



OPEN ACCESS

EDITED BY

Otilia Tica,
Emergency County Clinical Hospital of
Oradea, Romania

REVIEWED BY

Mauro Chiarito,
Humanitas University, Italy

*CORRESPONDENCE

Long Yun
✉ wwlyf@126.com
Liang Hao
✉ lianghao@hnucm.edu.cn

[†]These authors share first authorship

RECEIVED 30 July 2025

ACCEPTED 07 October 2025

PUBLISHED 27 October 2025

CITATION

Sidie H, Wen Z, Yidi Z, Yun L and Hao L (2025)
Risk stratification and survival prediction in
heart failure: from grades to scores.
Front. Cardiovasc. Med. 12:1676441.
doi: 10.3389/fcvm.2025.1676441

COPYRIGHT

© 2025 Sidie, Wen, Yidi, Yun and Hao. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Risk stratification and survival prediction in heart failure: from grades to scores

Huang Sidie^{1†}, Zhang Wen^{2†}, Zeng Yidi¹, Long Yun^{3*} and
Liang Hao^{1*}

¹Institute of TCM Diagnostics, Hunan University of Chinese Medicine, Changsha, Hunan, China,

²Department of Geriatrics, The First Hospital of Hunan University of Chinese Medicine, Changsha,

Hunan, China, ³Department of Cardiology, The First Hospital of Hunan University of Chinese Medicine, Changsha, Hunan, China

Heart failure (HF) continues to pose a significant global health burden, necessitating accurate prognostic tools to guide patient management. This mini-review presents grading systems, frailty scales, and scoring models, followed by challenges and future directions. We trace the evolution of stratification and prognostic assessments in HF, beginning with the foundational NYHA functional classification and progressing to the advanced prognostic scores currently in use. We examine the historical significance and clinical relevance of NYHA grades, which have long been pivotal in evaluating HF severity. The review then shifts focus to contemporary prognostic scores, including the Seattle Heart Failure Model (SHFM), the Heart Failure Survival Score (HFSS), and emerging tools leveraging machine learning (ML) and big data. We explore specific challenges encountered in current clinical practice and outline future directions. By highlighting the strengths and limitations of these tools, this mini-review aims to provide a critical appraisal of stratification and scoring models for HF to inform their optimal application in clinical practice, ultimately enhancing patient care and outcomes in HF.

KEYWORDS

heart failure, risk stratification, prognosis, frailty, machine learning

1 Introduction

Heart failure (HF) represents a critical global health challenge, affecting millions of individuals and placing substantial burdens on healthcare systems worldwide (1). Based on left ventricular ejection fraction (LVEF), HF was divided into heart failure with reduced ejection fraction (HFrEF, LVEF $\leq$ 40%) and heart failure with preserved ejection fraction (HFpEF, LVEF $\geq$ 50%) (2). Despite significant advancements in therapeutic interventions, the prognosis for patients with HF remains variable (3), underscoring the necessity for precise prognostic tools to guide clinical decision-making and enhance patient management.

Stratification and Prognostication in HF has evolved from rudimentary symptom-based classifications to sophisticated multivariate models. The New York Heart Association (NYHA) functional classification, introduced in 1928 (4), revolutionized clinical practice by categorizing patients into four grades of symptom severity. However, its subjectivity and poor correlation with objective biomarkers or mortality risk have driven demand for multidimensional risk stratification (5). Contemporary tools—such as the Seattle Heart Failure Model (SHFM) and Heart Failure Survival Score (HFSS)—integrate

demographic, hemodynamic, and biomarker data to quantify individual mortality risk. More recently, machine learning (ML) algorithms harness big data to predict outcomes with unprecedented granularity. This review seeks to highlight the strengths and weaknesses of these tools, providing insights into their optimal application in clinical practice to improve patient care and outcomes in HF.

2 Grading approaches for stratification and prognosis of HF

Historically, risk stratification in HF has relied on grading systems that categorize disease severity based on clinical presentation, hemodynamic status, or functional capacity. These systems, while foundational, exhibit distinct conceptual frameworks and limitations. The ACC/AHA HF staging system classifies HF into four progressive stages (A–D), focusing on disease evolution from risk factors (stage A) to refractory symptoms (stage D) (6). This approach emphasizes prevention and early intervention but lacks granularity for dynamic symptom assessment, rendering it less responsive to short-term clinical changes. In contrast, the New York Heart Association (NYHA) classification evaluates symptom severity during daily activities (classes I–IV) and remains a cornerstone for routine clinical decision-making due to its simplicity (7). However, its subjective nature introduces interobserver variability, and it poorly discriminates HF patients across the spectrum of functional impairment (5, 8).

For acute hemodynamic evaluation, the Killip/Forrester classification stratifies patients with acute myocardial infarction (MI)-induced HF into four classes based on signs of pulmonary congestion and peripheral hypoperfusion. While valuable in acute MI settings, its utility diminishes in chronic HF or non-ischemic etiologies (9). The Weber classification employs cardiopulmonary exercise testing (CPET)-derived peak oxygen consumption (VO_2) to categorize HF into classes A–D, offering a quantitative assessment of exercise tolerance. This system excels in prognostication for advanced HF, particularly in identifying candidates for ventricular assist devices or transplantation (10). However, its reliance on CPET limits widespread applicability, especially in resource-constrained settings (11).

A comparative analysis reveals key trade-offs (Table 1). ACC/AHA staging and NYHA classification serve complementary roles—the former guiding long-term management and the latter monitoring daily symptom fluctuations. Yet both overlook

comorbidities and non-cardiac contributors to prognosis. Killip/Forrester excels in acute MI but lacks relevance in chronic HF, while Weber's objectivity is counterbalanced by logistical challenges. Importantly, these systems are largely unidimensional, neglecting multidimensional risk factors such as renal function, biomarkers, or frailty, which significantly influence outcomes.

3 Frailty scales for stratification and prognosis of heart failure

Frailty scales have emerged as a critical advancement in HF stratification, addressing the limitations of traditional grading systems by integrating multidimensional assessments of physiological vulnerability. Unlike conventional approaches that focus narrowly on cardiac-specific metrics, frailty scales evaluate systemic functional decline, incorporating both subjective and objective measures of physical performance, cognitive status, and comorbidities (12). This holistic framework enhances risk stratification by capturing the interplay between HF severity and age-related multisystem impairments, which are strong predictors of mortality, hospitalization, and quality of life.

The Fried frailty phenotype (13) and Short Physical Performance Battery (SPPB) (14) are among the most widely validated tools. The Fried criteria define frailty as the presence of ≥ 3 components (unintentional weight loss, exhaustion, low physical activity, slow gait, and weak grip strength), while the SPPB quantifies lower extremity function through balance, gait speed, and chair-stand tests. Both scales demonstrate prognostic value in HF populations, with frail individuals exhibiting 2–3 times higher risks of adverse outcomes compared to non-frail counterparts (15–17). However, these tools require time-consuming physical measurements, limiting their practicality in routine clinical workflows (18). To address this, simplified questionnaires like the SARC-F (assessing strength, assistance walking, rising from a chair, climbing stairs, and falls) (19) and Clinical Frailty Scale (CFS) (20, 21) have gained traction. The SARC-F, for instance, correlates strongly with death or hospitalization in HF patients (OR 1.55, 95% CI 1.03–2.35) (22) and can be administered rapidly at bedside.

