEI/ARB, %	96
Beta blockers, %	100
MRA, %	96
SGLT2i, %	70
Diuretic, %	96
Ca-antagonist, %	9
Digoxin, %	15

LVEF, left ventricular ejection fraction; ARNI, angiotensin receptor antagonist neprylisin inhibitor; ACEI, angiotensin convertase enzyme inhibitor; ARB, angiotensin receptor blocker; MRA, mineralocorticoid receptor blocker; SGLT2i, sodium glucose transporter 2 inhibitor; Ca, calcium.

the PCGs, feature extraction, and training and evaluation of the ML models.

Patient selection

We performed a prospective nonrandomized cohort study that included 37 consecutive patients hospitalized for worsening heart failure at the Advanced Heart Failure Center, Dept. of Cardiology, UMC Ljubljana. Inclusion criteria were as follows: chronic heart failure of ischemic or non-ischemic etiology, hospitalization for worsening heart failure <24 h, age >18 years; We did not consider patients with severe valvular disease, artificial valves, patients with acute myocardial infarction and/or de-novo acute heart failure, patients in cardiogenic shock, on vasoactive and/or inotropic support, on mechanical ventilation or on short- or long-term mechanical circulatory support for this analysis or patients that were hospitalized for worsening heart failure >24 h for this analysis. Clinical, biochemical and medical therapy data were collected for all the patients at the time of the initial heart sound sampling.

All patients included in this analysis were recompensated using levosimendan, followed by the intravenous diuretic therapy. In all study participants, heart sounds were recorded before the infusion of levosimendan (decompensated phase) and upon reaching the optivolemic phase (recompensated phase).

Preprocessing

The first part of the preprocessing step was filtering. Although heart sounds have frequencies of up to 800 Hz (see Table 1), the most dominant frequencies are in the frequency range of 20–400 Hz (19). The mean spectral roll-off frequency (frequency below which 85% of the total spectral energy lies) of our dataset is 49.9 ± 9.7 and 304.2 ± 99.4 Hz for the PCGs recorded by the first and the second experimental setting, respectively. To reduce the effects of different recording settings and to reduce noise, the PCGs were filtered with a bandpass Butterworth filter of order 4 and a frequency range from 25 to 400 Hz.

As the PCGs obtained by the two experimental settings were also recorded at different amplitudes, the next preprocessing step was heart sound signal normalization. We used the root mean square (RMS) normalization with the target amplitude of -20 dBFS. As opposed to the peak normalization, which normalizes the signal based on the highest peak, the RMS normalization normalizes the signal based on the average power level by calculating the average value of all peaks.

The Springer's modification (20) of Schmidt's method (21) was used to split the heart sound into separate cardiac cycles and to find the four main states of each segment (RR): S1, systole, S2, and diastole. This algorithm uses a hidden semi-Markov model and Viterbi decoding and provides a state-of-the-art method for segmenting heart sounds. Segmentation allows us to extract the features of the possible abnormal sounds from the corresponding heart sound states. In manually reviewing the segmented PCGs, we found that seven (9%) of the recordings either consisted of a significant number of segments that were not correctly determined, or the recording itself was so unclear that it was impossible to tell whether the segments were correct or not. The two most common reasons for the segmentation error were that one of the main sounds (S1 or S2) was not detected by the PCG recorder, resulting in a segment that was too

TABLE 4 Results of classification of a representative subset of our dataset by the medical experts.

PCG	Class	Expert 1	Expert 2	Expert 3	Accuracy
1	1	0	1	0	0.33
2	0	1	1	1	0
3	1	1	0	0	0.33
4	1	1	1	1	1
5	0	0	0	0	1
6	0	0	0	1	0.67
7	0	1	1	1	0
8	1	1	0	0	0.33
9	1	0	1	0	0.33
10	0	0	0	1	0.67
11	0	0	0	0	1
12	1	0	1	0	0.33
Overall accuracy		0.58	0.67	0.25	0.5

Classes 0 and 1 are recompensated and decompensated CHF phases, respectively.

TABLE 5 Results of the models' performance.

