The role of artificial intelligence in the diagnosis, risk stratification, and treatment of heart failure: A narrative review

Julia Resch[A,D,E], Małgorzata Lelonek[E,F]

Department of Noninvasive Cardiology, Medical University of Lodz, Poland

A – research concept and design; B – collection and/or assembly of data; C – data analysis and interpretation;
D – writing the article; E – critical revision of the article; F – final approval of the article

Advances in Clinical and Experimental Medicine, ISSN 1899–5276 (print), ISSN 2451–2680 (online) *Adv Clin Exp Med*. 2026

Address for correspondence
Julia Resch
E-mail: julia.resch@umed.lodz.pl

Funding sources
None declared

Conflict of interest
None declared

Received on November 16, 2025
Reviewed on January 23, 2026
Accepted on February 25, 2026

Published online on July 3, 2026

Cite as
Resch J, Lelonek M. The role of artificial intelligence in the diagnosis, risk stratification, and treatment of heart failure: A narrative review [published online as ahead of print on July 3, 2026]. *Adv Clin Exp Med*. 2026. doi:10.17219/acem/218578

DOI
10.17219/acem/218578

Abstract

Heart failure (HF) is a complex, multifactorial, and difficult-to-treat syndrome. Over the past years, a concerning increase in its global prevalence, mortality, costs, and burden on the healthcare system has been observed. The recent development of machine learning (ML), especially unsupervised and deep learning (DL) algorithms, offers a potential way to facilitate diagnosis, enable more precise treatment, and reduce both mortality and costs of HF patients. Especially unsupervised ML and DL present new opportunities for increased efficiency in clinical practice in cardiology. Unsupervised ML, e.g., enables novel phenogrouping of HF patients into high-risk groups and disease outcome prediction. Deep learning algorithms can enhance echocardiographic analysis by improving image quality and ECG interpretation, and by providing assistance and guidance to inexperienced cardiologists. However, substantial challenges related to generalizability, external validation, overfitting, model explainability, and ethical considerations currently severely limit the implementation of AI-based tools in real-world clinical practice. This review critically evaluates current AI models in HF, focusing on their roles in diagnosis, risk stratification, and treatment personalization, as well as the major challenges that restrict their application in clinical practice.

Key words: heart failure, machine learning, deep learning, artificial intelligence, precision medicine

Highlights

- Artificial intelligence and machine learning are transforming heart failure management by improving diagnosis, risk stratification, and personalized treatment strategies.
- Unsupervised machine learning enables advanced phenogrouping of heart failure patients, helping to identify high-risk populations and predict clinical outcomes.
- Deep learning enhances cardiology imaging and ECG interpretation, supporting echocardiographic analysis and assisting less experienced clinicians in clinical decision-making.
- Key barriers to clinical AI implementation in heart failure include limited generalizability, overfitting, poor explainability, and ethical concerns, highlighting the need for robust validation and transparent models.

Introduction

Heart failure (HF) is a complex, multifactorial syndrome characterized by impaired cardiac pump function and/or structural or functional abnormalities of the heart, resulting in inadequate metabolic supply to the body.[1]

In 2021, the European Society of Cardiology (ESC) defined HF as "a clinical syndrome consisting of cardinal symptoms (e.g., breathlessness, ankle swelling, and fatigue) that may be accompanied by signs (e.g., elevated jugular venous pressure, pulmonary crackles, and peripheral edema). It is due to a structural and/or functional abnormality of the heart that results in elevated intracardiac pressures and/or inadequate cardiac output at rest and/or during exercise".[2] Traditionally, HF is classified into 3 phenotypes: HF with preserved ejection fraction (HFpEF), defined as left ventricular ejection fraction (LVEF) >50%; HF with mildly reduced ejection fraction (HFmrEF), defined as LVEF 41–49%; and HF with reduced ejection fraction (HFrEF), defined as ejection fraction (LVEF) <40%.[2]

The prevalence of HF has been steadily increasing over recent years, reaching more than 64 million cases worldwide.[1] Together with an estimated annual cost of up to €25,000 per patient and a 1-year mortality rate of up to 23.1% among patients with acute HF, this represents a substantial burden on the global healthcare system.[1]

In Poland, the number of patients with HF increased byapprox. 34% over the past 10 years, accompanied by a concerning decline in 1-year survival from 86% in 2014 to 76% in 2021.[3] A recent study conducted in Poland highlighted the importance of HF nurses, particularly the impact of their education on the diagnostic and therapeutic management of patients with HF. In 2021, a Polish online educational platform for nurses was introduced, including a certification course for HF nurses. According to the survey findings, this educational initiative led to a significant improvement in the quality of care and management of patients with HF.[4] Major challenges in HF include its complexity, multifactorial etiology, and substantial variability in pathomechanisms (Fig. 1), which not only make the diagnostic process highly time- and cost-intensive but may also result in suboptimal or ineffective treatment.[1]

Artificial intelligence (AI) is a rapidly evolving technology with considerable potential in medicine. It is expected to facilitate and accelerate diagnostic processes, individualize treatment, reduce clinicians' workload, and minimize healthcare expenditure, particularly in complex diseases such as HF.[5]

Objectives

This narrative review provides an overview of the most important AI models in the field of HF, their role in precision medicine, and the potential limitations of their clinical implementation.

Materials and methods

An electronic database search was conducted using PubMed. The database was searched for AI models developed and tested to support the diagnosis of HF, improve treatment precision in patients with HF, and assist clinical decision-making in HF management. Particular emphasis was placed on randomized controlled trials (RCTs) that included internal and/or external validation of the evaluated AI algorithms, as this is an important factor in assessing their clinical utility and potential implementation.