The Heart Failure Association of the European Society of Cardiology (HFA-ESC) designed a new HF frailty assessment score that encompasses four domains (23): clinical (comorbidities, weight), functional (impairment in activities of daily living, mobility and/or balance), psycho-cognitive

TABLE 1 Comparison of different grading approaches for stratification and prognosis of heart failure.

Systems	Parameters assessed	Key applications	Prognostic value	Strengths	Limitations
ACC/AHA stages	Risk factors, structural changes	Chronic HF risk prevention	Identifies pre-symptomatic disease	Guides preventive strategies	Static, insensitive to acute changes
NYHA class	Symptom severity	Routine clinical assessment	Quick bedside tool	Simple, widely adopted	Subjective, variable interpretation
Killip/Forrester	Hemodynamic stability	AMI with acute HF	Predicts in-hospital mortality	Rapid risk stratification	Limited to ischemic HF
WEBER classification	Exercise capacity (VO_2)	Chronic HF	Strong mortality correlation	Objective, prognostic power	Requires specialized testing

(cognitive impairment, dementia and/or depression), and social (social support, institutionalization and/or the lack of support.) domains. A study (24) has shown that frailty assessment based on the HFA-ESC frailty domains demonstrated a high prevalence of frailty among HF patients and successfully identified individuals at elevated risk for adverse events (AUC = 0.64, 95% CI 0.60–0.68).

Recent studies highlight the superiority of frailty scales over traditional HF grading systems in identifying high-risk subgroups. For example, in the FRAIL-HF cohort, among patients hospitalized with HF, frail patients (biological phenotype criteria) showed higher 1-year all-cause mortality [HR: 2.13, 95% CI: 1.07–4.23] even after adjusting for NYHA class and ejection fraction (25). Similarly, the GUIDE-IT trial demonstrated that a higher frailty (frailty index criteria) burden was associated with a significantly higher risk of HF hospitalization or death [HR: 1.76, 95% CI: 1.20–2.58], adjusted for LVEF, NYHA class, NT-proBNP and etc. (26). These findings underscore frailty's role as a modifier of HF trajectory, particularly in aging populations with multimorbidity.

Despite their utility, frailty scales face challenges in standardization and implementation (12). Heterogeneity in assessment tools complicates cross-study comparisons, while dynamic changes in frailty status necessitate repeated evaluations. Furthermore, few scales account for HF-specific variables such as fluid retention or arrhythmia burden, which may transiently impair physical performance. Villani ER et al. (27) reported that the prevalence of frailty in AF patients ranged from 4.4%–75.4% while AF prevalence in the frail population ranged from 48.2%–75.4%. Indicators reflecting fluid retention and arrhythmia, such as edema, shortness of breath, and chest tightness, can be incorporated into the Fried frailty to provide a more comprehensive assessment of frailty in HF patients. These indicators help detect temporary functional decline caused by fluid retention or arrhythmia in HF patients, thereby enabling a more accurate evaluation of their condition. Future efforts should focus on harmonizing definitions, validating HF-tailored frailty indices, and integrating these tools into electronic health records (EHR) for automated risk alerts.

4 Scoring models for stratification and prognosis of HF

Risk prediction models in HF have evolved from simplistic clinical grading systems to sophisticated multivariate tools that integrate demographic, biochemical, imaging, and therapeutic data. These models aim to quantify mortality risk, guide treatment decisions, and optimize resource allocation across both chronic and acute HF populations.

4.1 Acute HF risk models

In the context of emergency and critical care, acute HF models are primarily designed for short-term prognostic risk stratification and

optimal allocation of medical resources. These models rely on rapidly obtainable pathophysiological parameters at admission (such as blood pressure, serum creatinine, and NT-proBNP) to accurately predict in-hospital or 30-day mortality. This provides an evidence-based foundation for prioritizing triage of critically ill patients, determining eligibility for higher levels of monitoring, and guiding intensive intervention strategies during the acute phase.

For acute HF, the Emergency Heart Failure Mortality Risk Grade (EHMRG) (28, 29) and Multiple Estimation of risk based on the Emergency department Spanish Score In patients with Acute Heart Failure (MEESSI-AHF) (30, 31) emerged as frontline tools. EHMRG, validated in >12,000 emergency department patients, uses seven variables (e.g., systolic blood pressure, troponin) to predict 7-day mortality (AUC 0.79). MEESSI-AHF, incorporating NT-proBNP and potassium levels, outperforms EHMRG in 30-day risk stratification (AUC 0.85 vs. 0.80). Both models prioritize rapid risk assessment but overlook longitudinal outcomes beyond 30 days.

The GWTG-HF risk score (32) predicts in-hospital mortality using commonly available clinical variables such as age, systolic blood pressure, blood urea nitrogen, heart rate, sodium levels, chronic obstructive pulmonary disease, and non-Black race. It applies to a wide range of HF patients, including case with preserved left ventricular systolic function. Yasuyuki et al. (33) found that GWTG-HF risk score can show good discrimination and calibration in Japanese AHF patients (c-statistic, 0.763; 95% CI, 0.700–0.826), and the discriminative ability of the model was significantly improved with the addition of BNP levels (c statistic, 0.818; 95% CI, 0.771–0.865). Although the GWTG-HF score was originally developed for in-hospital patients, it also demonstrates good discrimination for 1-year mortality in a heterogeneous cohort of CICU patients (34).

4.2 Chronic HF risk models

The purpose of chronic HF models is to facilitate long-term risk assessment and the development of personalized advanced treatment strategies. These models integrate multidimensional variables reflecting long-term homeostasis of cardiac structure and function—such as left ventricular ejection fraction, peak oxygen consumption, and QRS duration—and are designed to predict all-cause mortality on an annualized basis. Their core clinical utility lies in providing objective, quantified criteria for patient selection and prioritization for scarce and high-risk end-stage therapies, including cardiac transplantation and left ventricular assist device (LVAD) implantation.

The HFSS (35), introduced in 1997, was among the first models to incorporate non-invasive variables—ischemic etiology, resting heart rate, left ventricular ejection fraction (LVEF), mean arterial pressure, QRS duration, serum sodium, and peak oxygen consumption (VO_2)—to stratify heart transplant candidates. Validated in cohorts with advanced HF, HFSS demonstrated moderate discrimination (c-statistic 0.56–0.79) but faced limitations in the β -blocker era, as it excluded pharmacotherapy effects. Subsequent studies confirmed its retained prognostic

value in β -blocker-treated patients, albeit with reduced sensitivity for low-risk identification (36). However, the Zugck et al. reported that HFSS was inferior to a two-variable model containing only LVEF and either peak oxygen uptake (peak VO₂) or 6-min walk test (6'WT) (37).