Classifier	Accuracy	Precision	Recall	F1	ROC AUC
LR	**0.72 (0.15; 0.61–0.83)**	0.73 (0.17; 0.61–0.86)	**0.73 (0.22; 0.56–0.90)**	**0.71 (0.16; 0.59–0.84)**	**0.74 (0.18; 0.61–0.88)**
LGBM	0.70 (0.16; 0.58–0.82)	0.68 (0.29; 0.47–0.90)	0.63 (0.29; 0.41–0.85)	0.63 (0.27; 0.43–0.84)	0.71 (0.16; 0.59–0.83)
SVC	0.68 (0.16; 0.56–0.79)	**0.75 (0.23; 0.58–0.92)**	0.60 (0.31; 0.37–0.84)	0.62 (0.22; 0.45–0.78)	0.71 (0.15; 0.59–0.83)
RF	0.68 (0.18; 0.54–0.81)	0.63 (0.28; 0.42–0.84)	0.68 (0.35; 0.42–0.94)	0.63 (0.28; 0.41–0.84)	0.66 (0.23; 0.48–0.83)
DT	0.62 (0.15; 0.51–0.73)	0.60 (0.25; 0.41–0.79)	0.66 (0.29; 0.44–0.87)	0.60 (0.23; 0.42–0.78)	0.66 (0.16; 0.53–0.78)
GB	0.61 (0.16; 0.48–0.73)	0.61 (0.25; 0.42–0.80)	0.63 (0.27; 0.43–0.83)	0.59 (0.23; 0.42–0.77)	0.65 (0.17; 0.52–0.77)
XGB	0.61 (0.19; 0.46–0.75)	0.59 (0.27; 0.39–0.80)	0.60 (0.31; 0.37–0.84)	0.58 (0.26; 0.38–0.77)	0.67 (0.18; 0.53–0.81)
KN	0.58 (0.16; 0.47–0.70)	0.60 (0.19; 0.45–0.75)	0.58 (0.23; 0.40–0.75)	0.57 (0.18; 0.43–0.71)	0.61 (0.17; 0.48–0.74)
SGD	0.58 (0.07; 0.52–0.63)	0.45 (0.23; 0.28–0.62)	0.71 (0.41; 0.40–1.02)	0.54 (0.29; 0.32–0.76)	0.63 (0.13; 0.53–0.73)
GNB	0.54 (0.14; 0.44–0.65)	0.52 (0.21; 0.37–0.68)	0.55 (0.27; 0.35–0.75)	0.52 (0.21; 0.36–0.68)	0.57 (0.23; 0.39–0.74)

The scores are given as mean (SD; 95% CI).
The results are calculated from 10-fold cross-validation.
The highest version of the individual metric is marked as bold.

long (longer than one RR interval), or that the high-amplitude noise was detected as one of the two main sounds, resulting in a segment that was too short (shorter than one RR interval).

As some of the features are calculated based on the characteristics of the S1 and S2 sounds, the segments where either of the sounds was not present (based on the signal envelope) or the signal-to-noise ratio was too high were excluded in the analysis. On average, 3.3 2.6% segments per PCG recording were removed. **Figure 1** shows an example of a clear segment, an example of a segment that was removed because S2 is missing, and an example of a segment that was removed because it is too noisy.

The normalization was performed using Python 3.7 (Python Programming Language, RRID:SCR_008394) and the library Pydub 0.25.1 (22), while the filtering and segmentation was performed using Matlab R2021a (MATLAB, RRID:SCR_001622) (23).

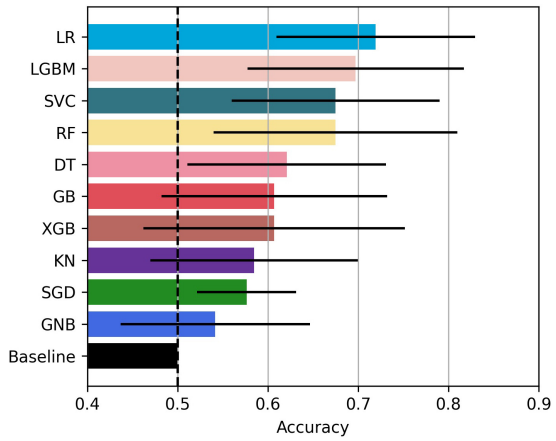


FIGURE 2
Bar plot of the model accuracies. Black horizontal lines represent 95% CI.

Feature extraction

A total of 177 features were extracted from each segment. These included features in the time domain, frequency domain, statistical features, and features generated by a 4-level wavelet decomposition. The complete list of features can be found in **Table 2**.

Features were extracted from the entire PCG signal data and from each of the four "*_states.*" In some cases, the features were calculated as a ratio of the features of the states. The "*_ratio*" features include S1/RR, Sys/RR, S2/RR, Dia/RR, S1/S2, Sys/Dia, Sys/S1, and Dia/S2. To extract the frequency domain features, the segments were transformed from the time domain to the frequency domain using a fast Fourier transform with a Hanning window of 64 milliseconds in length and a stride of 16 milliseconds. For the power spectral density frequency features, we selected frequency "*_bands*" of 25–40, 40–60, 60–80, 80–100, 100–120, 120–140, 140–160, 160–180, 180–200, 200–250, 250–300, and 300–400 Hz. The selected frequency bands are similar to those selected by the authors of Potes et al. (24). Although Mel-frequency cepstrum coefficients (25) (MFCCs) were developed to mimic human perception and are widely used in speech recognition, they have been shown to work for heart sound analysis as well (24, 26–28). We extracted the first 13 coefficients from each of the four states. The "*2–5*" bands of the spectral contrast features include: 25–50, 50–100, 100–200, and 200–400 Hz. Daubechies 4 wavelet was used as a basis for the discrete wavelet transform features.