The search strategy included all possible combinations of the following terms: heart failure AND (artificial intelligence OR machine learning OR supervised learning OR unsupervised learning OR deep learning). As the field of AI extends beyond medicine, the following exclusion criteria were applied: 1) AI models not directly related to HF; 2) AI models without internal and/or external validation; 3) substudies describing the same algorithm; and 4) lack of freely available full text. The inclusion criteria were as follows: 1) AI models directly related to HF; 2) AI models with internal and/or external validation; 3) RCTs published within the last 5 years; and 4) freely available full text. One author (J.R.) screened the titles and abstracts of the identified articles according to the inclusion and exclusion criteria (Fig. 2). The selected articles were subsequently assessed

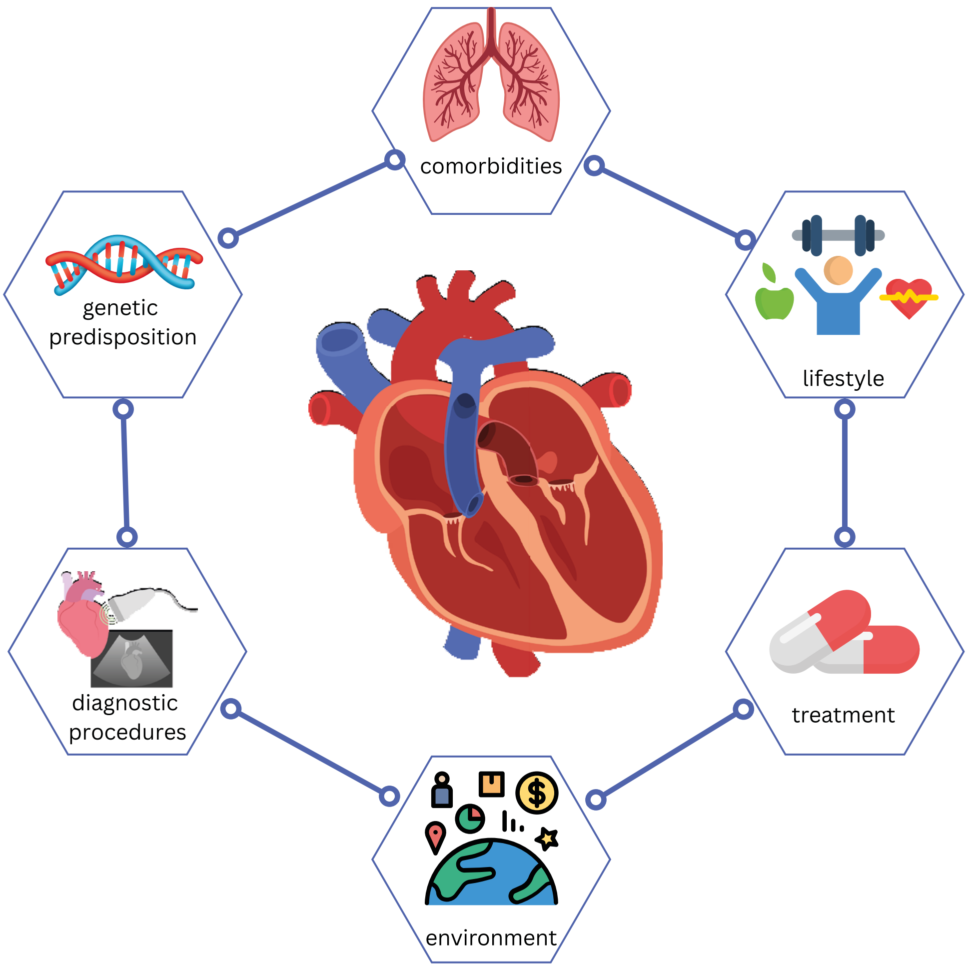

Fig. 1. Key factors influencing heart failure development and management

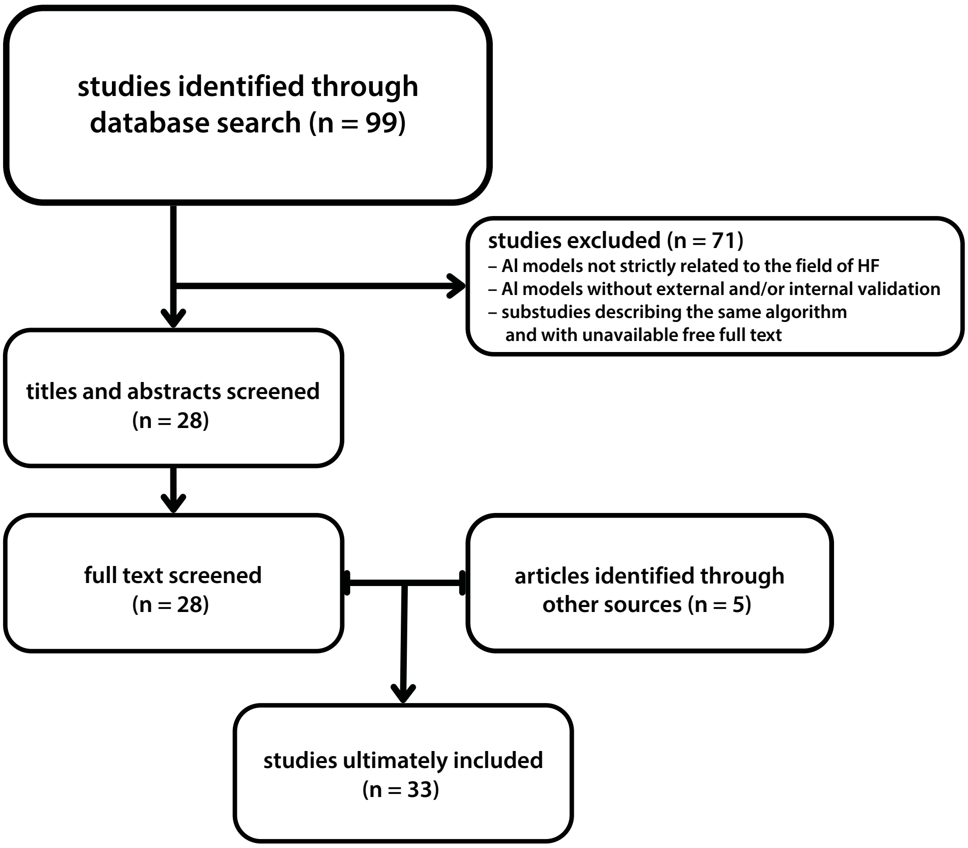

Fig. 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart illustrating the article screening process

HF – heart failure.

for potential selection bias and grouped into 2 categories: unsupervised learning AI algorithms and supervised learning AI algorithms. The most important studies in each category are presented in Table 1.[6–9] and Table 2.[10–12]

Techniques and methods of AI

AI is a branch of computer science focused on developing algorithms capable of performing tasks at or above human level.[5] In general, AI operates by mathematically transforming input data, so-called "features," into formats that can be processed by machine learning (ML) algorithms, typically using vector representations. Features are quantifiable data that may, in their simplest form, consist of clinical variables stored in a patient's health record (Table 3).[13]

Machine learning has been used in clinical applications since the 1970s and is based on algorithms that identify linear and nonlinear relationships between input data and labeled output data.[13] A further distinction can be made between supervised ML, unsupervised ML, and deep learning (DL), as shown in Fig. 3.