A paradigm shift occurred in 2006 with the SHFM (38), which integrated 24 variables, including medications (β -blockers, ACE inhibitors) and devices (ICDs), enabling dynamic survival estimation. Derived from the PRAISE I clinical trial database and validated in 14 cohorts ($n=16,057$), SHFM predicts 1–3-year survival with c-statistics of 0.63–0.81 (39). Its unique feature is simulating survival gains from guideline-directed therapies, such as adding sacubitril/valsartan or CRT-D. However, SHFM underestimates risk in HF with preserved ejection fraction (HFpEF) and relies on trial-derived cohorts (40, 41), limiting generalizability to real-world populations with multimorbidity. Incorporating diastolic function parameters (e.g., E/E' ratio, IVRT) or diastolic stress biomarkers (BNP, IL-6, etc.) could enhance risk prediction accuracy, improving clinical decision-making and patient management.

The Meta-Analysis Global Group in Chronic Heart Failure (MAGGIC) score (42, 43), developed in 2013, addressed heterogeneity by pooling individual patient data from 39,372 subjects across 30 studies. This 13-variable model (e.g., age, creatinine, LVEF) predicts 1- and 3-year mortality (c-statistic 0.73) and excels in applicability across HF subtypes, including HFpEF. However, it lacks granularity in capturing acute decompensation markers or device therapy impacts.

The Metabolic Exercise test data combined with Cardiac and Kidney Indexes (MECKI) score (44), developed for chronic HF, uniquely integrates CPET parameters (e.g., VO₂, VE/VCO₂ slope) with renal function and LVEF. Validated in 2,715 patients, it predicts 3-year survival with superior accuracy (AUC 0.83) compared to SHFM (AUC 0.76). However, its reliance on CPET limits routine clinical application.

Figure 1 summarizes the selection of these models based on clinical context for both acute and chronic HF, while Table 2 outlines their key trade-offs. Frailty scales reflect a patient's

physiological reserve and vulnerability, whereas scoring models primarily quantify cardiac-specific risk. An integrated approach that combines frailty scales with scoring models can provide a more comprehensive risk profile, thereby better guiding personalized treatment decisions. Few tools address the distinct pathophysiology of HFpEF, where SHFM and MAGGIC underperform; Discrepancies in biomarker assays (e.g., NT-proBNP vs. BNP) and EHR documentation practices hinder cross-institutional applicability; Operational inertia: Complex models like SHFM struggle with EHR integration, whereas oversimplified tools (e.g., ADHERE's 3-variable model) (45) sacrifice granularity.

5 Challenges and future prospects

Despite significant advancements in risk prediction models for HF, several challenges persist that limit their clinical utility and generalizability.

5.1 Limitations of prognostic models in HFpEF

Current models fail to adequately differentiate between HF subtypes, particularly HFpEF and HFrEF. Most models were derived from cohorts dominated by HFrEF patients, leading to poor calibration in HFpEF populations where distinct pathophysiological drivers such as systemic inflammation, metabolic dysregulation, and myocardial fibrosis disproportionately influence outcomes (46, 47).

Notably, most existing HFpEF-specific tools like the H2FPEF Score (48, 49) (originally designed for diagnostic probability stratification) and drug trial data (e.g., I-PRESERVE, TOPCAT) (50–53) primarily focus on diagnostic confirmation or therapeutic response rather than prognostic modeling. The HFA-PEFF score (54, 55), despite integrating echocardiographic parameters and NT-proBNP levels, still relies on static variables and fails to

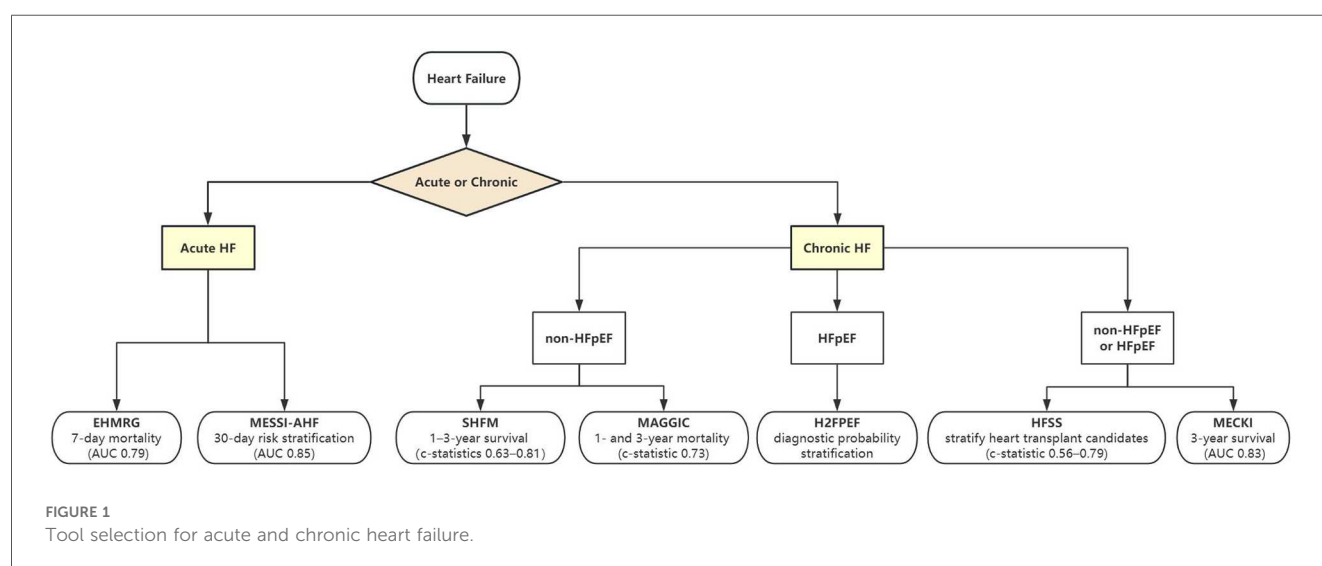


TABLE 2 Comparison of different scoring models for stratification and prognosis of heart failure.

