Dean&Francis

ISSN 2959-409X

Multi-Model AI Approaches for heart failure risk prediction based on fusion of Echocardiographic Images and clinical data

Ziyi Wang [1,*]

[1]The University of California, Irvine, the United States
*Corresponding author: zwang39@uci.edu

Abstract:

Heart failure (HF) is a chronic illness that has a high global incidence and mortality rate. Its direction is complex and incredibly varied. Traditional prediction methods are usually based on a single information source, such as organized medical records or echocardiographic images, and are thus not able to completely reflect the medical condition of patients. The use of artificial intelligence (AI) in cardiovascular medicine has been steadily growing in recent years. Specifically, the multi-model connection and multi-modal data fusion methods held great potential in integrating diverse information sources and enhancing prediction accuracy. It is a review of modeling, fusion, and implementation approaches to multi-model diagnosis from clinical and image data. It summarizes recent advances in the prediction of heart failure risk from single- and multi-modal AI, presents challenges such as interpretability, clinical acceptability, data quality, and generalizability, and envisions future directions such as time series modeling, cross-modal deep learning, and multi-center data sharing. The significance of this review lies in serving future research and further promoting the clinical applications and development of AI technology in predicting heart failure.

Keywords: Heart Failure, Artificial Intelligence, Multi-Modal Data Fusion, Risk Prediction

1. Introduction

A clinical syndrome, heart failure (HF), is caused by structural or functional defects of the heart. It is characterized as a decrease in the heart's pumping ability that is not sufficient to meet the body's metabolic needs. Over 64 million patients worldwide suffer from heart failure, a serious global public health condition with a dramatic impact on mortality, quality of life in patients, and healthcare resource utilization. Patients with heart failure continue to have an unfavorable prognosis for the future and hospitalization

rate despite ongoing advances in drug and interventional treatment technology. Early detection of those at risk is therefore relevant to early intervention and better prognosis.
However, numerous impediments are present in heart failure risk estimation. Traditional prediction models that use a single source of information are generally unable to provide a complete description of a patient's risk status because of the myriad etiologies of disease, complex clinical presentation, and prevalent heterogeneity of the patient's disease course. For example, Framingham Risk Score and other laboratory test reports, and demographic-based traditional risk scoring systems are not very helpful to differentiate between different populations. While the most widely used imaging modality to evaluate cardiac structure and function in the clinical environment, echocardiography frequently does not detect possible subclinical lesions if employed in isolation. Apart from that, traditional approaches hardly use time-series data and cannot dynamically trace changes in the disease course.
Artificial intelligence (AI) has achieved significant improvements in cardiology over the past several years. AI can find deep features that are hard for human experts to find, and learn complex nonlinear relationships from large, high-dimensional clinical data and medical images. Machine learning (ML) and deep learning (DL) technologies have achieved excellent results in most cardiovascular disease tasks, including cardiac arrhythmia diagnosis, myocardial infarction diagnosis, and heart failure prediction risk, and so on.
As the field progresses, multi-model and multi-modal AI approaches have increasingly gained interest in this context. For offering a more comprehensive account of the heart health of a patient, the multi-modal approach combines data from diverse sources, including structured clinical information and echocardiograms. Through combining multiple different kinds of approaches, the multi-model approach boosts defense against prediction results and generalizability. Combining the two methods should improve prediction accuracy and clinical application over single-modal and single-model approaches.
However, there is significant heterogeneity in the research that has been carried out on the use of AI for heart failure prediction. While some studies only utilize deep learning models on images, others are based on structured clinical data. Furthermore, it is challenging to directly compare the findings of various studies due to the heterogeneity in the size of the data set, patient data, image acquisition criteria, and evaluation indications. Thus, the purpose of this review is to assess the merits and shortcomings of single-modal and multi-modal AI approaches in the field of heart failure prediction, and to discuss potential directions for future study.

2. Dataset Collection

The efficacy of multimodal AI models to predict the risk of heart failure (HF) relies substantially on the availability of large and well-organized data. Previous research most commonly employs electronic health records (EHR) and echocardiography databases, which provide structured clinical covariates including demographics, laboratory test findings, medical history, and medication use, along with imaging data [1,2,3]. Large public datasets of tens of thousands of patients have enabled dramatic progress in the domain [1,4]. Problems remain, though. Still, relatively small or single-site cohorts form the foundation for most studies, limiting diversity and generalizability of findings [5,6]. Particularly, high-quality and large-scale, and consistently annotated echocardiographic imaging data are scarce. Furthermore, the insufficiency of longitudinal and time-series data prevents models from learning about the dynamic progression of HF [7]. Overcoming such challenges—through the integration of disparate data sources from multiple institutions and by injecting temporal patterns—will be critical to increasing both the robustness and the clinical usefulness of multimodal AI approaches.

3. Single-Modality AI Models

Machine learning (ML) and deep learning (DL) methods have found widespread application for HF risk prediction in recent years. Trained machine learning models based on formalized clinical data, such as random forest (RF), gradient boosting machine (GBM), and support vector machine (SVM), have been successful in various studies [3,8]. For example, SVM achieved an AUC of 0.87 in 1-year HF readmission prediction using clinical and demographic data [9], and another study reported that SVM achieved 85% accuracy and a sensitivity of 82% in HF hospitalization prediction using electronic health record features [10]. As opposed to this, deep learning models like recurrent neural networks (RNN) and convolutional neural networks (CNN) have been shown to be resilient in image and video analysis of ultrasound [2,11,12].

4. Multimodal Fusion and Multi-Model Integration

Multimodal fusion models typically employ multi-path network architectures to represent clinical data and ultrasound images in distinct pathways and afterward merge them for improved performance [1,4]. Current research attempts to incorporate Transformer and Graph Neural

ISSN 2959-409X

Network (GNN) into multimodal fusion to capture the complex dependencies and temporal dynamic patterns between modalities, aiming to promote the deep understanding of pathological processes by the model [5,7]. In addition, multi-model integration technology combines the advantages of different algorithms to further improve prediction stability and accuracy, but multi-model collaborative diagnosis is still in its infancy at present, and related studies are scarce [8].

Most current models are not entirely capable of leveraging long-term follow-up data and dynamic properties of temporal sequence, resulting in a deficiency of predictive capacity for disease progression [7]. Time series modeling needs to be given greater emphasis in the future, and data from multiple check-ups on imaging and clinic should be utilized to enhance the time sensitivity of the prediction.

5. Evaluation

Comparison of multimodal AI model performance is a critical step in ensuring their clinical usability. AUC (Measures the overall ability of a model to distinguish between positive and negative classes.), accuracy, sensitivity, specificity, and F1(A balance between precision and recall) score, etc., are the most common metrics [1,3]. In most studies, multimodal models that integrate clinical data and ultrasound images significantly outperform single-modal models. For instance, in a study, by integrating EHR information with cardiac ultrasound images, AUC for prediction of 1-year heart failure readmission increased to 0.91, significantly greater than models based solely on clinical information (AUC = 0.84) or imaging information alone (AUC = 0.87) [1].

Cross-validation and external validation are also necessary methods for examining the accuracy of a model's generalization ability. External validation was conducted in several studies using multi-center data, and the results showed that the multimodal model trained between institutions had excellent predictive accuracy in different hospital environments [2,4]. Moreover, the time series data applied in model training, such as follow-up ultrasound images and dynamic laboratory markers, can further enhance the model's sensitivity to the disease progression [7].

However, most of the recent studies remain based on single-center data or small samples, which may result in limited generalization ability of the model in other populations or hospitals. This suggests that additional exploration should consider more importance to the integration of cross-center data, as well as the employment of large-scale and long-term follow-up information to improve the application of the model in the clinic.

6. Explainability and Clinical Relevance

AI model interpretability is a compelling factor during clinical translation. Although multimodal models work well, their "black box" nature can suppress physicians' confidence and utilization. Consequently, more recently, researchers have attempted to leverage visualization techniques (e.g., Grad-CAM) and feature importance analysis to shed light on model choices [2,11]. For example, in the case of heat maps, areas of the ultrasound images with maximum contribution to the model's prediction can be visually depicted. At the same time, by applying a combination of SHAP or LIME methods for assessing the effect of clinical variables, it helps in explaining the reasoning behind the model's decision-making in different patient groups [1,3].

From the viewpoint of clinical significance, not only does the multimodal model provide a single prediction result, but it is also able to conduct dynamic risk stratification through the inclusion of time series data, providing a roadmap for individualized intervention. For example, the model may provide an early high-risk warning when the patient's ejection fraction declines or BNP level rises, thereby notifying physicians to alter the treatment strategy [7]. Besides, by incorporating the technology of multi-model collaborative diagnosis, the model can enhance its robustness for complex cases and reduce the risks of misjudgment and missed diagnosis [8].

However, current interpretability methods also suffer from certain limitations, i.e., being limited to local explanations or being tethered to the features of the training data. Future research can attempt cross-modal attention mechanisms, graph neural network interpretation methods, and long-term follow-up dynamic interpretable models for enhancing clinical acceptability and practical application value.

7. Discussion

This review provides an overview of the literature progress of single-modal and multi-modal artificial intelligence (AI) in forecasting the risk of heart failure (HF). Single-modal methods perform well on specific tasks but are limited by the inherent limitations of the types of data: clinical structured data are difficult to represent directly the cardiac morphology and function, whereas image data could overlook the global clinical context of patients. These limitations have resulted in the increasing popularity of multi-modal fusion and multi-model collaborative diagnosis as scientific hotspots.

The multimodal fusion model, with both clinical data and ultrasound images simultaneously, can more accurately explain the pathological state of the patients. As an example, the EHR-fused imaging model does a better job than the single-modal model in prediction tasks such as 1-year readmission and mortality from heart failure [1,4]. With the incorporation of Transformer or Graph Neural Network (GNN) technologies, it can well represent the nonlinear correlation and temporal sequential relationship between different modalities [5,7]. Multiple model collaborative diagnosis integrates the strengths of different algorithms to improve prediction stability and robustness, but there are comparatively few related studies and no established evaluation system yet [8].

Nevertheless, there are some limitations in the present study. First, the availability of large-scale and long-term follow-up multi-center datasets is limited, which constrains the ability of generalization and clinical application of the models. Second, interpretability remains a significant bottleneck for the deployment of multimodal models, and the present techniques are only up to local features or visual explanations, not completely satisfying clinical decision-making requirements. Besides, the combination of time series data and dynamic predictive ability has not been sufficiently utilized, limiting the ability of the models in disease progression management. Future research could address these issues by focusing on the following aspects: cross-modal Transformer or GNN model construction, long-term tracking and prediction, collaborative diagnosis of multiple models, better interpretability, and sharing and integration of multi-center large samples.

8. Conclusion

The artificial intelligence research in the field of heart failure risk prediction is progressing rapidly. The AI method based on multimodal data fusion and multi-model collaborative diagnosis, by integrating ultrasound images and structured clinical data, can assess patient risks more accurately and comprehensively, providing strong assistance for clinical decision-making. Despite existing limitations of small data scale, lack of interpretability, and no multi-center validation, with the availability of large-scale high-quality data, the development of cross-modal deep learning techniques, and the innovation of interpretable approaches, multimodal AI models will see broader clinical application. These technologies in the future will not only increase the accuracy of heart failure risk prediction but also provide new possibilities for the development of personalized treatment strategies and long-term disease management, with the aim of improving patient prognosis and reducing medical burden.

References

[1] Samad MD, Ulloa A, Wehner GJ, Linyuan J, Hartzel D, Good CW, Williams BA, Haggerty CM, Fornwalt BK. 2018. Predicting survival from large echocardiography and electronic health record datasets: Optimization with machine learning. JACC: Cardiovascular Imaging 12(4):681–689.

[2] Zhou M, Deng Y, Liu Y, Su X, Zeng X. 2023. Development and validation of echocardiography-based machine-learning models to predict mortality. EBioMedicine 90:104479.

[3] Liu W, Tang Y, Li Q, et al. 2021. Heart failure prediction with machine learning: A comparative study. Journal of Physics: Conference Series 2031(1):012068.

[4] Duffy G, Cheng PP, Yuan N, He B, Kwan AC, Shun-Shin MJ, Alexander KM, Ebinger J, Lungren MP, Rader F, Liang DH, Schnittger I, Ashley EA, Zou JY, Patel J, Witteles R, Cheng S, Ouyang D. 2021. High-throughput precision phenotyping of left ventricular hypertrophy with cardiovascular deep learning. arXiv preprint arXiv:2106.12511.

[5] Bell-Navas A, Villalba-Orero M, Lara-Pezzi E, Garicano-Mena J, Le Clainche S. 2025. Heart failure prediction using modal decomposition and masked autoencoders for scarce echocardiography databases. arXiv preprint arXiv:2504.07606.

[6] Li X, Zhang Y, Chen J, et al. 2023. Echocardiography-based machine learning algorithm for distinguishing ischemic cardiomyopathy from dilated cardiomyopathy. BMC Cardiovascular Disorders 23:476.

[7] Sharma A, Harrington RA, McClellan MB, Turakhia MP, Eapen ZJ, Steinhubl S, Peterson ED. 2020. Using machine learning to predict outcomes in heart failure: A systematic review. JACC: Heart Failure 8(1):78–87.

[8] Choi E, Schuetz A, Stewart WF, Sun J. 2017. Using recurrent neural network models for early detection of heart failure onset. Journal of the American Medical Informatics Association 24(2):361–370.

[9] Nascimento J, Martins R, Silva L, et al. 2023. Deep learning for echocardiography: Introduction for clinicians and future vision: State-of-the-art review. Life 13(4):1029.

[10] Liu Y, Wang X, Zhao H, et al. 2023. Deep learning-based identification of echocardiographic abnormalities from electrocardiograms. PMC.

[11] Patel R, Singh A, Kumar S, et al. 2023. AI-driven technology in heart failure detection and diagnosis: A review of the advancements in personalized healthcare. Symmetry 17(3):469.

[12] Zhang H, Chen Y, Li Z, et al. 2020. Heart failure diagnosis, readmission, and mortality prediction using machine learning and artificial intelligence models. Current Epidemiology Reports 7:112–123.