The most common type of machine learning is supervised ML, in which models are trained on labeled data. Input data are paired with corresponding output labels, enabling the model to learn the relationship between them and predict the output label for previously unseen data.[14] A common clinical application of supervised ML is the detection of electrocardiographic (ECG) abnormalities. A model is trained using a large dataset of ECG recordings labeled with diagnostic annotations corresponding to specific pathologies, such as atrial fibrillation or tachycardia. At the end of the training process, the model can predict the presence of previously learned pathological patterns in new ECGs. It is important to note that the model can only recognize patterns on which it has been trained; clinical validation by human experts remains essential.[15]

In contrast, unsupervised ML is primarily used for large and heterogeneous datasets. It analyzes large volumes of unlabeled data to identify recurring patterns within the input dataset.[16] In cardiology, unsupervised ML is used, e.g., for phenotyping patients with HF based on clinical variables, helping to identify high-risk individuals and support treatment individualization.[17]

Deep learning is a class of machine learning models based on artificial neural networks with multiple hidden layers, designed to enable automated learning of hierarchical data representations. It is a subset of ML and can be implemented using both supervised and unsupervised learning approaches.[14] Data flow through multiple neural network layers, where each neuron performs non-linear transformations, enabling learning with minimal manual feature engineering. Compared with traditional ML, DL is better suited for processing large and complex datasets, such as images and videos, and for performing highly complex predictive tasks with high accuracy. In the context of HF,

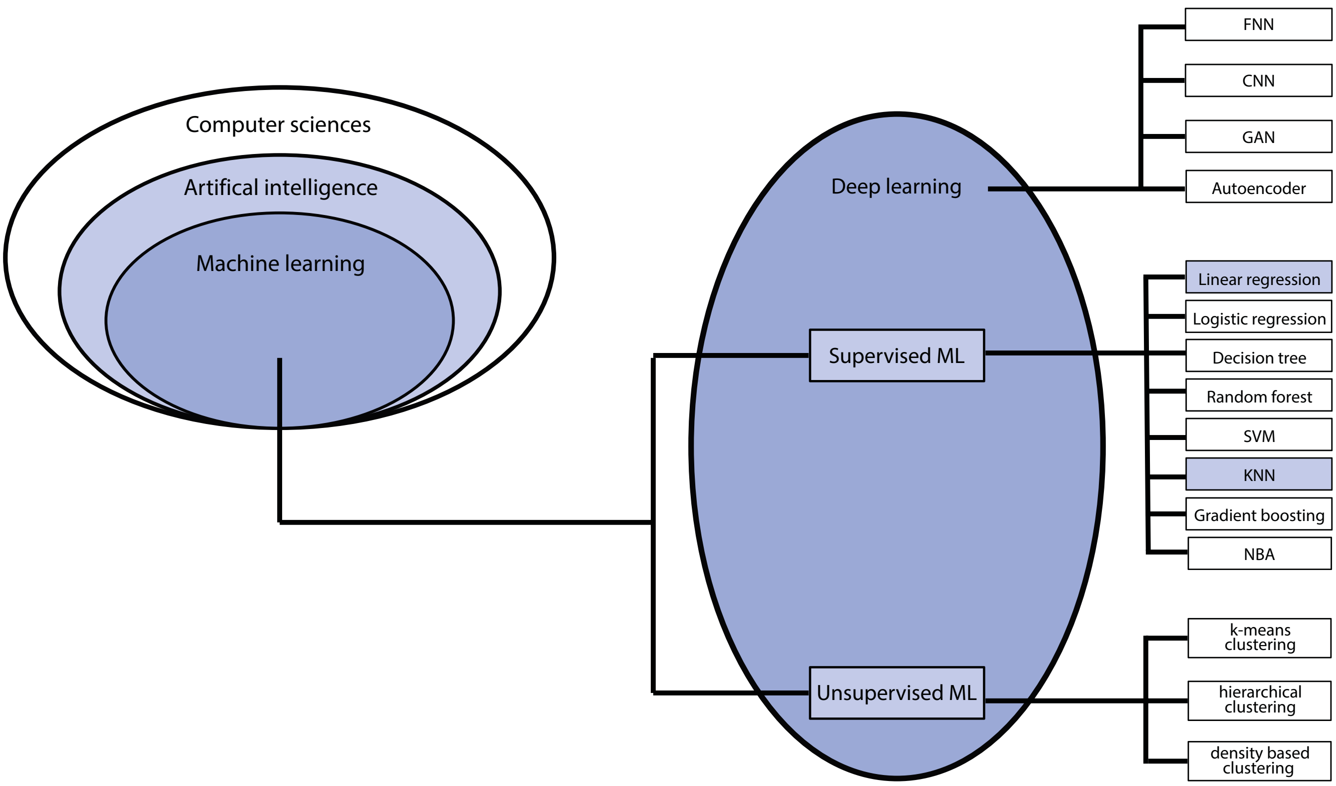

Fig. 3. Conceptual overview of artificial intelligence (AI), including machine learning (ML), deep learning, supervised learning, and unsupervised learning algorithms

CNN – convolutional neural network; FNN – feedforward neural network; GAN – generative adversarial network; KNN – k-nearest neighbour; NBA – naïve Bayes algorithm; SVM – support vector machine.

Table 1. Overview of the most recent and relevant studies regarding the use of unsupervised ML in the diagnosis of HF