Model	Population	Derivation cohort	Variables	Endpoint	Discrimination/calibration	External validation	Intended use	Key references
EHRG	Acute HF	Patients with HF who visited the emergency departments (ED) and fulfilled the Framingham criteria for HF	Age, Transport by EMS, SBP, Heart rate, Oxygen saturation, Creatinine, Potassium, Elevated troponin level, Active cancer, Metolazone at home	7-day mortality	c-statistic = 0.803 (95% CI, 0.763 to 0.840)	c-index (7-day mortality) = 0.73 (95% CI, 0.71–0.76); c-index (30-day mortality) = 0.71 (95% CI, 0.70–0.73)	Predict acute mortality among patients with HF who present to the ED	Prediction of Heart Failure Mortality in Emergent Care, External Validation and Refinement of Emergency Heart Failure Mortality Risk Grade Risk Model in Patients With Heart Failure in the Emergency Department
MEESSI-AHF	Acute HF	Patients presenting with AHF to ED	Barthel Index score, systolic blood pressure, age, NT-proBNP level, potassium level, elevated cardiac troponin T (cTnT) level, NYHA class IV disease, respiratory rate, low-output symptoms, oxygen saturation, episode associated with an acute coronary syndrome, left ventricular hypertrophy on electrocardiogram, and creatinine level	30-day mortality	c-statistic = 0.828 (95% CI, 0.802 to 0.853)	c-statistic = 0.80 (95% CI, 0.76 to 0.84)	Estimate individual risk for death within 30 days for patients with AHF who are admitted to the ED	Predicting 30-Day Mortality for Patients With Acute Heart Failure in the Emergency Department, External Validation of the MEESSI Acute Heart Failure Risk Score: A Cohort Study
GWTFG-HF risk score	Acute HF	Data from the American Heart Association's (AHA) Get With The Guidelines-Heart Failure module (GWTG-HF)	age, systolic blood pressure, blood urea nitrogen, heart rate, sodium levels, chronic obstructive pulmonary disease, non-Black race	In-hospital mortality	c-index = 0.75	c-statistic = 0.763 (95% CI, 0.700–0.826)	Predict the risk of in hospital mortality for patients hospitalized with HF	A Validated Risk Score for In-Hospital Mortality in Patients With Heart Failure From the American Heart Association Get With the Guidelines Program, Validation of the Get With The Guidelines-Heart Failure risk score in Japanese patients and the potential improvement of its discrimination ability by the inclusion of B-type natriuretic peptide level
HFSS	Chronic HF	Data on 80 clinical characteristics from 268 ambulatory patients with advanced heart failure	Ischemic cardiomyopathy, resting heart rate, LVEF, IVCD (QRS duration ≥ 0.12 s of any cause), mean resting blood pressure, peak V o ₂ , and serum sodium	Three risk strata (high-risk stratum, medium-risk stratum, low-risk stratum)	c-statistic = 0.56–0.79	Medium risk: HR 2.65, 95% CI 1.75–4.02, $p < 0.001$; high risk: HR 5.51, 95% CI 3.64–8.33, $p < 0.001$	Stratify AHF risk of adverse outcome and select candidates for cardiac transplantation	Development and Prospective Validation of a Clinical Index to Predict Survival in Ambulatory Patients Referred for Cardiac Transplant Evaluation, Heart failure survival score continues to predict clinical outcomes in patients with heart failure receiving β -blockers

(Continued)

TABLE 2 Continued

Model	Population	Derivation cohort	Variables	Endpoint	Discrimination/calibration	External validation	Intended use	Key references
SHFM	Chronic HF	The PRAISE I clinical trial database	Clinical characteristics [age, gender, NYHA class, weight, LVEF, systolic blood pressure (sBP), ischemic etiology], medications (angiotensin-converting enzyme inhibitor, angiotensin receptor blocker, β -blocker, statin, aldosterone blocker, loop diuretic equivalent dose, allopurinol), device therapy (implantable cardioverter-defibrillator, cardiac resynchronization therapy) and laboratory data (lymphocyte percentage and serum sodium, hemoglobin, uric acid, total cholesterol)	Mean, 1-, 2-, and 3-year survival	c-statistics = 0.63–0.81	c-statistics = 0.75	Estimate survival of heart failure patients	The Seattle Heart Failure Model: Validation and Recalibration of Seattle Heart Failure Model in Japanese Acute Heart Failure Patients
MAGGIC score	Chronic HF	Data on 39,372 patients with HF from 30 cohort studies	Age, lower EF, NYHA class, serum creatinine, diabetes, not prescribed beta-blocker, lower systolic BP, lower body mass, time since diagnosis, current smoker, chronic obstructive pulmonary disease, male gender, not prescribed ACE-inhibitor or angiotensin-receptor blockers	1- and 3-year mortality		c-statistic = 0.73	Readily quantifies individual patient mortality risk	Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies, Heart failure in younger patients: the Meta-analysis Global Group in Chronic Heart Failure (MAGGIC)
MECKI score	Chronic HF	2,716 systolic HF patients followed in 13 Italian centers	hemoglobin, Na ⁺ , MDRD, LVEF, peak VO2 (% predicted), and VE/VCO2 slope	0%–100%	AUC values of 0.804 (0.754–0.852) at 1year, 0.789 (0.750–0.828) at 2years, 0.762 (0.726–0.799) at 3years and 0.760 (0.724–0.796) at 4years.	AUC values of 0.81 $\pm$ 0.04 (0.73–0.89) at 1year, 0.76 $\pm$ 0.04 (0.68–0.84) at 2years, 0.80 $\pm$ 0.03 (0.73–0.86) at 3years	Identify the risk of cardiovascular death + urgent heart transplant	Metabolic exercise test data combined with cardiac and kidney indexes, the MECKI score: A multiparametric approach to heart failure prognosis, The metabolic exercise test data combined with Cardiac And Kidney Indexes (MECKI) score and prognosis in heart failure. A validation study

capture dynamic biomarker trajectories or phenotypic heterogeneity (e.g., cardiometabolic vs. elderly frail subtypes).

The modified EFFECT score (56) can be used to assess the 28-day and 1-year mortality risk in hospitalized patients with HFpEF and ADHF (AUC: 0.76 for 28-day, and 0.72 for one-year mortality). By incorporating mortality-related indicators such as age, systolic blood pressure (SBP), blood urea nitrogen (BUN), sodium, cerebrovascular disease [defined as stroke/transient ischemic attack (TIA) in ARIC], chronic obstructive pulmonary disease (COPD), and hemoglobin, it enables better identification of high-risk patients and guides clinical decision-making, including early triage, in-hospital monitoring, treatment, and early post-discharge follow-up. However, there is still a lack of external validation cohorts to confirm its generalizability.

Additionally, traditional models depend on baseline variables (e.g., LVEF, serum sodium) and neglect temporal risk modifiers such as fluctuating NT-proBNP levels with limited prognostic value in HFpEF (30% of cases show levels <125 pg/mL) (57, 58), treatment adherence patterns, or evolving comorbidities. The reliance on drug trial data further introduces selection bias, as participants often exclude HFpEF-dominant populations—such as elderly patients with multimorbidity (≥ 3 comorbidities in 67.4% of Asian cohorts) or underrepresented racial groups—thereby limiting real-world applicability.