To smooth out the outliers, we generated another set of 2×177 features representing the mean and standard deviation of the features taken as a sliding window with window size six across the segments of each PCG. To ensure that each segment was equally represented, windowing was performed

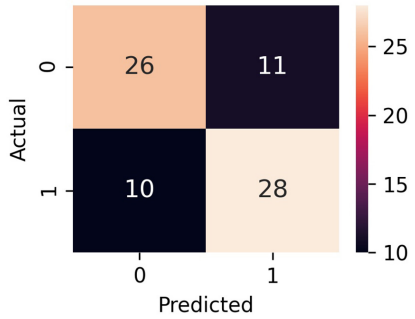


FIGURE 3
Confusion matrix of the LR model. Classes 0 and 1 are recompensated and decompensated CHF phases, respectively.

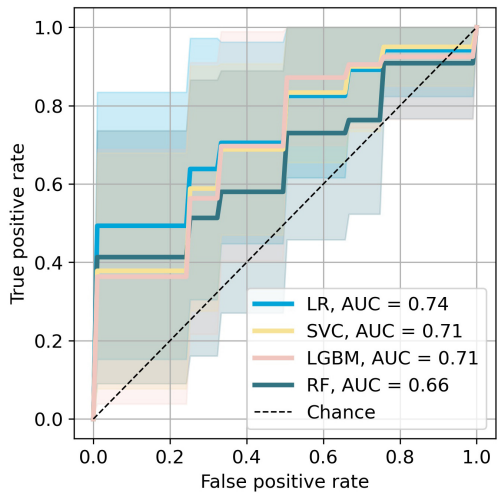


FIGURE 4
Receiver operating characteristic curves of the most accurate four models. The colored areas represent the 95% CI.

cyclically. The resulting 354 features were then used for the models' evaluation.

The features were extracted and calculated using Python 3.7 and libraries Librosa 0.9.1 (29), Scipy 1.5.2 (SciPy, RRID:SCR_008058) (30) and Numpy 1.18.5 (NumPy, RRID:SCR_008633) (31).

Experimental pipeline

We implemented 10 ML models. The decision tree classifier (DT) is a model that uses a tree diagram for decision making, where each branch is partitioned based on a threshold for a feature. The gradient boosting classifier (GB), the extreme gradient boosting classifier (XGB), the light gradient boosting machine classifier (LGBM), and the random forest classifier (RF) are ensemble methods that combine the predictions of multiple DTs. The C-support vector classifier (SVC) finds a hyperplane in the feature space that spatially separates the classes.

TABLE 6 *P*-values of McNemar's tests between the ML models.

	LR	LGBM	SVC	RF	DT	GB	XGB	KN	SGD	GNB
LR	/	0.91	0.88	0.81	0.79	0.74	0.76	0.72	0.83	0.78
LGBM	0.91	/	0.85	0.9	0.91	0.9	0.9	0.78	0.84	0.75
SVC	0.88	0.85	/	0.88	0.86	0.88	0.88	0.86	0.82	0.78
RF	0.81	0.9	0.88	/	1	0.85	0.9	0.86	0.85	0.86
DT	0.79	0.91	0.86	1	/	1	1	0.95	0.92	0.91
GB	0.74	0.90	0.88	0.85	1	/	0.95	0.90	0.86	0.85
XGB	0.76	0.9	0.88	0.9	1	0.95	/	0.85	0.85	0.85
KN	0.72	0.78	0.86	0.86	0.95	0.9	0.85	/	0.85	0.89
SGD	0.83	0.84	0.82	0.85	0.92	0.86	0.85	0.85	/	0.91
GNB	0.78	0.75	0.78	0.86	0.91	0.85	0.85	0.89	0.91	/

The K-neighbors classifier (KN) looks for closest neighbors in the features space to determine the class. The Gaussian naive Bayes (GNB) utilizes Bayes' theorem and makes the assumption that the features are independent and described by a Gaussian distribution. The logistic regression (LR) uses logistic function to map a linear combination of the features to a value between 0 and 1. The stochastic gradient descent classifier (SGD) takes iterative steps to minimize the cost function. All of the implemented models are probabilistic, meaning they assign probabilities for each class. The selected decision threshold for all models was 0.5. Each PCGs final decision was selected as the majority vote of the segments' predictions. The models were implemented using Python 3.7 and Scikit 0.24.2 (scikit-learn, RRID:SCR_002577) (32) and lightgbm 3.3.1 (LightGBM, RRID:SCR_021697) (33) libraries.

