Risk Management and Healthcare Policy

Dovepress
Taylor & Francis Group

Open Access Full Text Article

ORIGINAL RESEARCH

Exploring Mortality and Prognostic Factors of Heart Failure with In-Hospital and Emergency Patients by Electronic Medical Records: A Machine Learning Approach

Cheng-Sheng Yu[1–4,*], Jenny L Wu[5,*], Chun-Ming Shih[6–8], Kuan-Lin Chiu[9], Yu-Da Chen[9,10], Tzu-Hao Chang[5,11]

[1]Graduate Institute of Data Science, College of Management, Taipei Medical University, New Taipei City, 235603, Taiwan; [2]Clinical Data Center, Office of Data Science, Taipei Medical University, New Taipei City, 235603, Taiwan; [3]Fintech Innovation Center, Nan Shan Life Insurance Co., Ltd., Taipei, 11049, Taiwan; [4]Beyond Lab, Nan Shan Life Insurance Co., Ltd., Taipei, 11049, Taiwan; [5]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, New Taipei City, 235603, Taiwan; [6]Department of Internal Medicine, School of Medicine, College of Medicine, Taipei Medical University, Taipei, 11031, Taiwan; [7]Cardiovascular Research Center, Taipei Medical University Hospital, Taipei, 11031, Taiwan; [8]Taipei Heart Institute, Taipei Medical University, Taipei, 11031, Taiwan; [9]Department of Family Medicine, Taipei Medical University Hospital, Taipei, 11031, Taiwan; [10]School of Medicine, College of Medicine, Taipei Medical University, Taipei, 11031, Taiwan; [11]Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei, 11031, Taiwan

*These authors contributed equally to this work

Correspondence: Tzu-Hao Chang; Yu-Da Chen, Email kevinchang@tmu.edu.tw; 153072@h.tmu.edu.tw

Purpose: As HF progresses into advanced HF, patients experience a poor quality of life, distressing symptoms, intensive care use, social distress, and eventual hospital death. We aimed to investigate the relationship between morality and potential prognostic factors among in-patient and emergency patients with HF.

Patients and Methods: A case series study: Data are collected from in-hospital and emergency care patients from 2014 to 2021, including their international classification of disease at admission, and laboratory data such as blood count, liver and renal functions, lipid profile, and other biochemistry from the hospital's electrical medical records. After a series of data pre-processing in the electronic medical record system, several machine learning models were used to evaluate predictions of HF mortality. The outcomes of those potential risk factors were visualized by different statistical analyses.

Results: In total, 3871 hF patients were enrolled. Logistic regression showed that intensive care unit (ICU) history within 1 week (OR: 9.765, 95% CI: 6.65, 14.34; p-value < 0.001) and prothrombin time (OR: 1.193, 95% CI: 1.098, 1.296; <0.001) were associated with mortality. Similar results were obtained when we analyzed the data using Cox regression instead of logistic regression. Random forest, support vector machine (SVM), Adaboost, and logistic regression had better overall performances with areas under the receiver operating characteristic curve (AUROCs) of >0.87. Naïve Bayes was the best in terms of both specificity and precision. With ensemble learning, age, ICU history within 1 week, and respiratory rate (BF) were the top three compelling risk factors affecting mortality due to HF. To improve the explainability of the AI models, Shapley Additive Explanations methods were also conducted.

Conclusion: Exploring HF mortality and its patterns related to clinical risk factors by machine learning models can help physicians make appropriate decisions when monitoring HF patients' health quality in the hospital.

Keywords: mortality, risk factor, cardiovascular disease, multivariate statistical analysis, machine learning, artificial intelligence

Introduction

Heart failure (HF) is a global public health priority due to its high prevalence, high costs, and poor prognoses.[1,2] According to one estimation, HF affects more than 60 million individuals worldwide.[2] It is a complex clinical syndrome caused by the malfunction or structural impairment of the ventricle which fails to fulfill the blood requirements of tissues to maintain normal functions.[3] Once a patient is hospitalized for HF, they will have higher chances of readmission and mortality.[4] As HF

Received: 3 September 2024
Accepted: 30 December 2024
Published: 9 January 2025

progresses into advanced HF, patients experience a poor quality of life (QoL), distressing symptoms, intensive care use, social distress, and eventual hospital death.[5] Moreover, the cost of HF patients is also one of the major issues not only for patients themselves but also for healthcare systems. In the United States, the average cost for an HF patient was $10,995, yet it can increase to as much by $293,575 when HF is exacerbated and requires additional therapies such as circulatory support and even heart transplantation.[6] Palliative care is patient- and family-centered care that aims to improve the QoL of patients and their families that are experiencing life-threatening illness such as HF.[5] Many studies supported the beneficial effects of palliative care interventions on patients' QoL.[7,8] It also helps reduce healthcare spending by appropriate referral to hospice care and reducing the length of stay and number of interventions near end of life.[9–13] Early palliative care intervention has also been recommended to advanced HF patients to fulfill the needs such as future care planning.[14] Therefore, a highly accurate HF mortality prediction model can help physicians introduce palliative care to patients with limited-life nature with appropriate timing to provide a better end of life and save healthcare costs.

With improvement of information technology, the electronic medical records (EMRs) system has been widely adopted in current healthcare systems for it not only could protect information confidentiality but also improve the quality of documentation and accessibility of data.[15,16] Artificial intelligence (AI) has been applied in various aspects of healthcare research.[17–23] It helps reveal underlying patterns among massive amounts of medical data such as EMRs and medical images. Through machine learning (ML), AI can approach complex clinical problems with a higher efficacy and assist physicians in improving their current practice and protocols.[24,25] This technique has also been widely utilized in HF mortality prediction models and building risk scores and has better performances than those traditional methods such as the Cox regression.[26–30] However, existing ML models require statistical methods that specify different HF subtypes.[31] Moreover, more and more HF comorbidities or coexist syndrome such as frailty and cognitive impairment have been identified as novel risk factors for poor prognosis outcomes and much research has been discussing the importance of new HF risk model.[32,33] Therefore, it is essential that to construct a new HF model with novel methods. Despite many HF models have been invented, most of them are based on the United States or European countries. HF is also an urgent public health issue in Taiwan where it was ranked second in causes of death.[34] To the best of our knowledge, there is no heart failure mortality prediction study using explainable ML model in Taiwanese population. Therefore, the aim of this study was to investigate HF mortality and related risk factors by comparing various ML algorithms including explainable ML model and statistical method-based prediction in HF patients.

Materials and Methods

Study Design

In this study, we utilized data from the Taipei Medical University (TMU) Clinical Research Database (TMUCRD). Since 2015, the database has accumulated over 4.1 million patient EMRs. EMRs contain both structured and unstructured data from a total of 3000 beds. Structured data include basic demographics, cause of death, laboratory test results, inpatient nutritional assessments, vital signs, and medical devices. Unstructured data include image examinations, physicians' notes, and radiology and pathology reports. All data were preprocessed and validated before being appended into the database and complied with all relevant data protection and privacy regulations.

The TMUCRD covers data from 1998 to 2021.[35]

Participant Inclusion and Exclusion Criteria

The HF patients included adult patients (aged ≥18 years) enrolled into one of the hospitals in the TMU system between January 2014 and December 2018 with International Classification of Disease-9 (ICD-9) and ICD-10 codes of HF incidence recorded in the EMRs. All patients will be follow-up to at most 3 years. The latest follow-up period ended in December 2021. If patients survived more than 3 years after enroll date, they will be considered as survived observations. Complete ICD-9 and ICD-10 codes are listed in Supplementary Table 1. Once patients were enrolled in the study, their basic demographic information, including age and gender, was collected. Regular laboratory test data collected within 14 days from first visit were also extracted and the average values were calculated for model building. Laboratory test items included white blood test (basophils, eosinophils, lymphocytes, monocytes, neutrophils, and white blood cell (WBC) count), serum albumin, activated partial thromboplastin time (APTT), blood urea nitrogen (BUN), creatine kinase-myoglobin-binding (CKMB) test, creatine

phosphokinase (CPK), creatinine, the estimated glomerular filtration rate (eGFR), glucose ante cibum (AC), serum glutamic oxaloacetic transaminase (GOT), serum glutamic-pyruvic transaminase (GPT), hematocrit (HCT), hemoglobin (HGB), serum potassium (K), mean corpuscular hemoglobin (MCH), MCH concentration (MCHC), mean corpuscular volume (MCV), serum sodium (Na), platelet count (PLT), red blood cell (RBC) count, prothrombin time and international normalized ratio (PT_INR), and cardiac troponin I (troponin I). Moreover, patients' vital signs, past medical history (ICU visits and comorbidities) were included in model construction (Figure 1). The ICU history contained information on the ICU visit history of a patient from the past 1 year (ICU 1yr) from the index date to the future 1 week from the index date (ICU 1week). Comorbidities were included, and ICD-9 codes used for identification were based on the Charlson Comorbidity Index (CCI).[36] Patients with only outpatient visit records and sepsis incidence in EMRs were not included in the study. The complete patient collection process and timeline are illustrated in Figures 1 and 2. Before machine learning techniques were included, data were divided into training and validation sets with 80% and 20%, respectively. Both datasets are independent when the performance of machine learning algorithms is evaluated.