5.2 Challenges in the AI era

Practical implementation barriers also hinder widespread adoption. Many scoring systems require manual data entry, which is time-consuming and prone to errors. The evolution of HF prediction into the AI era (Figure 2) with higher accuracy has witnessed three technological waves: 1) ML, have the potential to improve classification performance over traditional

statistical tools by taking into account nonlinear impacts of variables to arrive at an accurate prediction (59); 2) deep learning (DL) – a branch of ML, leveraging convolutional neural networks (CNN) and recurrent neural network (RNN) for risk prediction (60, 61), outperformed traditional ML models; 3) and large language models (LLMs) capable of parsing multimodal data from EHRs and wearable devices (62).

Zhao H et al. (63) employed ML techniques such as RF and LASSO regression to construct alternative risk models. The models, which improved the accuracy of risk prediction and uncovered novel relationships between risk factors and outcomes, demonstrated good performance in predicting mortality and readmission among HFmrEF patients. Li et al. (64) developed ML algorithms to predict mortality of HF patients within ICU settings, with XGBoost demonstrating superior performance. While DL model like CNN showed a high discriminatory ability in categorizing HFpEF and control patients, achieving an AUC of 0.92 on the blinded test set, with a sensitivity of 0.98 and specificity of 0.6327 (61). HFmeRisk model, a DL model developed by Zhao X et al. (60), used both 5 clinical features and 25 DNA methylation loci to provides innovative insights into early risk assessment for HFpEF. The model underwent internal validation through tenfold cross-validation to ensure its generalization capability, and external validation with 38 samples demonstrated its reliable predictive performance (AUC=0.82). However, due to the limited sample size, these samples may not fully represent real-world patient populations. Additionally, as the FHS cohort used in the study primarily consisted of Caucasian and a small number of East Asian individuals, the model's applicability to other ethnic groups remains unclear.

While predictive accuracy improves with model complexity—DL shows 15% or higher compared to traditional ML (65–67)—interpretability progressively declines. DL's attention mechanisms

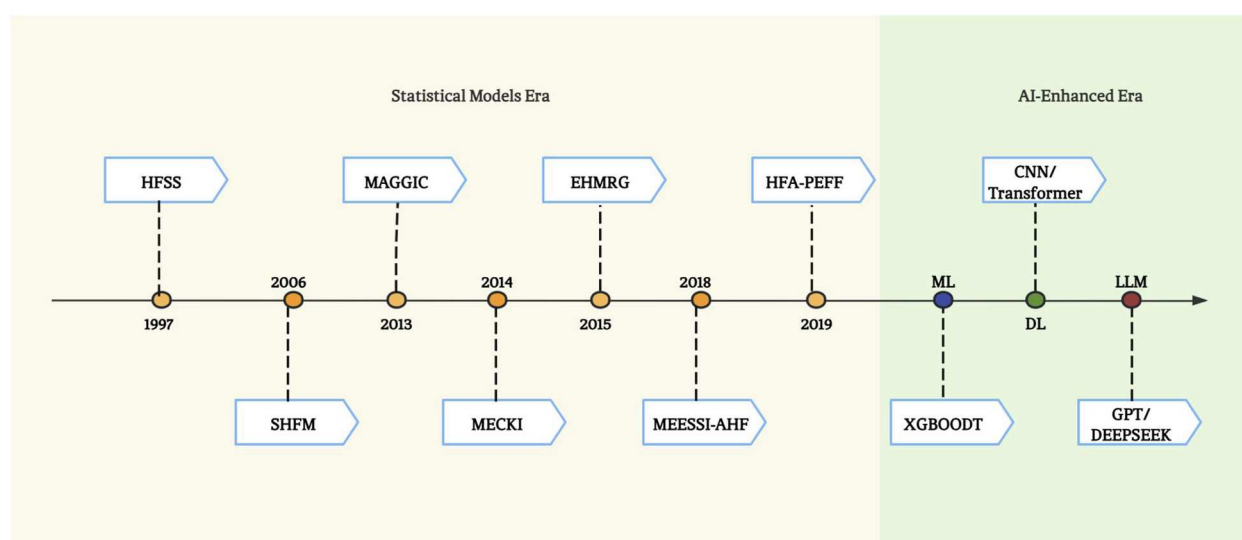


FIGURE 2
The evolution of risk stratification and survival prediction tools for heart failure.

and LLM's transformer architectures create nested decision layers that obscure clinical reasoning pathways (68). For instance, transformer-based heart language models analyzing electrocardiogram reports achieved F1 score of 93.33% to detect atrial fibrillation (69), yet their self-attention weights remain clinically uninterpretable. Clinicians face a precision-transparency tradeoff: gradient boosting models reveal feature importance through SHapley Additive exPlanations (SHAP) values but fail to explain temporal models (70), while LIME (Local Interpretable Model-agnostic Explanations) provides local approximations at the cost of global coherence (71). This “black box” dilemma persists despite hybrid approaches like model distillation that compress neural networks into rule-based surrogates with 20% accuracy loss, ultimately restricting trust and routine integration. A more comprehensive understanding may be achieved by combining multiple interpretation techniques (such as integrating SHAP's global perspective with LIME's local insights). Concurrently, developing intrinsically interpretable models or designing time-series architectures optimized for explainability can embed transparency directly into the model design phase. These efforts aim to progressively bridge the “black box” dilemma and enhance clinical trust.

Some models are trained solely on data from a single institution, and their extrapolation efficacy requires further validation through multi-center studies. It is advisable to test these models using more diverse datasets from different regions and research institutions to assess their generalization capability and mitigate potential prediction biases arising from sociodemographic factors. The lack of open-source and data availability for many AI models poses a significant obstacle, as their performance can be neither independently verified nor replicated, ultimately hindering scientific progress and the widespread adoption of technology. Integration with electronic health systems or wearable devices, along with the development of mobile applications, could facilitate tighter incorporation of artificial intelligence models into clinical practice.

5.3 Future prospects

Future research would prioritize four key directions to address these gaps. First, HF subtype-specific models are urgently needed. HFpEF, now representing over 50% of HF cases, demands distinct predictors (e.g., atrial fibrillation burden, diastolic stress biomarkers) compared to HFrEF. Second, EHR-integrated automated scoring systems could enhance practicality by leveraging structured data (e.g., lab results, medication lists) and natural language processing to extract unstructured clinical notes. For instance, integrating SHFM variables into EHRs could enable real-time risk alerts, though this requires standardization of data formats across institutions. To further enhance data comprehensiveness, socioeconomic factors linked to survival rates—such as healthcare access or education level—can also be incorporated into the EHR. Third, multimodal data fusion—combining genomics, proteomics, and imaging-derived radiomics—may uncover novel prognostic signatures. Wearable devices enabling continuous monitoring of physiological parameters (e.g., daily step count, nocturnal heart rate variability) could further refine dynamic risk prediction. Finally,

causal inference frameworks are needed to distinguish causation from correlation in longitudinal datasets, particularly when evaluating the impact of interventions like sacubitril/valsartan or cardiac resynchronization therapy.