Models were evaluated with a subject-wise 10-fold cross validation using stratified folds with respect to the two different setups for data acquisition. We found that the models, trained only on the PCGs that were correctly segmented, perform significantly better. Thus, for each training set, we removed subjects that correspond to one of the seven PCGs we manually determined are segmented incorrectly.

To keep the models explainable and as transparent as possible and to avoid overfitting, we performed feature selection, retaining only a subset of the features used as model input. Although the selected features can depend on the model (e.g., decision tree based models can calculate the importance of the features according to their ability to increase the pureness of the levels), we selected our features independently of the models. This means that all of the models used the same selected features. Features were selected by calculating the mutual information (34) between each feature and the outcome variable. The mutual information between two variables is zero if the two variables are independent, and higher values indicate greater dependence. Each training fold was divided into five stratified subfolds, and 40 features that had the highest mutual information with the outcome variable on average across the five subfolds were used for training.

Results

Patient clinical characteristics are outlined in Table 3. Our final dataset included 898 decompensated and 908 recompensated (1,806 in total) PCG segments with 354 features. We used accuracy, precision, recall, F1-score, and area under receiver operating curve (ROC AUC) as the evaluation metrics, with accuracy as the main metric of performance evaluation. Formulas for calculation of accuracy, precision, recall, and F1-score are given in Equations (1–4). TP, FP, TN, and FN denote true positive, false positive, true negative, and false negative, respectively.

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$precision = \frac{TP}{TP + FP} \quad (2)$$

$$recall = \frac{TP}{TP + TN} \quad (3)$$

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (4)$$

Classification by the experts

The baseline of our method was determined by three cardiologists experts who were each asked to independently listen to a representative subset of 12 PCG recordings and classify them as decompensated or recompensated. Importantly, no other clinical data on the CHF patients were available to the clinicians at that time. This subset included three decompensated and three recompensated recordings from each of the two data acquisition setups. The results are given in Table 4.

The experts' classification accuracies were 58, 67, and 25%, averaging at 50%, which coincides with the dataset class distribution, meaning the cardiac auscultation alone contributes little to the experts' recognition of CHF decompensation episode.

Evaluation the models' performance

The results of the models' performance evaluation are shown in Table 5. The results are given along with standard deviation (SD) and t-distribution 95% confidence interval (CI).

The bar plot of the models' accuracies is shown in Figure 2. All of the 10 implemented models outperform the baseline, while six of them outperform the baseline with the 95% CI. The best performing model is LR, which achieved accuracy (SD; 95% CI) of 0.72 (0.15; 0.61–0.83). Additionally, the LR model also achieved highest performance in recall with the score of 0.73 (0.22; 0.56–0.90), F1-score of 0.71 (0.16; 0.59–0.84), and ROC AUC with the score of 0.74 (0.18; 0.61–0.88). The confusion matrix of the LR is shown in Figure 3. The best performing model according to the precision metric was SVC, which achieved a score of 0.75 (0.28; 0.58–0.92). The models with accuracies comparable to that of the LR model are LGBM, which achieved the score of 0.70 (0.16; 0.59–0.82), SVC, which achieved the score of 0.68 (0.16; 0.56–0.79), and RF, which achieved the score of 0.68 (0.18; 0.51–0.73). The ROC curves of the four most accurate models are shown in Figure 4.

To test whether the difference in the models' predictive accuracy was statistically significant, we calculated the *p*-values from McNemar's test (35). This test is used on contingency tables of the two models' predictions. The results are given in Table 6. We see that all of the models provided similar predictions, since the *p*-values are all close to 1.

The list of the top 40 features is found in Table 7, with the most important features being time domain features (10/40), power spectral density features for different frequency bands (17/40), and MFCCs (9/40). The most important heart sound seems to be diastole, as 21/40 of the most important features were extracted from diastole.

TABLE 7 Top 40 best predictor features according to their mutual information with the outcome.

