REVIEW

Clinical Application of Machine Learning-Based Artificial Intelligence in the Diagnosis, Prediction, and Classification of Cardiovascular Diseases

Songren Shu, MD; Jie Ren, MD; Jiangping Song, MD, PhD

With the rapid development of artificial intelligence (AI) and machine learning (ML), as well as the arrival of the big data era, technological innovations have occurred in the field of cardiovascular medicine. First, the diagnosis of cardiovascular diseases (CVDs) is highly dependent on assistive examinations, the interpretation of which is time consuming and often limited by the knowledge level and clinical experience of doctors; however, AI could be used to automatically interpret the images obtained in auxiliary examinations. Second, some of the predictions of the incidence and prognosis of CVDs are limited in clinical practice by the use of traditional prediction models, but there may be occasions when AI-based prediction models perform well by using ML algorithms. Third, AI has been used to assist precise classification of CVDs by integrating a variety of medical data from patients, which helps better characterize the subgroups of heterogeneous diseases. To help clinicians better understand the applications of AI in CVDs, this review summarizes studies relating to AI-based diagnosis, prediction, and classification of CVDs. Finally, we discuss the challenges of applying AI to cardiovascular medicine.

Key Words: Artificial intelligence; Big data; Cardiovascular diseases; Machine learning; Precision medicine

Cardiovascular diseases (CVDs) are the leading cause of death in humans, currently accounting for approximately one-third of all deaths worldwide.[1] Although considerable progress has been made in the management of CVDs in recent years, many challenges remain in clinical practice. First, the diagnosis of CVDs is highly dependent on electrocardiograms (ECGs) and/or cardiovascular imaging, the interpretation of which is time consuming and requires experience.[2] Second, most currently used prediction models of CVDs are based on traditional statistical methods, limiting prediction performance.[3] Third, the "one-size-fits-all" management concept in the clinic is not helpful for prognosis because it ignores the heterogeneity of CVDs.[4]

Artificial intelligence (AI) is regarded as a revolutionary frontier technology, and it has become a global research interest in the field of medicine, especially in cardiovascular medicine. AI could help clinicians overcome the aforementioned 3 challenges by automatically interpreting ECGs and/or cardiovascular imaging results, building more powerful prediction models, and characterizing subgroups of CVDs.

In this review we provide an overview of the applications of machine learning (ML)-based AI in the diagnosis, prediction, and classification of CVDs. The objective of this review is to help practicing clinicians better understand 3 key questions: (1) what are AI, ML, and algorithms; (2) what tasks can ML-based AI perform in cardiovascular clinical practice; and (3) what is the general workflow to conduct a study associated with AI-based diagnosis, prediction, or classification in CVDs?

AI, ML, and Algorithms

AI is a field of computer science that aims to mimic human thought processes, learning capacity, and knowledge storage.[5] ML, one of the methods to realize AI, usually refers to the process by which a system obtains information from data through algorithms. ML can be roughly divided into supervised and unsupervised learning. There are 2 main differences between supervised and unsupervised learning. First, supervised learning uses data that have been tagged with 1 or more labels, like properties, characteristics, or classifications, whereas unsupervised learning uses data that have not been tagged.[5] Second, supervised learning is focused on classification, which involves classifying an observation into several subsets (e.g., classifying an ECG into atrial fibrillation (AF), sinus rhythm, or other), and

Table 1. Artificial Intelligence-Assisted Diagnosis of CVDs Based on Electrocardiograms

Study	Objectives	Sample size	Algorithm/ software	ML model performance
Cai et al[8]	Detect AF	16,557 12-lead ECGs	NN	Sensitivity 0.9919, specificity 0.9944, accuracy 0.9935
Attia et al[9]	Detect AF under sinus rhythm	649,931 12-lead ECGs	CNN	AUC 0.87, sensitivity 0.79, specificity 0.795, accuracy 0.794
Wasserlauf et al[10]	Detect AF based on ECGs using smartwatch	7,500 patients	CNN	Sensitivity 0.975
Tison et al[11]	Detect AF based on ECGs using smartwatch	9,750 patients	DNN	Sensitivity 0.980, specificity 0.902
Hannun et al[12]	Classify cardiac rhythms into 12 rhythm classes	91,231 single-lead ECGs	DNN	AUC 0.97, F1 0.837
Ribeiro et al[13]	Recognize 6 types of heart rhythm abnormalities	2 million 12-lead ECGs	DNN	F1 >0.8, specificity >0.99
Kwon et al[14]	Detect left ventricular hypertrophy	21,286 patients with 12-lead ECGs	DNN+CNN	Internal AUC 0.880, external AUC 0.868
Attia et al[15]	Detect cardiac contractile dysfunction	44,959 patients with 12-lead ECGs	CNN	AUC 0.93, sensitivity 0.863, specificity 0.857, accuracy 0.857
Attia et al[16]	Detect cardiac contractile dysfunction	16,056 patients with 12-lead ECGs	CNN	AUC 0.918
Noseworthy et al[17]	Detect cardiac contractile dysfunction	97,829 patients with 12-lead ECGs	CNN	AUC >0.93
Sengupta et al[18]	Detect abnormal myocardial relaxation	188 patients with 12-lead ECGs	RF	AUC 0.91
Ko et al[19]	Diagnose HCM	12-lead ECGs of 3,060 HCM patients and 63,941 controls	CNN	AUC 0.96, sensitivity 0.87, specificity 0.90
Kwon et al[20]	Diagnose HF	55,163 12-lead ECGs of 22,765 patients	DNN	Internal AUC 0.843, external AUC 0.889
Kwon et al[21]	Diagnose mitral regurgitation	70,709 12-lead ECGs of 38,241 patients	CNN	Internal AUC 0.816, external AUC 0.877
Kwon et al[22]	Diagnose aortic stenosis	56,689 12-lead ECGs of 43,051 patients	MLP+CNN	Internal AUC 0.884, external AUC 0.861
Kwon et al[23]	Diagnose pulmonary hypertension	70,709 12-lead ECGs of 38,241 patients	DNN+CNN	Internal AUC 0.859, external AUC 0.902

AF, atrial fibrillation; AUC, area under the curve; CNN, convolutional neural network; CVDs, cardiovascular diseases; DCNN, deep convolutional neural network; DNN, deep neural network; ECGs, electrocardiograms; NN, neural network; HCM, hypertrophic cardiomyopathy; HF, heart failure; MLP, multilayer perceptron; RF, random forest.

prediction, which involves estimating an unknown variable (e.g., predicting whether a patient will die in 5 years). In contrast, unsupervised learning is focused on discovering underlining patterns and relationships among the unlabeled dataset.[3] A typical representative example of unsupervised learning is clustering analysis, which involves subgrouping objects based on their similarity. ML is based on algorithms, such as decision trees, random forests (RF), support vector machines (SVM), neural networks (NN), and deep learning (DL). These algorithms have been reviewed in detail elsewhere.[6]

To clearly illustrate the role of ML-based AI in CVDs, in this review we summarize the applications of AI from 3 aspects: diagnosis, prediction, and prognosis.
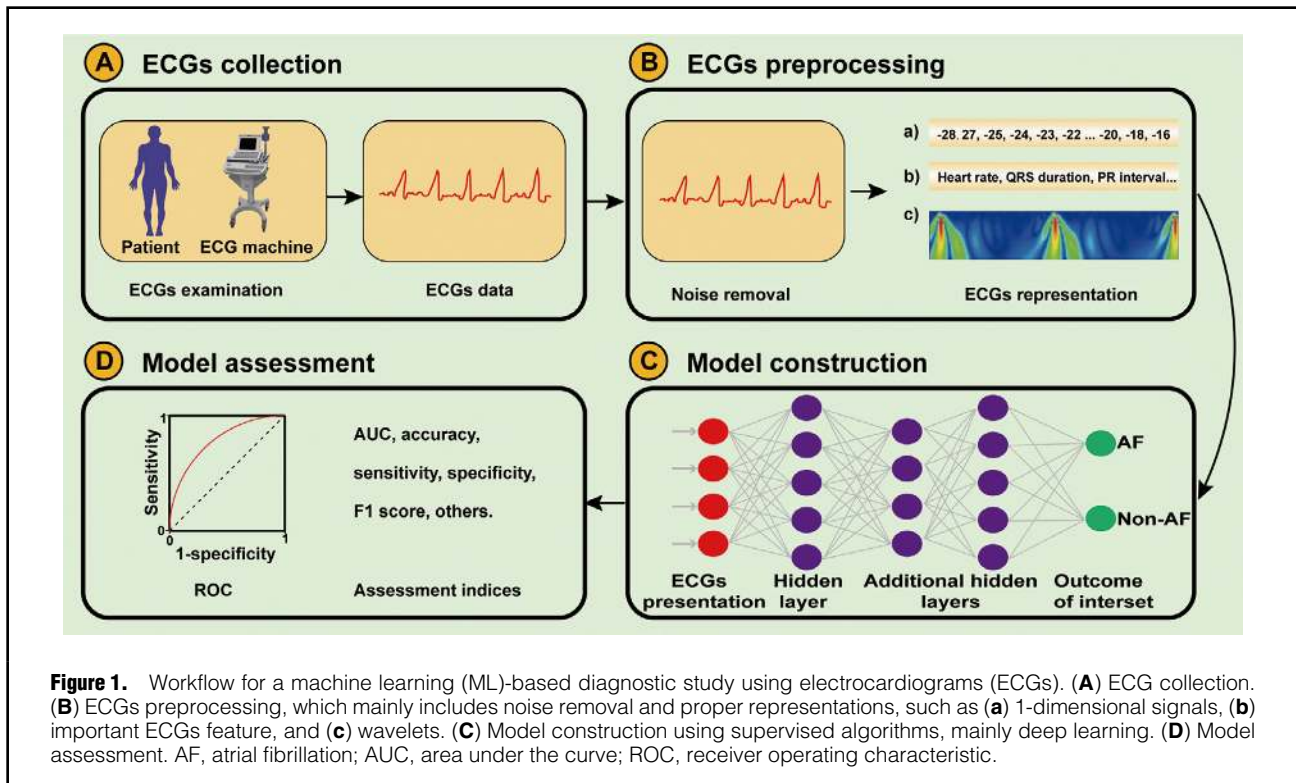
AI-Assisted Diagnosis of CVDs

The diagnosis of most CVDs is highly dependent on ECG and/or cardiovascular imaging examinations. However, the interpretation of medical images is time consuming and inter-rater variations may not be negligible. With the application of AI, image interpretation could be automated, saving clinicians much time, improving the detection rate, and reducing the rate of misdiagnosis and missed diagnosis. In this section, due to space limitations, we only focus on ECGs (**Table 1**).

As a cheap and non-invasive clinical tool, the ECG plays an important role in the diagnosis of CVDs. Although computer-aided ECG interpretation has been widely used in clinical practice, there is still considerable possibility of misinterpretation.[7] In recent years, progress in computer algorithms and the use of big data have significantly improved the accuracy of automatic ECG interpretation. To date, ECGs have been used as inputs to construct ML models to mainly perform 4 tasks: automatic recognition of heart rhythm, detection of cardiac structural abnormalities, detection of cardiac functional abnormalities, and detection of CVDs.

Recognition of Heart Rhythm

AI has been successfully used to recognize heart rhythms, especially AF. There has been increased interest in detecting AF due to its increasing incidence, as well as the possibility of preventing AF-related strokes. With 16,557 annotated 12-lead ECGs, an NN model was trained to diagnose AF, achieving an overall accuracy >0.99.[8] To further explore the ability of AI to detect AF during sinus rhythm, a convolutional neural network (CNN) was trained with 649,931 annotated 12-lead ECGs from 180,922 patients with sinus rhythm, with the ML model able to diagnose AF with an

Figure 1. Workflow for a machine learning (ML)-based diagnostic study using electrocardiograms (ECGs). (**A**) ECG collection. (**B**) ECGs preprocessing, which mainly includes noise removal and proper representations, such as (**a**) 1-dimensional signals, (**b**) important ECGs feature, and (**c**) wavelets. (**C**) Model construction using supervised algorithms, mainly deep learning. (**D**) Model assessment. AF, atrial fibrillation; AUC, area under the curve; ROC, receiver operating characteristic.

area under the receiver operating characteristic curve (AUC) of 0.87.[9] Furthermore, NN models that could detect AF have been inserted in smartwatches, and this technology may benefit thousands of people due to the widespread use of smartwatches.[10,11]

Importantly, AI could be used to recognize other types of heart rhythms in addition to AF. Using 91,232 single-lead ECGs from 53,549 patients who used a single-lead ambulatory ECG monitoring device, Hannun et al developed a deep neural network (DNN) model that was able to classify 12 cardiac rhythm classes, including 10 arrhythmias, sinus rhythm, and noise.[12] The mean F1 score (i.e., the harmonic mean of the positive predictive value and sensitivity) for the DNN exceeded that of average cardiologists (0.837 vs. 0.780).[12] Similarly, another DNN model was trained with >2 million labeled 12-lead ECGs, and the model was able to recognize 6 types of heart rhythm abnormalities, with F1 scores >0.80.[13]

Detection of Cardiac Structural or Functional Abnormalities
Apart from heart rhythm recognition, AI has also been used to detect abnormalities in cardiac structure or function. An ensemble neural network (ENN) model was constructed with ECGs to diagnose left ventricular hypertrophy, with the model significantly outperforming the cardiologist during internal validation.[14] This result was confirmed external validation (sensitivity: 0.454 vs. 0.284). The datasets used for the internal validation and ML model training were derived from the same hospital and cohort, whereas external validation involved an independent dataset derived from another hospital and cohort.[14] To detect cardiac contractile dysfunction based on ECG data alone, Attia et al trained a CNN with 44,959 annotated 12-lead ECGs; this model performed well in screening patients with cardiac

contractile dysfunction, with an AUC of 0.93.[15] Subsequently, the ability of this CNN to detect cardiac contractile dysfunction was validated in a prospective cohort of 16,056 patients in a single cardiovascular center, achieving an AUC of 0.918.[16] Furthermore, it was proved that this CNN could be applied to patients of different races and ethnicities.[17] Similarly, to detect abnormal myocardial relaxation, Sengupta et al developed an RF model.[18] But, unlike the study of Attia et al,[15] which directly used raw 12-lead ECGs, the ECGs used in the study of Sengupta et al were processed using continuous wavelet transform mathematics before being used to train the RF model.[18] Such image preprocessing actually amplified the ECG signal, and is a useful way to reduce sample size.[18]

Detection of CVDs
In addition to classifying cardiac rhythms and evaluating cardiac structure and function, AI has been used to directly diagnose CVDs, such as hypertrophic cardiomyopathy,[19] heart failure (HF),[20] mitral regurgitation,[21] aortic stenosis,[22] and pulmonary hypertension.[23] The outstanding performance of the ML models in diagnosing these diseases based on ECGs indicates that many CVDs may cause subtle abnormalities in ECGs that cannot be easily recognized by human eyes, but can be detected by AI.

Workflow
Generally speaking, the workflow of an ML-based diagnosis study using ECGs usually involves the collection of ECGs, preprocessing, model construction, and assessment. To enhance the generalization of ML models, it is better to use ECG machines that are widely used in clinics nationwide or worldwide (**Figure 1A**). ECG preprocessing primarily includes removal of noise and proper representations, such

Table 2. Artificial Intelligence-Assisted Prediction of CVDs

Diseases, events, or interventions at baseline	Predicted disease or event	Time from baseline	Sample size
Incidence prediction			
No CVDs	Death, stroke, CHD, CVDs, HF, and AF	12 years	6,814 subjects, 66.6% for training (3-fold cross-validation), 33.3% for testing
No CVDs	CVDs	13 years	Training: 3,230 subjects (2-fold cross-validation); internal validation: 3,229 subjects; external validation: 1,348 subjects
No VF	VF	30 s	27 cases and 28 controls (10-fold cross-validation)
Prognosis prediction			
CAD	Death	5 years	10,030 patients (10-fold cross-validation)
HFpEF	Death and hospitalization	3 years	1,767 patients (5-fold cross-validation)
Hypertension	Composite end point events	33 months	508 young patients with hypertension (10-fold cross-validation)
OHCA	In-hospital death	–	39,566 patients; 90% for training, 10% for testing
OHCA	Poor functional outcome	180 days	932 patients; 90% for training, 10% for testing (5-fold cross-validation)
CRT	Death	1, 2, 3, 4, and 5 years	Training: 1,510 patients (10-fold cross-validation); testing: 158 patients
PCI	In-hospital death	–	11,709 patients with 14,349 PCIs (8-fold cross-validation)
PCI	HF readmission	30 days	11,709 patients with 14,349 PCIs (8-fold cross-validation)
PCI	Death	180 days	11,709 patients with 14,349 PCIs (8-fold cross-validation)

Diseases, events, or interventions at baseline	Parameters used to construct ML models	Algorithm	AUC	Reference
Incidence prediction				
No CVDs	20 variables from imaging, non-invasive tests, questionnaires, and biomarker panels	RF	0.84 for death, 0.75 for stroke, 0.80 for CHD, 0.80 for CVD, 0.84 for HF, and 0.75 for AF	24
No CVDs	9 variables (age, sex, ethnicity, TC, HDL-C, SBP, treatment for hypertension, diabetes, and smoking)	SVM	Internal AUC: 0.94; external AUC: 0.95	26
No VF	4 variables from 120-s ECG signals	ANN	0.99	27
Prognosis prediction				
CAD	19 clinical and 35 CCTA parameters	Logit-Boost	0.79	28
HFpEF	86 clinical, laboratory, and ECG variables	RF	0.72 for death, 0.76 for hospitalization	29
Hypertension	11 clinical, laboratory, and echocardiographic variables	XGBoost	0.757	30
OHCA	46 clinical and laboratory parameters	GBM	0.87	32
OHCA	54 clinical and laboratory parameters	ANN	0.891	33
CRT	33 pre-implant clinical variables	RF	0.768, 0.793, 0.785, 0.776, 0.803 for 1-, 2-, 3-, 4-, and 5-year mortality prediction, respectively	34
PCI	52 admission variables	RF	0.92	35
PCI	358 discharge variables	RF	0.90	35
PCI	358 discharge variables	RF	0.87	35

ANN, artificial neural network; CAD, coronary artery diseases; CCTA, coronary computed tomography angiography; CHD, congenital heart disease; CRT, cardiac resynchronization therapy; GBM, gradient boosting machine; HDL-C, high-density lipoprotein cholesterol; HFpEF, heart failure with preserved heart failure; OHCA, out-of-hospital cardiac arrest; PCI, percutaneous coronary intervention; SBP, systolic blood pressure; SVM, support vector machine; TC, total cholesterol; VF, ventricular fibrillation. Other abbreviations as in Table 1.

as 1-dimensional signals (suitable for CNN[9,10,14–17,19,21–23]), important ECG features (suitable for DNN[11–14,20,23]), and wavelets (suitable for RF;[18] **Figure 1B**). Next, ML models can be constructed with supervised algorithms to perform classification tasks, like classifying the input ECG as AF or non-AF. The most commonly used algorithm is DL, especially CNN (**Figure 1C**). Finally, model performance is assessed by receiver operating characteristic (ROC) curve analysis and indices like the AUC, accuracy, sensitivity, specificity, and F1 score (**Figure 1D**).
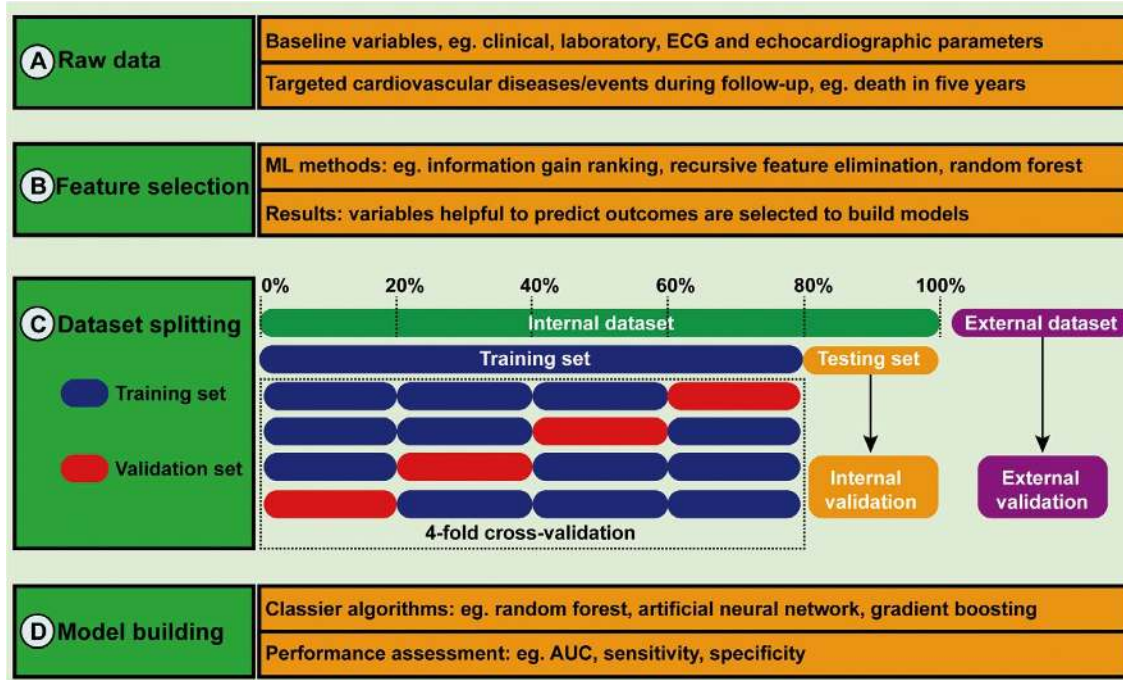
Figure 2. Workflow for the construction of a machine learning (ML)-based prediction model. (**A**) Raw data is collected at baseline and during follow-up. (**B**) More informative and non-redundant variables are selected from the available baseline variables to construct ML models. (**C**) The internal dataset is divided into a training set and a testing set (e.g., 4 : 1, as in the figure), which are used to develop the ML model and assess its generalizability, respectively. Furthermore, cross-validation is usually used to enhance the model's performance. The example of 4-fold cross-validation is shown in the figure, in which the training set is divided into 4 equal-sized groups, with one of the groups used as a validation set and the other 3 groups used as training sets at each iteration. In some studies, an external dataset is used to independently assess the model's performance. (**D**) Classifier algorithms are used to build ML models and then assess their performance. AUC, area under the curve; ECG, electrocardiogram.

AI-Assisted Prediction of CVDs

Apart from performing diagnostic tasks, AI performs well in prediction tasks, including predictions of incidence and prognosis (see **Table 2**).

Incidence Prediction

Most CVDs have a subclinical phase, during which the patients exhibit no clinical symptoms. Importantly, disease progression can be slowed down or even prevented if interventions occur during this phase. Therefore, the prediction of the incidence of CVDs is of great significance for the asymptomatic population.

To explore the ability of ML models to predict the incidence of CVDs, Ambale-Venkatesh et al evaluated 6,814 participants who were initially free of CVDs from the Multi-Ethnic Study of Atherosclerosis (MESA).[24] Baseline data for 735 variables, including imaging, non-invasive tests, questionnaires, and biomarkers, were collected for these participants, and the RF technique was used to identify the top 20 variables for each of the 6 outcomes of death, stroke, coronary artery disease (CAD), CVDs, HF, and AF. These 20 variables were then used to construct 6 RF models with 3-fold cross-validation to predict the incidence of each of the 6 outcomes in 12 years. Results showed that the concordance indices, a generalization of the AUC and a useful parameter to evaluate the performance of the

predictive model,[25] for all these RF models were not less than 0.75.[24] To further compare the prediction abilities of ML models and the traditional American College of Cardiology and American Heart Association (ACC/AHA) risk calculator, Kakadiaris et al trained SVM models using the MESA cohort and 2-fold cross-validation.[26] The SVM models, constructed with the same 9 traditional risk variables used by the ACC/AHA Risk Calculator, performed better than ACC/AHA Risk Calculator (AUC 0.94 vs. 0.72) and this was verified by using an external validation cohort (AUC 0.95 vs. 0.71).[26]

In addition to predicting long-term events, ML models can also predict the incidence of cardiovascular events in the short term. For example, an NN model was constructed with 4 QRS complex shape features from baseline 120-s ECGs of 27 ventricular fibrillation cases and 28 controls.[27] This model was trained with 10-fold cross-validation and was able to predict the incidence of ventricular fibrillation in 30s, with an AUC of 0.99. Although 30s is a short period, it is of importance to save patients' lives in the clinic.[27]

Prognosis Prediction

Prediction of the prognosis of CVDs is critical in clinical practice because an accurate prognosis prediction model could inform clinicians of each patient's prognosis, helping with decision making, the use of disease management programs, and in discussing end-of-life preferences.

ML models have been used to predict the prognosis of chronic CVDs, like HF, CAD, and hypertension. The classification algorithm Logit-Boost was used to predict 5-year mortality of patients with CAD.[28] Before ML model building, features were selected using information gain ranking, and only those variables helpful in predicting outcomes (information gain >0) were selected for model building. Although data on 25 clinical plus 44 coronary computed tomography angiography (CCTA) parameters were collected, only 19 clinical and 35 CCTA variables were selected for model building. The prediction model was trained and validated using 10-fold cross-validation. This ML model (AUC 0.79) outperformed the Framingham risk score (AUC 0.61) and CCTA severity scores (AUC 0.62–0.64).[28]

Similarly, an ML model was constructed to predict death and hospitalization in 3 years for patients with HF with preserved ejection fraction (HFpEF) based on data at discharge. The model used 86 clinical, laboratory, and ECG variables from 1,767 patients, and the RF algorithm and 5-fold cross-validation. This model achieved an AUC of 0.72 and 0.76 for death and hospitalization, respectively.[29]

To achieve accurate prognosis prediction for young patients with hypertension, Wu et al collected 58 variables at baseline and the 33-month follow-up for 508 patients.[30] Then, features were selected using recursive feature elimination, with only 11 variables finally selected to build the prediction model. A classifier algorithm called extreme gradient boosting, as well as 10-fold cross-validation, was used to train and validate the ML model. This model did well in predicting composite endpoint events and achieved an AUC of 0.757, higher than the recalibrated Framingham risk score model (AUC 0.529).[30]

Similarly, prognosis prediction for patients with acute CVDs can be achieved using ML models. Out-of-hospital cardiac arrest (OHCA) is an acute cardiovascular event with over 300,000 cases among adults in the US.[31] The prediction of adverse events for patients after OHCA is critical because it could inform clinicians and the patients' families of the prognosis and then guide intervention. ML models have been used to predict short- and long-term outcomes for this population. To predict the in-hospital mortality of OHCA patients, Nanayakkara et al collected 43 clinical and laboratory parameters from 39,566 OHCA patients within the first 24 h after OHCA.[32] Then, 90% of the total data set was used to train the prediction model using a classifier algorithm called gradient boosting machine, after which the model was tested with the remaining 10% of data and showed great ability to predict in-hospital mortality, with an AUC of 0.87.[32] To further predict the long-term functional outcome of OHCA patients, a dataset composed of 54 clinical and laboratory parameters at admission and 180-day follow-up records for 932 OHCA patients was collected.[33] The outcome prediction model was trained with 90% of the data with 5-fold cross-validation, with the remaining 10% of data used for testing. An NN algorithm was used during model development, and the model performed well in predicting poor functional outcomes, comprising dependence, coma, or vegetative state, and death within 180 days after OHCA (AUC=0.891).[33]

ML models can also be used to predict prognosis for patients after cardiovascular interventions, such as cardiac resynchronization therapy (CRT) and percutaneous coronary intervention (PCI). To build ML models capable of predicting long-term prognosis after CRT, a database of 1,510 patients undergoing CRT implantation was used.[34]

A total of 33 pre-implantation clinical variables was collected to train an RF model, and 10-fold cross-validation was performed. The model was then tested on an independent cohort of 158 patients and did well in predicting 1-, 2-, 3-, 4-, and 5-year mortality, with an AUC >0.75.[34] To predict the short- and long-term prognosis after PCI, Zack et al built RF prediction models by analyzing 11,709 patients with 14,349 PCIs.[35] In all, 52 clinical parameters at admission were used to predict in-hospital mortality, whereas 358 variables at discharge were used to predict 30-day HF readmission and all-cause death. Eight-fold cross-validation was used in the RF models, and all 3 models achieved an AUC >0.85.[35]

The studies mentioned above show that ML models may be superior to standard linear regression models[29,32,33,35] and currently used clinical risk scoring systems[28,30,32,34] in performing prediction tasks. This is mainly because ML is not only able to incorporate a larger number of variables, but it can also analyze the possibly complex interactions and nonlinear effects of the variables.[36]

Workflow

The workflow of an ML-based prediction study usually involves the collection of raw data, feature selection, dataset splitting, and model building. There are mainly 2 types of raw data: variables available at baseline, and whether a subject experience targeted CVDs or events based on follow-up records (**Figure 2A**). Feature selection is used to select more informative and non-redundant variables from the available variables, and the selected variables are used to construct ML models (**Figure 2B**). The whole dataset is typically divided into a training set and a testing set. The former is used to develop the ML model, whereas the latter is used to assess its generalizability. The training set is usually randomly divided into several equal-sized groups; one of the groups is used as a validation set, whereas the other groups are used as training sets at each iteration, a process called "cross-validation", a useful way to avoid overfitting. Briefly, overfitting indicates that models perform well on the training set but poorly on unseen datasets. Details regarding the reasons for overfitting and how to avoid it have been reviewed elsewhere.[37] Sometimes an external dataset is used to further test a model's generalizability (**Figure 2C**).

The prediction of incidence or prognosis is actually a classification task, so classifier algorithms, such as RF, NN, and gradient boosting, are usually required in the development of prediction ML models. Once the prediction model is built, indices like the AUC, sensitivity, and specificity are calculated to quantify the performance of the ML model (**Figure 2D**).

AI-Assisted Classification of CVDs

Most CVDs are heterogeneous,[38] which means patients with the same CVD may have distinct etiologies, clinical characteristics, auxiliary examination results, outcomes, and therapeutic responses. Therefore, there is an urgent need to integrate data from different sources to make the classification of a disease more accurate. Such accurate classification could guide risk stratification, prognostic prediction, and even the choice of treatment, so it is discussed in this section.

Table 3. Artificial Intelligence-Assisted Classification of CVDs

Disease	Sample size	Parameters used in unsupervised ML methods	No. subgroups	Differences among subgroups	Reference
HFpEF	Discovery cohort: 397; validation cohort: 107	46 clinical, laboratory, ECG, and echocardiographic parameters	3	Clinical characteristics, cardiac structure/function, invasive hemodynamics, and outcomes	40
HFpEF	Discovery cohort: 654; internal validation cohort: 1,113; external validation cohort 216	61 clinical, laboratory, ECG, and echocardiographic parameters	3	Clinical characteristics and long-term outcomes	41
HFpEF	320	32 echocardiographic and 11 clinical/laboratory parameters	6	Clinical characteristics and outcomes, as well as concentrations of inflammatory and cardiovascular plasma proteins	42
HFpEF	177	8 resting and post-exercise echocardiographic parameters	2	Left ventricular systolic reserve and prognosis	43
PAH	Discovery cohort: 281; validation cohort: 104	Circulating proteomic panel of 48 cytokines, chemokines, and factors	4	Blood proteomic immune profiles, clinical risk, and long-term outcomes	44
PMR	122	64 clinical and echocardiographic variables	3	Clinical characteristics, prognosis, and therapeutic response to surgery (mitral valve repair or replacement)	45
AC	Discovery cohort: 60; validation cohort: 92	18 parameters derived from pathological images of explanted AC hearts	4	Genetic background, echocardiographic and ECG parameters	46
HF	1,106	50 clinical, laboratory, ECG, and echocardiographic parameters	4	Clinical characteristics, biomarker values, ventricular structure/function, and therapeutic response to CRT	47

AC, arrhythmogenic cardiomyopathy; ECG, electrocardiography; PAH, pulmonary arterial hypertension; PMR, primary mitral regurgitation. Other abbreviations as in Tables 1,2.

HF With Preserved Ejection Fraction (HFpEF)

HFpEF is an acknowledged phenotypically heterogeneous disease with a high prevalence and no proven useful medical therapies.[39] Accurate classification of HFpEF may be a critical step for the design of a clinical trial and the development of useful therapies for specific HFpEF subgroups. To this end, several research groups have used AI to identify phenotypically distinct HFpEF categories. For example, Shah et al prospectively collected 67 phenotypic variables from 397 HFpEF patients, generated a correlation matrix of phenotypic variables, and filtered out variables that were correlated at a correlation coefficient of >0.6, leaving 46 continuous variables for the final clustering analyses.[40] Three clusters were determined using the 46 identified variables. Surprisingly, the 3 subgroups differed significantly not only in clinical characteristics, but also survival. These results were validated in another prospective cohort of 107 HFpEF patients.[40] Using different HFpEF cohorts but similar study strategies, Segar et al also identified 3 mutually exclusive subgroups of HFpEF patients with distinct clinical characteristics and long-term outcomes.[41] Hedman et al used 32 echocardiographic and 11 clinical and laboratory variables to perform ML-based clustering and identified 6 phenotype-based groups.[42] Importantly, the results of that study revealed differential characteristics and outcomes, as well as different levels of inflammatory and cardiovascular plasma proteins across the newly identified subgroups.[42] In another study, instead of inputting several different types of medical data, Przewlocka-Kosmala et al used only resting and postexercise echocardiographic parameters and divided HFpEF patients into 2 subgroups.[43] One

of the subgroups was characterized by a relatively isolated impairment of left ventricular systolic reserve and a better prognosis, whereas the other showed abnormal longitudinal deformation, ventricular-arterial coupling, and cardiac output responses to exercise.[43] All the studies described above proved the feasibility of ML-based clustering analysis to define HFpEF subgroups with different clinical characteristics and prognoses, but further studies are required to determine whether these subgroups respond differently to specific therapies and whether there are optimal therapeutic targets for each of the subgroups (**Table 3**).

Other CVDs

Unsupervised clustering analysis has also been applied in other CVDs besides HFpEF. Sweatt et al used unsupervised ML to classify pulmonary arterial hypertension (PAH) patients into 4 clusters based on blood proteomic profiles that included 48 inflammation- or autoimmunity-related molecules.[44] These 4 PAH clusters were distinct in terms of proteomic immune profiles, clinical risk, and long-term outcomes. That study was valuable because it identified possible immunotherapy targets for PAH.[44]

Primary mitral regurgitation (PMR) is another heterogeneous clinical disease, with considerable differences in prognosis among patients after valve surgery. To identify phenotypically distinct categories of PMR patients, Pimor et al performed unsupervised clustering analysis using 64 clinical and echocardiographic variables of PMR patients before valve surgery.[45] These patients were then classified into 3 phenotypes that differed markedly in terms of clinical characteristics and post-surgery prognosis. The ML
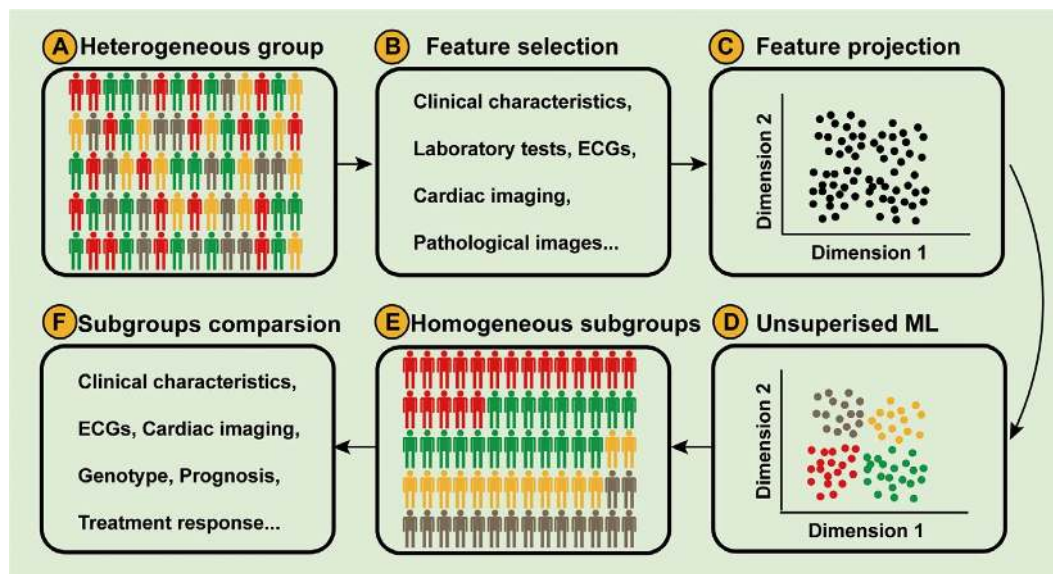
Figure 3.   Workflow to conduct a classification study of cardiovascular diseases (CVDs) using machine learning (ML). (**A**) Most CVDs are heterogeneous. (**B**,**C**) Dimensionality reduction consists of 2 important processes, namely feature selection (**B**) and feature projection (**C**). (**D**) Unsupervised ML. (**E**) Homogeneous subgroups. (**F**) Comparisons among different subgroups. ECG, electrocardiogram.

model could be used to guide cardiac surgeons to identify the high-risk subgroup, and these patients could be carefully monitored and may even be treated earlier.[45]

Arrhythmogenic cardiomyopathy (AC) is an inherited cardiomyopathy that is heterogeneous in the overall distribution of fibrofatty infiltration in the heart. We have previously used unsupervised clustering to classify AC patients into 4 subgroups based on 18 parameters derived from pathological images of 60 explanted AC hearts, and these 4 subgroups had distinct genetic backgrounds, echocardiographic variables, and ECG parameters.[46] That study established a novel pathological classification with distinct genotypes indicating different potential mechanisms in the pathogenesis of AC.[46]

HF is a heterogeneous clinical syndrome with a substantial proportion of patients who do not respond to CRT. To identify patients who are likely to respond to CRT, Cikes et al used unsupervised ML to categorize 1,106 HF patients who were randomized to either receive CRT or not.[47] Fifty baseline clinical and echocardiographic variables were used in the ML method, and 4 phenogroups were identified. Surprisingly, 2 of these phenogroups were found to be likely to benefit from CRT by comparing the HF-free survival rate after treatment in each of the phenogroups.[47] This finding may guide cardiologists to identify patients who are most likely to respond to CRT (**Table 3**).

Workflow

Most CVDs are heterogeneous (**Figure 3A**). To classify the heterogeneous population into several homogenous subgroups, information is collected for the available variables, such as clinical characteristics, cardiac imaging, ECGs, laboratory tests, and even pathological images. Then, dimensionality reduction, including feature selection and feature projection, is performed. Feature selection involves using

algorithms to select more valuable features for classification, and is critical to improve the performance of algorithms by reducing redundant features (**Figure 3B**). Feature projection involves projecting the selected features into a 2-dimensional space, which helps visualization (**Figure 3C**). After dimensionality reduction, unsupervised ML is used to define homogeneous subgroups (**Figure 3D,E**). Finally, a comparison among different subgroups is performed (**Figure 3F**). Unsupervised learning was used to achieve clustering analysis in the most of the relevant studies, and the determination of different clusters is based on the similarity of patients' input data.

Summary and Perspectives

To help clinicians better understand AI and conduct related studies, we have described some basic knowledge about AI, ML, and algorithms, and then summarized reported studies associated with AI-based diagnosis, prediction, and classification in CVDs (**Tables 1–3**), after which the general workflow of each of the 3 applications was illustrated (**Figures 1–3**).

There are still some obstacles in using ML-based AI in cardiovascular practice. First, data availability limits the generalizability of ML algorithms. The data used for the training of ML models are typically acquired from 1 or several laboratories, health centers, or hospitals, and the algorithms are therefore likely to fail when applied to different populations.[6] Second, obtaining large quantities of high-quality labeled data, which are essential for the training of supervised learning algorithms, is labor intensive and often performed manually.[48] Third, the "black box" property of DL, which means the inner mechanisms and processes of DL models, cannot be explained and is not accepted by many clinicians.[49]

Disclosures

None declared.

References

1. Joseph P, Leong D, McKee M, Anand SS, Schwalm JD, Teo K, et al. Reducing the global burden of cardiovascular disease, part 1: The epidemiology and risk factors. *Circ Res* 2017; **121:** 677–694.
2. Vieillard-Baron A, Millington SJ, Sanfilippo F, Chew M, Diaz-Gomez J, McLean A, et al. A decade of progress in critical care echocardiography: A narrative review. *Intensive Care Med* 2019; **45:** 770–788.
3. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, et al. Artificial intelligence in cardiology. *J Am Coll Cardiol* 2018; **71:** 2668–2679.
4. Baumgartner H, Bonhoeffer P, De Groot NMS, de Haan F, Deanfield JE, Galie N, et al. ESC guidelines for the management of grown-up congenital heart disease (new version 2010). *Eur Heart J* 2010; **31:** 2915–2957.
5. Krittanawong C, Zhang H, Wang Z, Aydar M, Kitai T. Artificial intelligence in precision cardiovascular medicine. *J Am Coll Cardiol* 2017; **69:** 2657–2664.
6. Trayanova NA, Popescu DM, Shade JK. Machine learning in arrhythmia and electrophysiology. *Circ Res* 2021; **128:** 544–566.
7. Schlapfer J, Wellens HJ. Computer-interpreted electrocardiograms: Benefits and limitations. *J Am Coll Cardiol* 2017; **70:** 1183–1192.
8. Cai W, Chen Y, Guo J, Han B, Shi Y, Ji L, et al. Accurate detection of atrial fibrillation from 12-lead ECG using deep neural network. *Comput Biol Med* 2020; **116:** 103378.
9. Attia ZI, Noseworthy PA, Lopez-Jimenez F, Asirvatham SJ, Deshmukh AJ, Gersh BJ, et al. An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: A retrospective analysis of outcome prediction. *Lancet* 2019; **394:** 861–867.
10. Wasserlauf J, You C, Patel R, Valys A, Albert D, Passman R. Smartwatch performance for the detection and quantification of atrial fibrillation. *Circ Arrhythm Electrophysiol* 2019; **12:** e006834.
11. Tison GH, Sanchez JM, Ballinger B, Singh A, Olgin JE, Pletcher MJ, et al. Passive detection of atrial fibrillation using a commercially available smartwatch. *JAMA Cardiol* 2018; **3:** 409–416.
12. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* 2019; **25:** 65–69.
13. Ribeiro AH, Ribeiro MH, Paixão GMM, Oliveira DM, Gomes PR, Canazart JA, et al. Automatic diagnosis of the 12-lead ECG using a deep neural network. *Nat Commun* 2020; **11:** 1760.
14. Kwon JM, Jeon KH, Kim HM, Kim MJ, Lim SM, Kim KH, et al. Comparing the performance of artificial intelligence and conventional diagnosis criteria for detecting left ventricular hypertrophy using electrocardiography. *Europace* 2020; **22:** 412–419.
15. Attia ZI, Kapa S, Lopez-Jimenez F, McKie PM, Ladewig DJ, Satam G, et al. Screening for cardiac contractile dysfunction using an artificial intelligence-enabled electrocardiogram. *Nat Med* 2019; **25:** 70–74.
16. Attia ZI, Kapa S, Yao X, Lopez-Jimenez F, Mohan TL, Pellikka PA, et al. Prospective validation of a deep learning electrocardiogram algorithm for the detection of left ventricular systolic dysfunction. *J Cardiovasc Electrophysiol* 2019; **30:** 668–674.
17. Noseworthy PA, Attia ZI, Brewer LC, Hayes SN, Yao X, Kapa S, et al. Assessing and mitigating bias in medical artificial intelligence: The effects of race and ethnicity on a deep learning model for ECG analysis. *Circ Arrhythm Electrophysiol* 2020; **13:** e007988.
18. Sengupta PP, Kulkarni H, Narula J. Prediction of abnormal myocardial relaxation from signal processed surface ECG. *J Am Coll Cardiol* 2018; **71:** 1650–1660.
19. Ko WY, Siontis KC, Attia ZI, Carter RE, Kapa S, Ommen SR, et al. Detection of hypertrophic cardiomyopathy using a convolutional neural network-enabled electrocardiogram. *J Am Coll Cardiol* 2020; **75:** 722–733.
20. Kwon JM, Kim KH, Jeon KH, Kim HM, Kim MJ, Lim SM, et al. Development and validation of deep-learning algorithm for electrocardiography-based heart failure identification. *Korean Circ J* 2019; **49:** 629–639.
21. Kwon JM, Kim KH, Akkus Z, Jeon KH, Park J, Oh BH. Artificial intelligence for detecting mitral regurgitation using electrocardiography. *J Electrocardiol* 2020; **59:** 151–157.
22. Kwon JM, Lee SY, Jeon KH, Lee Y, Kim KH, Park J, et al. Deep learning-based algorithm for detecting aortic stenosis using electrocardiography. *J Am Heart Assoc* 2020; **9:** e014717.
23. Kwon JM, Kim KH, Medina-Inojosa J, Jeon KH, Park J, Oh BH. Artificial intelligence for early prediction of pulmonary hypertension using electrocardiography. *J Heart Lung Transplant* 2020; **39:** 805–814.
24. Ambale-Venkatesh B, Yang X, Wu CO, Liu K, Hundley WG, McClelland R, et al. Cardiovascular event prediction by machine learning: The Multi-Ethnic Study of Atherosclerosis. *Circ Res* 2017; **121:** 1092–1101.
25. Tripepi G, Jager KJ, Dekker FW, Zoccali C. Statistical methods for the assessment of prognostic biomarkers (Part I): Discrimination. *Nephrol Dial Transplant* 2010; **25:** 1399–1401.
26. Kakadiaris IA, Vrigkas M, Yen AA, Kuznetsova T, Budoff M, Naghavi M. Machine learning outperforms ACC/AHA CVD Risk Calculator in MESA. *J Am Heart Assoc* 2018; **7:** e009476.
27. Taye GT, Shim EB, Hwang HJ, Lim KM. Machine learning approach to predict ventricular fibrillation based on QRS complex shape. *Front Physiol* 2019; **10:** 1193.
28. Motwani M, Dey D, Berman DS, Germano G, Achenbach S, Al-Mallah MH, et al. Machine learning for prediction of all-cause mortality in patients with suspected coronary artery disease: A 5-year multicentre prospective registry analysis. *Eur Heart J* 2017; **38:** 500–507.
29. 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:** 12–21.
30. Wu X, Yuan X, Wang W, Liu K, Qin Y, Sun X, et al. Value of a machine learning approach for predicting clinical outcomes in young patients with hypertension. *Hypertension* 2020; **75:** 1271–1278.
31. Benjamin EJ, Virani SS, Callaway CW, Chamberlain AM, Chang AR, Cheng S, et al. Heart disease and stroke statistics – 2018 update: A report from the American Heart Association. *Circulation* 2018; **137:** e67–e492.
32. Nanayakkara S, Fogarty S, Tremeer M, Ross K, Richards B, Bergmeir C, et al. Characterising risk of in-hospital mortality following cardiac arrest using machine learning: A retrospective international registry study. *PLoS Med* 2018; **15:** e1002709.
33. Johnsson J, Björnsson O, Andersson P, Jakobsson A, Cronberg T, Lilja G, et al. Artificial neural networks improve early outcome prediction and risk classification in out-of-hospital cardiac arrest patients admitted to intensive care. *Crit Care* 2020; **24:** 474.
34. Tokodi M, Schwertner WR, Kovács A, Tősér Z, Staub L, Sárkány A, et al. Machine learning-based mortality prediction of patients undergoing cardiac resynchronization therapy: The SEMMELWEIS-CRT score. *Eur Heart J* 2020; **41:** 1747–1756.
35. Zack CJ, Senecal C, Kinar Y, Metzger Y, Bar-Sinai Y, Widmer RJ, et al. Leveraging machine learning techniques to forecast patient prognosis after percutaneous coronary intervention. *JACC Cardiovasc Interv* 2019; **12:** 1304–1311.
36. Goldstein BA, Navar AM, Carter RE. Moving beyond regression techniques in cardiovascular risk prediction: Applying machine learning to address analytic challenges. *Eur Heart J* 2017; **38:** 1805–1814.
37. Al'Aref SJ, Anchouche K, Singh G, Slomka PJ, Kolli KK, Kumar A, et al. Clinical applications of machine learning in cardiovascular disease and its relevance to cardiac imaging. *Eur Heart J* 2019; **40:** 1975–1986.
38. Kao D, Purohit S, Jhund P. Therapeutic futility and phenotypic heterogeneity in heart failure with preserved ejection fraction: What is the role of bionic learning? *Eur J Heart Fail* 2020; **22:** 159–161.
39. Roh J, Houstis N, Rosenzweig A. Why don't we have proven treatments for HFpEF? *Circ Res* 2017; **120:** 1243–1245.
40. Shah SJ, Katz DH, Selvaraj S, Burke MA, Yancy CW, Gheorghiade M, et al. Phenomapping for novel classification of heart failure with preserved ejection fraction. *Circulation* 2015; **131:** 269–279.
41. Segar MW, Patel KV, Ayers C, Basit M, Tang WHW, Willett D, et al. Phenomapping of patients with heart failure with preserved

ejection fraction using machine learning-based unsupervised cluster analysis. *Eur J Heart Fail* 2020; **22:** 148–158.

42. Hedman ÅK, Hage C, Sharma A, Brosnan MJ, Buckbinder L, Gan LM, et al. Identification of novel pheno-groups in heart failure with preserved ejection fraction using machine learning. *Heart* 2020; **106:** 342–349.

43. Przewlocka-Kosmala M, Marwick TH, Dabrowski A, Kosmala W. Contribution of cardiovascular reserve to prognostic categories of heart failure with preserved ejection fraction: A classification based on machine learning. *J Am Soc Echocardiogr* 2019; **32:** 604–615.e606.

44. Sweatt AJ, Hedlin HK, Balasubramanian V, Hsi A, Blum LK, Robinson WH, et al. Discovery of distinct immune phenotypes using machine learning in pulmonary arterial hypertension. *Circ Res* 2019; **124:** 904–919.

45. Pimor A, Galli E, Vitel E, Corbineau H, Leclercq C, Bouzille G,

et al. Predictors of post-operative cardiovascular events, focused on atrial fibrillation, after valve surgery for primary mitral regurgitation. *Eur Heart J Cardiovasc Imaging* 2019; **20:** 177–184.

46. Chen L, Song J, Chen X, Chen K, Ren J, Zhang N, et al. A novel genotype-based clinicopathology classification of arrhythmogenic cardiomyopathy provides novel insights into disease progression. *Eur Heart J* 2019; **40:** 1690–1703.

47. Cikes M, Sanchez-Martinez S, Claggett B, Duchateau N, Piella G, Butakoff C, et al. Machine learning-based phenogrouping in heart failure to identify responders to cardiac resynchronization therapy. *Eur J Heart Fail* 2019; **21:** 74–85.

48. Krittanawong C, Zhang H, Wang Z, Aydar M, Kitai T. Artificial intelligence in precision cardiovascular medicine. *J Am Coll Cardiol* 2017; **69:** 2657–2664.

49. Castelvecchi D. Can we open the black box of AI? *Nature* 2016; **538:** 20–23.