Title	Authors	Year	Type of ML used	Type of study	Number of patients	Objective	Method	Outcome	External validation (Y = yes, N = no)	Conclusion
Clinical phenogroups are more effective than left ventricular ejection fraction categories in stratifying heart failure outcomes	Gevaert et al.[6]	2021	unsupervised ML (cluster analysis)	RCT	1693	Identification of clinical phenogroups in HF patients and prediction of clinical outcomes based on patient-specific clinical parameters.	Patients' parameters were extracted from the most recent examinations or the last 5 years' clinical records. A clustering algorithm was created and used to identify 6 phenogroups of HF patients. These were observed over the next 12 months, and clinical outcomes were measured. The predictive accuracy of this model was compared to LVEF-based groups.	Primary outcome: all-cause death or rehospitalization in the next 6 and 12 months. Secondary outcome: HF-related death or rehospitalization in the next 6–12 months.	N	The ML-based phenogroup classification provided superior prognostic information for all-cause and CV endpoints at 6 and 12 months compared with classification based on 3- or 6-LVEF categories, suggesting that comorbidities are major drivers of prognosis in HF patients.
The role of deep learning-based echocardiography in the diagnosis and evaluation of the effects of routine anti-heart-failure Western medicines in elderly patients with acute keft heart failure	Chen et al.[7]	2021	unsupervised ML (DCNN)	RCT	80	To evaluate whether ML-enhanced echocardiography increases accuracy and predictive strength compared to conventional echocardiography in patients with acute left heart failure.	The patients were divided into observation and control groups (n = 40). A DCNN algorithm was created for image processing, including binarized threshold segmentation for denoising and illumination processing to balance image brightness, which was used in the observational group. Both groups received standard pharmacological treatment for ALHF for 5 months.	Primary outcome: diagnostic accuracy of ML-enhanced echocardiography vs conventional. Secondary outcome: Hospitalization status of the patients 5 months after drug treatment.	N	DCNN-based echocardiography has high diagnostic accuracy, could reduce the risk of cardiovascular events in patients with heart failure, and could decrease mortality, diagnostic, and treatment costs.
Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy	Cikes et al.[8]	2018	unsupervised ML (multiple kernel learning and K-means clustering)	RCT	1106	To determine whether an ML-based algorithm that utilizes both complex echocardiographic data and clinical parameters can identify patients with a beneficial response to cardiac resynchronization therapy.	The patients' baseline parameters were collected, and they were divided into 2 groups: one receiving CRT-D and the other receiving ICD. Responder phenogroups were created using unsupervised ML.	Primary outcome: all-cause death or HF even. Secondary outcome: Development of an ML-based algorithm identifying responder phenogroups.	N	This analysis confirms the utility of unsupervised ML as a novel approach to integrating complex echocardiographic data with clinical parameters to phenotype patients with HF and may help optimize the rate of responders to certain treatments.
Machine learning-derived echocardiographic phenotypes predict heart failure incidence in asymptomatic individuals	Kobayashi et al.[9]	2021	unsupervised ML (cluster analysis)	RCT	827	To identify homogenous echocardiographic phenotypes of cardiac structure in community-based cohorts and assess their association with circulating biomarkers and CV outcomes.	Echocardiographic phenotypes were identified in asymptomatic patients using K-means clustering in the first generation of the STANISLAS Cohort, and their associations with vascular function and circulating biomarkers were assessed using ML-based algorithms.	Primary outcome: a composite of initial hospitalization for HF or cardiovascular death. Secondary outcome: a composite of cardiovascular mortality and hospitalization (HF, arrhythmia, or CAD) and biomarkers.	Y	ML-based echocardiographic phenotypes capture clinically meaningful variation in cardiac structure and function, are linked to specific biomarker profiles, and provide prognostic information beyond traditional echocardiographic definitions of diastolic dysfunction.

RCT – randomized controlled trial; ML – machine learning; HF – heart failure; LVEF – left ventricular ejection fraction; CV – cardiovascular; DCNN – deep convolutional neural network; ALHF – acute left heart failure; CRT – cardiac resynchronization therapy; CRT-D – cardiac resynchronization therapy defibrillator; ICD – implantable cardioverter-defibrillator; CAD – coronary artery disease.

Table 2. Overview of the most recent and relevant studies regarding the use of supervised ML in the diagnosis of HF

Title	Authors	Year	Type of ML used	Type of study	Number of patients	Objective	Method	Outcome	External validation (Y = yes, N = no)	Conclusion
Machine learning predicting atrial fibrillation as an adverse event in the warfarin and aspirin in reduced cardiac ejection fraction (WARCEF trial) (substudy of the WARCEF trial)	Gue et al.[10]	2023	supervised ML (AdaBoost, random forest)	randomized double-blind trial	2305	To evaluate the performance of machine learning in identifying the development of ischemic stroke in HF patients with reduced ejection fraction but without previous atrial fibrillation.	Patient data were retracted from the WARCEF trial. Data from patients who developed AF were compared with those who did not after warfarin and aspirin therapy, using ML-based models to identify factors predicting the development of atrial fibrillation in patients with heart failure.	Primary outcome: diagnosis of AF after warfarin and aspirin therapy. Secondary outcome: Development of an ML-based algorithm to identify characteristics for AF development.	N	Social factors may disproportionately increase the risk of atrial fibrillation in the underrepresented non-White patient groups with heart failure.
Development and validation of a prediction model based on machine learning algorithms for predicting the risk of heart failure in middle-aged and older US people with prediabetes or diabetes	Wang et al.[11]	2023	supervised ML (random forest)	RCT	3527	To develop and validate a machine learning-based prediction model for the risk of heart failure in patients with prediabetes or diabetes.	Patients with prediabetes or diabetes were selected, and baseline parameters were collected. Five ML algorithms were created to predict the risk of HF in those patients.	Primary outcome: identification of risk factors for the development of HF in patients with diabetes or prediabetes. Secondary outcome: Identification of the most accurate ML-based algorithm.	N	The risk of HF in middle-aged and elderly patients with prediabetes or diabetes can be accurately predicted using ML models, especially random forest.
Point-of-care screening for heart failure with reduced ejection fraction using artificial intelligence during ECG-enabled stethoscope examination	Bachtiger et al.[12]	2022	supervised ML (DL)	prospective, observational study	1050	To test an AI algorithm applied to a single-lead ECG, recorded during ECG-enabled stethoscope examination, to validate a potential point-of-care screening tool for LVEF ≤ 40%.	A 15-s single-lead ECG was performed on patients undergoing TTE. The ECG was performed in 4 standard anatomical positions for cardiac auscultation using a convolutional neuronal network (AI-ECG) enabled stethoscope. TTE was also performed, and the LVEF percentage was compared with the single-lead AI-ECG results.	Primary outcome: performance of AI-ECG in classifying reduced LVEF (LVEF ≤ 40%).	Y	AI-ECG could identify patients with reduced LVEF (≤40%) from single-lead ECG inputs.

ML – machine learning; HF – heart failure; AF – atrial fibrillation; RCT – randomized controlled trial; DL – deep learning; AI – artificial intelligence; ECG – electrocardiogram; LVEF – left ventricular ejection fraction; TTE – transthoracic echocardiograph.