Rank	Feature	MI	Rank	Feature	MI
1	m_BPM	0.16	21	m_Dur_Ratio_S1RR	0.1
2	m_Dur_Ratio_DiaRR	0.15	22	m_PSD_Dia_140_160Hz	0.1
3	m_Dur_Dia	0.14	23	sd_PSD_Sys_250_300Hz	0.1
4	sd_PSD_Dia_200_250Hz	0.14	24	sd_PSD_Sys_200_250Hz	0.1
5	sd_BPM	0.14	25	m_mfcc6_Dia	0.1
6	m_Dur_Ratio_SysS1	0.13	26	m_ZC_Dia	0.1
7	m_mfcc4_Dia	0.13	27	sd_PSD_Sys_140_160Hz	0.1
8	m_Dur_Ratio_SysDia	0.13	28	m_PSD_Dia_300_400Hz	0.1
9	sd_PSD_Dia_250_300Hz	0.13	29	m_PSD_Sys_300_400Hz	0.1
10	m_mfcc2_Sys	0.12	30	m_SpecContrast5_Dia	0.09
11	sd_PSD_Dia_180_200Hz	0.12	31	m_mfcc1_Sys	0.09
12	sd_PSD_Dia_160_180Hz	0.12	32	m_SpecCentroid_Dia	0.09
13	sd_PSD_Sys_120_140Hz	0.12	33	m_SpecBandwidth_Dia	0.09
14	m_mfcc2_Dia	0.12	34	m_mfcc6_S1	0.09
15	sd_Dur_Ratio_SysS1	0.11	35	sd_PSD_Sys_300_400Hz	0.09
16	m_mfcc1_Dia	0.11	36	m_PSD_Dia_250_300Hz	0.09
17	sd_PSD_Dia_140_160Hz	0.11	37	m_mfcc4_Sys	0.09
18	sd_PSD_Sys_180_200Hz	0.11	38	m_Dur_Ratio_SysRR	0.09
19	m_PSD_Dia_160_180Hz	0.11	39	m_Dur_Sys	0.09
20	m_mfcc6_Sys	0.11	40	sd_PSD_Dia_100_120Hz	0.08

The prefixes "m_" and "sd_" correspond to the mean and standard deviation of the features taken as a sliding window with window size six across the segments of each PCG. BPM, beats per minute; Dur, duration; PSD, power spectral density; mfcc, Mel-frequency cepstral coefficients; ZC, zero crossings; SpecContrast, spectral contrast; SpecCentroid, spectral centroid; SpecBandwidth, spectral bandwidth.

PhysioNet dataset experiments

To test the robustness of our pipeline, we tested it against dataset A of the PhysioNet (PhysioNet, RRID:SCR_007345) (36) public database of heart sound recordings. The dataset A contains PCGs from 117 normal and 292 abnormal hearts recorded from children and adults. Both healthy subjects and patients contributed between one and six PCGs. The recordings lasted between 9 and 36 s. The models were compared in a 10-fold cross validation with folds stratified with respect to the class. The model with the best accuracy was SVC, which achieved a score of 0.80 (0.06; 0.76–0.85), better than that of the majority, which was 0.71. The other best performing models were LGBM, GB, and XGB, with the accuracy scores of 0.80 (0.07; 0.74–0.85), 0.78 (0.08; 0.72–0.84), and 0.78 (0.06; 0.73–0.83), respectively. The results obtained with this method are very similar to those obtained with our previous approach in (11). It should be noted that the results may be somewhat positively biased as the subjects are not labeled and the recordings of the same subjects may be included in both the training and test sets.

Discussion

In this study, we used 10 ML models to classify decompensation episodes in CHF using a dataset of heart sound recordings from 37 CHF patients. We used 40 domain predictor features, extracted from the four states of heart sounds. All models outperformed the classification performed independently by three cardiology experts, which averaged at 50%. Logistic regression proved to be the best model in terms of accuracy, reaching 72 (15; 61–83)%. Power spectral density features, time domain features, and Mel-frequency cepstrum coefficients were found to be the most important predictors. Most of these features were extracted from diastole. From the medical perspective, this is reasonable, since the sounds produced in the diastole originate from the heart chambers refilling by blood. The heart of a CHF patient is more rigid than a healthy heart and will thus vibrate differently. Another observation is that several of the important features are related to heart rate (BMP). Again, this is relevant from the medical point of view, as patients in the decompensated phase have a faster pulse than those that are not decompensated. Our method was additionally tested on a public dataset of normal/abnormal heart sounds where it achieved exemplary results, although a direct comparison is not possible because the public dataset was heavily unbalanced.

In view of early detection of decompensation episodes of CHF and thus preventing decompensation from occurring or to occur in milder forms that would not lead to hospitalization, the results are promising, as they demonstrate that the ML algorithms can substantially outperform a human expert solely based on the heart sounds. It is important to stress that the algorithms were trained on data from the two extreme phases of CHF, indicating that the results obtained likely represent the upper limit on the accuracy such an approach can achieve. Thus, using only heart sounds for detection of decompensation is not sufficient, however, it can represent a valuable component of a decision-support system that takes into account additional patient data, with patients performing daily/weekly self-recording and self-assessment.