Missing Values

Data with missing values were regulated by the K-Nearest Neighbor (KNN) algorithm because of its robustness and straightforward practice in high-dimensional distribution spaces.[37]

While some amount of missing data is expected, missing data reduce the power of databases. However, such cannot eliminate the potential bias. More attention should be paid to the missing data in the design and performance of the studies and in the analysis of the resulting data. Since the potential bias cannot be eliminated even a well-developed

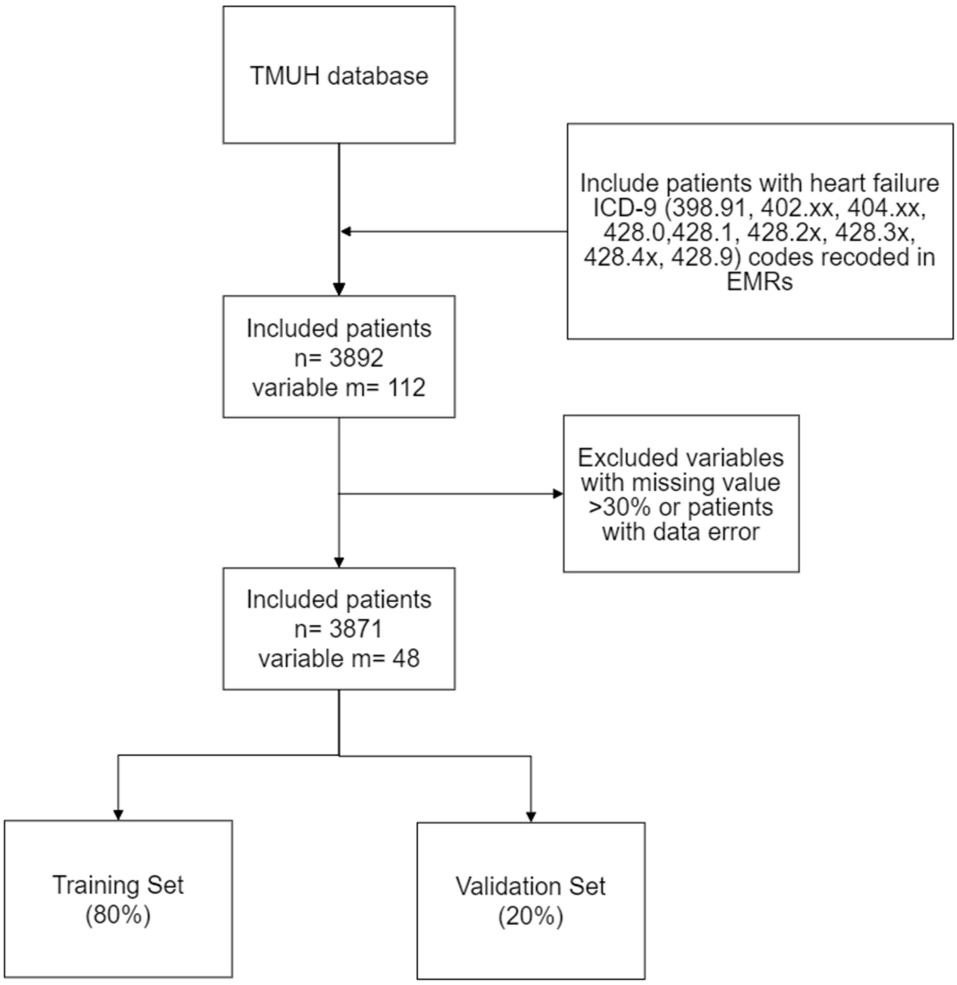

Figure 1 A flowchart of the step-by-step procedure from collection and pre-processing of heart failure electronic medical records database to machine learning datasets.

https://doi.org/10.2147/RMHP.S488159

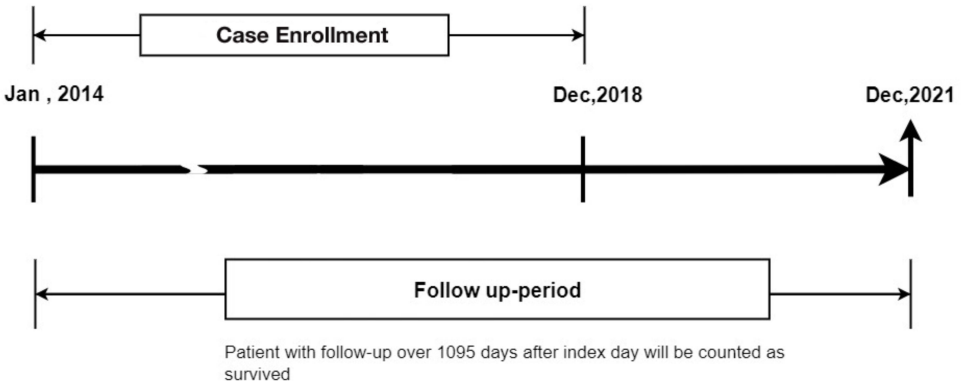

Figure 2 Patient enrollment and follow-up period timeline.

missing value imputation method included in the study, more attention should be paid to the missing data in the study design and analysis.[38]

After deleting variables with missing values of >30% of the sample size, 48 variables were retained. The complete list of variables has been added in supplement as Supplementary table 2. The pre-processing procedure is illustrated in Figure 1.

Data Cleaning and Statistical Analysis

Serial data pre-processing was conducted by SAS Enterprise Guide 8.3 (SAS, Cary, NC, USA). Both in-hospital and emergency medical records were combined and summarized by year into an analyzable medical database. Machine learning algorithms and statistical analyses were conducted using R (vers. 4.2.2) or SPSS (vers. 18.0) software (SPSS, IBM, Armonk, NY, USA[39]). Baseline characteristics of enrollees were described, and the p value denotes comparison between surviving and dead HF participants. Categorical variables were examined using the Chi-squared test, while the nonparametric Mann–Whitney U-test was applied to continuous variables by comparing the rank of median values. A multivariate logistic regression and Cox regression were both conducted to assess the significance of risk factors with a stepwise procedure. In all statistical analysis, p values denote whether the variable was statistically significant at <0.05, which was accepted as statistical significance.

Machine Learning Techniques

Several types of machine learning models are presented to predict mortality due to HF. The typical statistical learning model, logistic regression, can classify patients based on multiple predictive risk factors. Naïve Bayes is a non-linear classification algorithm using Bayes' Theorem with an independent definition among risk factors. The support vector machine (SVM) can efficiently perform both non-linear and linear data classification since the SVM algorithm finds the hyperplane that best separates two groups of patients among a high-dimensional space constructed by the risk vari ables.[40]

In statistical theorem, let $p(C_k|x_1,\ldots,x_n)$ for each of the K possible outcomes or classes C_k given a problem instance to be classified, represented by a vector $\mathbf{x} = (x_1,\ldots,x_n)$ encoding some n features in Naïve Bayes classifiers. Using Bayes' theorem, the conditional probability can be decomposed as follows:

$$p(C_k|\mathbf{x}) = \frac{p(C_k)p(\mathbf{x}|C_k)}{p(\mathbf{x})} \tag{1}$$

The corresponding classifier, a Bayes classifier, is the

$$\hat{y} = \underset{k \in \{1,\ldots,.K\}}{argmax} \prod_{i=1}^{n} p(x_i|C_k). \tag{2}$$

A hinge loss function is as following:

$$\max(0, 1 - \mathrm{y_i}(\mathrm{W^T x_i} - \mathrm{b})). \tag{3}$$

while y_i is the i-th target (ie, in this case, 1 or −1), and the $\left(\mathrm{W^T x}_i - \mathrm{b}\right)$ is the i-th output.

In SVM, the optimization is to minimize

$$\lambda||\mathrm{w}|| = \frac{1}{n}\sum_{i=1}^{n} \max\left(0, 1 - y_i\left(\mathrm{W^T x_i} - \mathrm{b}\right)\right) \tag{4}$$

where the parameter λ>0 determines the trade-off between increasing the margin size and ensuring that the $\mathrm{x}_i$lie on the correct side of the margin.

A decision tree is a single classification model with entropy or information gain as its classifying technique to depict a clear tree-based inductive algorithm to differentiate two classes of patients. In this study, classification and regression trees (CARTs) used the Gini index as a display of information gain, while C5.0, derived from the Iterative Dichotomiser 3 and C4.5, uses entropy to divide those individuals at each branch into different groups of leaves.[41,42]

In decision tree, the Gini impurity is computed by summing pairwise products of these probabilities for each class label:

$$\mathrm{I}_G(p) = \sum_{i=1}^{J}(p_i \sum_{k\neq i} p_k) = 1 - \sum_{i=1}^{J}\left(p_i^2\right). \tag{5}$$

for a set of items with J classes and relative frequencies $p_i, i \in \{1, 2, \ldots, J\}$,, the probability of choosing an item with label i is p_i

While information gain is based on the concept of entropy and information content from information theory. Entropy is defined as below

$$\mathrm{H}(T) = \mathrm{I}_E(p_1, p_2, \ldots, p_J) = -\sum_{i=1}^{J} p_i \log_2 p_i \tag{6}$$

where $p_1, p_2, \ldots$ are fractions that add up to 1 and represent the percentage of each class present in the child node that results from a split in the tree.

$$IG = -\sum_{i=1}^{J} p_i \log_2 p_i - \sum_{i=1}^{J} \Pr(i|a) \log_2 p_i \Pr(i|a) \tag{7}$$

Ensemble learning aggregates multiple classifiers to enhance the overall predictive performance. Two types of ensemble algorithm were applied here. Random forest uses a parallel ensemble method for classification, while Adaptive boosting (AdaBoost) adopts a sequential ensemble method that trains a base learner in series. Both algorithms enable a better prediction than a single model.[43,44]

With ensemble learning techniques, after training procedure, the predictions of random forest for unseen samples x' can be made by averaging the predictions from all the individual regression trees on x'

$$\hat{f} = \frac{1}{B}\sum_{b=1}^{B} f_b(\, x') \tag{8}$$

where B is the bagging times repeatedly, and f_b is the regression tree. Meanwhile, Adaboost uses weak learner to produce an output hypothesis $\boldsymbol{h}$ which fixes a prediction $h(x_i)$ for each sample in the training set. At each iteration $\boldsymbol{t}$, a weak learner is selected and assigned a coefficient α_t such that the total training error E_t of the resulting $\boldsymbol{t}$-stage boosted classifier is minimized.

$$E_t = \sum_i E[F_{t-1}(x_i) + \alpha_t h(x_i)] \tag{9}$$

Here $F_{t-1}(x)$ is the boosted classifier that has been built up to the previous stage of training and $f_t(x) = \alpha_t h(x)$ is the weak learner that is being considered for addition to the final classifier.

https://doi.org/10.2147/RMHP.S488159

All these machine learning algorithms applied to this study had been executed under R programming. The details of each algorithm and its related package are shown in Supplementary table 3, including their hyperparameters and function settings.

Evaluation

Receiver operating characteristic (ROC) curves were used to illustrate the diagnostic ability of machine learning classification. Several criteria such as the accuracy, sensitivity, specificity, F1-score, precision, and area under the ROC curve (AUROC) are shown in both a table and a graphical plot.[18] SHAP (SHapley Additive exPlanations) will also be used to evaluate ensemble learning machine learning model performance. It is derived from cooperative game theory's Shapley value proposed by Lloyd Stowell Shapley. By assuming each data point as a player in game and the prediction as the payout, Shapley value would then show the solution of fair distribution of payoffs (prediction) to all players (data points).[29,45]

Results

In total, 3871 hF patients were enrolled after serial data pre-processing as illustrated in Figure 1. After data cleaning, baseline characteristics of the enrollees were described, and the *p* value denotes a comparison between surviving and dead HF participants in Table 1 (continuous variables) and Table 2 (categorical variables). Table 1 describes clinical risk biomarkers, and most of them were significant with a *p* value of <0.001 except for CKMB and MCH. Glucose AC was also significant with a *p* value of 0.038, but it was not extremely significant. Statistical outcomes of clinical factors and comorbidities with mortality due to HF are presented in Table 2. The ICU history within 1 week and some comorbidities were extremely significant with *p* values of <0.001.

Both categorical and continuous variables were analyzed by a multivariate logistic regression and Cox regression. Biomarkers that had prominent effects on mortality due to HF are listed in Table 3. In Table 3, all risk factors were

Table 1 Descriptive Statistics and Hypothesis Testing of Biomarkers in the Electronic Medical Record (EMR) Database with Mortality Due to Heart Failure

	Cases	Survival	Death	p-value
	n= 3871	n= 1971	n= 1900	
	Median (IQR)	Median (IQR)	Median (IQR)	
Age, year-old	77 (65–86)	70 (60–80)	83 (74–89)	<0.001**
Basophil, %	0.5 (0.3–0.7)	0.6 (0.4–0.8)	0.4 (0.2667–0.6)	<0.001**
Eosinophil, %	1.5 (0.7–2.7)	1.8 (1–2.971)	1.1 (0.4667–2.2542)	<0.001**
Lymphocyte, %	15.9 (9.9–22.8)	19.75 (13.992–26.4)	11.912 (7.644–17.821)	<0.001**
Monocyte, %	7.96 (6.3–9.8)	8.3 (6.9–10)	7.5 (5.675–9.518)	<0.001**
Neutrophil, %	72.8 (64.7–80.4)	68.35 (60.66–74.97)	77.7 (70.1–83.85)	<0.001**
Albumin, g/dL	3.4 (3.2–3.5)	3.456 (3.365–3.6)	3.3 (3–3.5)	<0.001**
APTT, sec	40.6 (35.7–45.5)	38.7 (35–43.53)	42.1 (37.4–48.09)	<0.001**
BUN, mg/dL	27.26 (17.71–44.99)	21.13 (15.4–32.09)	36.5 (23.8–56.02)	<0.001**
CKMB, ng/mL	20.07 (16.6–26.40)	20.06 (17.16–25)	20.07 (16–28.23)	0.565
CPK, U/L	126 (71.94–201.07)	145.3 (91–202.7)	103.8 (56–199.5)	<0.001**
Creatinine, mg/dL	1.1217 (0.9–2.333)	1.05 (0.8–1.6)	1.569 (0.9857–3.25)	<0.001**
CRP, mg/dL	3.48 (1.852–6.101)	2.998 (1.71–4.376)	4.733 (2.009–8.264)	<0.001**
eGFR, mL/min/1.73 m^2	55.67 (27.5–82)	68 (41.55–88)	41.75 (18.84–71)	<0.001**
Glucose AC, mg/dL	142.4 (118–165.3)	143.5 (114–165.7)	141.6 (121.7–165)	0.038†
GOT, IU/L	32.5 (21–53.66)	29 (20–47.76)	35.5 (22.5–64.24)	<0.001**
GPT, IU/L	22 (14.5–37.76)	21 (15–33)	24 (14–42.81)	<0.001**
HCT, %	33.4 (28.66–38.6)	36.5 (31.91–40.9)	30.5 (26.97–35)	<0.001**
HGB, g/dL	11.27 (9.65–13.1)	12.4 (10.8–13.95)	10.25 (9.125–11.736)	<0.001**
K, mEq/L	4.043 (3.7–4.42)	4 (3.7–4.35)	4.067 (3.7–4.516)	0.001†

(*Continued*)

Table 1 (Continued).

	Cases	Survival	Death	p-value
	n= 3871	n= 1971	n= 1900	
	Median (IQR)	Median (IQR)	Median (IQR)	
MCH, pg	30.7 (29.13–32.19)	30.7 (29.16–32.05)	30.71 (29.1–32.35)	0.196
MCHC, g/dL	33.88 (33.2–34.47)	34.05 (33.4–34.6)	33.7 (33–34.3)	<0.001**
MCV, fL	90.3 (86.4–94.2)	89.9 (86.02–93.4)	90.87 (86.8–95.4)	<0.001**
Na, mEq/L	138.2 (135.5–140.7)	139 (136.6–140.7)	137.5 (134–140.6)	<0.001**
PLT, $x10^3$ /uL	188 (142.6–237.1)	200.5 (161.67–244.83)	171 (126.9–225.2)	<0.001**
Prothrombin Time INR	1.17 (1.050–1.343)	1.1 (1.01–1.25)	1.23 (1.11–1.417)	<0.001**
RBC, $x10^6$/uL	3.759 (3.19–4.39)	4.12 (3.585–4.63)	3.408 (2.997–3.928)	<0.001**
Troponin_I, ng/mL	0.0765 (0.03832–0.154)	0.06173 (0.028–0.1151)	0.09779 (0.05–0.21915)	<0.001**
WBC, $x10^3$/uL	8.11 (6.31–10.393)	7.57 (6.077–9.345)	8.813 (6.694–11.55)	<0.001**
BF, time/min	18.294 (17.565–19.5)	17.92 (17.38–18.55)	19 (17.952–20.625)	<0.001**

Note: ** Indicates $p < 0.001$, † indicates $p < 0.05$.

Abbreviations: APTT, activated partial thromboplastin time; BUN, blood urea nitrogen; CKMB, creatine kinase-myoglobin binding test; CPK, creatine phosphokinase; eGFR, creatinine, estimated glomerular filtration rate; Glucose AC, glucose ante cibum; GOT, serum glutamic oxaloacetic transaminase; GPT, serum glutamic-pyruvic transaminase; HCT, hematocrit; HGB, hemoglobin; K, serum potassium; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; Na, serum sodium; PLT, platelet count; RBC, red blood cell count; PT_INR, prothrombin time and international normalized ratio; Troponin_I, cardiac troponin I; WBC, white blood cell count; BF, respiratory rate.

Table 2 Statistics of Comorbidities with Mortality Due to Heart Failure

Factors	Survival	Death	p-value
	n_1 = 1971 No. (%)	n_2 = 1900 No. (%)	
Sex			
Female	877(44.5%)	935(49.2%)	0.003†
Male	1094(55.5%)	965(50.8%)	
ICU 1yr			
No	1940(98.4%)	1830(96.3%)	<0.001**
Yes	31(1.6%)	70(3.7%)	
ICU 1week			
No	1928(97.8%)	1376(72.4%)	<0.001**
Yes	43(2.2%)	524(27.6%)	
Cerebrovascular Disease			
No	1880(95.4%)	1661(87.4%)	<0.001**
Yes	91(4.6%)	239(12.6%)	
Peripheral Vascular Disease			
No	1969(99.9%)	1894(99.7%)	0.172
Yes	2(0.01%)	6(0.3%)	
Dementia Disease			
No	1964(99.6%)	1845(97.1%)	<0.001**
Yes	7(0.4%)	55(2.9%)	
Chronic Pulmonary Disease			
No	1912(97%)	1738(91.5%)	<0.001**
Yes	59(3%)	162(8.5%)	
Rheumatic Disease			
No	1966(99.7%)	1887(99.5%)	0.059
Yes	5(0.3%)	13(0.5%)	

(*Continued*)

https://doi.org/10.2147/RMHP.S488159

Table 2 (Continued).

Factors	Survival	Death	p-value
	n_1 = 1971 No. (%)	n_2 = 1900 No. (%)	
Peptic Ulcer Disease			
No	1943(98.6%)	1860(97.9%)	0.113
Yes	28(1.4%)	40(2.1%)	
Mild Liver Disease			
No	1962(99.5%)	1876(98.7%)	0.008†
Yes	9(0.5%)	24(1.3%)	
Diabetes without Chronic Disease			
No	1832(92.9%)	1655(87.1%)	<0.001**
Yes	139(7.1%)	245(12.9%)	
Diabetes with Chronic Disease			
No	1970(99.9%)	1889(99.4%)	0.003†
Yes	1(0.01%)	11(0.6%)	
Hemiplegia Paraplegia Disease			
No	1969(99.9%)	1898(99.9%)	0.999
Yes	2(0.1%)	2(0.01%)	
Malignancy Disease			
No	1941(98.5%)	1757(92.5%)	<0.001**
Yes	30(1.5%)	143(7.5%)	
Moderate or Severe Liver Disease			
No	1970(99.9%)	1892(99.6%)	0.019†
Yes	1(0.01%)	8(0.04%)	
HIV Infection Disease			
No	1970(99.9%)	1895(99.7%)	0.118
Yes	1(0.01%)	5(0.03%)	
AIDS Disease			
No	1967(99.8%)	1868(98.3%)	<0.001**
Yes	4(0.02%)	32(1.7%)	
Renal Disease			
No	1831(92.9%)	1567(82.5%)	<0.001**
Yes	140(7.1%)	333(17.5%)	

Notes: ** Indicates $p < 0.001$, † indicates $p < 0.05$. ICU 1yr, whether or not a patient had been admitted to the ICU within 1 year before the index date; ICU 1week, whether or not a patient had been admitted to the ICU within 1 week after the index date.

significant with p values of <0.05, while the odds ratios (ORs) of some of the factors were paramount with ORs of >10, such as diabetes with chronic disease, moderate or severe liver disease, and human immunodeficiency virus (HIV) infectious disease. In the Cox regression, hazard ratios (HRs) were not immensely variant with 10-fold differences as the ORs in Table 3, but PT_INR, diabetes without chronic disease, malignancy disease, moderate or severe liver disease, renal disease, ICU 1week, and BF had positive HRs of >1.1 as shown in Table 3.

Figures 3–6 and Table 4 are data visualizations and provide the overall performance of machine learning predictions. For every machine learning model, the efficiency of different criteria is listed in Table 4. Random forest, SVM, Adaboost, and logistic regression had better overall performances with AUROC values of >0.87 as shown in Figure 3, while both decision tree models were not powerful in predicting mortality due to HF. Both ensemble learning algorithms had better performance on F1-scores, and the random forest had the highest value of sensitivity. In addition, Naïve Bayes was the best in terms of both specificity and precision.

In Figure 4, the ranking of all risk factors for HF mortality between both ensemble learning algorithms are depicted. Age, ICU history within 1 week, and BF were the top three compelling risk factors of mortality due to HF. In Figure 5,

Table 3 Important Markers in the Stepwise Multivariate Logistic Regression and Stepwise Cox Regression Analyses for Mortality Due to Heart Failure

a. Stepwise Multivariate Logistic Regression			
Factors	p-value	OR	95% CI of OR
Sex	0.002†	1.335	(1.115, 1.599)
Age, year-old	<0.001**	1.058	(1.051, 1.066)
Neutrophil, %	<0.001**	1.026	(1.017, 1.035)
Albumin, g/dL	<0.001**	0.397	(0.311, 0.505)
BUN, mg/dL	<0.001**	1.021	(1.015, 1.026)
CKMB, ng/mL	<0.001**	1.006	(1.004, 1.009)
CPK, U/L	0.016†	1	(1, 1)
Glucose AC, mg/dL	0.027†	1.002	(1, 1.004)
HGB, %	<0.001**	0.657	(0.521, 0.828)
K, mEq/L	0.018†	0.812	(0.683, 0.965)
MCHC, g/dL	<0.001**	0.698	(0.612, 0.795)
MCV, fL	<0.001**	1.082	(1.048, 1.117)
Na, mEq/L	<0.001**	0.924	(0.905, 0.943)
Prothrombin Time INR	<0.001**	1.589	(1.262, 2.002)
RBC, $x10^6$/uL	0.029†	2.149	(1.08, 4.277)
Cerebrovascular Disease	<0.001**	2.137	(1.545, 2.955)
Diabetes with Chronic Disease	0.066	10.027	(0.861, 116.845)
Malignancy Disease	<0.001**	2.588	(1.604, 4.176)
Moderate or Severe Liver Disease	0.054	11.025	(0.963, 126.167)
HIV Infection Disease	0.038†	20.343	(1.188, 348.224)
Renal Disease	0.002†	1.628	(1.205, 2.2)
ICU 1week	<0.001**	9.765	(6.65, 14.341)
BF	<0.001**	1.236	(1.176, 1.3)
b. Stepwise Cox Regression			
Factors	p-value	HR	95% CI of HR
Age, year-old	<0.001**	1.03	(1.026, 1.034)
Neutrophil, %	<0.001**	1.022	(1.016, 1.027)
Albumin, g/dL	<0.001**	0.635	(0.566, 0.713)
BUN, mg/dL	<0.001**	1.008	(1.006, 1.01)
CKMB, ng/mL	<0.001**	1.003	(1.002, 1.003)
HGB, %	<0.001**	0.902	(0.878, 0.926)
MCHC, g/dL	<0.001**	0.788	(0.753, 0.826)
MCV, fL	<0.001**	1.026	(1.02, 1.032)
Na, mEq/L	<0.001**	0.96	(0.952, 0.969)
PLT, $x10^3$ /uL	0.001†	0.999	(0.998, 1)
Prothrombin Time INR	<0.001**	1.193	(1.098, 1.296)
Diabetes Without Chronic Disease	0.025†	1.171	(1.02, 1.345)
Malignancy Disease	<0.001**	1.858	(1.563, 2.208)
Moderate or Severe Liver Disease	0.007†	2.646	(1.312, 5.337)
Renal Disease	0.046†	1.145	(1.002, 1.307)
ICU 1week	<0.001**	2.679	(2.386, 3.007)
BF, time/min	<0.001**	1.135	(1.116, 1.154)

Notes: ** Indicates $p < 0.001$, † indicates $p < 0.05$. ICU 1week, whether or not a patient had been admitted to the ICU within 1 week after the index date; OR, odds ratio.

Abbreviations: HR, hazard ratio; BF, respiratory rate; BUN, blood urea nitrogen; CKMB, creatine kinase-myoglobin binding test; CPK, creatine phosphokinase; Glucose AC, glucose ante cibum; HGB, hemoglobin; INR, international normalized ratio; K, serum potassium; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; Na, serum sodium; PLT, platelet count; RBC, red blood cell count.

https://doi.org/10.2147/RMHP.S488159

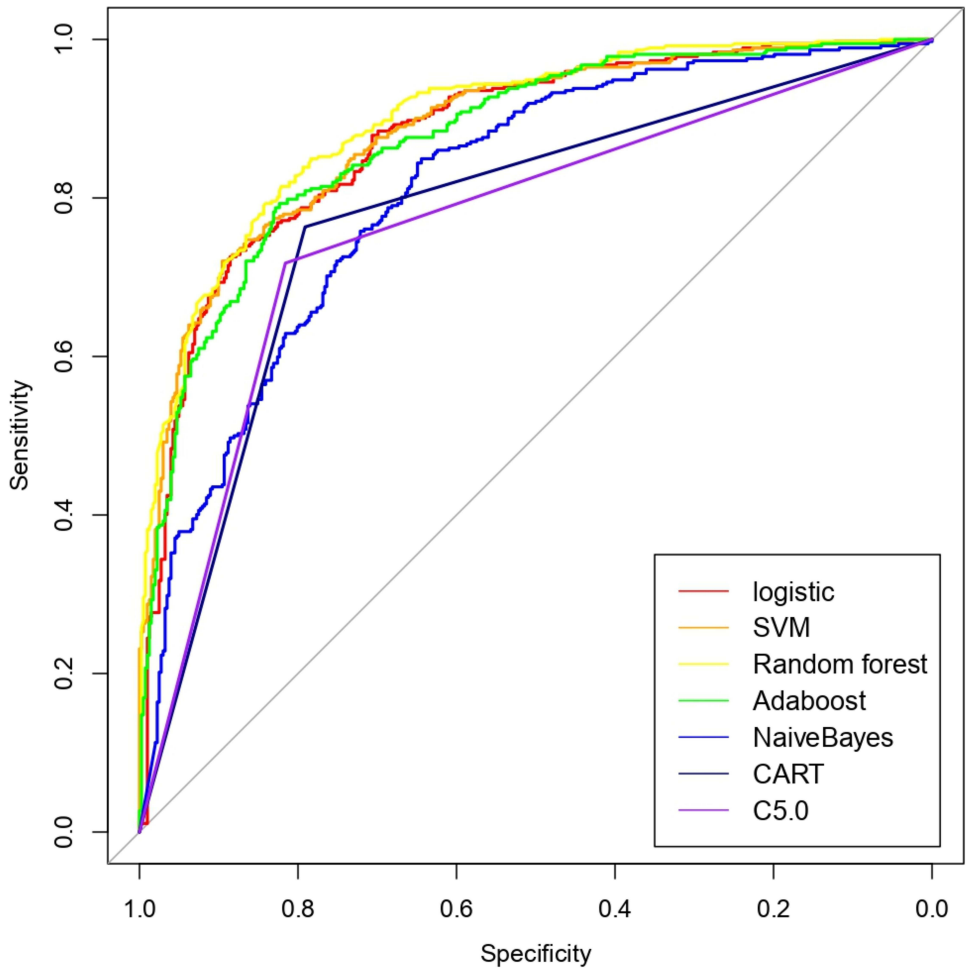

Figure 3 Area under the receiver operating characteristic (ROC) curve of machine learning models.

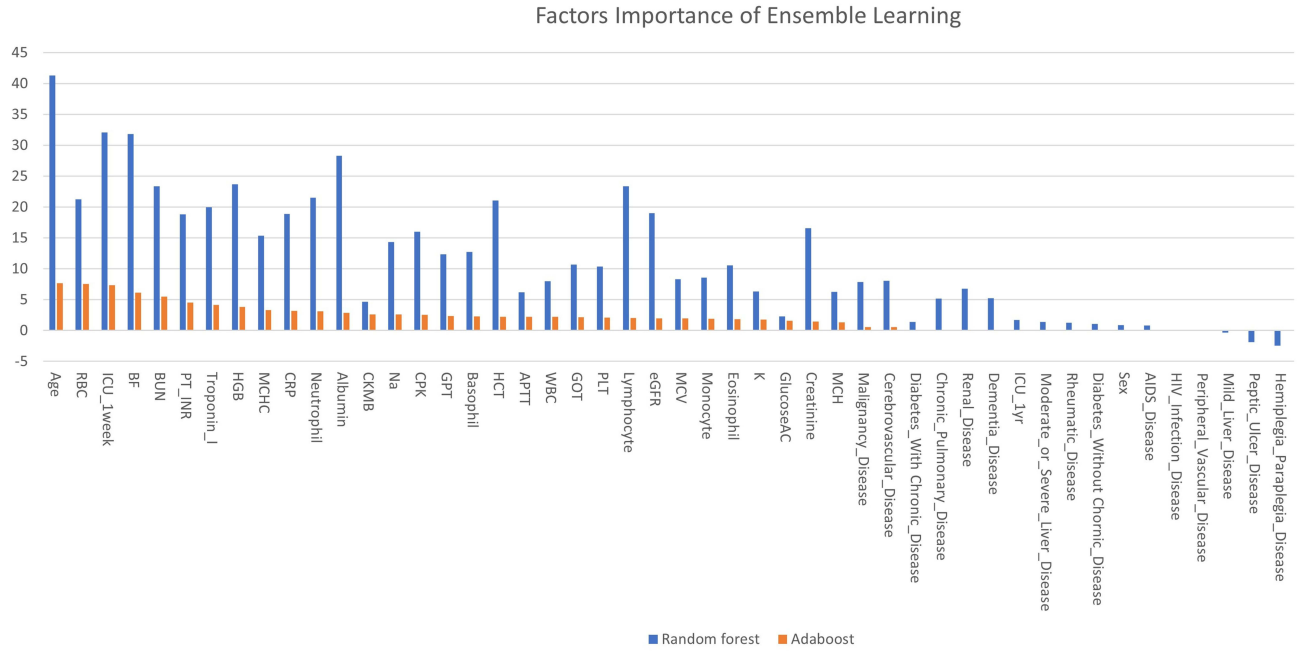

Figure 4 The variable importance of ensemble learning.

https://doi.org/10.2147/RMHP.S488159

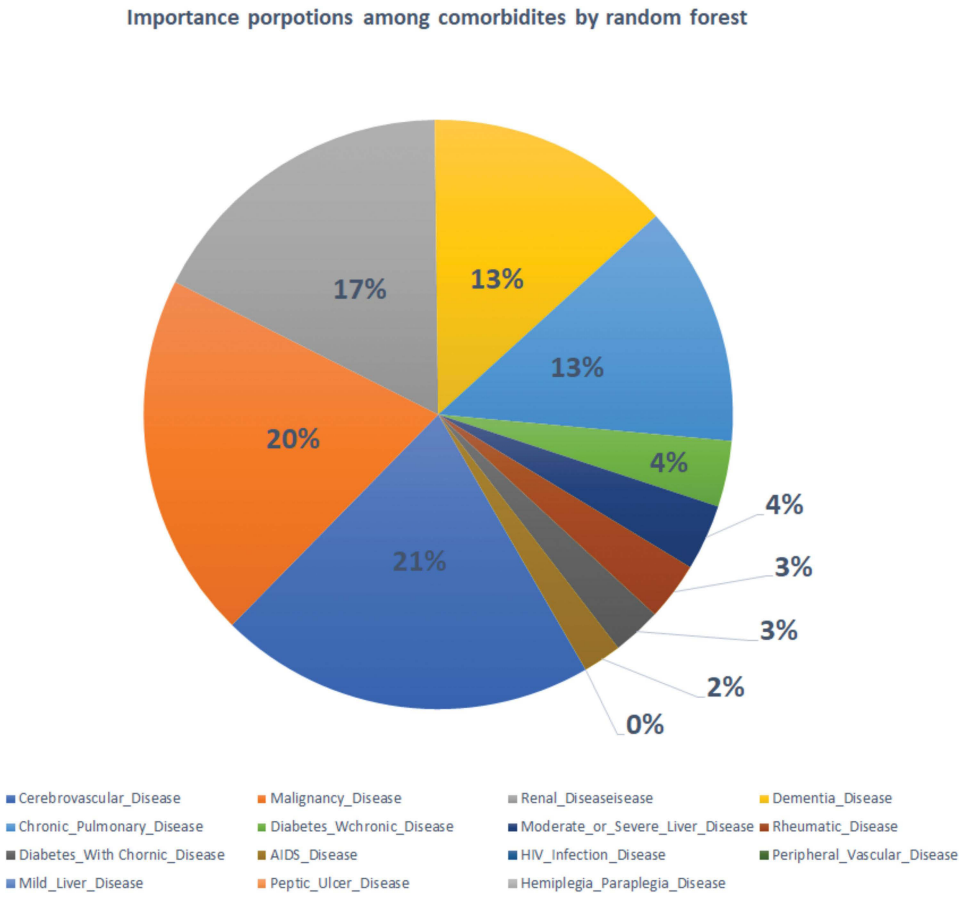

Figure 5 Pie chart of the variable importance of the random forest for comorbidities only.

proportions of comorbidity importance by the random forest are presented in a pie chart. Cerebrovascular disease and malignant disease were both >20%, while HIV infection, peripheral vascular disease, mild liver disease, peptic ulcer disease, and hemiplegia paraplegia disease had negligible predictive importance with nearly 0% proportion in the random forest. In Figure 6, proportions of other risk biomarkers without comorbidities are shown. In addition to the top three risk factors mentioned above, albumin, HGB, and RBC were also influential in ensemble learning predictions. An explanation of ensemble learning method by SHAP methods is depicted in Figure 7. And more explanations for the sampling are provided in Supplement Figure 1. Briefly, the results of SHAP methods also show that ICU history within 1 week, age, and albumin were the top three compelling risk factors of mortality due to HF.

Discussion

First of all, the overall performance of predicting HF mortality was effective and splendid with machine learning. From those data visualization outcomes, some clinical factors such as age, RBC, ICU history within 1 week, BF, and BUN were more powerful than comorbidities, even though some comorbidities had strong ORs in the traditional regression analysis. Second, different machine learning models had their pros and cons, but no one model had the best performance for each criterion in the ROC curves. Third, significant risk factors, which appeared in both traditional statistical analysis and machine learning predictions, were indicative risk factors as either clinical biomarkers or comorbidities. Finally, physicians can focus on those outcomes to monitor HF patients' health care in the hospital. A suggestion for palliative care may be considered since patients may survive for more than 30 days, but those principal risk factors will not improve or possibly change like age or comorbidity records. Both the ranking of variable importance and explanation of SHAP method show that ICU history within 1 week, BF and albumin are the top risk factors, while age is also important with a high negative phi value in SHAP.

https://doi.org/10.2147/RMHP.S488159

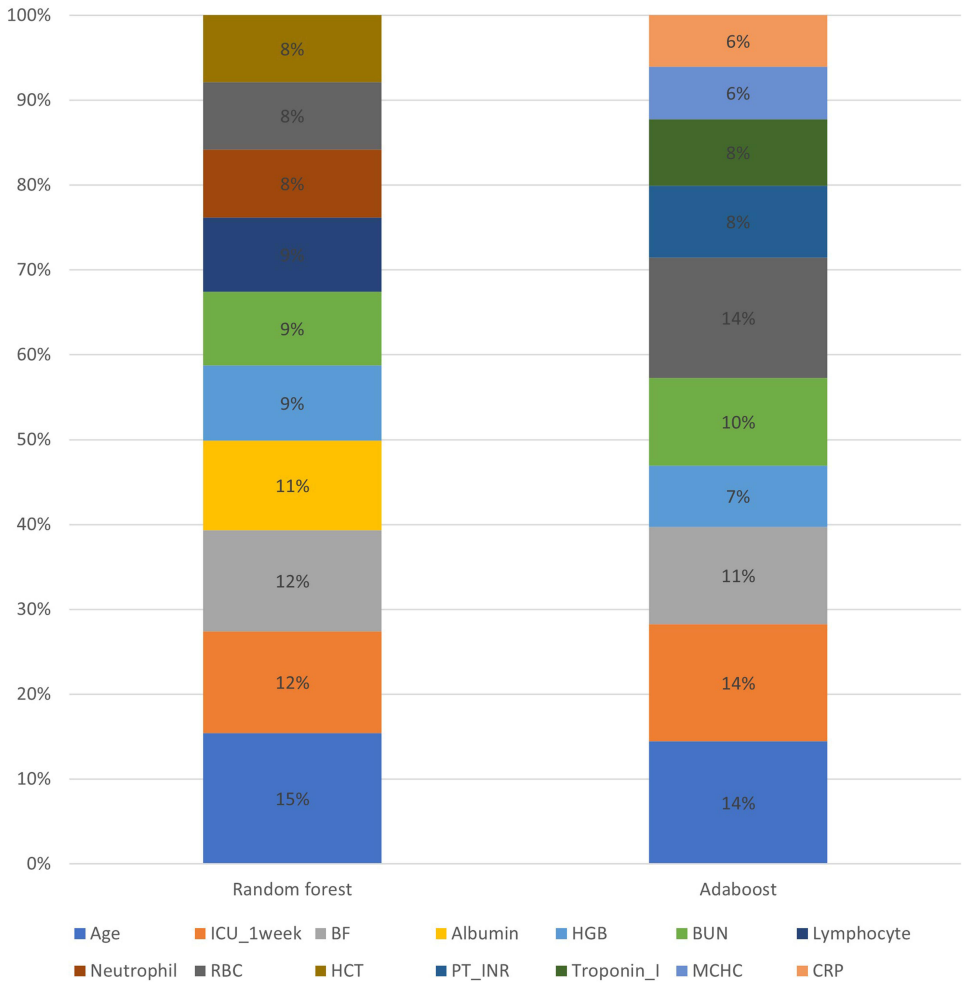

Figure 6 The top 10 important variables with proportions among the ten factors for each ensemble learning model.

Furthermore, there has been research on outcomes of age stratification among HF patients. Although aging is not a direct factor in the progression or causation of HF, its related factors, such as hypertension and age-related deterioration of cardiac functions such as hypertrophy of smooth muscles and fragmentation of internal elastic thick membrane in the arterial walls, amplify the HF risk in older adults.[46–48] Moreover, hypoalbuminemia is commonly observed in HF patients, especially in elderly groups and is recognized as an independent predictor for cardiovascular mortality.[49,50] One possible explanation for abnormal albumin levels among HF patients is a result of comorbidities, such as cachexia or

Table 4 Performances of Different Machine Learning Models on Predicting Mortality Due to Heart Failure

Model	Accuracy	Sensitivity	Specificity	F1-Score	Precision	AUC
Logistic	0.7972	0.7994	0.7952	0.7852	0.7715	0.8778
Naïve Bayes	0.6744	0.3790	0.9478	0.5281	0.8704	0.8111
SVM	0.8036	0.7634	0.8408	0.7889	0.8161	0.8852
CART	0.7778	0.7634	0.7910	0.7676	0.7717	0.7772
C5.0	0.7687	0.7177	0.8159	0.7489	0.7830	0.7668
Adaboost	0.8075	0.7903	0.8234	0.7972	0.8055	0.8741
Random Forest	0.8165	0.8118	0.8209	0.8097	0.8075	0.898

Abbreviation: AUC, area under the receiver operating characteristic curve.

https://doi.org/10.2147/RMHP.S488159

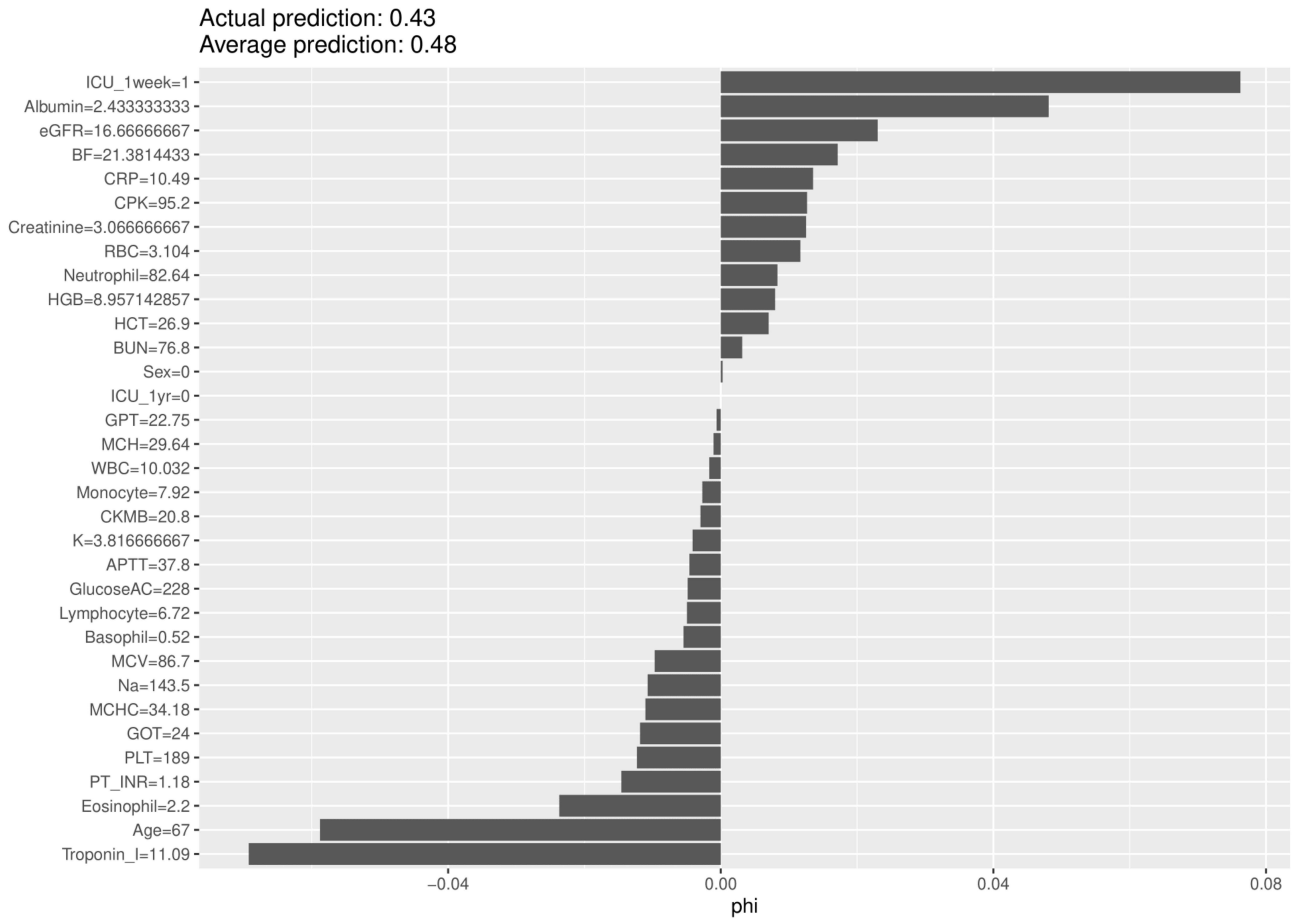

Figure 7 A visualization of output explanation by SHAP method.

liver dysfunction, that decrease albumin synthesis.[51–54] Besides liver-related diseases, renal impairment is also frequently observed in HF patients and was related to increased probabilities of morbidity and mortality.[55,56] Therefore, indicators of renal function, such as eGFR and BUN, also had great prognostic power in our predictive model.[57] Another significant factor in our model was the respiration rate (BF), which may have been due to changes in respiratory function during HF progression.[58] As the heart and lungs coexist in the same enclosed thoracic cavity, they are intimately linked.[59] When the cardiovascular system fails during HF exacerbation, the lungs' fluid balance is disturbed. In response, the respiratory system has to remodel by increasing ventilatory demands or the respiratory rate for compensation.[60] PT_INR, as a well-validated indicator for measuring coagulation abnormalities, was also found to be significant in the HF mortality assessment. Many studies showed that HF can initiate coagulation processes which increase the chances of thromboembolic events such as stroke or pulmonary embolisms.[61,62] Declines of coagulation factor concentrations were also observed in severe chronic HF patients, which indicates an increase in risk of adverse events.[63]

Moreover, cardiac markers (CKMB, CPK, and troponin I) for myocardial infraction were also shown to be highly predictive for HF mortality. Some studies showed the relationship between increasing troponin levels and the risk of mortality in cardiac patients.[64,65] A possible mechanism for elevated troponin in HF patients may be myocardial damage caused by subclinical cardiac events such as left ventricular remodeling, subendocardial ischemia, or coronary microvascular dysfunction.[66,67] Like troponin-I, elevation of CK and CK-MB levels showed strong relations with mortality in cardiac patients. One study suggested that CK-MB and cardiac troponin I levels were correlated with the severity of HF, for they reflect the progression of myocardial failure.[68] Two WBC-related metrics reflect an inflammatory condition, viz., lymphocytes and neutrophils, which were included in our models. This result was consistent with those from previous studies which indicated that inflammatory factors can reflect HF progression.[69,70] In addition to WBC measurements, two

https://doi.org/10.2147/RMHP.S488159

RBC indices of HGB and HCT had clinical significance in identifying anemia, which is commonly comorbid with HF.[71,72] Studies indicated that anemia is independently related to the risk stratification of mortality in both acute and chronic HF.[73–75] Therefore, HGB and HCT may have been important in model feature selection due to the relationship between anemia and HF severity. Contrary to previous studies, RBCs were presented as a significant indicator in our predictive model.[76] A potential explanation may be side effects of chronic kidney disease, which is also commonly comorbid with HF. The most significant predictive variable in the model was ICU-1week which is reasonable, for it is a direct sign of severe exacerbation of HF progression.

One important discovery about disadvantages of EMR database is as follows. The data of comorbidities are according to the ICD9 and 10 recorded in patients' EMRs before their first enrollment time in this study. This may have underestimation of comorbidities prevalence in the study. While the prevalence of diabetes mellitus in this study is about 10.2%, which is the lower bound of the general prevalence reported in previous heart failure studies from 10% to 47%. But the prevalence of DM is higher in patients hospitalized with HF, with some reports of >40%.[77] This is the bias and drawbacks in the Taiwan National Health Insurance Database because electronic health databases may encounter coding errors and intentional "upcoding", even when data of comorbidities are according to the ICD9 and 10 recorded in patients' EMRs before their first enrollment time. For example, healthcare providers may upcode diagnoses to more severe ones to prevent reimbursement refusal by the national health insurance. Misclassification bias may occur if the diagnosis code for heart failure has not been properly validated. This is a defect if heart failure patients were identified using ICD codes by EMR databases.

Limitation

There are also several limitations in this study. First of all, some HF indicators, such as the ejection fraction, were not included in the study due to data restriction of IRB. The number of patients with ejection fraction (EF) measured in total population is only 643/3871. The general statistic about HF patients with EF measurement is described in Supplementary Figure 2. The percentage of subgroup patients with HFrEF, HFmrEF, and HFpEF is 28%, 16% and 56%, respectively. The mortality of each EF subgroup is 20%, 14% and 17%, respectively. However, only a quarter of patients with EF measured are classified as HFrEF subgroup. Moreover, the difference in mortality among these subgroups is less than 6%. Our study focuses on the general HF patients instead of the specific HF patients. Secondly, there are limitations inherent in the use of EMRs. Variables with missing value >30% are excluded, and these missing data can be due to a lack of collection or a lack of documentation. Having more variables can add to the predictive power of the model. In addition, patients included in this study are those who visit either inpatient or emergency department due to HF according to their EMRs, and patients containing only outpatients' records are excluded. While the data of comorbidities are according to the ICD9 and 10 recorded in patients' EMRs before their first enrollment time in this study, this may have underestimation of comorbidities prevalence in the study. The prevalence of diabetes mellitus in this study is about 10.2%, which is significantly lower than the 30–40% prevalence reported in previous heart failure studies.[77] We evaluated patients' glucose AC records on average and found out 67% of patients may be considered as diabetes (Glucose AC ≥126 mg/dL) and 21% of patients may be classified as prediabetes (Glucose AC between 100 mg/dl and 125 mg/dl). Although 67% may be overestimated because of averages in glucose AC, this finding matches our result since most of the comorbidities are not significant predictors in our study. Thirdly, this study did not include medication history in model building, and thus the effects of therapies or medications such as Ivabradine (Corlanor®) and Valsartan/sacubitril (Entresto®) were ignored. Future study should investigate whether medication history add to the predictive power of model. Fourthly, this study focused on the mortality of general HF in the Taiwanese population and may need to be repeated and validated for other HF subtypes and populations. Moreover, in the IRB approved study design, the unexposed (control) population was not included. The results of this study need to be validated with a more stringent prospective cohort design. Finally, our models are internally validated, and our results should be validated using an external population.

Conclusion

Exploring HF mortality and its patterns related to clinical risk factors by machine learning models may aid physicians in deciding therapeutic strategies for HF patients. In the future, prediction models of various HF subtypes in Taiwanese patients can be constructed when more databases for each subtype are collected for further study.

Ethics Statement

The studies involving human participants were reviewed and approved by Ethics Committee of Taipei Medical University (N202204033). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Acknowledgments

This study was supported by the Ministry of Science and Technology Grant (NSTC111-2314-B-038-163) and Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan (DP2-111-21121-01- A-10). No funding bodies had any role in study design, data collection, or analysis; decision to publish; or preparation of the manuscript.

Disclosure

The author(s) report no conflicts of interest in this work.

References

1. Bradley J, Schelbert EB, Bonnett LJ, et al. Predicting hospitalisation for heart failure and death in patients with, or at risk of, heart failure before first hospitalisation: a retrospective model development and external validation study %J Lancet Digital Health. *Lancet Digital Health*. 2022;4(6):e445–e454. doi:10.1016/S2589-7500(22)00045-0
2. 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
3. Heidenreich PA, Bozkurt B, Aguilar D, et al. AHA/ACC/HFSA guideline for the management of heart failure: a report of the American College of Cardiology/American Heart Association joint committee on clinical practice guidelines %J circulation. *Circulation*. 2022;145(18):895–1032.
4. Lan T, Liao Y-H, Zhang J, et al. Mortality and readmission rates after heart failure: a systematic review and meta-analysis. *Ther Clin Risk Manag*. 2021;17:1307–1320. doi:10.2147/TCRM.S340587
5. Maciver J, Ross HJ. A palliative approach for heart failure end-of-life care. *Curr Opin Cardiol*. 2018;33(2):202–207. doi:10.1097/HCO.0000000000000484
6. Kwok CS, Abramov D, Parwani P, et al. Cost of inpatient heart failure care and 30-day readmissions in the United States. *Int J Cardiol*. 2021;329:115–122. doi:10.1016/j.ijcard.2020.12.020
7. Zhuang H, Ma Y, Wang L, Zhang H. Effect of early palliative care on quality of life in patients with non-small-cell lung cancer. *Curr Oncol*. 2018;25(1):54–58. doi:10.3747/co.25.3639
8. Vanbutsele G, Van Belle S, Surmont V, et al. The effect of early and systematic integration of palliative care in oncology on quality of life and health care use near the end of life: a randomised controlled trial. *Eur J Cancer*. 2020;124:186–193. doi:10.1016/j.ejca.2019.11.009
9. Morrison RS, Dietrich J, Ladwig S, et al. Palliative care consultation teams cut hospital costs for Medicaid beneficiaries. *Health Aff*. 2011;30(3):454–463. doi:10.1377/hlthaff.2010.0929
10. Laribi S, Aouba A, Nikolaou M, et al. Trends in death attributed to heart failure over the past two decades in Europe. *Eur J Heart Fail*. 2012;14(3):234–239. doi:10.1093/eurjhf/hfr182
11. Tse DMW, Chan KS, Lam WM, Leu K, Lam PT. The impact of palliative care on cancer deaths in Hong Kong: a retrospective study of 494 cancer deaths. *J Palliat Med*. 2007;21(5):425–433. doi:10.1177/0269216307079825
12. Taylor J, Donald H, Ostermann J, Van Houtven CH, Tulsky JA, Steinhauser K. What length of hospice use maximizes reduction in medical expenditures near death in the US Medicare program? *Soc Sci Med*. 2007;65(7):1466–1478. doi:10.1016/j.socscimed.2007.05.028
13. Lin YJ, Chen RJ, Tang JH, et al. Machine-learning monitoring system for predicting mortality among patients with noncancer end-stage liver disease: retrospective study. *JMIR Med Info*. 2020;8(10):e24305. doi:10.2196/24305
14. Gonzalez-Jaramillo V, Maessen M, Luethi N, et al. Unmet needs in patients with heart failure: the importance of palliative care in a heart failure clinic. *Front Cardiovasc Med*. 2022;9. doi:10.3389/fcvm.2022.866794
15. Noraziani K, Nurul'Ain A, Azhim M, et al. An overview of electronic medical record implementation in healthcare system: lesson to learn. *World Appl Sci J*. 2013;25(2):323–332.
16. Cowie MR, Blomster JI, Curtis LH, et al. Electronic health records to facilitate clinical research. *Clin Res Cardiol*. 2017;106(1):1–9. doi:10.1007/s00392-016-1025-6
17. Zhou L-Q, Wang J-Y, Yu S-Y, et al. Artificial intelligence in medical imaging of the liver. *World J Gastroenterol*. 2019;25(6):672–682. doi:10.3748/wjg.v25.i6.672
18. Shimizu H, Nakayama KI. Artificial intelligence in oncology. *Cancer Sci*. 2020;111(5):1452–1460. doi:10.1111/cas.14377
19. Nensa F, Demircioglu A, Rischpler C. Artificial intelligence in nuclear medicine. *J Nucl Med*. 2019;60(Suppl):2.
20. Yu CS, Chen YD, Chang SS, Tang JH, Wu JL, Lin CH. Exploring and predicting mortality among patients with end-stage liver disease without cancer: a machine learning approach. *Eur J Gastroenterol Hepatol*. 2021;33(8):1117–1123. doi:10.1097/MEG.0000000000002169
21. Chang HH, Chiang JH, Wang CS, et al. Predicting mortality using machine learning algorithms in patients who require renal replacement therapy in the critical care unit. *J Clin Med*. 2022;11(18):5289. doi:10.3390/jcm11185289
22. Yu CS, Chang SS, Lin CH, Lin YJ, Wu JL, Chen RJ. Identify the characteristics of metabolic syndrome and non-obese phenotype: data visualization and a machine learning approach. *Front Med*. 2021;8:626580. doi:10.3389/fmed.2021.626580
23. Yu CS, Chang SS, Chang TH, et al. A COVID-19 pandemic artificial intelligence-based system with deep learning forecasting and automatic statistical data acquisition: development and implementation study. *J Med Internet Res*. 2021;23(5):e27806. doi:10.2196/27806
24. Buch VH, Ahmed I, Maruthappu M. Artificial intelligence in medicine: current trends and future possibilities. *Br J Gen Pract*. 2018;68(668):143. doi:10.3399/bjgp18X695213

25. Kriksciuniene D, Sakalauskas V, Ognjanović I, Šendelj R. Discovering Healthcare Data Patterns by Artificial Intelligence Methods. In: *Intelligent Systems for Sustainable Person-Centered Healthcare %J SpringerLink*. Cham, Switzerland: Springer; 2022:185–210.
26. Lee MD DS, Austin PC, Rouleau JL, Liu PP, Naimark D, Tu JV. Predicting mortality among patients hospitalized for heart failure: derivation and validation of a clinical JAMA. *JAMA*. 2003;290(19):2581–2587. doi:10.1001/jama.290.19.2581
27. Ahmad T, Lund LH, Rao P, 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(8):e008081. doi:10.1161/JAHA.117.008081
28. Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Med Inf Decis Making*. 2020;20(1):1–16.
29. Lee KH, Chu YC, Tsai MT, et al. Artificial intelligence for risk prediction of end-stage renal disease in sepsis survivors with chronic kidney disease. *Biomedicines*. 2022;10(3):546.
30. Guo A, Pasque M, Loh F, Mann DL, Payne PRO. Heart failure diagnosis, readmission, and mortality prediction using machine learning and artificial intelligence models. *Curr Epidemiol Rep*. 2020;7(4):212–219. doi:10.1007/s40471-020-00259-w
31. Rahimi K, Bennett D, Conrad N, et al. Risk prediction in patients with heart failure: a systematic review and analysis. *JACC Heart Fail*. 2014;2(5):440–446. doi:10.1016/j.jchf.2014.04.008
32. Niu Q, Liu W, Wang F, Tian L, Dong Y. The utility of cognitive screening in asian patients with heart failure: a systematic review. *Front Psych*. 2022;13:930121.
33. Uchmanowicz I, Nessler J, Gobbens R, et al. Coexisting frailty with heart failure. *Front Physiol*. 2019;10:791.
34. Ministry of Health and Welfare. Population health and welfare quality indicators report. *Ministry of Health* and *Welfare*. 2017:11–13.
35. Office of data science-taipei medical university. Taipei Medical University Clinical Research Database (TMUCRD) 2022 Available from https://ods.tmu.edu.tw/portal_c3_cnt.php?owner_num=c3_69585&button_num=c3&folder_id=4165. Accessed, 2024.
36. Glasheen WP, Cordier T, Gumpina R, Haugh G, Davis J, Renda A. Charlson comorbidity index: ICD-9 update and ICD-10 translation. *Am Health Drug Benefits*. 2019;12(4):188–197.
37. Keerin P, Kurutach W, Boongoen T. Cluster-based KNN missing value imputation for DNA microarray data. Paper presented at: 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)2012.
38. Kang H. The prevention and handling of the missing data. *Korean J Anesthesiol*. 2013;64(5):402–406. doi:10.4097/kjae.2013.64.5.402
39. Statistics PJI. *PASW Statistics Version 18.0*. Chicago, IL: SPSS; 2009.
40. Karatzoglou A, Meyer D, KJJoss H. Support vector machines in R. *J Stat Software*. 2006;15:1–28.
41. Quinlan JR. Program for machine learning. *C4 5*. 1993; 1993:1.
42. Kuhn M, Weston S, Culp M, Coulter N, Quinlan R. Package 'C50' 2022 Available from https://cran.r-project.org/web/packages/C50/C50.pdf. Accessed, 2024.
43. Freund Y, Schapire RE. A decision-theoretic generalization of on-line learning and an application to boosting. *J Computer System Sci*. 1997;55(1):119–139. doi:10.1006/jcss.1997.1504
44. Alfaro E, Gamez M, Garcia NA. An R package for classification with boosting and bagging. *J Stat Software*. 2013;54:1–35. doi:10.18637/jss.v054.i02
45. Molnar C, Schratz P. Package 'Iml' 2022 Available from https://cran.r-project.org/web/packages/iml/iml.pdf. Accessed, 2024.
46. Mosterd A, AWJh H. Clinical epidemiology of heart failure. *Heart*. 2007;93(9):1137–1146. doi:10.1136/hrt.2003.025270
47. Strait JB, Lakatta EG. Aging-associated cardiovascular changes and their relationship to heart failure. *Heart Failure Clinics*. 2012;8(1):143–164. doi:10.1016/j.hfc.2011.08.011
48. Joyner MJ. Effect of exercise on arterial compliance. *Circulation*. 2000;102(11):1214–1215. doi:10.1161/01.CIR.102.11.1214
49. Ancion A, Allepaerts S, Oury C, Gori A-S, Piérard LA, Lancellotti P. Serum albumin level and hospital mortality in acute non-ischemic heart failure. *ESC Heart Failure*. 2017;4(2):138–145. doi:10.1002/ehf2.12128
50. Ozcan S, Cakmak HA, Ikitimur B, et al. The prognostic significance of narrow fragmented QRS on admission electrocardiogram in patients hospitalized for decompensated systolic heart failure. *Clin Cardiol*. 2013;36(9):560–564. doi:10.1002/clc.22158
51. Araújo JP, Lourenço P, Rocha-Gonçalves F, Ferreira A, Bettencourt P. Nutritional markers and prognosis in cardiac cachexia. *Int J Cardiol*. 2011;146(3):359–363. doi:10.1016/j.ijcard.2009.07.042
52. Horwich TB, Kalantar-Zadeh K, MacLellan RW, Fonarow GC. Albumin levels predict survival in patients with systolic heart failure. *Am Heart J*. 2008;155(5):883–889. doi:10.1016/j.ahj.2007.11.043
53. Arques S. Serum albumin and heart failure: recent advances on a new paradigm. *Annales de cardiologie et d'angeiologie*. 2011;60(5):272–278. doi:10.1016/j.ancard.2011.07.006
54. Carr JG, Stevenson LW, Walden JA, Heber D. Prevalence and hemodynamic correlates of malnutrition in severe congestive heart failure secondary to ischemic or idiopathic dilated cardiomyopathy. *Am J Cardiol*. 1989;63(11):709–713. doi:10.1016/0002-9149(89)90256-7
55. Damman K, Testani JM. The kidney in heart failure: an update. *Eur Heart J*. 2015;36(23):1437–1444. doi:10.1093/eurheartj/ehv010
56. de Silva R, Nikitin NP, Witte KK, et al. Incidence of renal dysfunction over 6 months in patients with chronic heart failure due to left ventricular systolic dysfunction: contributing factors and relationship to prognosis. *Eur Heart J*. 2006;27(5):569–581. doi:10.1093/eurheartj/ehi696
57. Foundation. NK. K/DOQI clinical practice guidelines for chronic kidney disease: evaluation, classification, and stratification. *Ame J Kidney Dis*. 2002;39(2 Suppl 1):S1–266.
58. Dimopoulou I, Daganou M, Tsintzas OK, Tzelepis GE. Effects of severity of long-standing congestive heart failure on pulmonary function. *Respir Med*. 1998;92(12):1321–1325. doi:10.1016/S0954-6111(98)90136-6
59. Ceridon M, Wanner A, Johnson BD. Does the bronchial circulation contribute to congestion in heart failure? *Med Hypotheses*. 2009;73(3):414–419. doi:10.1016/j.mehy.2009.03.033
60. Cross TJ, Kim C-H, Johnson BD, Lalande S. The interactions between respiratory and cardiovascular systems in systolic heart failure. *J Appl Physiol*. 2020;128(1):214–224. doi:10.1152/japplphysiol.00113.2019
61. Cugno M, Mari D, Meroni PL, et al. Haemostatic and inflammatory biomarkers in advanced chronic heart failure: role of oral anticoagulants and successful heart transplantation. *Br J Haematol*. 2004;126(1):85–92. doi:10.1111/j.1365-2141.2004.04977.x
62. Siniarski A, GĄSecka A, Borovac JA, et al. Blood coagulation disorders in heart failure: from basic science to clinical perspectives. *J Card Fail*. 2023;29(4):517–526. doi:10.1016/j.cardfail.2022.12.012

63. Alehagen U, Dahlström U, Lindahl TL. Low plasma concentrations of coagulation factors II, VII and XI indicate increased risk among elderly with symptoms of heart failure. *Blood Coagulation Fibrinol*. 2010;21(1):62–69. doi:10.1097/MBC.0b013e328332aa2b
64. Campbell AR, Rodriguez AJ, Larson DM, et al. Resource utilization and outcome among patients with selective versus nonselective troponin testing. *Am Heart J*. 2018;199:68–74. doi:10.1016/j.ahj.2018.01.010
65. Roos A, Bandstein N, Lundbäck M, Hammarsten O, Ljung R, MJJJotAcoc H. Stable high-sensitivity cardiac troponin T levels and outcomes in patients with chest pain. *J Ame College Cardiol*. 2017;70(18):2226–2236. doi:10.1016/j.jacc.2017.08.064
66. Taqueti VR, Solomon SD, Shah AM, et al. Coronary microvascular dysfunction and future risk of heart failure with preserved ejection fraction. *Eur Heart J*. 2018;39(10):840–849. doi:10.1093/eurheartj/ehx721
67. Logeart D, Beyne P, Cusson C, et al. Evidence of cardiac myolysis in severe nonischemic heart failure and the potential role of increased wall strain. *Ame Heart J*. 2001;141(2):247–253. doi:10.1067/mhj.2001.111767
68. Yilmaz A, Yalta K, Turgut OO, et al. Clinical importance of elevated CK-MB and troponin I levels in congestive heart failure. *Adv ther*. 2006;23(6):1060–1067. doi:10.1007/BF02850226
69. Silvestre-Roig C, Braster Q, Ortega-Gomez A, Soehnlein O. Neutrophils as regulators of cardiovascular inflammation. *Nat Rev Cardiol*. 2020;17(6):327–340. doi:10.1038/s41569-019-0326-7
70. Hermansen SE, Kalstad T, How OJ, Myrmel T. Inflammation and reduced endothelial function in the course of severe acute heart failure. *Transl Res*. 2011;157(3):117–127. doi:10.1016/j.trsl.2010.12.002
71. Billett HH. Hemoglobin and Hematocrit. In: *Clinical Methods: The History, Physical, and Laboratory Examinations*. *3rd* ed. Butterworths; 1990.
72. Ather S, Chan W, Bozkurt B, et al. Impact of noncardiac comorbidities on morbidity and mortality in a predominantly male population with heart failure and preserved versus reduced ejection fraction. *J Ame College Cardiol*. 2012;59(11):998–1005. doi:10.1016/j.jacc.2011.11.040
73. Migone de Amicis M, Chivite D, Corbella X, Cappellini MD, Formiga FJI, medicine E. Anemia is a mortality prognostic factor in patients initially hospitalized for acute heart failure. *Int Emerg med*. 2017;12:749–756. doi:10.1007/s11739-017-1637-5
74. Groenveld HF, Januzzi JL, Damman K, et al. Anemia and mortality in heart failure patients: a systematic review and meta-analysis. *J Ame College Cardiol*. 2008;52(10):818–827. doi:10.1016/j.jacc.2008.04.061
75. Xia H, Shen H, Cha W, Lu QJFi CM. The prognostic significance of anemia in patients with heart failure: a meta-analysis of studies from the last decade. *Front Cardiovasc Med*. 2021;8:632318. doi:10.3389/fcvm.2021.632318
76. Hosseinpour M, Hatamnejad MR, Montazeri MN, et al. Comparison of the red blood cell indices based on accuracy, sensitivity, and specificity to predict one-year mortality in heart failure patients. *BMC Cardiovasc Disord*. 2022;22(1):532. doi:10.1186/s12872-022-02987-x
77. Dunlay SM, Givertz MM, Aguilar D, et al. Type 2 diabetes mellitus and heart failure: a scientific statement from the American heart association and the heart failure society of America: this statement does not represent an update of the 2017 ACC/AHA/HFSA heart failure guideline update. *Circulation*. 2019;140(7):e294–e324. doi:10.1161/CIR.0000000000000691