Table 3. Overview over the different types of ML relevant in clinical medicine

Type of ML model	Explanation	Classification/ prediction	Goals	Examples
Supervised machine learning (ML)				
Linear regression	The model learns from labeled continuous input data, assuming a linear relationship between the input and output. It uses the most accurate regression for data analysis.	P	To predict unlabeled data	Interpretation of ECGs
Logistic regression	The model learns from labeled binary input data, with outputs in one of 2 classes. It uses sigmoid functions to map the input data to probabilities between 0 and 1.	C	To predict the probability of a value belonging to one class	Estimating the probability of having a disease
Decision tree	A diagram showing different options (branches) for solving a problem. The model tries to find the best option.	P + C	To assist in finding the best solution for a problem	Development of decision-making assisting systems in hospitals
Random forest	An assemblage of decision trees that combines (averages) predictions of multiple trees to increase its accuracy.	P + C	To assist in finding the best solution for a problem	Development of decision-making assisting systems in hospitals
Support vector machine (SVM)	The model tries to find the best hyperplane (boundary) between different classes in data.	P + C	To find the best decision boundary	Development of models to predict and detect heart diseases
K-nearest neighbors (KNN)	The model tries to find the closest neighboring data points (k) to the input data and makes a decision based on the most common class of the neighbors.	P + C	Predicts unlabeled data according to the closest neighboring points	Development of models for assisting in the diagnosis and individualizing treatment of heart diseases
Gradient boosting	This model combines multiple weak learner algorithms to subsequently reduce errors.	P + C	To reduce errors and increase the accuracy of algorithms	Development of models for mortality prediction of HF patients
Naïve Bayes algorithm (NBA)	The model predicts the outcome based solely on the input data features, independent of other data.	C	Prediction which category one data point belongs to	Risk prediction models in HF patients
Unsupervised ML				
K-means clustering	The model categorizes unlabeled data into K clusters based on similarity, then groups new data around the centroids.	N	Clustering of unlabeled data	Phenogrouping of patients with HF
Hierarchical clustering	The model clusters unlabeled data by using dendrograms, where each dataset is initially seen as its own cluster and later associated with the most similar data.	N	Clustering of unlabeled data	Phenogrouping of patients with HF
Density-based clustering	The model clusters unlabeled data according to density and reliably identifies noise and outliers.	N	Clustering of unlabeled data and detection of noise/outliers	Phenogrouping of patients with HF
Deep learning (DL)				
Feedforward neural networks (FNN)	A neural network in which information flows only unidirectionally (from input to output layer).	C	Pattern recognition tasks	Analysis of ECG pictures for hidden features
Convolutional neural network (CNN)	An advanced neural network that specializes in extracting data from grid-like matrices, like images or videos.	N	Convolution of grid matrices	Enhancement of echocardiographic images
Generative adversarial networks (GAN)	Two networks, a generator and a discriminator, help machines generate new, creative data from examples.	N	Generation of new data (images)	Image processing in cardiac-related processes
Autoencoder	A neural network specialized in compressing data.	N	Reducing noise, error reduction and feature detection	Predicting outcomes of cardiac diseases

P – prediction; C – classification; N – none; ECG – electrocardiography.

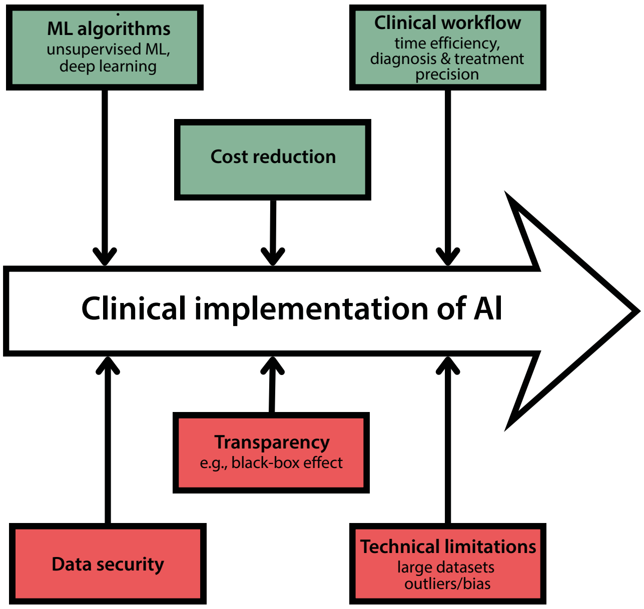

Fig. 4. Schematic representation of the advantages and limitations of implementing artificial intelligence (AI) in clinical practice

ML - machine learning.

DL algorithms are used, e.g., to automate echocardiographic analysis, identify predictive patterns in imaging data associated with future cardiovascular events, and support diagnosis and outcome prediction in patients with HF.[18]

AI and heart failure

The recent emergence of unsupervised ML and DL has opened new avenues for the diagnosis and management of HF.[19] AI-driven methodologies offer the potential to overcome current limitations, such as multifactorial pathophysiology, imprecise classification, and suboptimal therapeutic approaches. Table 1 and Table 2[6–12] provide an overview of the most recent advances in ML-based algorithms for HF.

Phenogrouping of patients with heart failure

To date, patients with HF have been classified primarily according to LVEF. However, this classification does not adequately reflect the complexity, multifactorial etiology, and burden of comorbidities in individual patients with HF. As a result, current management often focuses predominantly on symptom control rather than addressing the underlying causes of HF.

Several studies have proposed unsupervised ML models to identify novel phenogroups among patients with HF. These models incorporate clinical characteristics, comorbidities, and underlying pathophysiological mechanisms rather than relying solely on LVEF, offering the potential for more individualized outcome prediction and treatment strategies.