Recently, various approaches to automatic detection of heart disease have been successfully implemented for numerous data set modalities such as clinical features, images, and electrocardiograms (ECG) (37). Although the reported accuracies are very decent, some data are very difficult and/or expensive to obtain. Future plan for our system to support patients with CHF is to incorporate data that are easy to obtain and relatively inexpensive, such as clinical data, self-reported data, ECG data, daily activity data, and possibly others. In addition, our models could be integrated into a virtual coaching system (38) that tracks the patient's cardiac status and overall well-being and promotes medication use and/or physical activity to prevent deterioration of the condition.

Limitations

This study has the following limitations. First, the recorded patients are at different stages of CHF so a decompensated phase of a relatively healthy CHF patient can be similar to a recompensated phase of a patient with a later stage of CHF. In addition, there are different subtypes of CHF, which we did not consider in model building. Models trained separately for each stage/subtype would most likely provide better results. Second, the PCG is recorded when an individual is admitted to/discharged from the hospital, and not on a regular basis with the intention of capturing the deterioration of the condition. Deterioration is unpredictable, and therefore data collection starting with a CHF patient in good condition and then waiting until the situation deteriorates is not practical. Third, since we are using a dataset collected by ourselves, we cannot directly compare the accuracy of our method with related work, but only by testing it against a public dataset. Fourth, although ML models outperform the experts' classification, the inputs to the models are computer-extracted sound features most of which are not intuitive to the experts. Therefore, the models do not really provide the experts with additional knowledge to help them make decisions while listening to the heart, but can only be used as a component of stand-alone decision-making tools.

Conclusion

This study demonstrates that in chronic heart failure patients machine learning algorithms may

outperformcardiologists in detecting decompensation episodes based on heart sounds alone. The key predictor features are derived from diastole and come both from time and frequency domains. Although the results are promising, showing that machine learning algorithms perform better than cardiology experts, the use of heart sound data alone is not sufficient for early detection of decompensation. Therefore, additional clinical data must be added to the protocol before considering the integration of this method into a decision-support system. The inclusion of additional predictor variables such as weight, self-reported data, and electrocardiogram falls within the scope of future work.

Data availability statement

Raw data will be made available upon request.

Ethics statement

The studies involving human participants were reviewed and approved by Republic of Slovenia National Medical Ethics Committee. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

GP and AG designed the study. GP collected the data. DS developed the algorithms and performed the analysis. All authors participated in the analysis of the results and wrote the manuscript.

Funding

The authors acknowledge the funding from the Slovenian Research Agency (ARRS), Grant (PR-10495) (DS) and Basic core funding P2-0209 (AG).

Acknowledgments

We thank Matjaž Gams and Vito Janko for interesting and useful discussions.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J.* (2021) 42:3599–726. doi: 10.1093/eurheartj/ehab368

2. ClinicalTrials.gov. *CardioMEMS Heart Sensor Allows Monitoring of Pressure to Improve Outcomes in NYHA Class III Heart Failure Patients. ClinicalTrials.gov identifier: NCT00531661.* (2015). Available online at: https://clinicaltrials.gov/ct2/show/study/NCT0053166 (accessed July 25, 2022).

3. Jahmunah V, Oh SL, Wei JKE, Ciaccio EJ, Chua K, San TR, et al. Computer-aided diagnosis of congestive heart failure using ECG signals – A review. *Phys Med.* (2019) 62:95–104. doi: 10.1016/j.ejmp.2019.05.004

4. Bhurane AA, Sharma M, San-Tan R, Acharya UR. An efficient detection of congestive heart failure using frequency localized filter banks for the diagnosis with ECG signals. *Cogn Syst Res.* (2019) 55:82–94. doi: 10.1016/j.cogsys.2018.12.017

5. Tripoliti EE, Papadopoulos TG, Karanasiou GS, Naka KK, Fotiadis DI. Heart failure: diagnosis, severity estimation and prediction of adverse events through machine learning techniques. *Comput Struct Biotechnol J.* (2017) 15:26–47. doi: 10.1016/j.csbj.2016.11.001

6. Liu C, Springer D, Li Q, Moody B, Juan RA, Chorro FJ, et al. An open access database for the evaluation of heart sound algorithms. *Physiol Meas.* (2016) 37:2181–213. doi: 10.1088/0967-3334/37/12/2181

7. Clifford GD, Liu C, Moody B, Millet J, Schmidt S, Li Q, et al. Recent advances in heart sound analysis. *Physiol Meas.* (2017) 38:10–25. doi: 10.1088/1361-6579/aa7ec8

8. Sidra G, Ammara N, Taimur H, Bilal H, Ramsha A. Fully automated identification of heart sounds for the analysis of cardiovascular pathology. In: Khan F, Jan MA, Alam M editors. *Applications of Intelligent Technologies in Healthcare.* Cham: Springer (2018). p. 117–29.