Author contributions

HS: Conceptualization, Data curation, Writing – original draft. ZW: Methodology, Formal analysis, Writing – original draft. ZY: Data curation, Resources, Visualization, Supervision, Validation, Writing – review & editing. LY: Funding acquisition, Visualization, Validation, Writing – review & editing. LH: Supervision, Validation, Conceptualization, Funding acquisition, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This work was supported by the grants of the National Natural Science Foundation of China (No. 82274411), Key Project of Joint Fund of Hunan University of Chinese Medicine and Hospitals (No. 2024XYLH339), Hunan Provincial Health Commission (No. D202303019470), Scientific Research Project of Hunan Provincial Department of Education (No. 23C0162) and Science and Technology Innovation Program of Hunan Province (No. 2022RC1021).

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.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

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

- Khan MS, Shahid I, Bennis A, Rakisheva A, Metra M, Butler J. Global epidemiology of heart failure. *Nat Rev Cardiol.* (2024) 21(10):717–34. doi: 10.1038/s41569-024-01046-6
- Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: the task force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur J Heart Fail.* (2016) 18(8):891–975. doi: 10.1002/ehf.592
- Emmons-Bell S, Johnson C, Roth G. Prevalence, incidence and survival of heart failure: a systematic review. *Heart.* (2022) 108(17):1351–60. doi: 10.1136/heartjnl-2021-320131
- Fisher JD. New York Heart association classification. *Arch Intern Med.* (1972) 129(5):836. doi: 10.1001/archinte.1972.00320050160023
- Caraballo C, Desai NR, Mulder H, Alhanti B, Wilson FP, Fiuzat M, et al. Clinical implications of the New York heart association classification. *JAHA.* (2019) 8(23):e014240. doi: 10.1161/JAHA.119.014240
- Ammar KA, Jacobsen SJ, Mahoney DW, Kors JA, Redfield MM, Burnett JC, et al. Prevalence and prognostic significance of heart failure stages. *Circulation.* (2007) 115(12):1563–70. doi: 10.1161/CIRCULATIONAHA.106.666818
- Miller-Davis C, Marden S, Leidy NK. The New York heart association classes and functional status: what are we really measuring? *Heart Lung.* (2006) 35(4):217–24. doi: 10.1016/j.hrtlung.2006.01.003
- Raphael C, Briscoe C, Davies J, Whinnett ZI, Manisty C, Sutton R, et al. Limitations of the New York heart association functional classification system and self-reported walking distances in chronic heart failure. *Heart.* (2007) 93(4):476–82. doi: 10.1136/hrt.2006.089656
- El-Menayar A, Zubaid M, AlMahmeed W, Sulaiman K, AlNabti A, Singh R, et al. Killip classification in patients with acute coronary syndrome: insight from a multicenter registry. *Am J Emerg Med.* (2012) 30(1):97–103. doi: 10.1016/j.ajem.2010.10.011
- Guazzi M, Myers J, Abella J, Peberdy MA, Bensimhon D, Chase P, et al. The added prognostic value of ventilatory efficiency to the weber classification system in patients with heart failure. *Int J Cardiol.* (2008) 129(1):86–92. doi: 10.1016/j.ijcard.2007.05.028
- Rostagno C, Galanti G, Comeglio M, Boddi V, Olivo G, Serner GGN. Comparison of different methods of functional evaluation in patients with chronic heart failure. *Eur J Heart Fail.* (2000) 2(3):273–80. doi: 10.1016/S1388-9842(00)00091-X
- McDonagh J, Martin L, Ferguson C, Jha SR, Macdonald PS, Davidson PM, et al. Frailty assessment instruments in heart failure: a systematic review. *Eur J Cardiovasc Nurs.* (2018) 17(1):23–35. doi: 10.1177/1474515117708888
- Fried LP, Tangen CM, Walston J, Newman AB, Hirsch C, Gottdiener J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci.* (2001) 56(3):M146–56. doi: 10.1093/gerona/56.3.m146
- Guralnik JM, Simonsick EM, Ferrucci L, Glynn RJ, Berkman LF, Blazer DG, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol.* (1994) 49(2):M85–94. doi: 10.1093/geronj/49.2.M85
- Zhang Y, Yuan M, Gong M, Tse G, Li G, Liu T. Frailty and clinical outcomes in heart failure: a systematic review and meta-analysis. *J Am Med Dir Assoc.* (2018) 19(11):1003–1008.e1. doi: 10.1016/j.jamda.2018.06.009
- Konishi M. Scoring the physical frailty phenotype of patients with heart failure. *J Cachexia Sarcopenia Muscle.* (2022) 13(1):5–7. doi: 10.1002/jcsm.12883
- Kitai T, Shimogai T, Tang WHW, Iwata K, Xanthopoulos A, Otsuka S, et al. Short physical performance battery vs. 6-minute walking test in hospitalized elderly patients with heart failure. *Eur Heart J Open.* (2021) 1(1):oeab006. doi: 10.1093/ehjopen/oeab006
- Talha KM, Pandey A, Fudim M, Butler J, Anker SD, Khan MS. Frailty and heart failure: state-of-the-art review. *J Cachexia Sarcopenia Muscle.* (2023) 14(5):1959–72. doi: 10.1002/jcsm.13306
- Noda T, Kamiya K, Hamazaki N, Nozaki K, Ichikawa T, Yamashita M, et al. SARC-F predicts poor motor function, quality of life, and prognosis in older patients with cardiovascular disease and cognitive impairment. *Exp Gerontol.* (2023) 171:112021. doi: 10.1016/j.exger.2022.112021
- Kanenawa K, Isotani A, Yamaji K, Nakamura M, Tanaka Y, Hirose-Inui K, et al. The impact of frailty according to clinical frailty scale on clinical outcome in patients with heart failure. *ESC Heart Fail.* (2021) 8(2):1552–61. doi: 10.1002/ehf2.13254
- Sunaga A, Hikoso S, Yamada T, Yasumura Y, Uematsu M, Tamaki S, et al. Prognostic impact of clinical frailty scale in patients with heart failure with preserved ejection fraction. *ESC Heart Fail.* (2021) 8(4):3316–26. doi: 10.1002/ehf2.13482
- Somech J, Joshi A, Mancini R, Chetrit J, Michel C, Sheppard R, et al. Comparison of questionnaire and performance-based physical frailty scales to predict survival and health-related quality of life in patients with heart failure. *J Am Heart Assoc.* (2023) 12(6):e026951. doi: 10.1161/JAHA.122.026951
- Vitale C, Jankowska E, Hill L, Piepoli M, Doehner W, Anker SD, et al. Heart failure association of the European Society of Cardiology position paper on frailty in patients with heart failure. *Eur J Heart Fail.* (2019) 21(11):1299–305. doi: 10.1002/ehf.1611
- Villaschi A, Chiarito M, Pagnesi M, Stolfo D, Baldetti L, Lombardi CM, et al. Frailty according to the 2019 HFA-ESC definition in patients at risk for advanced heart failure: insights from the HELP-HF registry. *Eur J Heart Fail.* (2024) 26(6):1399–407. doi: 10.1002/ehf.3234
- Vidán MT, Blaya-Novakova V, Sánchez E, Ortiz J, Serra-Rexach JA, Bueno H. Prevalence and prognostic impact of frailty and its components in non-dependent elderly patients with heart failure. *Eur J Heart Fail.* (2016) 18(7):869–75. doi: 10.1002/ehf.518
- Khan MS, Segar MW, Usman MS, Singh S, Greene SJ, Fonarow GC, et al. Frailty, guideline-directed medical therapy, and outcomes in HFrEF: from the GUIDE-IT trial. *JACC Heart Fail.* (2022) 10(4):266–75. doi: 10.1016/j.jchf.2021.12.004
- Villani ER, Tummolo AM, Palmer K, Gravina EM, Vetrano DL, Bernabei R, et al. Frailty and atrial fibrillation: a systematic review. *Eur J Intern Med.* (2018) 56:33–8. doi: 10.1016/j.ejim.2018.04.018
- Lee DS, Stitt A, Austin PC, Stukel TA, Schull MJ, Chong A, et al. Prediction of heart failure mortality in emergent care. *Ann Intern Med.* (2012) 156(11):767–75. doi: 10.7326/0003-4819-156-11-201206050-00003
- Lee DS, Lee JS, Schull MJ, Borgundvaag B, Edmonds ML, Ivankovic M, et al. Prospective validation of the emergency heart failure mortality risk grade for acute heart failure. *Circulation.* (2019) 139(9):1146–56. doi: 10.1161/CIRCULATIONAHA.118.035509
- Miró Ò, Rossello X, Gil V, Martín-Sánchez FJ, Llorens P, Herrero-Puente P, et al. Predicting 30-day mortality for patients with acute heart failure in the emergency department. *Ann Intern Med.* (2017) 167(10):698–705. doi: 10.7326/M16-2726
- Wussler D, Kozuharov N, Sabti Z, Walter J, Strebel I, Scholl L, et al. External validation of the MEESI acute heart failure risk score. *Ann Intern Med.* (2019) 170(4):248–56. doi: 10.7326/M18-1967
- Peterson PN, Rumsfeld JS, Liang L, Albert NM, Hernandez AF, Peterson ED, et al. A validated risk score for in-hospital mortality in patients with heart failure from the American Heart Association get with the guidelines program. *Circ Cardiovasc Qual Outcomes.* (2010) 3(1):25–32. doi: 10.1161/CIRCOUTCOMES.109.854877
- Shiraishi Y, Kohsaka S, Abe T, Mizuno A, Goda A, Izumi Y, et al. Validation of the get with the guideline–heart failure risk score in Japanese patients and the potential improvement of its discrimination ability by the inclusion of B-type natriuretic peptide level. *Am Heart J.* (2016) 171(1):33–9. doi: 10.1016/j.ahj.2015.10.008
- Lyle M, Wan S, Murphree D, Bennett C, Wiley BM, Barsness G, et al. Predictive value of the get with the guidelines heart failure risk score in unselected cardiac intensive care unit patients. *J Am Heart Assoc.* (2020) 9(3):e012439. doi: 10.1161/JAHA.119.012439
- Aaronson KD, Schwartz JS, Chen TM, Wong KL, 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(12):2660–7. doi: 10.1161/01.CIR.95.12.2660
- Koelling TM, Joseph S, Aaronson KD. Heart failure survival score continues to predict clinical outcomes in patients with heart failure receiving β -blockers. *J Heart Lung Transplant.* (2004) 23(12):1414–22. doi: 10.1016/j.healun.2003.10.002
- Zugck C, Krüger C, Kell R, Körber S, Schellberg D, Kübler W, et al. Risk stratification in middle-aged patients with congestive heart failure: prospective comparison of the heart failure survival score (HFSS) and a simplified two-variable model. *Eur J Heart Fail.* (2001) 3(5):577–85. doi: 10.1016/S1388-9842(01)00167-2
- Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, et al. The Seattle heart failure model. *Circulation.* (2006) 113(11):1424–33. doi: 10.1161/CIRCULATIONAHA.105.584102
- Goda A, Williams P, Mancini D, Lund LH. Selecting patients for heart transplantation: comparison of the heart failure survival score (HFSS) and the Seattle heart failure model (SHFM). *J Heart Lung Transplant.* (2011) 30(11):1236–43. doi: 10.1016/j.healun.2011.05.012
- Canepa M, Fonseca C, Chioncel O, Laroche C, Crespo-Leiro MG, Coats AJS, et al. Performance of prognostic risk scores in chronic heart failure patients enrolled in the European Society of Cardiology heart failure long-term registry. *JACC Heart Fail.* (2018) 6(6):452–62. doi: 10.1016/j.jchf.2018.02.001
- Jia YY, Cui NQ, Jia TT, Song JP. Prognostic models for patients suffering a heart failure with a preserved ejection fraction: a systematic review. *ESC Heart Fail.* (2024) 11(3):1341–51. doi: 10.1002/ehf2.14696

42. Wong CM, Hawkins NM, Petrie MC, Jhund PS, Gardner RS, Ariti CA, et al. Heart failure in younger patients: the meta-analysis global group in chronic heart failure (MAGGIC). *Eur Heart J*. (2014) 35(39):2714–21. doi: 10.1093/eurheartj/ehu216
43. Sartipy U, Dahlström U, Edner M, Lund LH. Predicting survival in heart failure: validation of the MAGGIC heart failure risk score in 51 043 patients from the Swedish heart failure registry. *Eur J Heart Fail*. (2014) 16(2):173–9. doi: 10.1111/ehf.32
44. Agostoni P, Corrà U, Cattadori G, Veglia F, La Gioia R, Scardovi AB, et al. Metabolic exercise test data combined with cardiac and kidney indexes, the MECKI score: a multiparametric approach to heart failure prognosis. *Int J Cardiol*. (2013) 167(6):2710–8. doi: 10.1016/j.ijcard.2012.06.113
45. Fonarow GC, Adams KF, Abraham WT, Yancy CW, Boscardin WJ. ADHERE scientific advisory committee SG and investigators, for the. Risk stratification for in-hospital mortality in acutely decompensated heart failure classification and regression tree analysis. *JAMA*. (2005) 293(5):572–80. doi: 10.1001/jama.293.5.572
46. Palazzuoli A, Beltrami M. Are HFpEF and HFmrEF so different? The need to understand distinct phenotypes. *Front Cardiovasc Med*. (2021) 8. doi: 10.3389/fcvm.2021.676658
47. Simmonds SJ, Cuijpers I, Heymans S, Jones EAV. Cellular and molecular differences between HFpEF and HFmrEF: a step ahead in an improved pathological understanding. *Cells*. (2020) 9(1):242. doi: 10.3390/cells9010242
48. Sun Y, Wang N, Li X, Zhang Y, Yang J, Tse G, et al. Predictive value of H2FPEF score in patients with heart failure with preserved ejection fraction. *ESC Heart Failure*. (2021) 8(2):1244–52. doi: 10.1002/ehf2.13187
49. Sueta D, Yamamoto E, Nishihara T, Tokitsu T, Fujisue K, Oike F, et al. H2FPEF score as a prognostic value in HFpEF patients. *Am J Hypertens*. (2019) 32(11):1082–90. doi: 10.1093/ajh/hpz108
50. Komajda M, Carson PE, Hetzel S, McKelvie R, McMurray J, Ptaszynska A, et al. Factors associated with outcome in heart failure with preserved ejection fraction: findings from the Irbesartan in Heart Failure with Preserved Ejection Fraction Study (I-PRESERVE). *Circ Heart Fail*. (2011) 4(1):27–35. doi: 10.1161/CIRCHEARTFAILURE.109.932996
51. Carson PE, Anand IS, Win S, Rector T, Haass M, Lopez-Sendon J, et al. The hospitalization burden and post-hospitalization mortality risk in heart failure with preserved ejection fraction: results from the I-PRESERVE trial (irbesartan in heart failure and preserved ejection fraction). *JACC Heart Fail*. (2015) 3(6):429–41. doi: 10.1016/j.jchf.2014.12.017
52. Pfeffer MA, Claggett B, Assmann SF, Boineau R, Anand IS, Clausell N, et al. Regional variation in patients and outcomes in the treatment of preserved cardiac function heart failure with an aldosterone antagonist (TOPCAT). *Trial. Circulation*. (2015) 131(1):34–42. doi: 10.1161/CIRCULATIONAHA.114.013255
53. Silverman DN, Plante TB, Infeld M, Callas PW, Juraschek SP, Dougherty GB, et al. Association of β -blocker use with heart failure hospitalizations and cardiovascular disease mortality among patients with heart failure with a preserved ejection fraction: a secondary analysis of the TOPCAT trial. *JAMA Netw Open*. (2019) 2(12):e1916598. doi: 10.1001/jamanetworkopen.2019.16598
54. Pieske B, Tschöpe C, de Boer RA, Fraser AG, Anker SD, Donal E, et al. How to diagnose heart failure with preserved ejection fraction: the HFA-PEFF diagnostic algorithm: a consensus recommendation from the heart failure association (HFA) of the European Society of Cardiology (ESC). *Eur Heart J*. (2019) 40(40):3297–317. doi: 10.1093/eurheartj/ehz641
55. Gao YP, Liu HY, Bi XJ, Sun J, Zhu Y, Zhou W, et al. H2FPEF and HFA-PEFF scores for heart failure risk stratification in hypertrophic cardiomyopathy patients. *ESC Heart Fail*. (2025) 12(3):2225–38. doi: 10.1002/ehf2.15247
56. Thorvaldsen T, Claggett BL, Shah A, Cheng S, Agarwal SK, Wruck LM, et al. Predicting risk in patients hospitalized for acute decompensated heart failure and preserved ejection fraction: the atherosclerosis risk in communities study heart failure community surveillance. *Circ Heart Fail*. (2017) 10(12):e003992. doi: 10.1161/CIRCHEARTFAILURE.117.003992
57. Dsouza G, Sharma M. NT-proBNP in heart failure with preserved ejection fraction: a comprehensive review. *Indian J Clin Cardiol*. (2024) 5(4):372–84. doi: 10.1177/26324636241261422
58. 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(15):1182–9. doi: 10.1136/heartjnl-2018-314173
59. Angraal S, Mortazavi BJ, Gupta A, Khera R, Ahmad T, Desai NR, et al. Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. *JACC Heart Fail*. (2020) 8(1):12–21. doi: 10.1016/j.jchf.2019.06.013
60. Zhao X, Sui Y, Ruan X, Wang X, He K, Dong W, et al. A deep learning model for early risk prediction of heart failure with preserved ejection fraction by DNA methylation profiles combined with clinical features. *Clin Epigenet*. (2022) 14(1):11. doi: 10.1186/s13148-022-01232-8
61. Wang Z, Chen X, Tan X, Yang L, Kannapur K, Vincent JL, et al. Using deep learning to identify high-risk patients with heart failure with reduced ejection fraction. *JHEOR*. (2021) 8(2):6–13. doi: 10.36469/jheor.2021.25753
62. Cheema B, Pandit J. AI and heart failure: present state and future with multimodal large language models. *JACC Adv*. (2024) 3(9):101029. doi: 10.1016/j.jacadv.2024.101029
63. Zhao H, Li P, Zhong G, Xie K, Zhou H, Ning Y, et al. Machine learning models in heart failure with mildly reduced ejection fraction patients. *Front Cardiovasc Med*. (2022) 9:1042139. doi: 10.3389/fcvm.2022.1042139
64. Li J, Liu S, Hu Y, Zhu L, Mao Y, Liu J. Predicting mortality in intensive care unit patients with heart failure using an interpretable machine learning model: retrospective cohort study. *J Med Internet Res*. (2022) 24(8):e38082. doi: 10.2196/38082
65. Aggarwal R, Sounderajah V, Martin G, Ting DSW, Karthikesalingam A, King D, et al. Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis. *npj Digit Med*. (2021) 4(1):65. doi: 10.1038/s41746-021-00438-z
66. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. *NPJ Digital Med*. (2018) 1(1):18. doi: 10.1038/s41746-018-0029-1
67. Litjens G, Sánchez CI, Timofeeva N, Hermesen M, Nagtegaal I, Kovacs I, et al. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. *Sci Rep*. (2016) 6(1):26286. doi: 10.1038/srep26286
68. Raiaan MAK, Mukta MSH, Fatema K, Fahad NM, Sakib S, Mim MMJ, et al. A review on large language models: architectures, applications, taxonomies, open issues and challenges. *IEEE Access*. (2024) 12:26839–74. doi: 10.1109/ACCESS.2024.3365742
69. Tudjarski S, Gusev M, Kanoulas E. Transformer-based heart language model with electrocardiogram annotations. *Sci Rep*. (2025) 15(1):5522. doi: 10.1038/s41598-024-84270-x
70. Sood A, Craven M. Feature importance explanations for temporal black-box models. *Proc AAAI Conf Artif Intell*. (2022) 36(8):8351–60. doi: 10.1609/aaai.v36i8.20810
71. Kalusivalingam AK, Sharma A, Patel N, Singh V. Leveraging SHAP and LIME for enhanced explainability in AI-driven diagnostic systems. *Int J AI ML*. (2021) 2(3). Available online at: <https://cognitivecomputingjournal.com/index.php/IJAIML-V1/article/view/81> (Accessed March 31, 2025).