Gaevert et al., e.g., used AI-based cluster analysis to identify 6 phenogroups of patients with HF and stratify outcomes into risk groups. Each phenogroup was characterized by a predominant comorbidity: coronary heart disease (CAD), valvular heart disease, atrial fibrillation (AF), sleep apnea, chronic obstructive pulmonary disease (COPD), or few comorbidities. All groups were independent of LVEF. A 12-month follow-up analysis of the primary endpoints suggested that this comorbidity-based classification could predict HF outcomes more accurately than the traditional LVEF-based classification.[6]

A similar study was performed by Urban et al., who developed an unsupervised ML-based model identifying 6 phenogroups of patients with acute HF based on 63 different parameters. The mortality rate of each phenogroup was evaluated over the following year, and the researchers observed statistically significant differences in 1-year mortality ($p = 0.002$), suggesting potential value for future risk stratification in patients with HF.[17]

In addition to phenotyping patients who have already developed HF, researchers have sought to identify individuals at increased risk of developing HF, thereby enabling more effective preventive strategies in high-risk populations. For example, a study from the USA analyzed prediabetic and diabetic patients using an AI-based random forest model to predict individual risk of developing HF (area under the curve (AUC) = 0.978 in the training set and AUC = 0.865 in the test set). The study identified age, poverty-to-income ratio, prior myocardial infarction, CAD, chest pain, and glucose-lowering medication use as independent predictors of HF ($p < 0.05$).[11]

Although these models show promising potential for improving outcome prediction in patients with HF and supporting treatment individualization, most phenogrouping models remain hypothesis-generating and have not been implemented in clinical practice. This is due to several critical limitations, including limited generalizability, overfitting, algorithmic bias, and a lack of prospective validation studies. For example, hierarchical clustering algorithms, which are commonly used in phenogrouping models, may prioritize categorical variables over continuous ones, leading to inappropriate weighting of certain variables.

AI clinical support system

Choi et al. developed an AI-based clinical decision support system (AI-CDSS) for HF diagnosis and evaluated its diagnostic accuracy. The concordance rate between the AI-CDSS and HF specialists was 98%, whereas the concordance rate between non-HF specialists and HF specialists was 76%. These findings suggest that AI-based models may be useful in identifying HF, particularly in settings with limited access to specialists and constrained healthcare resources. However, it is important to note that the AI-CDSS is a supportive tool and has only seen regional use due to limited generalizability.[20]

Transthoracic echocardiography

Echocardiography is a time-consuming and costly tool for the diagnosis of HF, with a substantial risk of variability due to differences in operator expertise and subjective interpretation of echocardiographic findings. AI-powered models, particularly ML-based approaches, may reduce inter-observer variability and provide a cost-effective, automated screening tool.[16] In 2021, a deep convolutional neural network (DCNN) model was developed in China to generate more stable, noise-reduced real-time echocardiographic images during conventional echocardiographic examination. Transthoracic echocardiography (TTE) supported by an ML-based program demonstrated higher diagnostic accuracy than conventional echocardiography alone, potentially reducing the risk of cardiovascular events in patients with HF while lowering mortality, diagnostic costs, and treatment-related expenditures.[18]

AGILE-Echo has developed an innovative program designed to assist less experienced physicians in performing TTE assessments by integrating AI-based guidance during the examination. AI-guided TTE demonstrated superior accuracy compared with non-AI-assisted examinations, highlighting its potential to enhance the quality and consistency of cardiac imaging. This advancement holds significant promise for improving the timely and accurate diagnosis of HF and valvular heart disease, particularly in rural, remote, or resource-limited settings.

Despite the positive results of these studies, neither program has yet been implemented in routine clinical practice due to limited generalizability, logistical constraints, lack of long-term follow-up data, and insufficient time-to-event data. Moreover, neither model has undergone external validation while all carry a risk of overfitting, and therefore cannot currently be considered generalizable.[21]

ECG and heart tones

Kagiyama et al. recently published a noteworthy study describing the development of an AI-assisted program that analyzes ECG data to identify early left ventricular diastolic dysfunction (LVDD) by quantifying myocardial relaxation. The model accurately predicted LVEF across HF categories, with AUCs of 0.84, 0.80, and 0.81. This approach has the potential to serve as an ECG-free, cost-effective screening tool for the early detection of LVDD. However, as the model was based on a limited set of clinical parameters and lacked follow-up data analysis, further refinement is required before clinical implementation.[22]

Another model, based on a convolutional neural network (CNN), was developed to detect early LVDD by analyzing recorded heart sounds. The model achieved high performance metrics (accuracy: 0.987, sensitivity: 0.986, specificity: 0.988). Nevertheless, several limitations – including single-center data collection, sensitivity to recording conditions, lack of external validation, its adjunctive nature, and the limited interpretability of DL models – currently hinder its clinical implementation.[23]

Invasive treatment methods

In 2018, an unsupervised ML-based model was developed to support patient selection for cardiac resynchronization therapy (CRT). The model demonstrated superior predictive performance compared with traditional selection criteria when evaluated against clinical outcomes.[8] Another model using unsupervised ML was developed to identify positive and negative risk factors for HF development in elderly patients undergoing coronary rotational atherectomy (CRA).[24] Although both models demonstrated promising predictive performance, they were developed retrospectively using small datasets and have not undergone external validation, limiting their potential clinical implementation.

Pharmacological treatment

Machine learning-based algorithms offer considerable potential for identifying high-risk patients and supporting treatment individualization. A recent study by Bayes-Genis et al. used artificial neural network (ANN) algorithms to identify the most probable mechanism of action of empagliflozin, a sodium-glucose co-transporter 2 inhibitor (SGLT2i), and subsequently linked these findings to the gene expression profiles of patients with HFpEF. The identified mechanism of action of empagliflozin may contribute to a better understanding of the clinical benefits of SGLT2i in HFpEF and help guide more individualized treatment strategies. However, a major limitation of this study is its small sample size, which may result in an inaccurate representation of the relative contribution of different mechanisms of action of SGLT2i. Further research with larger sample sizes is needed.[25]

The HOMAGE trial developed a model to identify high-risk patients with HF based on echocardiographic features and evaluated their response to spironolactone, finding that a phenogroup characterized by a high proportion of women and elevated blood pressure appeared to benefit from the anti-remodeling effects of spironolactone. However, as this was a retrospective analysis, the findings should be considered hypothesis-generating.[26]

AI in clinical practice

Clinical implementation and regulatory approval

Despite their considerable potential, the implementation of AI-based models in real-world clinical practice remains in its early stages. Supervised and unsupervised ML models have experimentally demonstrated the potential to outperform conventional diagnostic and risk stratification methods, improve clinical workflow and efficiency, and provide clinician support, while generally receiving positive acceptance among physicians.[27,28] Moreover, AI-based models may reduce costs and lessen the burden on the healthcare system (Fig. 4).[29] However, the number of AI-based models approved for and implemented in clinical practice remains low compared with the number of experimentally developed models. Currently, regulatory approval of AI-based tools in HF is largely limited to screening, automated imaging quantification, and clinical decision support functions. Examples of approved tools in these areas include EchoGo® Heart Failure (cleared in 2022), an AI-based tool designed to assist in the diagnosis of HFpEF using echocardiographic features.[30]

Additional examples include Eko Low EF AI (cleared in 2024), an AI tool integrated into a stethoscope that analyzes ECG signals and heart sounds during auscultation,[31] and Anumana ECG-AI LEF (cleared in 2023), medical software that analyzes 12-lead ECGs to assess LVEF and detect EF < 40%.[19,32]

Regulatory approval of AI tools in HF is subject to strict requirements: human oversight is always required, and AI tools cannot replace standard HF diagnostic processes; a narrowly defined intended use must be specified, and only locked algorithms may be used, meaning that the algorithm cannot continue learning during clinical deployment. These models are approved as adjunctive tools for clinicians rather than autonomous diagnostic or therapeutic systems. It is important to note that regulatory approval of AI models does not necessarily translate into improved patient outcomes.

Limitations of AI models in heart failure

Limited generalizability is a major barrier to the clinical implementation of AI-based models. These models are often developed and trained in single-center settings and evaluated only in internal test environments rather than in real-world clinical practice. As a result, models may demonstrate high predictive accuracy in controlled test settings but substantially lower performance when applied to broader patient populations. This is particularly problematic in countries such as the USA, where certain ethnic groups, including Hispanic and Black populations, remain underrepresented in clinical trials, thereby increasing the risk of perpetuating bias in ML-based algorithms.

Moreover, there is concern that entire countries or regions may be excluded from AI research due to financial disparities.[33] External validation is essential to demonstrate a model's generalizability; however, in real-world settings, it is often limited by the difficulty of obtaining sufficiently large and heterogeneous datasets.[30]

Another major challenge is the risk of overfitting in ML-based models. These models require large and diverse datasets to ensure robust pattern recognition. Particularly in complex, multifactorial medical problems, insufficient training data may lead algorithms to capture noise, spurious associations, and random fluctuations rather than true underlying relationships. This results in impaired model performance and violates the principle of parsimony.[34]

Limited explainability, often referred to as the "black box" nature of algorithms such as deep neural networks (DNNs), represents another major barrier to the implementation of AI models. This term refers to opaque and difficult-to-interpret internal decision-making processes, which may reduce trust and acceptance among physicians, patients, and legal stakeholders.[35]

Outliers may also adversely affect AI-driven models, as extreme values can distort pattern recognition and reduce predictive performance. Increasing dataset size and diversity may mitigate the impact of outliers, thereby improving overall model robustness.[34]

Future directions

Despite the initial clinical implementation of AI-based models for HF screening, image analysis, and clinical decision support, several major challenges remain. Current AI-based models cannot replace complex therapeutic decision-making and therefore serve primarily as adjunctive tools. Multimorbidity is not yet adequately captured by existing AI models, limiting the scope of their clinical implementation. Furthermore, ethical considerations, including patient preferences and data security, remain unresolved. Addressing these issues should be a priority in future model development and validation.[36]

Conclusions

AI shows considerable potential for enhancing HF care by improving diagnostic accuracy, risk stratification, treatment individualization, and workflow efficiency, while reducing costs and the burden on the healthcare system. However, currently implemented AI tools remain adjunctive and cannot replace clinical judgment, comprehensive decision-making, or individualized care in complex, multimorbid patients. Clinicians should regard AI as a supportive tool that can supplement, but not replace, specialist clinical assessment and decision-making, while remaining mindful of its limitations regarding generalizability and ethical considerations.

Use of AI and AI-assisted technologies

Not applicable.

ORCID iDs

Julia Resch https://orcid.org/0009-0003-2867-3506
Małgorzata Lelonek https://orcid.org/0000-0003-0756-5541

References

1. Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GMC, Coats AJS. Global burden of heart failure: A comprehensive and updated review of epidemiology. *Cardiovasc Res*. 2023;118(17):3272–3287. doi:10.1093/cvr/cvac013
2. McDonagh TA, Metra M, Adamo M, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J*. 2021;42(36):3599–3726. doi:10.1093/eurheartj/ehab368
3. Bohdan M, Kowalczys A, Nessler J, Straburzyńska-Migaj E, Gruchała M, Lelonek M. Heart failure in Poland: A 20-year epidemiological perspective. *Medicina (Kaunas)*. 2025;61(8):1472. doi:10.3390/medicina61081472
4. Bohdan M, Furman-Niedziejko A, Kałużna-Oleksy M, et al. Importance of an online educational platform for heart failure nurses in Poland: The results of the survey. *Cardiol J*. 2025;32(6):607–614. doi:10.5603/cj.105494
5. Xie Y, Zhang L, Sun W, et al. Artificial intelligence in diagnosis of heart failure. *J Am Heart Assoc*. 2025;14(8):e039511. doi:10.1161/JAHA.124.039511
6. Gevaert AB, Tibebu S, Mamas MA, et al. Clinical phenogroups are more effective than left ventricular ejection fraction categories in stratifying heart failure outcomes. *ESC Heart Fail*. 2021;8(4):2741–2754. doi:10.1002/ehf2.13344
7. Chen J, Gao Y. The role of deep learning-based echocardiography in the diagnosis and evaluation of the effects of routine anti-heart-failure Western medicines in elderly patients with acute left heart failure. *J Healthcare Eng*. 2021;2021:4845792. doi:10.1155/2021/4845792
8. Cikes M, Sanchez-Martinez S, Claggett B, et al. Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy. *Eur J Heart Fail*. 2019;21(1):74–85. doi:10.1002/ejhf.1333
9. Kobayashi M, Huttin O, Magnusson M, et al. Machine learning-derived echocardiographic phenotypes predict heart failure incidence in asymptomatic individuals. *JACC Cardiovasc Imaging*. 2022;15(2):193–208. doi:10.1016/j.jcmg.2021.07.004
10. Gue Y, Correa E, Thompson JLP, Homma S, Qian M, Lip GYH. Machine learning predicting atrial fibrillation as an adverse event in the Warfarin and Aspirin in Reduced Cardiac Ejection Fraction (WARCEF) trial. *Am J Med*. 2023;136(11):1099–1108.e2. doi:10.1016/j.amjmed.2023.07.019
11. Wang Y, Hou R, Ni B, Jiang Y, Zhang Y. Development and validation of a prediction model based on machine learning algorithms for predicting the risk of heart failure in middle-aged and older US people with prediabetes or diabetes. *Clin Cardiol*. 2023;46(10):1234–1243. doi:10.1002/clc.24104
12. Bachtiger P, Petri CF, Scott FE, et al. Point-of-care screening for heart failure with reduced ejection fraction using artificial intelligence during ECG-enabled stethoscope examination in London, UK: A prospective, observational, multicentre study. *Lancet Digit Health*. 2022;4(2):e117–e125. doi:10.1016/S2589-7500(21)00256-9
13. Feeny AK, Chung MK, Madabhushi A, et al. Artificial intelligence and machine learning in arrhythmias and cardiac electrophysiology. *Circ Arrhytm Electrophysiol*. 2020;13(8):e007952. doi:10.1161/CIRCEP.119.007952
14. Komuro J, Kusumoto D, Hashimoto H, Yuasa S. Machine learning in cardiology: Clinical application and basic research. *J Cardiol*. 2023;82(2):128–133. doi:10.1016/j.jjcc.2023.04.020
15. Sraitih M, Jabrane Y, Hajjam El Hassani A. An automated system for ECG arrhythmia detection using machine learning techniques. *J Clin Med*. 2021;10(22):5450. doi:10.3390/jcm10225450
16. Khan MS, Arshad MS, Greene SJ, et al. Artificial intelligence and heart failure: A state-of-the-art review. *Eur J Heart Fail*. 2023;25(9):1507–1525. doi:10.1002/ejhf.2994
17. Urban S, Błaziak M, Jura M, et al. Novel phenotyping for acute heart failure: Unsupervised machine learning-based approach. *Biomedicines*. 2022;10(7):1514. doi:10.3390/biomedicines10071514
18. Lau ES, Di Achille P, Kopparapu K, et al. Deep learning-enabled assessment of left heart structure and function predicts cardiovascular outcomes. *J Am Coll Cardiol*. 2023;82(20):1936–1948. doi:10.1016/j.jacc.2023.09.800
19. Yasmin F, Shah SMI, Naeem A, et al. Artificial intelligence in the diagnosis and detection of heart failure: The past, present, and future. *Rev Cardiovasc Med*. 2021;22(4):1095. doi:10.31083/j.rcm2204121
20. Choi DJ, Park JJ, Ali T, Lee S. Artificial intelligence for the diagnosis of heart failure. *NPJ Digit Med*. 2020;3(1):54. doi:10.1038/s41746-020-0261-3
21. Soh CH, Wright L, Baumann A, et al. Use of artificial intelligence-guided echocardiography to detect cardiac dysfunction and heart valve disease in rural and remote areas: Rationale and design of the AGILE-echo trial. *Am Heart J*. 2024;277:11–19. doi:10.1016/j.ahj.2024.08.004
22. Kagiyama N, Piccirilli M, Yanamala N, et al. Machine learning assessment of left ventricular diastolic function based on electrocardiographic features. *J Am Coll Cardiol*. 2020;76(8):930–941. doi:10.1016/j.jacc.2020.06.061
23. Yang Y, Guo XM, Wang H, Zheng YN. Deep learning-based heart sound analysis for left ventricular diastolic dysfunction diagnosis. *Diagnostics (Basel)*. 2021;11(12):2349. doi:10.3390/diagnostics11122349
24. Zhang L, Zhou X, Cao J. Establishment and validation of a heart failure risk prediction model for elderly patients after coronary rotational atherectomy based on machine learning. *PeerJ*. 2024;12:e16867. doi:10.7717/peerj.16867
25. Bayes-Genis A, Iborra-Egea O, Spitaleri G, et al. Decoding empagliflozin's molecular mechanism of action in heart failure with preserved ejection fraction using artificial intelligence. *Sci Rep*. 2021;11(1):12025. doi:10.1038/s41598-021-91546-z
26. Kobayashi M, Huttin O, Ferreira JP, et al. A machine learning-derived echocardiographic algorithm identifies people at risk of heart failure with distinct cardiac structure, function, and response to spironolactone: Findings from the HOMAGE trial. *Eur J Heart Fail*. 2023;25(8):1284–1289. doi:10.1002/ejhf.2859
27. Karatzia L, Aung N, Aksentijevic D. Artificial intelligence in cardiology: Hope for the future and power for the present. *Front Cardiovasc Med*. 2022;9:945726. doi:10.3389/fcvm.2022.945726
28. Biondi-Zocca G. Artificial intelligence in cardiology: General perspectives and focus on interventional cardiology. *Anatol J Cardiol*. 2025;29(4):152–163. doi:10.14744/AnatolJCardiol.2025.5237
29. Bayne J, Garry J, Albert MA. Brief review: Racial and ethnic disparities in cardiovascular care with a focus on congenital heart disease and precision medicine. *Curr Atheroscler Rep*. 2023;25(5):189–195. doi:10.1007/s11883-023-01093-3
30. Akerman AP, Al-Roub N, Angell-James C, et al. External validation of artificial intelligence for detection of heart failure with preserved ejection fraction. *Nat Commun*. 2025;16(1):2915. doi:10.1038/s41467-025-58283-7
31. Guo L, Pressman GS, Kieu SN, et al. Automated detection of reduced ejection fraction using an ECG-enabled digital stethoscope. *JACC Adv*. 2025;4(3):101619. doi:10.1016/j.jacadv.2025.101619
32. Lopez-Jimenez F, Alger HM, Attia ZI, et al. A multicenter pragmatic implementation study of AI-ECG-based clinical decision support software to identify low LVEF: Clinical trial design and methods. *Am Heart J Plus*. 2025;54:100528. doi:10.1016/j.ahjo.2025.100528
33. Voermans AM, Mewes JC, Broyles MR, Steuten LMG. Cost-effectiveness analysis of a procalcitonin-guided decision algorithm for antibiotic stewardship using real-world U.S. hospital data. *OMICS*. 2019;23(10):508–515. doi:10.1089/omi.2019.0113
34. Yin J, Ngiam KY, Teo HH. Role of artificial intelligence applications in real-life clinical practice: Systematic review. *J Med Internet Res*. 2021;23(4):e25759. doi:10.2196/25759
35. Amann J, Blasimme A, Vayena E, Frey D, Madai VI. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Med Inform Decis Mak*. 2020;20(1):310. doi:10.1186/s12911-020-01332-6
36. Szczepanowski R, Uchmanowicz I, Pasieczna-Dixit AH, et al. Application of machine learning in predicting frailty syndrome in patients with heart failure. *Adv Clin Exp Med*. 2024;33(3):309–315. doi:10.17219/acem/184040