9. Dwivedi AK, Imtiaz SA, Rodriguez-Villegas E. Algorithms for automatic analysis and classification of heart sounds–a systematic review. *IEEE Access.* (2019) 7:8316–45. doi: 10.1109/ACCESS.2018.2889437

10. Gjoreski M, Simjanoska M, Gradišek A, Peterlin A, Gams M, Poglajen G. Chronic heart failure detection from heart sounds using a stack of machine-learning classifiers. In: Jason J editor. *Proceedings of the 13th International Conference on Intelligent Environments; 2017 Aug 23-25.* Seoul: IEEE (2017). p. 14–9.

11. Gjoreski M, Gradišek A, Budna B, Gams M, Poglajen G. Machine learning and end-to-end deep learning for the detection of chronic heart failure from heart sounds. *IEEE Access.* (2020) 8:20313–24. doi: 10.1109/ACCESS.2020.2968900

12. Gao S, Zheng Y, Guo X. Gated recurrent unit-based heart sound analysis for heart failure screening. *Biomed Eng.* (2020) 19:3. doi: 10.1186/s12938-020-0747-x

13. Liu Y, Guo X, Zheng Y. An automatic approach using ELM classifier for HFpEF identification based on heart sound characteristics. *J Med Syst.* (2019) 43:285. doi: 10.1007/s10916-019-1415-1

14. Zheng Y, Guo X, Qin J, Xiao S. Computer-assisted diagnosis for chronic heart failure by the analysis of their cardiac reserve and heart sound characteristics. *Comput Methods Programs Biomed.* (2015) 112:372–83. doi: 10.1016/j.cmpb.2015.09.001

15. Susič D, Poglajen G, Gradišek A. Machine learning models for detection of decompensation in chronic heart failure using heart sounds. In: Humberto H, Valera A, Luštrek M editors. *Proceedings of the Workshops at 18th International Conference on Intelligent Environments (IE2022)*. Amsterdam: IOS Press (2022). p. 340–9.

16. Gjoreski M, Gradišek A, Budna B, Gams M, Poglajen G. Toward early detection and monitoring of chronic heart failure using heart sounds. In: Muñoz A, Ouhbi S, Minker W, Echabbi L, Navarro-Cía M editors. *Proceedings of the 15th International Conference on Intelligent Environments in conjunction with the 15th International Conference on Intelligent Environments (IE19); 2019 Jun 24-27; Rabat, Morocco (Ambient intelligence and smart environments, 26)*. Amsterdam: IOS Press (2019). p. 336–43.

17. Littmann Electronic Stethoscopes 3M United States. *Littmann Electronic Stethoscopes*. (2022). Available online at: https://www.littmann.com/3M/en_US/littmann-stethoscopes/products/~{}/3M-Littmann-Stethoscopes/Electronic-Stethoscopes/?N=5142935+8711017+8727094+3294857497&rt=r3 (accessed July 13 2022).

18. Eko. *Eko DUO ECG + Digital Stethoscope*. (2022). Available online at: https://shop.ekohealth.com/products/duo-ecg-digital-stethoscope?variant=39350415655008 (accessed July 13, 2022).

19. McGee S. Auscultation of the heart: general principles. 5th ed. *Evidence-Based Physical Diagnosis*. Philadephia, PA: Elsevier (2022). p. 327–32.

20. Springer DB, Tarassenko L, Clifford GD. Logistic regression-HSMM-based heart sound segmentation. *IEEE Trans Biomed Eng.* (2016) 63:822–32. doi: 10.1109/TBME.2015.2475278

21. Schmidt SE, Holst-Hansen C, Graff C, Toft E, Struijk JJ. Segmentation of heart sound recordings by a duration-dependent hidden Markov model. *Physiol Meas.* (2010) 31:513–29. doi: 10.1109/CIC.2008.4749049

22. Robert J, Webbie M. *Pydub [Internet]*. San Francisco, CA: GitHub (2018)

23. Matlab. *version 9.10.0 (R2021a)*. Natick, MA: The MathWorks Inc (2021).

24. Potes C, Parvaneh S, Rahman A, Conroy B. Ensemble of feature-based and deep learning-based classifiers for detection of abnormal heart sounds. *Proceedings of the 2016 Computing in Cardiology Conference (CinC)*. Vancouver, BC: IEEE (2016). p. 621–4. doi: 10.22489/CinC.2016.182-399

