

REVIEW

Open Access



Machine learning and multi-omics integration: advancing cardiovascular translational research and clinical practice

Mingzhi Lin^{1†}, Jiuqi Guo^{1†}, Zhilin Gu¹, Wenyi Tang¹, Hongqian Tao¹, Shilong You¹, Dalin Jia^{1*}, Yingxian Sun^{1,2*} and Pengyu Jia^{1*} 

Abstract

The global burden of cardiovascular diseases continues to rise, making their prevention, diagnosis and treatment increasingly critical. With advancements and breakthroughs in omics technologies such as high-throughput sequencing, multi-omics approaches can offer a closer reflection of the complex physiological and pathological changes in the body from a molecular perspective, providing new microscopic insights into cardiovascular diseases research. However, due to the vast volume and complexity of data, accurately describing, utilising, and translating these biomedical data demands substantial effort. Researchers and clinicians are actively developing artificial intelligence (AI) methods for data-driven knowledge discovery and causal inference using various omics data. These AI approaches, integrated with multi-omics research, have shown promising outcomes in cardiovascular studies. In this review, we outline the methods for integrating machine learning, one of the most successful applications of AI, with omics data and summarise representative AI models developed that leverage various omics data to facilitate the exploration of cardiovascular diseases from underlying mechanisms to clinical practice. Particular emphasis is placed on the effectiveness of using AI to extract potential molecular information to address current knowledge gaps. We discuss the challenges and opportunities of integrating omics with AI into routine diagnostic and therapeutic practices and anticipate the future development of novel AI models for wider application in the field of cardiovascular diseases.

Keywords Translational medicine, Cardiovascular diseases, Machine learning, Artificial intelligence, Multi-omics

Introduction

Cardiovascular diseases (CVDs) are the leading cause of death worldwide, significantly contributing to health deterioration and increased healthcare cost [1]. The development of CVDs involves a complex interplay of genetic and environmental factors, resulting in a wide range of clinical features and outcomes among patients [2]. Given the limitations of traditional approaches in achieving precision medicine for CVDs, new methods need to be developed to deepen our understanding of the intricate regulatory mechanisms, to enhance the prediction and interpretation of disease progression.

Multi-omics approaches that integrate information from different omics provide an individualised and

[†]Mingzhi Lin and Jiuqi Guo have contributed equally to this work.

*Correspondence:

Dalin Jia
djia89@cmu.edu.cn
Yingxian Sun
yxsun@cmu.edu.cn
Pengyu Jia
pyjia@cmu.edu.cn

¹ Department of Cardiology, The First Hospital of China Medical University, 155 Nanjing North Street, Heping District, Shenyang 110001, People's Republic of China

² Key Laboratory of Environmental Stress and Chronic Disease Control and Prevention, Ministry of Education, China Medical University, Shenyang, Liaoning, China



comprehensive molecular picture of human beings from health to disease [3]. Current major omics technologies include genomics, epigenomics, transcriptomics, proteomics, metabolomics. Genomics provides innate inheritance and variation information of an organism. Epigenomics identifies genome modifications, contributing to expression regulation of genes. Transcriptomics explores the functions of RNA transcripts and the regulation by non-coding RNAs. Proteomics explains post-translational changes in the executive functions of proteins. Metabolomics quantifies a wide range of cellular metabolites, including amino acids, fatty acids, carbohydrates and other products of small molecule types [4]. Various omics technologies offer different perspectives for disease interpretation, and integrating these can reveal the complex biological processes and regulatory networks within organisms [5]. The volume of omics data for the cardiovascular system in public databases is growing exponentially annually, driven by reduced costs, updated instrumentation platforms, and improved technical protocols [6]. Based on the continuous update of genome-wide association study (GWAS) data of human whole genome sequencing, genetic omics data has reached a huge scale. Although proteomics or metabolomics used to contain only a few hundred molecules, with the update of technology platforms, companies such as Olink and Somalogic have recently launched products that can identify up to 5,000 analytes, and the curse of dimensionality is becoming a problem for relevant researchers [7]. Managing this explosion of data has become a significant challenge, requiring substantial computational efforts to ensure data quality control, extract clear biological significance, and apply it into CVDs research.

Artificial Intelligence (AI) is an umbrella term encompassing a wide range of technologies that share the common goal of computationally simulating human intelligence. Machine learning (ML), as a subgroup of AI, makes predictions by identifying idiosyncratic patterns in data using a mathematical framework. Complex tasks such as pattern recognition, anomaly detection, and predictive modelling have all become targets for ML applications [8]. In recent years, ML technologies have evolved and now include new methods represented by deep learning (DL) [9]. The use of ML to analyse such huge and high-dimensional datasets as multi-omics significantly improves the efficiency of mechanistic studies and clinical practice of CVDs. Previous reviews have emphasized the importance of integrating multi-omics data for CVDs research, summarised the platforms and processes for multi-omics research, and richly discussed different multi-omics features and technical challenges of multi-omics research [7, 10]. However, in such an era of rapid

iteration of AI and continuous updating of high-throughput omics information, the cross-application of AI and multi-omics is growing rapidly. ML is accelerating the integration of omics data, making it more widely involved in clinical diagnosis and treatment. It is necessary to summarise the current applications and approaches of ML in multi-omics research for clinical translation. We discuss the potential for AI-driven omics methods to be widely adopted in clinical applications, which could inspire researchers in the fields of CVDs and AI to foster closer interdisciplinary collaboration, thereby accelerating the advent of an era of AI in CVDs omics.

In this review, we systematically summarise ML models used in conjunction with omics analysis, with a focus on their combined application in exploring the entire continuum of CVDs, including stages of disease prevention, diagnosis, treatment, and prognosis (Fig. 1). We summarise and evaluate the models that have been representative of the research described above in recent years, for the benefit of scientists engaged in relevant research. Furthermore, we critically assess the limitations of ML models for CVDs research and highlights the challenges related to their clinical translation. Our review also aims to identify knowledge gaps in this field, emphasizing areas that require further study to advance the widespread application of ML combined with omics technologies in CVDs.

Methods for integrating ML with omics data

Interactive analysis, visualization, and deep phenotyping using ML technologies can simulate and extend various types of omics data, which facilitates the revelation of the link between omics variation and disease mechanisms [11]. Currently, the main ML methods include supervised learning, unsupervised learning, and reinforcement learning. DL, which primarily relies on artificial neural networks, represents a subset of ML methods that allows for automatic feature extraction from raw data through a multi-layer architecture. While traditional ML methods like Random Forest (RF) require hand-engineered features, DL leverages large-scale neural networks to learn these representations in an end-to-end manner. Transfer learning, a concept built upon ML and DL, further extends the adaptability of models across related domains (Fig. 2).

Traditional ML methods

Supervised learning requires representative benchmark datasets for model training and the selection of another reliable validation sets to assess model performance [12]. For example, researchers hope to use proteomics data from patients with myocardial infarction (MI) to predict the risk of poor prognosis. This first requires the selection

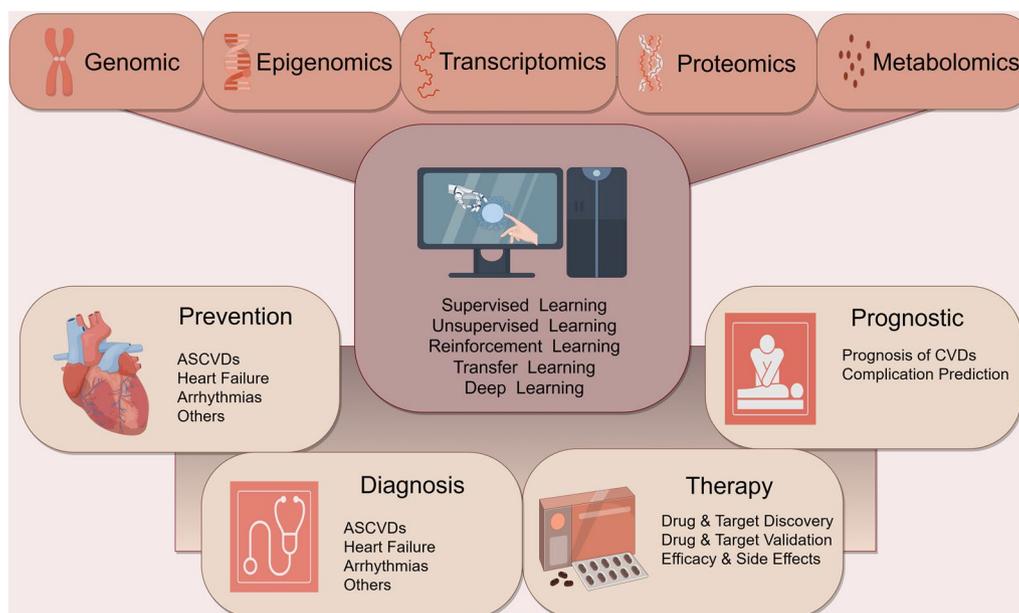


Fig. 1 ML to integrate multi-omics data (e.g., genomic, epigenomic, transcriptomics, proteomics, metabolomics) for cardiovascular research from bench to bedside

of appropriate architectural algorithms for constructing the model, such as RF, Support Vector Machines (SVM), etc. [8]. Next, the feature labelling of the proteomic data is done by setting the corresponding hyperparameters in the first training cycle of the model. The classifier calibration is performed (e.g., Platt scaling) and results of the training were assessed for performance [13]. Researchers typically tune the resulting terms parameters to ensure the reliability and robustness of the obtained parameters until the training results meet expectations and are put into prediction, producing a binary or probabilistic outcome of whether or not an infarcted patient has experienced a poor prognosis in the near future. Throughout the supervised learning process, researchers are required to have cross-disciplinary knowledge to combine domain expertise with ML understanding, annotate results meaningfully, and adjust model parameters effectively. This is particularly important in the early stages of training, where insights into the biological or clinical context guide the selection and tuning of model parameters, thus ensuring the predictions are both accurate and contextually relevant. Care needs to be taken during training to reduce noise from omics data as well as the overfitting effect of outliers in the model and to create a balance with underfitting [14].

Unsupervised learning does not require pre-training to label the dataset and calibrate the model [15]. The main methods of unsupervised learning are algorithms such as k-means, which performs the steps of dimensionality

reduction or clustering on omics data [16]. It is suitable to accomplish exploring hidden structures and directions in cardiovascular omics, such as discovering biological markers of MI, identifying unknown cellular subpopulations, etc. [17]. Due to the lack of labelling of the training dataset, the output of unsupervised learning is usually unknown. This property makes unsupervised learning often used to explore new possibilities in existing cardiovascular omics. Self-supervised learning methods have been proposed in recent years, and the main process of which is to automate the process of assigning pseudo-labels to the training dataset, saving costs through automated annotation [18].

Reinforcement learning, as a technique to improve models based on error feedback, achieves model performance enhancement through cumulative effects [19]. Currently, applications of reinforcement learning in cardiovascular research are focused on the design of drugs or proteins [20]. In simple terms, the model lengthens or bends the molecules provided by constantly correcting errors. In this way, researchers are making hundreds of AI-designed proteins and confirming feasibility of patent medicine in the laboratory [21].

Other ML methods

In addition to traditional ML methods, evolving DL processes information by mimicking the neural network of the human brain, which usually consists of a large number of computational neurons converging into layers and

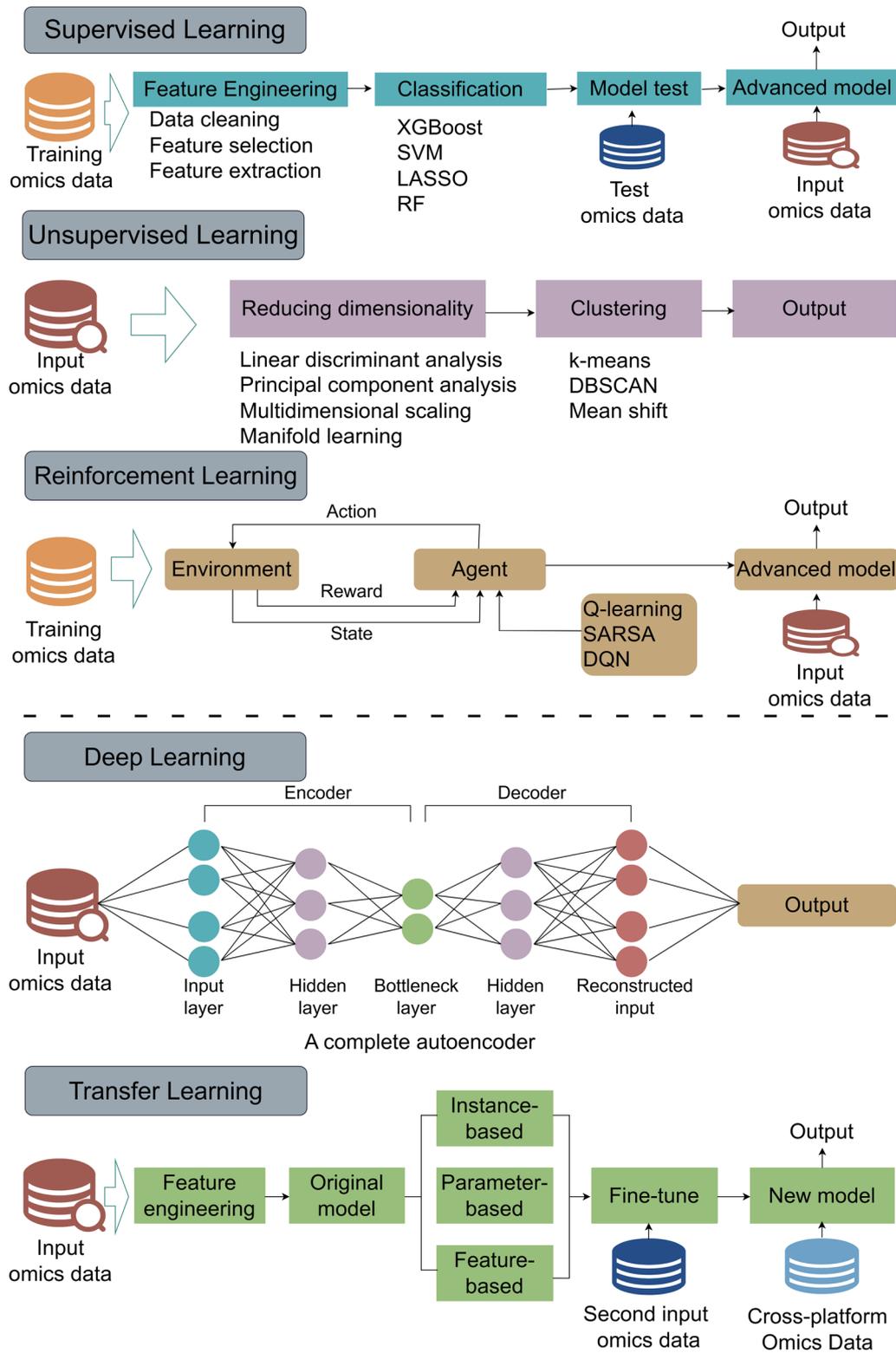


Fig. 2 Schematic of the ML pipeline in multi-omics studies

communicating with neurons in other layers [22]. With the recent use of Transformer-based large language models in omics, there has been a significant increase in read length for omics sequence fragments to predict long-range interactions and scarce data tasks [23]. However, the cost of annotating large volumes of data for DL, the problem of local versus extreme generalisation, the complexity of the training process, and the inability of the process to be interpreted remain obstacles to the creation of powerful models. Due to the high cost involved in training advanced models, researchers have come up with transfer learning, which has the ability to map a trained model to a research model for another research purpose. Using technologies such as instance-based, parameter-based and feature-based algorithms has been explored for cross-platform, cross-species integration of transcriptomics data [24]. Since transfer learning may have negative transfer events contrary to expectations when iterating from the source region to the target task, there is a need to perform research-context-based quality control and set reasonable transfer boundaries during the transfer learning process [25].

ML integrates multi-omics data

Different ML models often show great performance differences after integrating multi-omics data during external validation. The model construction framework, omics data quality, and upstream integration strategies of are the main factors affecting the quality of models [26]. The model construction framework is often selected according to the actual purpose of the research. For example, semi-supervised autoencoders utilise a limited number of labels to enhance the learning performance of complex datasets, whilst generalising labelled datasets to unlabelled examples. They depend on a small set of positioning labels to CVDs-related representations in an unbiased manner, which is essential for constructing predictive models for CVDs. However, it is important to acknowledge that traditional unsupervised autoencoders inherently avoid any biases introduced by labelling, as they do not rely on labels. This attribute is particularly advantageous for exploratory research aimed at discovering potential CVD-related targets that have not previously been documented in the literature, especially in complex datasets such as multi-omics. The strategy of using ML models to integrate multiple omics is equally important [14]. This includes early integration that directly connects data sets, mid-term integration that identifies common latent structures in data sets through methods such as joint matrix decomposition, or late integration that is applied to each data set separately and combined with predictions after analysis. In addition, there are strategies such as hybrid integration and hierarchical integration.

The choice of different strategies also leads to different results.

Prediction and prevention of CVDs

Advancements in omics technologies have facilitated the identification of high-risk populations for monogenic diseases such as long QT syndrome [27]. However, the regulatory mechanisms of CVDs are often complex and heterogeneous due to the influence of multiple genetic and environmental factors [28]. Traditional risk scores that focus on classifying risk groups based on clinical features (such as smoking, alcohol consumption, and diabetes history) can predict relative risk to some extent, but their predictive accuracy at the individual level remains limited. By building ML models to solve the existing complexity and heterogeneity of multi-omics data, researchers have been able to discover, with unprecedented efficiency, the risk ratios of gene expression combinations that traditional analysis methods have failed to focus on. So that clinicians can identify individuals at risk of developing CVDs in the future from healthy populations, and initiate individualised and precise preventive diagnosis and treatment in a timely manner (Table 1).

For the risk of future coronary heart disease and MI events prediction, GPS-mult, a model based on supervised learning, was trained by scoring genomic data from 116,649 multiracial individuals who had not yet developed atherosclerotic cardiovascular diseases (ASCVDs), was shown to outperform traditional risk scoring models in predicting the risk of coronary heart disease over the next decade or more, which more accurately distinguishes between high-risk and safe individuals in healthy populations [29]. Compared to genomics, the use of plasma proteomics to complete supervised learning training has demonstrated greater value in the prediction of short-term risk, particularly in predicting earlier myocardial infarction or adverse cardiovascular events [30]. Hoogeveen et al. used ML model and showed superior prediction of traditional clinical risk factors over Framingham risk scores by combining targeted proteomics with clinical datasets from a large European cohort [13]. In the face of changing lifestyles and medical interventions over the course of life, repetitive proteomic-based risk assessment as a "liquid health check" may help to further improve lifelong risk assessment. In addition, ML model training using plasma metabolomics could also serve the purpose of predicting short-term CVDs events, which still needs to be validated with higher quality datasets [31]. Complex multi-omics data, analysed using such ML models, can identify intricate non-linear associations and interactions among various biological layers. These methods allow us to capture complex patterns that go beyond the conventional

Table 1 ML models for the risk prediction and prevention of CVDs

Disease	Method	Data type	Code availability	Refs
ASCVDs	Integration	Multi-omics	https://github.com/noambar/ACStudy	[28]
ASCVDs	Integration	Genomics	https://www.nature.com/articles/s41591-023-02429-x	[29]
ASCVDs	DNN, RF	Multi-omics	https://github.com/thbuerg/NeuralCVD	[30]
ASCVDs	XGBoost	Proteomics	-	[13]
ASCVDs	BR	Multi-omics	https://github.com/xuyu-cam/atlas_genetic_scores_omic_traits	[32]
ASCVDs	RF	Proteomics	-	[34]
ASCVDs	SVM, RF, LR	Transcriptomics	-	[35]
ASCVDs	Integration	Genomics	https://github.com/LichtargeLab/EAML	[36]
ASCVDs	GCN	Genomics	-	[37]
HF	RF, SVM, LASSO	Genomics	-	[38]
HF	Integration	Epigenomics	-	[39]
HF, Age-related Disease	LASSO	Multi-omics	https://github.com/hamiltonoh/organage	[40]
CVDs	RNN	Metabolomics	https://github.com/thbuerg/MetabolomicsCommonDiseases	[41]
HF, AF, Other CVDs	RF	Multi-omics	-	[17]
Sudden Cardiac Death	Integration	Genomics	https://fabricgenomics.com/fabric-gem/	[42]
Arrhythmia	LR, MLP	Multi-omics	https://github.com/rachellea/medgenetics	[44]
AF	CNN	Genomics	-	[45]
Hypertension	XGBoost	Multi-omics	https://github.com/Tran031194/integrating_omics_BP	[46]
Hypertension	SVM	Multi-omics	https://github.com/Dongwon-Lee/bph2	[47]

ASCVDs, Atherosclerotic Cardiovascular Diseases; HF, Heart Failure; AF, Atrial Fibrillation; CVDs, Cardiovascular Diseases; GCN, Graph Convolutional Network; SVM, Support Vector Machine; FNN, Feedforward Neural Network; RF, Random Forest; LR, Logistic Regression; BR, Bayesian ridge Regression; LASSO, Least Absolute Shrinkage and Selection Operator; CNN, Convolutional Neural Network; MLP, Multilayer Perceptron; XGBoost, eXtreme Gradient Boosting; RNN, Recurrent Neural Network; DNN, Deep Neural Network

linear interpretations typically made by human analysis, revealing insights that are often inaccessible through traditional approaches [7]. A Bayesian network-based ML model involved 50,000 participants with three levels of omics data: transcriptomics, proteomics and metabolomics, which included six causal relationships between changes at the omics level and the eventual occurrence of ASCVDs [32]. However, the combination of multi-omics did not seem to significantly outperform the former single-omics model in predicting endpoints, which may be due to the complexity of the dataset related to the algorithms chosen, and a more advanced algorithmic framework such as the much further advanced AtheroNET may be needed [33]. Another model developed by Núñez et al. based on plasma proteomics achieved prediction of subclinical coronary atherosclerosis stage and proposed APOA, IGAH2 and HPT as a new combination of risk proteins [34]. In addition, AI has helped researchers discover new ASCVDs risk genes or mutations that can help refine prediction models as new parametric indicators [35–37].

For people potentially at risk of heart failure (HF), the AI model developed by Yang et al. is able to identify single nucleotide polymorphisms (SNPs) associated with HF risk in genomics for distinguishing high-risk populations that have not yet developed HF [38]. HFmeRisk used a

combination of DNA methylation data and clinical characteristics to predict the early risk of developing HF with preserved ejection fraction in the Framingham Heart Study cohort to help clinicians make decisions to prevent the disease in high-risk populations [39]. Hamilton SeHwee Oh et al. introduced a learning framework called Organage, which uses plasma proteomics to model organ health and biological ageing. In a 15-year follow-up of a population with no initially active disease or abnormalities in clinical biomarkers, for every 4.1 year of increasing cardiac age was associated with a nearly 2.5-fold increase in the risk of HF and a 23% increase in the risk of cardiac senescence per year, and the long-term risk of HF could be effectively estimated by organ-specific age predictions [40]. At the metabolomic level, Thore Buergele et al. used a multitasking residual neural network to explore MRI-derived metabolomics profiles. The discriminative improvements of predicting the development of HF and other cardiovascular diseases in the next 10 years is superior to traditional clinical predictors [41].

Malignant arrhythmias are often an important cause of sudden death in young people. Arrhythmias are most often predicted using cardiac electrophysiological data. However, most individuals who die suddenly do not show abnormal electrocardiograms. Multi-omics methods based on AI can combine genetic information and

potential disease phenotypes to provide a comprehensive and highly accurate prediction conclusion. Electrocardiographic changes in the absence of an attack are usually hidden, which makes it a difficult task to prevent sudden death. ML is useful for identifying multi-omics features of people at high risk of sudden cardiac death, which can help to identify potentially at-risk populations and avoid related tragedies in time, especially for the young people with unexplained sudden deaths [42]. Although the current accuracy for predicting sudden cardiac death is not yet at the level expected by researchers [43], genetically characterised ML models are able to predict arrhythmia-associated genetic variants with a high degree of accuracy, or to identify potentially causative genes, providing ideas for the study of disease mechanisms [44]. In another study utilising genome-wide data from an East Asian population for atrial fibrillation (AF) prediction, a deep neural network constructed based on the cumulative effect of SNPs and genetic interactions for feature selection also achieved $AUC > 0.75$ performance for prediction of an external multiethnic cohort. Although not reaching the high level of precision required at the individual level, this study presents another perspective on the primary prevention of AF by examining more stable genomic sequences [45]. Considering the difficulty of obtaining cardiac transcripts, a ML model constructed by performing transcriptomics on peripheral blood was used to assist in finding people at high risk for HF and atrial fibrillation (AF) [17]. Such applications can accordingly be extended to other CVDs to address the individualised needs of patients. The studies by Lee et al. and Louca et al. have successfully discovered features that regulate hypertension in high-risk populations using ML models, such genetic variation or key metabolites (cis-4-decenoyl carnitine, lactate and so on) [46, 47].

At present, from a single-omics level, DL models based on genomic or epigenomic data tend to predict CVDs risks in the longer term, while DL models trained by plasma proteomics and metabolomics provide ideas for short-term risk prediction. Long-term risk prediction performance depends on the continuously updated GWAS data and characteristic gene tags [48]. From a public health perspective, the long-term benefits of CVDs prevention in healthy people are more worthwhile, especially some genetic information indicates a higher probability of sudden cardiac death in young people. Likewise, breakthroughs are being made in the genetic characterization of plasma proteomics and plasma metabolomics. The ability to predict the short-term risk of CVDs in the population will also continue to improve [49, 50]. The choices made in terms of training sets for populations of CVDs highlight the need to consider the costs and benefits of adding an omics layer, as

adding data does not systematically improve model performance, but increases model complexity in the training set which may reduce the reproducibility of models in external cohorts [51]. Some risk prediction models attempt to skip feature selection and use unsupervised prediction methods, similar to probabilistic models of clustering effects, at the expense of personalized accuracy, but doing so can significantly reduce running costs [52]. The researchers applied ML model in retrospective studies to identify novel susceptibility genes related to CVDs, demonstrating the strong verifiability and scalability of ML in leveraging existing information. However, these genes identified by the model still require validation in prospective cohorts. The multi-omics information in the human body acts like a dynamic fingerprint of an individual's health status. With the assistance of ML, this information facilitates personalised assessment and prediction, which is crucial for guiding future prevention strategies.

Cardiovascular biomarkers and early disease diagnose

Classic biomarkers such as troponin and BNP can help clinicians diagnose MI and decompensated HF. However, the onset of CVDs, such as early coronary artery disease and preclinical symptoms of HF, is often insidious without changes in BNP. Supported by AI, multi-omics technologies, can paint an individualised picture of the early intra-organismal landscape of disease at a much faster rate, helping to advance early diagnosis and further typing of CVDs (Table 2).

Nurmohamed et al. used targeted proteomics to optimise existing clinical risk models such as SMART, Reynolds Risk Score and Framingham Risk Score. The AI model constructed on this basis predicted recurrent cardiovascular events in two large cohorts of diagnosed ASCVDs. This approach improved the AUC from 0.75 to 0.81 compared with the traditional risk score, enabling risk stratification for secondary prevention of ASCVDs and highlighting the ability to predict coronary heart disease in the non-inflammatory manifestation phase [53]. Another AI-enabled prediction model developed by Zhang et al. integrates peripheral blood leukocyte DNA methylation-regulated genes and transcriptome data from the Framingham Heart Study cohort, and constructed a coronary heart disease prediction model using five hub methylation-regulated genes as biomarkers of ASCVDs, which also proved superior to traditional phenotypic models [54]. ML model identified diagnostic markers associated with plaque instability in coronary atherosclerosis, enabling prediction of unstable plaques in both external datasets and clinical samples. This method is expected to replace invasive testing

Table 2 ML models for CVDs biomarker discovery and disease diagnosis

Disease	Method	Data type	Code availability	Refs
ASCVDs	XGBoost	Proteomics	–	[53]
ASCVDs	LightGBM, XGBoost, RF	Multi-omics	–	[54]
ASCVDs	Integration	Transcriptomics	–	[55]
ASCVDs	Integration	Proteomics	–	[56]
ASCVDs	LASSO, LR	Proteomics	–	[57]
HF	RF	Metabolomics	–	[58]
HF, DCM	LASSO, XGBoost, RF, NN, SVM	Multi-omics	https://www.frontiersin.org/articles/10.3389/fcell.2022.1089915/full	[59]
AF	NB, SMO, RF	Multi-omics	–	[60]
Hypertension	Integration	Multi-omics	https://doi.org/10.5281/zenodo.7033087	[61]
Cardiomyopathy	Integration	Transcriptomics	https://doi.org/10.5281/zenodo.3941331	[62]
DCM	LASSO, SVM	Transcriptomics	https://doi.org/s41598-022-19027-5	[63]
HCM	SVM, RF, KNN	Transcriptomics	https://doi.org/10.6084/m9.figshare.14650536	[64]
HCM	RF	Metabolomics	–	[65]
CVDs	Integration	Genomics	https://github.com/exomiser/Exomiser/	[66]

ASCVDs, Atherosclerotic Cardiovascular Diseases; HF, Heart Failure; DCM, Dilated Cardiomyopathy; AF, Atrial Fibrillation; HCM, Hypertrophic Cardiomyopathy; PAH, Pulmonary Hypertension; CVDs, Cardiovascular Diseases; LightGBM, Light Gradient Boosting Machine; XGBoost, eXtreme Gradient Boosting; RF, Random Forest; LASSO, Least Absolute Shrinkage and Selection Operator; LR, Logistic Regression; NN, Neural Networks; SVM, Support Vector Machine; NB, Native Bayes; SMO, Sequential Minimal Optimization; KNN, K-nearest Neighbors

as a convenient way of identifying plaque morphology to avoid the risk of patient-related invasive injury [55]. For example, the use of ML tools to predict the progression of subclinical calcified plaques avoids the radiation as well as contrast burden of repeated imaging with radiation [56]. The integrated AI-driven proteomics model developed by McCarthy et al., which successfully diagnoses obstructive coronary lesions in patients with suspected MI. This holds the potential to be useful in the diagnosis of acute MI earlier in the time window of troponin elevation, and overcomes the current limitations of limited serological testing capacity in the hyperacute phase of infarction [57].

In the early diagnosis of HF, a combination of ML model and untargeted metabolomics enables the prediction of early HF with reduced ejection fraction (HFrEF) by circulating metabolites, which appears to be comparable in diagnostic value to the more general BNP currently available and has particular application in specific clinical settings such as patients treated with BNP supplementation [58]. At the single cell level, Zhu et al. identified highly correlated genes expressed in both dilated cardiomyopathy (DCM) and HF as reliable markers for diagnosing early HF in patients with DCM by testing several ML methods and selecting the best performing RF model [59]. For patients with valvular heart disease, AF is the most common and insidious complication and a risk factor for future HF. Bayesian network-based ML model identifies biomarkers associated with AF pathogenicity in heart valve disease at transcriptomic and proteomic

levels and possesses high predictive value, providing possible targets for the study of the mechanisms of AF occurrence [60].

Multi-omics-based ML is expected to not only enable early diagnosis of diseases, but also help clinicians to classify subtypes of diseases. For example, identifying different causes and contributing factors of hypertension is key to targeted management and reduction of cardiovascular complications. Reel et al. used ML analysis of multi-omics in blood and urine to differentiate between primary and secondary hypertension, and further identified subtypes of secondary hypertension [61]. It can help clinicians to choose targeted anti-hypertensive strategies to avoid cardiovascular complications. Alimadadi et al. developed ML models capable of classifying types of human cardiomyopathy by identifying 50 highly correlated genes with high accuracy, distinguishing ischaemic from non-ischaemic cardiomyopathy with greater accuracy [62]. In addition, ML models enable early diagnosis of different types of cardiomyopathies through multi-omics [63–65].

The addition of DL models based on WES to the traditional diagnostic process reduces the time spent analysing gene sequencing in the diagnostic laboratory and greatly improves diagnostic efficiency [66]. ML has the potential to drastically shorten the diagnostic time window, improve diagnostic accuracy and speed of diagnosis by capturing imperceptible bio-signals in the early stages of a disease from an AI perspective and validating multiple markers against each other at multiple omics levels.

Therefore, omics data at multiple levels are important, whether it is genetics or metabolomics, and their integration is expected to further improve the accuracy of diagnosis. ML in a multi-omics paradigm should move away from binary yes-or-no conclusions and focus on refining disease typing, staging, and disease identification in CVDs. Taking the change in the diagnostic paradigm of hypertension as an example, the highly complex relationship between hypertension and multi-omics discovered by integrating multiple data sets in a DL framework transcends the traditional threshold and risk stratification blood pressure model, and provides multi-omics insights for the exploration of biological mechanisms of blood pressure values as continuous variables [5, 61]. Through multi-omics comprehensive analysis, the GWAS results of human blood pressure are converted into biological insights, and then high-priority blood pressure-related genes are mined and sorted to crack the code of differences in blood pressure values among the population [67, 68]. This reflects the trend of the future multi-omics hypertension landscape under the AI framework.

Cardiovascular therapy

The inefficiency of traditional experimental methods in identifying drug targets has consistently limited the progress of disease treatment. The integration of multi-omics research with (ML facilitates the discovery of previously unknown therapeutic targets and drugs. Additionally,

it enables the testing of proposed drug targets' effects through gene-protein regulatory networks (Table 3).

Drug and target discovery

Yang et al. used BAG3-deficient cardiomyocyte-derived pluripotent stem cells to construct an in vitro expanded cardiomyopathy model, and identified cardioprotective drugs from thousands of biological compounds through phenotypic screening as well as DL modelling, and demonstrated the possibility of accelerating drug discovery by incorporating DL in an in vivo experiment [69]. Geneformer was able to make predictions of downstream targets versus indirect targets with an extremely limited external queues dataset after extensive pre-training on large-scale corpus of nearly 30 million single cell transcriptomics data. And candidate drugs for cardiomyopathy predicted by this model were shown to improve myocardial function [70]. The multi-omics-based DL model developed by Iborra-Egea et al. started with previously identified genes mediating adverse remodelling to map the evolution of in vivo markers at different time points after MI, ultimately identifying IGF1R, RAF1, KPCA, JUN and PTN11 as regulators of cardiac remodelling, which identify potential targets for drug development [71]. A DL model with a new feature representation was used to identify bioactive peptides for the treatment of hypertension, using a new feature representation corresponding to dipeptides as binary numbers and then re-inputs it into the model for validation, achieving 99%

Table 3 ML models for treatment strategies in CVDs

Category	Function	Method	Data type	Code availability	Refs
Drug & Target Discovery	Candidate Target Discovery	Integration	Transcriptomics	https://huggingface.co/ctheodoris/Geneformer	[70]
	Candidate Target Discovery	ANN	Multi-omics	https://www.mdpi.com/article/10.3390/cells10123268/s1	[71]
Drug & Target Validation	Finding Drugs	CNN, GRU	Proteomics	http://ahttps.zhanglab.site/	[72]
	Drug Repurposing Predictions	Integration	Multi-omics	https://github.com/NCATSTranslator/Translator-All/wiki/Molecular-Data-Provider	[73]
Efficacy & Side Effects	Drug-target Interaction Validation	LASSO, RF	Proteomics	https://github.com/QUST-AIBBDR/LRF-DTIs/	[74]
	Drug-target Interaction Validation	Integration	Multi-omics	http://120.77.11.78/DeepMPF/	[75]
	Efficacy Verification	Integration	Genomic	–	[76]
Efficacy & Side Effects	Efficacy Verification	FNN, SVM	Transcriptomics	–	[77]
	Efficacy Verification	Integration	Genomic	https://www.hgsc.bcm.edu/software	[78]
	Prediction of Drug Side Effects	Integration	Transcriptomics	https://github.com/vauyeung38/sct-transcriptomics-waveforms	[79]
	Prediction of Drug Side Effects	Integration	Proteomics	http://github.com/yuliyi/MSDSE	[81]
	Prediction of Drug Side Effects	CNN, MLP	Multi-omics	http://github.com/OnurUner/DeepSide	[82]
	Drug Sensitivity Prediction	GNN	Multi-omics	https://github.com/kkayle/GCFMCL	[83]
	Optimization of Drugs	Integration	Multi-omics	https://github.com/dahjan/DMS_opt	[84]

ANN, Artificial Neural Network; CNN, Convolutional Neural Network; GRU, Gated Recurrent Unit; KNN, K-nearest Neighbors; LASSO, Least Absolute Shrinkage and Selection Operator; RF, Random Forest; DNN, Deep Neural Network; FNN, Feedforward Neural Network; SVM, Support Vector Machine; GNN, Graph Neural Network; MLP, Multilayer Perceptron

accuracy when validated against external data [72]. In addition, ML is utilised for indications for drugs that are already in use [73].

Drug and target validation

To validate the effects of drugs on the acting target, the method called LRF-DTI through the integration of multiple ML algorithms achieved an overall correct rate of 94.88% in predicting the effects of drug-target interactions of different types of receptors, including enzymes, ion channels, G protein-coupled receptors and nuclear receptors [74]. DL model called DEEPMPF constructed a protein-drug-disease heterogeneous network consisting of three entities. By calculating drug-target interaction probabilities through joint learning, the model demonstrated competitive predictive performance in screening proteins as bioactive compounds [75]. Although this type of research is not a substitute for *in vivo* experiments at this time, its cost-effectiveness represents a promising prospect for a wide range of applications.

Efficacy and side effects

In the translation of AI into clinical applications, a new framework is being used to predict drug efficacy in phase 3 clinical trials. The ML performed drug efficacy predictions for 24 HF drugs from 266 phase 3 clinical trials used to assess the efficacy of repurposing drugs to treat targeted CVDs [76]. For drugs or genes known to be potentially available, SVM neural networks have been used to predict the efficacy of three novel biomarkers, HBG1, SNCA and GYPB, for AF associated with stroke [77]. The Mayo-Baylor RIGHT 10 K study combined genomics with a DL approach to identify deleterious gene variants at the individual level for specific clopidogrel-resistant patients with ASCVDs using pharmacogenomic AI prediction, potentially improving patient care through dose adjustment or alternative treatments [78]. The models described above facilitate the conduct of large-scale cohort clinical trials of drugs, saving hundreds of millions of dollars in research and development.

Several studies have been conducted to develop corresponding ML models for adverse drug reactions. Phenotypic and transcriptomic data from physiologically relevant cardiac models of multiple cardiotoxic compounds through the application of ML approaches in the treatment of oncology drugs for improved guidance of structured cardiac toxicity of chemotherapeutic agents and the ability to identify potential target gene markers to aid subsequent targeted drug development [79, 80]. MSDSE learns and integrates multimodal features from local to global perspectives to cope with possible drug side effects in clinical trials [81]. The same pharmacogenomics-based DL has enabled the prediction of possible

adverse reactions to multiple drugs acting simultaneously in the human body, particularly those reported in the literature but missing in the ground truth side effect dataset [82]. In order to predict the sensitivity between miRNAs and drugs, a multi-view contrast learning model for graph collaborative filtering called GCFMCL is the first attempt to for predicting sensitivity between miRNAs and drugs, which may be promising for applications in overcoming drug resistance in humans [83]. In addition, DL from human omics data can enable optimisation of existing drugs, as demonstrated in antibody engineering [84].

De novo sequence design and drug validation through AI are in full swing. Interaction and docking models have contributed well to drug development. However, the current situation is that prediction models for phase II and III clinical trials are still relatively limited. In addition, the *in-vivo* multi-omics information after drug intervention used to describe the treatment response is still waiting to be improved. The ML model based on the integration of such multi-omics data (especially transcriptomics, proteomics and metabolomics) is expected to simulate the human body picture after specific drug treatment. In this way, it is expected to increase a round of AI drug screening experiments, thereby saving huge economic costs.

Prognostic prediction of CVDs

Individual prognosis is a crucial component of a clinician's assessment of treatment efficacy and strategic planning. Multi-omics can comprehensively and accurately reflect changes in the body in response to therapeutic interventions, allowing for timely adjustments to therapeutic strategies. The integration of ML with multi-omics allows for the projection of CVDs trajectories, anticipated adverse events, and survival outcomes (Table 4).

Prognosis of CVDs

A ML model developed by Wallentin et al. based on plasma proteomics newly identified 13 proteins associated with cardiovascular-related mortality in chronic coronary artery disease. The c-statistic of 0.71 and 0.79 was achieved in the prediction of two large cohorts [85]. Another model focusing on patients with chronic coronary artery disease undergoing secondary prevention or haemodialysis combined plasma proteomics to ultimately identify 8 biomarkers associated with cardiovascular mortality that could assist in the prediction of CVDs prognosis. However, the predictive ability for nonfatal cardiovascular events requires further research due to the lack of data on related events [86]. Some ML models have failed to achieve effective prediction of the occurrence of adverse events in patients with acute coronary syndromes, possibly owing to the fact that these models

Table 4 ML models for predicting prognosis in CVDs

Category	Disease	Method	Data type	Code availability	Refs
Prognosis of CVDs	ASCVDs	RF	Proteomics	http://www.clinicalstudydatarequest.com/	[85]
	ASCVDs	SVM,LightGBM,XGBoost	Multi-omics	https://doi.org/10.1038/s41397-021	[86]
	ASCVDs	LASSO	Proteomics	–	[87]
	ASCVDs	LASSO	Proteomics	–	[88]
	CVDs	LR, XGBoost, NN	Proteomics	https://doi.org/10.1016/j.jacc.2021.08.018	[93]
	HF	LASSO,DeepHit,XGBoost,RF	Proteomics	–	[94]
	HCM	RF	Proteomics	–	[95]
	CVDs	LASSO	Proteomics	–	[89]
	CVDs	DNN	Multi-omics	https://doi.org/10.1186/s12874	[91]
	Complications prediction	Heart transplant	Integration	Multi-omics	–
Heart transplant		Integration	LASSO, SVM-RFE	–	[97]
ASCVDs		RF, LASSO	Metabolomics	–	[98]
ASCVDs		LASSO	Proteomics	–	[99]
Cardiac arrest		SVM, RF, LASSO	Multi-omics	https://github.com/johncolby/SVM-RFE	[100]

CVDs, Cardiovascular Diseases; ASCVDs, Atherosclerotic Cardiovascular Diseases; HCM, Hypertrophic Cardiomyopathy; HF, Heart Failure; RF, Random Forest; SVM, Support Vector Machine; LightGBM, Light Gradient Boosting Machine; XGBoost, eXtreme Gradient Boosting; LASSO, Least Absolute Shrinkage and Selection Operator; CNN, Convolutional Neural Network; LR, Logistic Regression; NN, Neural Network; DNN, Deep Neural Network

have selected partial combinations of protein markers rather than the whole proteome data for modelling [87, 88]. Current omics prognostic models for patients with acute coronary syndromes are still limited, and the main reason for this may be the lack of high-quality omics data and sufficient number of accepted prognostics markers [89]. In the future, researchers need to further explore more biomarkers that are highly correlated with recovery and survival in patients with infarction, and can also try to combine other parameters for model iteration, such as combining electrocardiography with omics data [90, 91].

A large proportion of patients with CVDs may progress to irreversible end-stage HF. Predicting the risk of survival from HF has a significant impact on the assessment of patient readmission rates, the planning of rehabilitation training and the advancement of heart transplant programmes. Ouwkerk et al. used ML to integrate genomics, transcriptomics, proteomics and clinical data, predicted risk of all-cause mortality in a cross-validation cohort of patients with central decline with an AUC of 0.81, and it also identified four pathways that influence the progression of HF [92]. Another DL model developed by Unterhuber et al. enables improved prediction of all-cause mortality in patients with CVDs using high-throughput proteomics only compared to traditional clinical scoring and regression models, the single-omics approach sacrifices some performance, it reduces the difficulty of training and still manages to outperform traditional models (C-statistic increase 0.17–0.19) [93]. However, due to the lack of heart failure proteomics data, no significant improvement in the prediction of

HF outcomes was found in another performance validation of the proteomics multiple ML models [94]. In addition to predicting all-cause mortality in patients, the use of ML also allows the assessment of possible non-fatal events in patients. For example, Shimada et al. developed a proteomics-based model to predict possible future adverse cardiovascular events in patients with hypertrophic cardiomyopathy, In the predictive performance test of this model, its AUC reached 0.81 [95].

Complications prediction

For postoperative anti-rejection in cardiac transplant patients, the identification of specific genes associated with different subtypes of cardiac transplant rejection using ML models trained in multi-omics (e.g., Allomap and the EMB) outperforms traditional methods of monitoring cardiac transplant rejection in terms of predictive ability for acute transplant rejection [96, 97]. Song et al. construct a metabolomics model to predict the risk of recurrent angina in patients after PCI [98]. And it achieved over 89% accuracy in the prediction of all three large external prospective cohorts. In another study combining proteomics with clinical information, ML models were able to directly predict procedural kidney injury in coronary angiography patients, despite a limitation of only 79% cross-validation accuracy due to ignoring contrast dose [99]. However, limitations of the study such as failure to consider contrast dosage resulted in a cross-validation accuracy of only 79%. In addition, Li et al. used a gene co-expression network constructed with ML to interpret neurological damage in cardiac arrest patients

from a molecular perspective rather than clinical representations [100]. Although the researchers used microarray data with missing values and the correlation factors identified require further experimental validation, which reduces the model's credibility. Increasing the data volume could inspire new strategies for brain resuscitation following cardiopulmonary resuscitation.

For now, the performance of prognostic models has not met expectations for widespread use, and model construction remains a significant challenge. Variations in patient function, treatment strategies, dosage selection, and treatment response complicate the integration and attribution of multi-genomic data, limiting the models' generalizability to diverse patient populations. To address this issue, some studies have focused on a select few specific proteomics for prediction, which inevitably decreases prediction accuracy and can lead to negative conclusions. For the prediction of disease prognosis, due to the difficulty in obtaining transcriptomics of cardiac samples, it is more common to use plasma proteomics and metabolomics for ML modelling. However, the lack of data is still a problem that constrains the prognostic model [94]. For current omics research, even the multi-omics data of patients with CVDs who have received the most classical treatments are still incomplete, which makes it difficult to depict the prognostic picture after treatment. At this stage, ML is better suited to serve as an adjunct rather than a substitute for clinicians.

Challenges and prospects

Omics technologies are increasingly being integrated into various comprehensive CVDs studies, significantly impacting clinical decision-making and prediction. ML-based integrative algorithms have rapidly enhanced analytical efficiency and outperformed traditional risk prediction models. However, to achieve the widespread application of ML-based omics research in clinical practice, several challenges and limitations persist.

The first challenge is data collection and standardisation. Conventional blood samples do not accurately reflect regulatory activity in the heart due to the lack of tissue specificity and the ease of RNA degradation [101, 102]. Multi-omics studies often require cardiac samples to accurately map the regulatory networks within the system. In the process of extracting cardiomyocytes, due to the highly developed intercellular connections in myocardial tissue, the large size and myofibril structure of adult cardiomyocytes make them easily broken by mechanical shear force or enzymatic hydrolysis, resulting in low recovery rates [103]. This not only prevents the omics results from reflecting the real situation *in vivo*, but also leads to differences in the depth of sequencing and quality of data obtained

between different studies. Several technologies have been developed to address this challenge, such as using reversible hydrogels for sample preparation or the addition of (-)-blebbistatin and myosin II ATPase inhibitor to maximise the preservation of cardiomyocyte activity [104, 105]. In addition, Live-seq avoids transcriptional changes during cardiomyocyte inactivation and maximises the preservation of temporal cardiomyocyte transcriptional activity [106]. Effective and standardised experimental conditions and processes need to be established in the future to provide more widely accepted omics benchmark data for AI model training. Especially when using related technologies to transfer models (such as transfer learning) or replicate results, inconsistencies in protocols may lead to misinterpretation of conclusions. Since different platforms do not provide the same systematic analysis, even overlapping detections between different platforms may lead to different results, such as Olink and Somalogic proteomic datasets [7]. This variability can lead to differing associations with CVDs outcomes depending on the platform used. Such inconsistencies underscore the need for standardised protocols or, at a minimum, a thorough understanding of platform-specific differences within the scientific community. In addition, most of the current open-source omics data are retrospective, and it is challenging to repeatedly obtain cardiac samples for prospective prediction of disease progression in large populations. In recent years, several large cardiovascular cohorts, such as the Framingham Cardiovascular Study, have progressively incorporated omics data over several years of follow-up [48, 107]. Although current studies are not yet at an ideal scale, the iterative evolution of models across generations with the release of more open-access omics datasets is foreseeable.

The second challenge is the integration of multimodal data. With the application of ML, research in clinical electronic health records, electrophysiology, imaging and cardiac ultrasound is rapidly evolving. Integrating omics research with everyday multimodal medical information is essential for overall improvement in clinical practice [108]. In clinical environment, the integration of patient cardiac ultrasound and medical history with omics data requires not only consideration of the effects of platform noise, but also the use of autoencoders or deep generative models for feature extraction and dimensionality reduction of high-dimensional omics data [109]. Multimodal medical fusion networks have been developed to process multimodal medical information jointly [110]. However, there are still key challenges, such as exacerbating research disparities and overemphasizing outcomes when new predictive indicators are added to existing training models with low-weight key indicators. Researchers need

to balance the cost of retraining or re-evaluating performance for newly added information layers.

Furthermore, ethical considerations are critical when building AI integrated models. Unethical collection or use of omics data, such as involving race or gender-biased input subjects, can lead to biased model outputs [111]. AI predictions based on omics data for unauthorised genetic editing experiments also pose serious ethical issues. Therefore, strict oversight and registration of AI projects are urgently needed. As training datasets consist of large amounts of human genomic data, any research organisation must take data security seriously. When publishing research models online, appropriate privacy protocols must be included to protect omics data from the theft of important omics information by related criminals [112, 113]. To this end, we call on the industry to form internationalised standards or regulations, and future efforts should aim to fill this knowledge gap and ensure the proper development of AI.

There are also challenges in the AI development process. AI is often described as a "black box", wherein the underlying decision-making processes remain opaque. This opacity impedes distinguishing the impacts of different new parameters added to the model, such as age and gender, on the resulting predictions. It is unclear how the parameter weights associated with different features have changed, especially as the model complexity increases. Consequently, it is difficult to ascertain whether the addition of parameters has positive or negative consequences, often undermining confidence in the model's predictive accuracy [114]. To address the issue of indecipherability, researchers have proposed technologies such as neuralisation propagation, hidden state analysis, variable importance measures, and feature visualisation to explain the operating mechanisms of models, or strategies that combine multiple model frameworks to enhance credibility [115–118].

Lastly, the issue of generalisation in ML models emerge as a pivotal concern. In the context of CVDs research, due to the heterogeneity of patient data (including variations in genetics, lifestyle, and comorbidities) and the current insufficiency of open-source omics data, model training is more prone to capture noise in the training dataset, leading to overfitting and diminished predictive performance on external cohorts [119, 120]. One solution is to expand the dataset size for parameter training to avoid overfitting, which requires comprehensive consideration of patient data diversity and disease pattern complexity. Incorporating manual biological insights can significantly enhance feature selection within existing datasets. For instance, alongside potential biomarkers identified by the AI model, experimentally validated CVDs biomarkers from the literature can be integrated to construct more

robust and biologically informed models. Furthermore, regularization techniques can be applied during model training to mitigate overfitting by penalising excessive model complexity, thereby reducing the likelihood of learning noise. However, too many constraints may result in the performance degradation [121]. Another approach is the cross-validation technique, which divides various datasets into subsets for training and evaluation. The aim is to determine the optimal hyper-parameters for most generalised models [122]. Furthermore, transfer learning and domain adaptation approaches are also candidates to improve predictive potential across multiple datasets [123]. Future research should focus on developing models that are predictive, adaptable, and interpretable in different medical settings, possibly involving the integration of multi-omics data, longitudinal patient records, and environmental factors to create comprehensive integrated models.

In conclusion, despite those challenges, the joint application of ML and omics in CVDs research holds promising prospects. With the advent of the era of precision medicine, harnessing the efficient management capabilities of AI can effectively alleviate the heavy burden of CVDs research and management around the world. The convergence of ML and multi-omics technologies is an evolving field that is rapidly advancing the understanding of CVDs from a molecular perspective. A staged summary of what has been achieved in the collision of two rapidly evolving fields is conducive to new inspiration, while researchers and clinicians must weigh the uncertainties of AI results to avoid the potential pitfalls of overreliance on AI in clinical environments and care services.

Abbreviations

AI	Artificial intelligence
ASCVDs	Atherosclerotic cardiovascular diseases
AF	Atrial fibrillation
CVDs	Cardiovascular diseases
DCM	Dilated cardiomyopathy
DL	Deep learning
GWAS	Genome-wide association study
HF	Heart failure
HFrEF	Heart failure with reduced ejection fraction
ML	Machine learning
MI	Myocardial infarction
RF	Random Forest
SNPs	Single nucleotide polymorphisms

Acknowledgements

We would like to thank Dr. Shuzhi Yu for his professional advice on AI in this article.

Author contributions

Pengyu Jia, Yingxian Sun, Mingzhi Lin, and Jiuqi Guo conceived, designed and drafted the manuscript. Pengyu Jia, Yingxian Sun, Mingzhi Lin, and Jiuqi Guo, and Dalin Jia. revised it for important intellectual content. Zhilin Gu, Wenyi Tang, Hongqian Tao, and S.Y. made contributions to drafted and revised the manuscript. All authors approved the final version of the manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (No. 82100302) and China Postdoctoral Science Foundation (No. 2024MD754005).

Availability of data and materials

Not applicable.

Declarations**Ethics approval and consent to participate**

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 6 August 2024 Accepted: 25 March 2025

Published online: 02 April 2025

References

- Mensah G, Fuster V, Murray C, Roth G. Global burden of cardiovascular diseases and risks, 1990–2022. *J Am Coll Cardiol*. 2023;82:2350–473.
- Leopold JA, Loscalzo J. Emerging role of precision medicine in cardiovascular disease. *Circ Res*. 2018;122:1302–15.
- Hasin Y, Seldin M, Luskis A. Multi-omics approaches to disease. *Genome Biol*. 2017;18:83.
- Vandereyken K, Sifrim A, Thienpont B, Voet T. Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet*. 2023;24:494–515.
- Guo J, Guo X, Sun Y, Li Z, Jia P. Application of omics in hypertension and resistant hypertension. *Hypertens Res*. 2022;45:775–88.
- Kumar KR, Cowley MJ, Davis RL. Next-generation sequencing and emerging technologies. *Semin Thromb Hemost*. 2024;50:1026–38.
- Joshi A, Rienks M, Theofilatos K, Mayr M. Systems biology in cardiovascular disease: a multiomics approach. *Nat Rev Cardiol*. 2021;18:313–30.
- Vadapalli S, Abdelhalim H, Zeeshan S, Ahmed Z. Artificial intelligence and machine learning approaches using gene expression and variant data for personalized medicine. *Brief Bioinf*. 2022;23(5):191.
- Shin H, Roth H, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers R. Deep convolutional neural networks for computer-aided detection: cnn architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging*. 2016;35:1285–98.
- Leon-Mimila P, Wang J, Huertas-Vazquez A. Relevance of multi-omics studies in cardiovascular diseases. *Front Cardiovasc Med*. 2019;6:91.
- Greener J, Kandathil S, Moffat L, Jones D. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol*. 2022;23:40–55.
- Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, Wang Y, Dong Q, Shen H, Wang Y. Artificial intelligence in healthcare: past, present and future. *Stroke Vasc Neurol*. 2017;2:230–43.
- Hoogeveen RM, Pereira JPB, Nurmohamed NS, Zampoleri V, Bom MJ, Baragetti A, Boekholdt SM, Knaapen P, Khaw KT, Wareham NJ, et al. Improved cardiovascular risk prediction using targeted plasma proteomics in primary prevention. *Eur Heart J*. 2020;41:3998–4007.
- Picard M, Scott-Boyer MP, Bodein A, Périn O, Droit A. Integration strategies of multi-omics data for machine learning analysis. *Comput Struct Biotechnol J*. 2021;19:3735–46.
- Glielmo A, Husic BE, Rodriguez A, Clementi C, Noé F, Laio A. Unsupervised learning methods for molecular simulation data. *Chem Rev*. 2021;121:9722–58.
- McLellan MA, Skelly DA, Dona MSI, Squiers GT, Farrugia GE, Gaynor TL, Cohen CD, Pandey R, Diep H, Vinh A, et al. High-resolution transcriptomic profiling of the heart during chronic stress reveals cellular drivers of cardiac fibrosis and hypertrophy. *Circulation*. 2020;142:1448–63.
- Petersen TB, de Bakker M, Asselbergs FW, Harakalova M, Akkerhuis KM, Brugts JJ, van Ramshorst J, Lumbers RT, Ostroff RM, Katsikis PD, et al. HFrEF subphenotypes based on 4210 repeatedly measured circulating proteins are driven by different biological mechanisms. *EBioMedicine*. 2023;93:104655.
- Krishnan R, Rajpurkar P, Topol EJ. Self-supervised learning in medicine and healthcare. *Nat Biomed Eng*. 2022;6:1346–52.
- Wise T, Emery K, Radulescu A. Naturalistic reinforcement learning. *Trends Cogn Sci*. 2024;28:144–58.
- Lutz ID, Wang S, Norn C, Courbet A, Borst AJ, Zhao YT, Dosey A, Cao L, Xu J, Leaf EM, et al. Top-down design of protein architectures with reinforcement learning. *Science*. 2023;380:266–73.
- Tropsha A, Isayev O, Varnek A, Schneider G, Cherkasov A. Integrating QSAR modelling and deep learning in drug discovery: the emergence of deep QSAR. *Nat Rev Drug Discov*. 2024;23(2):141–55.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521:436–44.
- Chandra A, Tünnermann L, Löfstedt T, Gratz R. Transformer-based deep learning for predicting protein properties in the life sciences. *Elife*. 2023.
- Toseef M, Olayemi Petinrin O, Wang F, Rahaman S, Liu Z, Li X, Wong KC. Deep transfer learning for clinical decision-making based on high-throughput data: comprehensive survey with benchmark results. *Brief Bioinform*. 2023.
- Sun Q, Liu Y, Chen Z, Chua TS, Schiele B. Meta-Transfer Learning Through Hard Tasks. *IEEE Trans Pattern Anal Mach Intell*. 2022;44:1443–56.
- Drouard G, Mykkänen J, Heiskanen J, Pohjonen J, Ruohonen S, Pahkala K, Lehtimäki T, Wang X, Ollikainen M, Ripatti S, et al. Exploring machine learning strategies for predicting cardiovascular disease risk factors from multi-omic data. *BMC Med Inform Decis Mak*. 2024;24:116.
- Kaplinger JD, Tester DJ, Salisbury BA, Carr JL, Harris-Kerr C, Pollevick GD, Wilde AA, Ackerman MJ. Spectrum and prevalence of mutations from the first 2,500 consecutive unrelated patients referred for the FAMILION long QT syndrome genetic test. *Heart Rhythm*. 2009;6:1297–303.
- Talmor-Barkan Y, Bar N, Shaul AA, Shahaf N, Godneva A, Bussi Y, Lotan-Pompan M, Weinberger A, Shechter A, Chezar-Azerrad C, et al. Metabolomic and microbiome profiling reveals personalized risk factors for coronary artery disease. *Nat Med*. 2022;28:295–302.
- Patel AP, Wang M, Ruan Y, Koyama S, Clarke SL, Yang X, Tcheandjieu C, Agrawal S, Fahed AC, Ellinor PT, et al. A multi-ancestry polygenic risk score improves risk prediction for coronary artery disease. *Nat Med*. 2023;29:1793–803.
- Steinfeldt J, Buerger T, Loock L, Kittner P, Ruyoga G, Zu Belzen JU, Sasse S, Strangalies H, Christmann L, Hollmann N, et al. Neural network-based integration of polygenic and clinical information: development and validation of a prediction model for 10-year risk of major adverse cardiac events in the UK Biobank cohort. *Lancet Digit Health*. 2022;4:e84–94.
- Li Z, Gong R, Chu H, Zeng J, Chen C, Xu S, Hu L, Gao W, Zhang L, Yuan H, et al. A universal plasma metabolites-derived signature predicts cardiovascular disease risk in MAFLD. *Atherosclerosis*. 2024;392:117526.
- Xu Y, Ritchie SC, Liang Y, Timmers P, Pietzner M, Lannelongue L, Lambert SA, Tahir UA, May-Wilson S, Foguet C, et al. An atlas of genetic scores to predict multi-omic traits. *Nature*. 2023;616:123–31.
- Sopic M, Kararigas G, Devaux Y, Magni P. Call for participation in the AtheroNET COST action to implement multiomics in atherosclerotic cardiovascular disease research. *Eur Heart J*. 2023;44:2143–5.
- Núñez E, Fuster V, Gómez-Serrano M, Valdivielso JM, Fernández-Alvira JM, Martínez-López D, Rodríguez JM, Bonzon-Kulichenko E, Calvo E, Alfayate A, et al. Unbiased plasma proteomics discovery of biomarkers for improved detection of subclinical atherosclerosis. *EBioMed*. 2022;76:103874.
- Peng W, Sun Y, Zhang L. Construction of genetic classification model for coronary atherosclerosis heart disease using three machine learning methods. *BMC Cardiovasc Disord*. 2022;22:42.
- Shapiro D, Lee K, Asmussen J, Bourquard T, Lichtarge O. Evolutionary action-machine learning model identifies candidate genes associated with early-onset coronary artery disease. *J Am Heart Assoc*. 2023;12:e029103.
- Zhang T, Lin Y, He W, Yuan F, Zeng Y, Zhang S. GCN-GENE: a novel method for prediction of coronary heart disease-related genes. *Comput Biol Med*. 2022;150:105918.

38. Yang NI, Yeh CH, Tsai TH, Chou YJ, Hsu PW, Li CH, Chan YH, Kuo LT, Mao CT, Shyu YC, et al. Artificial intelligence-assisted identification of genetic factors predisposing high-risk individuals to asymptomatic heart failure. *Cells*. 2021;10:2430.
39. Zhao X, Sui Y, Ruan X, Wang X, He K, Dong W, Qu H, Fang X. A deep learning model for early risk prediction of heart failure with preserved ejection fraction by DNA methylation profiles combined with clinical features. *Clin Epigenetics*. 2022;14:11.
40. Oh HS, Rutledge J, Nachun D, Pálovics R, Abiose O, Moran-Losada P, Channappa D, Urey DY, Kim K, Sung YJ, et al. Organ aging signatures in the plasma proteome track health and disease. *Nature*. 2023;624:164–72.
41. Buerger T, Steinfeldt J, Ruyoga G, Pietzner M, Bizzarri D, Vojinovic D, Upmeyer Zu Belzen J, Looock L, Kittner P, Christmann L, et al. Metabolic profiles predict individual multidisease outcomes. *Nat Med*. 2022;28:2309–2320.
42. Puckelwartz MJ, Pesce LL, Hernandez EJ, Webster G, Dellefave-Castillo LM, Russell MW, Geisler SS, Kearns SD, Karthik F, Etheridge SP, et al. The impact of damaging epilepsy and cardiac genetic variant burden in sudden death in the young. *Genome Med*. 2024;16:13.
43. Barker J, Li X, Khavandi S, Koeckerling D, Mavilakandy A, Pepper C, Bountziouka V, Chen L, Kotb A, Antoun I, et al. Machine learning in sudden cardiac death risk prediction: a systematic review. *Europace*. 2022;24:1777–87.
44. Draelos RL, Ezekian JE, Zhuang F, Moya-Mendez ME, Zhang Z, Rosamilia MB, Manivannan PKR, Henao R, Landstrom AP. GENESIS: gene-specific machine learning models for variants of uncertain significance found in catecholaminergic polymorphic ventricular tachycardia and long QT syndrome-associated genes. *Circ Arrhythm Electrophysiol*. 2022;15:e010326.
45. Kwon OS, Hong M, Kim TH, Hwang I, Shim J, Choi EK, Lim HE, Yu HT, Uhm JS, Joung B, et al. Genome-wide association study-based prediction of atrial fibrillation using artificial intelligence. *Open Heart*. 2022.
46. Louca P, Tran TQB, Toit CD, Christofidou P, Spector TD, Mangino M, Suhre K, Padmanabhan S, Menni C. Machine learning integration of multimodal data identifies key features of blood pressure regulation. *EBioMedicine*. 2022;84:104243.
47. Lee D, Han SK, Yaacov O, Berk-Rauch H, Mathiyalagan P, Ganesh SK, Chakravarti A. Tissue-specific and tissue-agnostic effects of genome sequence variation modulating blood pressure. *Cell Rep*. 2023;42:113351.
48. Baysoy A, Bai Z, Satija R, Fan R. The technological landscape and applications of single-cell multi-omics. *Nat Rev Mol Cell Biol*. 2023;24(10):695–713.
49. Karjalainen MK, Karthikeyan S, Oliver-Williams C, Sliz E, Allara E, Fung WT, Surendran P, Zhang W, Jousilahti P, Kristiansson K, et al. Genome-wide characterization of circulating metabolic biomarkers. *Nature*. 2024;628:130–8.
50. Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrismisdottir EL, Gunnarsdottir K, Helgason A, Oddsson A, Halldorsson BV, et al. Large-scale integration of the plasma proteome with genetics and disease. *Nat Genet*. 2021;53:1712–21.
51. Islam MA, Majumder MZH, Miah MS, Jannaty S. Precision healthcare: a deep dive into machine learning algorithms and feature selection strategies for accurate heart disease prediction. *Comput Biol Med*. 2024;176:108432.
52. Kaur I, Ahmad T. A cluster-based ensemble approach for congenital heart disease prediction. *Comput Methods Programs Biomed*. 2024;243:107922.
53. Nurmohamed N, Belo Pereira J, Hoogeveen R, Kroon J, Kraaijenhof J, Waissi F, Timmerman N, Bom M, Hoefler I, Knaapen P, et al. Targeted proteomics improves cardiovascular risk prediction in secondary prevention. *Eur Heart J*. 2022;43:1569–77.
54. Zhang X, Wang C, He D, Cheng Y, Yu L, Qi D, Li B, Zheng F. Identification of DNA methylation-regulated genes as potential biomarkers for coronary heart disease via machine learning in the Framingham heart study. *Clin Epigenetics*. 2022;14:122.
55. Wang J, Kang Z, Liu Y, Li Z, Liu Y, Liu J. Identification of immune cell infiltration and diagnostic biomarkers in unstable atherosclerotic plaques by integrated bioinformatics analysis and machine learning. *Front Immunol*. 2022;13:956078.
56. Royer P, Björnson E, Adiels M, Álvarez MB, Fagerberg L, Bäckhed F, Uhlén M, Gummesson A, Bergström G. Plasma proteomics for prediction of subclinical coronary artery calcifications in primary prevention. *Am Heart J*. 2024;271:55–67.
57. McCarthy CP, Neumann JT, Michelhaugh SA, Ibrahim NE, Gaggin HK, Sörensen NA, Schaefer S, Zeller T, Magaret CA, Barnes G, et al. Derivation and external validation of a high-sensitivity cardiac troponin-based proteomic model to predict the presence of obstructive coronary artery disease. *J Am Heart Assoc*. 2020;9:e017221.
58. Marcinkiewicz-Siemion M, Kaminski M, Ciborowski M, Ptaszynska-Kopczynska K, Szpakowicz A, Lisowska A, Jasiewicz M, Tarasiuk E, Kretowski A, Sobkowicz B, Kaminski KA. Machine-learning facilitates selection of a novel diagnostic panel of metabolites for the detection of heart failure. *Sci Rep*. 2020;10:130.
59. Zhu Y, Yang X, Zu Y. Integrated analysis of WGCNA and machine learning identified diagnostic biomarkers in dilated cardiomyopathy with heart failure. In *Front Cell Dev Biol*, vol. 10, 2022/12/23 edition. pp. 1089915; 2022:1089915.
60. Liu Y, Bai F, Tang Z, Liu N, Liu Q. Integrative transcriptomic, proteomic, and machine learning approach to identifying feature genes of atrial fibrillation using atrial samples from patients with valvular heart disease. *BMC Cardiovasc Disord*. 2021;21:52.
61. Reel PS, Reel S, van Kralingen JC, Langton K, Lang K, Eric Z, Larsen CK, Amar L, Pamporaki C, Mulaturo P, et al. Machine learning for classification of hypertension subtypes using multi-omics: a multi-centre, retrospective, data-driven study. *EBioMedicine*. 2022;84:104276.
62. Alimadadi A, Manandhar I, Aryal S, Munroe PB, Joe B, Cheng X. Machine learning-based classification and diagnosis of clinical cardiomyopathies. *Physiol Genomics*. 2020;52:391–400.
63. Zhang F, Xia M, Jiang J, Wang S, Zhao Q, Yu C, Yu J, Xian D, Li X, Zhang L, et al. Machine learning and bioinformatics to identify 8 autophagy-related biomarkers and construct gene regulatory networks in dilated cardiomyopathy. *Sci Rep*. 2022;12:15030.
64. Xu J, Liu X, Dai Q. Integration of transcriptomic data identifies key hallmark genes in hypertrophic cardiomyopathy. *BMC Cardiovasc Disord*. 2021;21:330.
65. Park JK, Petrazzini BO, Saha A, Vaid A, Vy HMT, Márquez-Luna C, Chan L, Nadkarni GN, Do R. Machine learning identifies plasma metabolites associated with heart failure in underrepresented populations with the TTR V122I variant. *J Am Heart Assoc*. 2023;12:e027736.
66. O'Brien TD, Campbell NE, Potter AB, Letaw JH, Kulkarni A, Richards CS. Artificial intelligence (AI)-assisted exome reanalysis greatly aids in the identification of new positive cases and reduces analysis time in a clinical diagnostic laboratory. *Genet Med*. 2022;24:192–200.
67. Kamali Z, Keaton JM, Haghjooy Javanmard S, International Consortium Of Blood P, Million Veteran P, e QC, Bios C, Edwards TL, Snieder H, Vaez A. 2022. Large-scale multi-omics studies provide new insights into blood pressure regulation. *Int J Mol Sci*. 23:7557.
68. Drouard G, Ollikainen M, Mykkänen J, Raitakari O, Lehtimäki T, Kähönen M, Mishra PP, Wang X, Kaprio J. Multi-omics integration in a twin cohort and predictive modeling of blood pressure values. *OMICS*. 2022;26:130–41.
69. Yang J, Grafton F, Ranjbarvaziri S, Budan A, Farshidfar F, Cho M, Xu E, Ho J, Maddah M, Loewke KE, et al. Phenotypic screening with deep learning identifies HDAC6 inhibitors as cardioprotective in a BAG3 mouse model of dilated cardiomyopathy. *Sci Transl Med*. 2022;14:eab15654.
70. Theodoris CV, Xiao L, Chopra A, Chaffin MD, Al Sayed ZR, Hill MC, Mantione H, Brydon EM, Zeng Z, Liu XS, Ellinor PT. Transfer learning enables predictions in network biology. *Nature*. 2023;618:616–24.
71. Iborra-Egea O, Gálvez-Montón C, Prat-Vidal C, Roura S, Soler-Botija C, Revuelta-López E, Ferrer-Curriu G, Segú-Vergés C, Mellado-Bergillos A, Gomez-Puchades P, et al. Deep learning analyses to delineate the molecular remodeling process after myocardial infarction. *Cells*. 2021;10:3268.
72. Shi H, Zhang S. Accurate prediction of anti-hypertensive peptides based on convolutional neural network and gated recurrent unit. *Interdiscip Sci*. 2022;14:879–94.
73. Ma C, Zhou Z, Liu H, Koslicki D. KGML-xDTD: a knowledge graph-based machine learning framework for drug treatment prediction and mechanism description. *Gigascience*. 2022.

74. Shi H, Liu S, Chen J, Li X, Ma Q, Yu B. Predicting drug-target interactions using Lasso with random forest based on evolutionary information and chemical structure. *Genomics*. 2019;111:1839–52.
75. Ren ZH, You ZH, Zou Q, Yu CQ, Ma YF, Guan YJ, You HR, Wang XF, Pan J. DeepMPF: deep learning framework for predicting drug-target interactions based on multi-modal representation with meta-path semantic analysis. *J Transl Med*. 2023;21:48.
76. Zong N, Chowdhury S, Zhou S, Rajaganapathy S, Yu Y, Wang L, Dai Q, Bielinski SJ, Chen Y, Cerhan JR. Advancing Efficacy Prediction for EHR-based Emulated Trials in Repurposing Heart Failure Therapies. *medRxiv*, 2024, Nov 1:2023.05.25.23290531.
77. Wang X, Meng X, Meng L, Guo Y, Li Y, Yang C, Pei Z, Li J, Wang F. Joint efficacy of the three biomarkers SNCA, GYPB and HBG1 for atrial fibrillation and stroke: analysis via the support vector machine neural network. *J Cell Mol Med*. 2022;26:2010–22.
78. Wang L, Scherer S, Bielinski S, Muzny D, Jones L, Black J, Moyer A, Giri J, Sharp R, Matey E, et al. Implementation of preemptive DNA sequence-based pharmacogenomics testing across a large academic medical center: the Mayo-Baylor RIGHT 10K Study. *Genet Med*. 2022;24:1062–72.
79. Au Yeung VPW, Obrezanova O, Zhou J, Yang H, Bowen TJ, Ivanov D, Saffadi I, Carter AS, Subramanian V, Dillmann I, et al. Computational approaches identify a transcriptomic fingerprint of drug-induced structural cardiotoxicity. *Cell Biol Toxicol*. 2024;40:50.
80. Zhu Z, Wu R, Luo M, Zeng L, Zhang D, Hu N, Hu Y, Li Y. Two-dimensional deep learning frameworks for drug-induced cardiotoxicity detection. *ACS Sens*. 2024;9:3316–26.
81. Yu L, Xu Z, Qiu W, Xiao X. MSDSE: Predicting drug-side effects based on multi-scale features and deep multi-structure neural network. *Comput Biol Med*. 2024;169: 107812.
82. Uner OC, Kuru HI, Cinbis RG, Tastan O, Cicek AE. DeepSide: a deep learning approach for drug side effect prediction. *IEEE/ACM Trans Comput Biol Bioinform*. 2023;20:330–9.
83. Wei J, Zhuo L, Zhou Z, Lian X, Fu X, Yao X. GCFMCL: predicting miRNA-drug sensitivity using graph collaborative filtering and multi-view contrastive learning. *Brief Bioinf*. 2023.
84. Mason DM, Friedensohn S, Weber CR, Jordi C, Wagner B, Meng SM, Ehling RA, Bonati L, Dahinden J, Gainza P, et al. Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning. *Nat Biomed Eng*. 2021;5:600–12.
85. Wallentin L, Eriksson N, Olszowka M, Grammer TB, Hagström E, Held C, Kleber ME, Koenig W, März W, Stewart RAH, et al. Plasma proteins associated with cardiovascular death in patients with chronic coronary heart disease: a retrospective study. *PLoS Med*. 2021;18: e1003513.
86. Liu X, Xu H, Xu H, Geng Q, Mak WH, Ling F, Su Z, Yang F, Zhang T, Chen J, et al. New genetic variants associated with major adverse cardiovascular events in patients with acute coronary syndromes and treated with clopidogrel and aspirin. *Pharmacogenomics J*. 2021;21:664–72.
87. Eggers KM, Lindhagen L, Lindhagen L, Baron T, Erlinge D, Hjort M, Jernberg T, Marko-Varga G, Rezeli M, Spaak J, Lindahl B. Predicting outcome in acute myocardial infarction: an analysis investigating 175 circulating biomarkers. *Eur Heart J Acute Cardiovasc Care*. 2021;10:806–12.
88. Hjort M, Eggers KM, Lindhagen L, Baron T, Erlinge D, Jernberg T, Marko-Varga G, Rezeli M, Spaak J, Lindahl B. Differences in biomarker concentrations and predictions of long-term outcome in patients with ST-elevation and non-ST-elevation myocardial infarction. *Clin Biochem*. 2021;98:17–23.
89. Williams SA, Ostroff R, Hinterberg MA, Coresh J, Ballantyne CM, Matsushita K, Mueller CE, Walter J, Jonasson C, Holman RR, et al. A proteomic surrogate for cardiovascular outcomes that is sensitive to multiple mechanisms of change in risk. *Sci Transl Med*. 2022;14:eabj9625.
90. Chen L, Fu G, Jiang C. Deep learning-derived 12-lead electrocardiogram-based genotype prediction for hypertrophic cardiomyopathy: a pilot study. *Ann Med*. 2023;55:2235564.
91. Deng Y, Liu L, Jiang H, Peng Y, Wei Y, Zhou Z, Zhong Y, Zhao Y, Yang X, Yu J, et al. Comparison of state-of-the-art neural network survival models with the pooled cohort equations for cardiovascular disease risk prediction. *BMC Med Res Methodol*. 2023;23:22.
92. Ouwerkerk W, Belo Pereira JP, Maasland T, Emmens JE, Figarska SM, Tromp J, Koekemoer AL, Nelson CP, Nath M, Romaine SPR, et al. Multi-omics analysis provides novel pathways related to progression of heart failure. *J Am Coll Cardiol*. 2023;82:1921–31.
93. Unterhuber M, Kresoja KP, Rommel KP, Besler C, Baragetti A, Klötting N, Ceglarek U, Blüher M, Scholz M, Catapano AL, et al. Proteomics-enabled deep learning machine algorithms can enhance prediction of mortality. *J Am Coll Cardiol*. 2021;78:1621–31.
94. Xu D, Cunningham J, Marti-Castellote PM, Zhang L, Patel-Murray NL, Prescott MF, Chutkow W, Mendelson MM, Solomon SD, Claggett BL. Machine learning for proteomic risk scores in heart failure. *J Card Fail*. 2023;29:1583–5.
95. Shimada YJ, Raita Y, Liang LW, Maurer MS, Hasegawa K, Fifer MA, Reilly MP. Prediction of major adverse cardiovascular events in patients with hypertrophic cardiomyopathy using proteomics profiling. *Circ Genom Precis Med*. 2022;15: e003546.
96. Halloran PF, Reeve J, Mackova M, Madill-Thomsen KS, Demko Z, Olymbios M, Campbell P, Melenovsky V, Gong T, Hall S, Stehlik J. Comparing plasma donor-derived cell-free DNA to gene expression in endomyocardial biopsies in the trifecta-heart study. *Transplantation*. 2024.
97. Shi H, Yuan M, Cai J, Shi J, Li Y, Qian Q, Dong Z, Pan G, Zhu S, Wang W, et al. Exploring personalized treatment for cardiac graft rejection based on a four-archetype analysis model and bioinformatics analysis. *Sci Rep*. 2024;14:6529.
98. Cui S, Li L, Zhang Y, Lu J, Wang X, Song X, Liu J, Li K. Machine learning identifies metabolic signatures that predict the risk of recurrent angina in remitted patients after percutaneous coronary intervention: a multi-center prospective cohort study. *Adv Sci (Weinh)*. 2021;8:2003893.
99. Ibrahim NE, McCarthy CP, Shrestha S, Gaggin HK, Mukai R, Magaret CA, Rhyne RF, Januzzi JL Jr. A clinical, proteomics, and artificial intelligence-driven model to predict acute kidney injury in patients undergoing coronary angiography. *Clin Cardiol*. 2019;42:292–8.
100. Li Z, Qin Y, Liu X, Chen J, Tang A, Yan S, Zhang G. Identification of predictors for neurological outcome after cardiac arrest in peripheral blood mononuclear cells through integrated bioinformatics analysis and machine learning. *Funct Integr Genomics*. 2023;23:83.
101. Nikanjam M, Kato S, Kurzrock R. Liquid biopsy: current technology and clinical applications. *J Hematol Oncol*. 2022;15:131.
102. Lewandowski P, Gołowski M, Baron M, Reichman-Warmusz E, Wojnicz R. A systematic review of miRNA and cfDNA as potential biomarkers for liquid biopsy in myocarditis and inflammatory dilated cardiomyopathy. *Biomolecules*. 2022;12:2.
103. Pimpalwar N, Czuba T, Smith ML, Nilsson J, Gidlöf O, Smith JG. Methods for isolation and transcriptional profiling of individual cells from the human heart. *Heliyon*. 2020;6: e05810.
104. Komatsu J, Cico A, Poncin R, Le Bohec M, Morf J, Lipin S, Graindorge A, Eckert H, Saffarian A, Cathaly L, et al. RevGel-seq: instrument-free single-cell RNA sequencing using a reversible hydrogel for cell-specific barcoding. *Sci Rep*. 2023;13:4866.
105. Zhou B, Shi X, Tang X, Zhao Q, Wang L, Yao F, Hou Y, Wang X, Feng W, Wang L, et al. Functional isolation, culture and cryopreservation of adult human primary cardiomyocytes. *Signal Transduct Target Ther*. 2022;7:254.
106. Chen W, Guillaume-Gentil O, Rainer PY, Gäbelein CG, Saelens W, Gardeux V, Klaefer A, Dainese R, Zachara M, Zambelli T, et al. Live-seq enables temporal transcriptomic recording of single cells. *Nature*. 2022;608:733–40.
107. Andersson C, Naylor M, Tsao CW, Levy D, Vasan RS. Framingham heart study: JACC focus seminar, 1/8. *J Am Coll Cardiol*. 2021;77:2680–92.
108. Baltrusaitis T, Ahuja C, Morency LP. Multimodal machine learning: a survey and taxonomy. *IEEE Trans Pattern Anal Mach Intell*. 2019;41:423–43.
109. Soenksen LR, Ma Y, Zeng C, Boussioux L, Villalobos Carballo K, Na L, Wiberg HM, Li ML, Fuentes I, Bertsimas D. Integrated multimodal artificial intelligence framework for healthcare applications. *NPJ Digit Med*. 2022;5:149.
110. Han Y, Zeng X, Hua L, Quan X, Chen Y, Zhou M, Chuang Y, Li Y, Wang S, Shen X, et al. The fusion of multi-omics profile and multimodal EEG data contributes to the personalized diagnostic strategy for neurocognitive disorders. *Microbiome*. 2024;12:12.
111. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366:447–53.
112. Tian Y, Wang S, Xiong J, Bi R, Zhou Z, Bhuiyan MZA. Robust and privacy-preserving decentralized deep federated learning training: focusing

- on digital healthcare applications. *IEEE/ACM Trans Comput Biol Bioinf.* 2023.
113. Minssen T, Vayena E, Cohen IG. The challenges for regulating medical use of ChatGPT and other large language models. *JAMA.* 2023;330:315–6.
 114. Wang H, Fu T, Du Y, Gao W, Huang K, Liu Z, Chandak P, Liu S, Van Katwyk P, Deac A, et al. Scientific discovery in the age of artificial intelligence. *Nature.* 2023;620:47–60.
 115. Hamm CA, Baumgärtner GL, Biessmann F, Beetz NL, Hartenstein A, Savic LJ, Froböse K, Dräger F, Schallenberg S, Rudolph M, et al. Interactive explainable deep learning model informs prostate cancer diagnosis at MRI. *Radiology.* 2023;307: e222276.
 116. Kauffmann J, Esders M, Ruff L, Montavon G, Samek W, Müller KR. From clustering to cluster explanations via neural networks. *IEEE Trans Neural Netw Learn Syst.* 2024;35:1926–40.
 117. Wouters PC, van de Leur RR, Vessies MB, van Stipdonk AMW, Ghossein MA, Hassink RJ, Doevendans PA, van der Harst P, Maass AH, Prinzen FW, et al. Electrocardiogram-based deep learning improves outcome prediction following cardiac resynchronization therapy. *Eur Heart J.* 2023;44:680–92.
 118. Toussaint PA, Leiser F, Thiebes S, Schlesner M, Brors B, Sunyaev A. Explainable artificial intelligence for omics data: a systematic mapping study. *Brief Bioinf.* 2023.
 119. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In 2016 IEEE conference on computer vision and pattern recognition (CVPR); 27–30 June 2016. 2016: 770–778.
 120. Yu S, Tomasi C. Identity connections in residual nets improve noise stability. 2019.
 121. Vinga S. Structured sparsity regularization for analyzing high-dimensional omics data. *Brief Bioinform.* 2021;22:77–87.
 122. Mohr F, van Rijn JN. Fast and informative model selection using learning curve cross-validation. *IEEE Trans Pattern Anal Mach Intell.* 2023;45:9669–80.
 123. Bica I, Alaa AM, Lambert C, van der Schaar M. From real-world patient data to individualized treatment effects using machine learning: current and future methods to address underlying challenges. *Clin Pharmacol Ther.* 2021;109:87–100.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.