25. Rabiner L, Juang BH. *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall (1993).

26. Zabihi M, Rad AB, Kiranyaz S, Gabbouj M, Katsaggelos AK. Heart sound anomaly and quality detection using ensemble of neural networks without segmentation. *Proceedings of the 2016 Computing in Cardiology Conference (CinC)*. Vancouver, BC: IEEE (2016). p. 613–6. doi: 10.22489/CinC.2016.180-213

27. Chen TE, Yang SI, Ho LT, Tsai KH, Chen YH, Chang YF, et al. S1 and S2 heart sound recognition using deep neural networks. *IEEE Trans Biomed Eng.* (2017) 64:372–80. doi: 10.1109/TBME.2016.2559800

28. Nilanon T, Yao J, Hao J, Purushotham S, Liu Y. Normal / abnormal heart sound recordings classification using convolutional neural network. *Proceedings of the 2016 Computing in Cardiology Conference (CinC)*. Vancouver, BC: IEEE (2016). p. 585–8. doi: 10.22489/CinC.2016.169-535

29. McFee B, Metsai A, McVicar M, Balke S, Thomé C, Raffel C, et al. *Librosa/Librosa: 0.9.1 [Internet]*. (2022). Available online at: https://zenodo.org/record/6097378 (accessed February 15, 2022).

30. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods.* (2020) 17:261–72. doi: 10.1038/s41592-019-0686-2

31. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. *Nature.* (2020) 585:357–62. doi: 10.1038/s41586-020-2649-2

32. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res.* (2011) 12:2825–30.

33. Ke G, Meng Q, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree. In: Jordan MI, LeCun Y, Solla SA editors. *Advances in Neural Information Processing Systems*. Red Hook, NY: Curran Associates, Inc (2017). 30 p. doi: 10.1016/j.envres.2020.110363

34. Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. *Phys Rev E.* (2004) 69:066138.

35. McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika.* (1947) 12:153–7. doi: 10.1007/bf02295996

36. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circ.* (2000) 101:215–20. doi: 10.1161/01.cir.101.23.e215

37. Javeed A, Khan SU, Ali L, Ali S, Imrana Y, Rahman A. Machine learning-based automated diagnostic systems developed for heart failure prediction using different types of data modalities: a systematic review and future directions. *Comput Math Methods Med.* (2022) 2022:1–30. doi: 10.1155/2022/9288452

38. Tsiouris KM, Tsakanikas VD, Gatsios D, Fotiadis DI. A review of virtual coaching systems in healthcare: closing the loop with real-time feedback. *Front Digit Health.* (2020) 2:567502. doi: 10.3389/fdgth.2020.567502

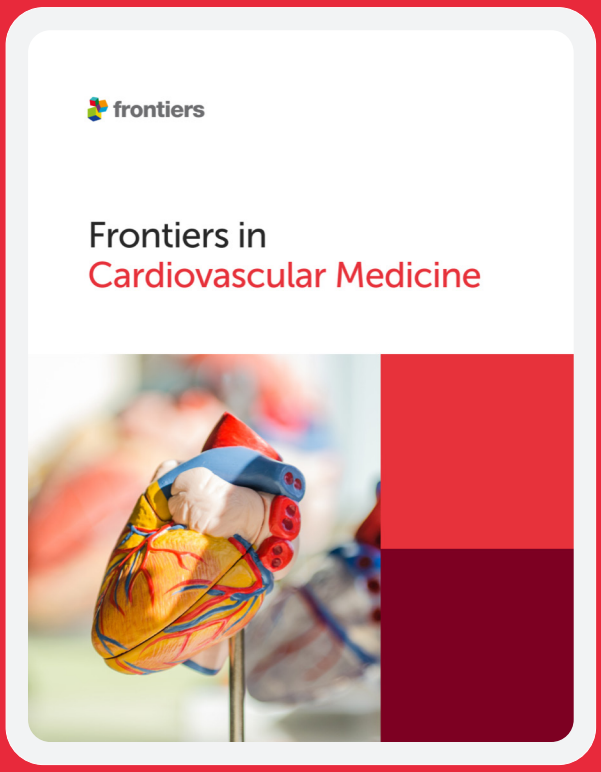
Frontiers in
Cardiovascular Medicine

Innovations and improvements in cardiovascular treatment and practice

Focuses on research that challenges the status quo of cardiovascular care, or facilitates the translation of advances into new therapies and diagnostic tools.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact