South China University of Technology

《统计学习与数据科学导论》课 程 设 计

(2023 -2024 学年 暑假学期)

论文题目 **Multimodal Fusion-Based Cardiac Disease Diagnosis System: Using the ECG and ECHO Data from MIMIC Databases**

学生姓名：叶朝阳

提交日期： 2024 年 8 月 29 日 学生签名：叶朝阳

学　　号	202130322339	座位编号	
学　　院	数学学院	专业班级	21 级信息与计算科学
课程名称	统计学习与数据科学导论课程设计	任课教师	夏立
教师评语：			
本论文成绩评定：______分			

Multimodal Fusion-Based Cardiac Disease Diagnosis System: Using the ECG and ECHO Data from MIMIC Databases

Zhaoyang YE
School of Mathematics, South China University of Technology

Abstract

This study develops a multimodal disease diagnosis system using ECG and ECHO data from the MIMIC-III and MIMIC-IV databases, which include detailed clinical records from ICU patients. Hypertrophic cardiomyopathy (HCM), the most common cardiac genetic disease with a prevalence of 1:500 to 1:200, poses significant diagnostic challenges due to its symptom overlap with hypertensive left ventricular hypertrophy (LVH), leading to misclassification rates as high as 30%. To address this, we employed convolutional neural networks (CNN) for ECG analysis and recurrent neural networks (RNN) with attention mechanisms for ECHO data, integrating them through advanced multimodal fusion techniques. Our system, validated across multiple cohorts, demonstrates that this multimodal approach significantly enhances diagnostic accuracy, particularly in distinguishing HCM from other causes of LVH. This highlights the potential of AI-driven multimodal data fusion in improving clinical diagnostics.

Keywords: ECG, ECHO, Deep Learning, Multimodal-Fusion, Heart Disease Diagnosis

Introduction

Hypertrophic cardiomyopathy (HCM) is a prevalent cardiac genetic disorder, affecting 1 in 500 to 1 in 200 individuals, and poses significant risks, including heart failure and sudden death. Accurate diagnosis is crucial, especially for family screening, yet it remains challenging due to its clinical overlap with hypertensive left ventricular hypertrophy (LVH), a condition present in up to 45% of U.S. adults[1]. Misclassification rates can be as high as 30%, often mistaking HCM for hypertension. This diagnostic complexity highlights the need for advanced tools like artificial intelligence (AI) to assist clinicians in distinguishing these conditions more effectively[2].

Recent advancements in AI have led to the development of deep learning models that focus on single modalities, such as ECG or echocardiograms. For example, Ko et al. utilized convolutional neural networks (CNN) for ECG-based HCM detection, demonstrating strong discriminatory power[3]. However, their model faced limitations when distinguishing HCM from hypertensive heart disease, particularly in cases where both conditions coexisted. Similarly, Zhang's approach, which employed 2D CNNs to analyze echocardiograms, was constrained by its inability to capture temporal dynamics, which are critical for accurate cardiac diagnosis[4].

The limitations of single-modality approaches become evident when tackling complex diseases like HCM, where both electrical and structural abnormalities are involved. ECG alone may miss structural insights, while echocardiograms might overlook subtle electrical patterns[5], [6]. These challenges underscore the need for a more integrated approach. Multimodal data analysis, which combines different types of clinical data, offers a more comprehensive view, potentially leading to more accurate and reliable diagnoses[7].

To address these limitations, our study introduces the fusion model, a multimodal diagnostic system that integrates Electrocardiogram (ECG) and echocardiogram (ECHO) data. Unlike previous single-modality methods, the fusion model leverages the strengths of both ECG and ECHO, providing a more holistic analysis of cardiac health. By incorporating temporal convolutions, our model can capture and analyze dynamic motion patterns in echocardiogram videos while simultaneously processing ECG signals[8], enabling it to detect complex patterns that may be missed when each modality is considered in isolation.

Furthermore, while earlier research demonstrated the benefits of combining ECG and ECHO data in a stepwise manner, our approach goes further by deeply integrating these modalities within a single framework. This integration not only improves diagnostic accuracy but also enhances the model's generalizability across different patient populations and clinical settings[9]. The superiority of multimodal approaches over traditional single-modality methods is increasingly recognized in medical research, particularly for complex cardiac conditions where a single data type is insufficient.

Despite the potential of multimodal deep learning, there remains a scarcity of studies that deeply integrate ECG and ECHO data using large-scale datasets[10]. This research aims to fill this gap by presenting a comprehensive multimodal diagnostic system validated across multiple cohorts. Our findings underscore the importance of multimodal AI in advancing clinical practice, offering a more robust and accurate tool for diagnosing cardiac diseases like HCM. Figure 1 below shows the overview of the ECG + ECHO Multimodal (EEMM) model.

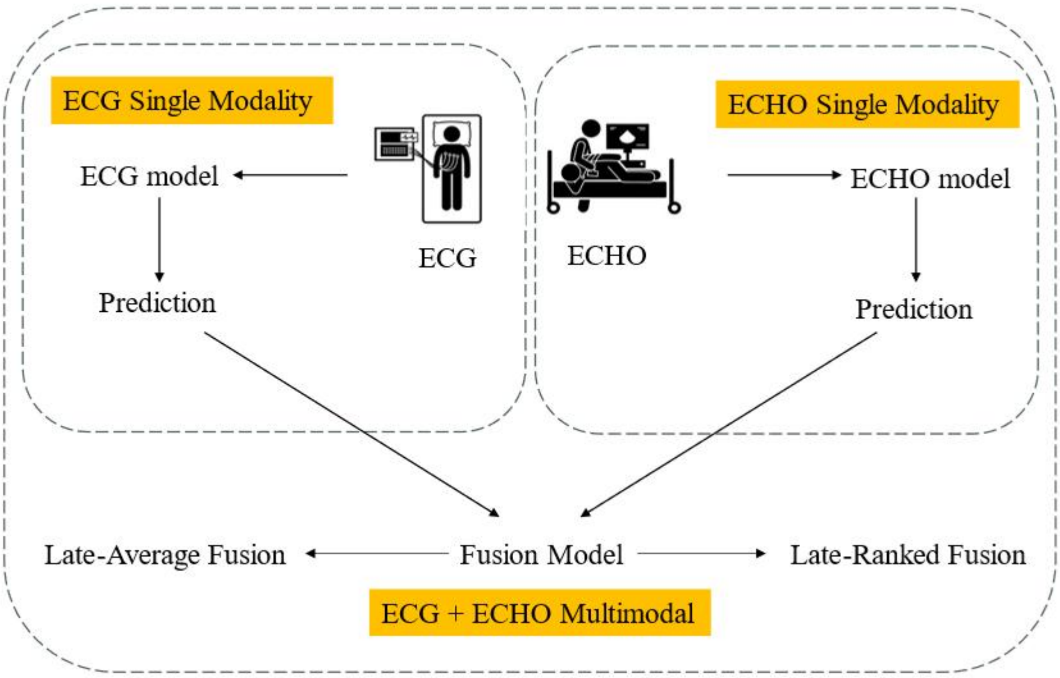


Figure 1 An Overview of EEMM

Method

Data Collection and Preprocessing

The study leverages the MIMIC-III and MIMIC-IV databases, which are rich in clinical and physiological data from ICU patients at Beth Israel Deaconess Medical Center. The focus is on extracting high-quality electrocardiogram (ECG) and echocardiogram (ECHO) data. Specific inclusion criteria are applied to ensure that only data meeting certain clinical relevance and quality thresholds are utilized. This selection process includes filtering out incomplete records, ensuring temporal alignment between ECG and ECHO data, and normalizing the data to account for differences in patient demographics, such as age and gender[11].

For the ECG data, signal denoising is performed using methods like wavelet transform or median filtering to remove noise and artifacts, ensuring a clean signal for analysis. The signals are then segmented into time windows of interest, and features such as QRS complex, heart rate variability (HRV), and ST-segment deviation are extracted. These features are essential for identifying various cardiac conditions.

For the ECHO data, frames are extracted from the video recordings at specific cardiac cycle phases, such as end-diastole and end-systole. Image preprocessing includes resizing, normalization, and data augmentation techniques (e.g., rotation, flipping) to increase the robustness of the model. Feature extraction involves calculating parameters such as ejection fraction, wall motion abnormalities, and chamber dimensions, which are critical indicators of cardiac function.

Model Architecture

ECG Analysis with Convolutional Neural Networks (CNN):

The ECG data is fed into a deep convolutional neural network. The CNN is designed with multiple convolutional layers followed by pooling layers to capture the spatial and temporal features of the ECG signals[12]. The mathematical formulation for the convolutional operation is as follows:

$$y_{i,j}^{(l)} = f\left(\sum_{m=1}^{M}\sum_{n=1}^{N} w_{m,n}^{(l)} \cdot x_{(i+m-1),(j+n-1)}^{(l-1)} + b^{(l)}\right) \tag{1.1}$$

Here, $y_{i,j}^{(l)}$ is the output of the l-th layer, x is the input feature map from the previous layer, w are the convolutional weights, b is the bias term, and f is the activation function (e.g., ReLU). The CNN captures complex patterns in the ECG data, such as arrhythmias and ischemic changes.

ECHO Analysis with Recurrent Neural Networks (RNN) and Attention Mechanisms:

The ECHO data, represented as sequences of frames, is analyzed using a recurrent neural network equipped with attention mechanisms. The RNN captures the temporal dependencies within the sequences, while the attention mechanism allows the model to focus on specific frames or phases that are most relevant to the diagnosis[4], [13]. The RNN's operation is governed by:

$$h_t = \sigma(W_h h_{t-1} + W_x x_t + b_h) \tag{1.2}$$

where h_t is the hidden state at time step t, W_h and W_x are weight matrices, x_t is the input at time t, and b_h is the bias term. The attention mechanism computes a context vector c_t as:

$$c_t = \sum_{i=1}^{T} \alpha_{t,i} h_i \tag{1.3}$$

where $\alpha_{t,i}$ are attention weights, learned during training, determining the importance of each hidden state h_i.

Multimodal Fusion

The integration of ECG and ECHO data is achieved through multimodal fusion, which combines the outputs from the CNN and RNN models into a unified framework. This fusion occurs at the decision level, where the attention mechanisms from both models generate weighted context vectors, which are then concatenated and passed through fully connected layers to produce the final diagnostic output[14]. The fusion process can be mathematically described as:

$$\text{Fusion\_output} = \mathrm{f}_{\text{fusion}}(\mathrm{c}_{\text{ECG}}, \mathrm{c}_{\text{ECHO}}) \tag{1.4}$$

where c_{ECG} and c_{ECHO} are the context vectors from the ECG and ECHO models, respectively, and $\mathrm{f}_{\text{fusion}}$ represents the fusion function, typically a series of fully connected layers followed by a softmax activation function to yield the final class probabilities.

Experimental Design and Validation

Development Cohort: This cohort is used for model training. The data is split into training and validation sets, and the model is trained using techniques such as cross-entropy loss minimization and Adam optimization. The model's performance is monitored using metrics like accuracy, precision, recall, and F1-score.

Internal Validation Cohort: This cohort consists of data from the same database but is not used during training. The internal validation assesses how well the model performs on unseen data from the same distribution, helping to tune hyperparameters and prevent overfitting.

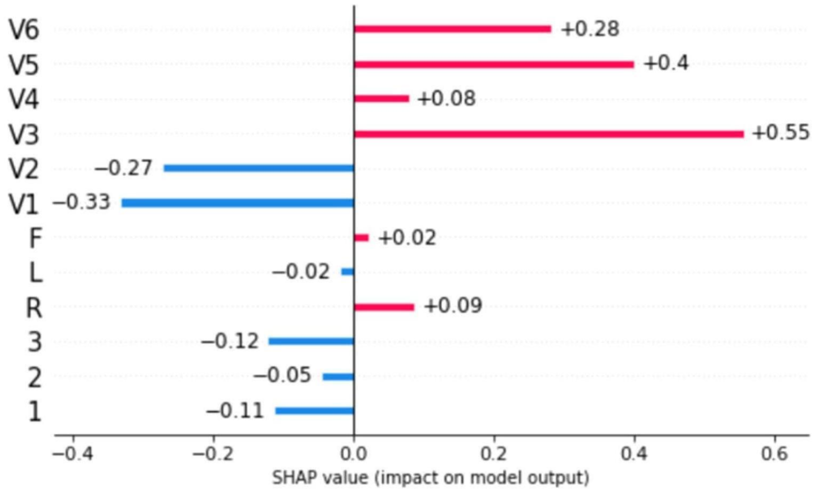
Results

In this study, both single-modal and multimodal neural network models were examined. Two different multimodal fusion architectures were explored, combining ECG and echocardiogram information in different ways: late-average fusion (LAF) and late-ranked fusion (LRF) models. In the late-average fusion model, soft voting is performed by computing the average probability for each class from the individual ECG and echocardiogram classifiers and predicting the class with maximal average probability. In the late-ranked fusion model, the probabilities for each class from the individual ECG and echocardiogram classifiers are ranked and a prediction is determined from the highest-ranked probability. The final result is shown in Table 1

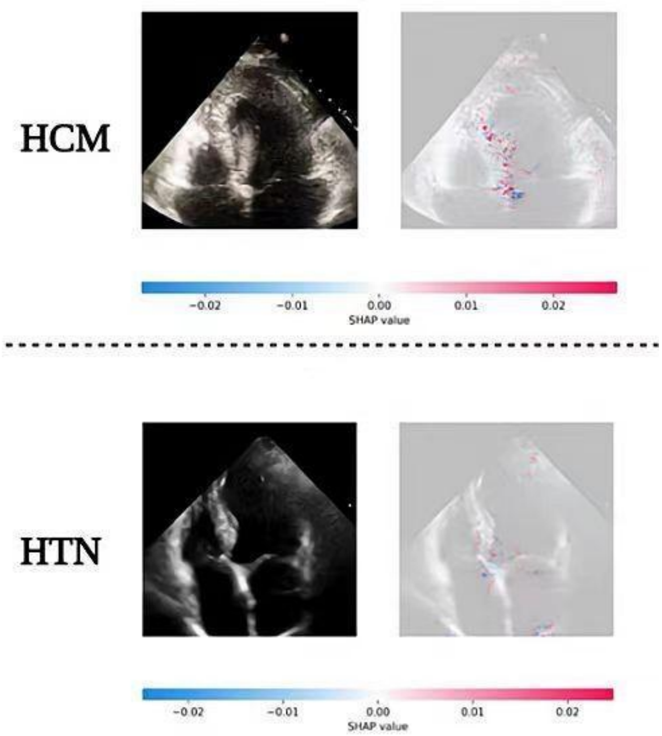
Models	auROC	auPRC	F1-score	Precision
EEMM(LAF)	**0.92**	**0.80**	**0.73**	**0.73**
EEMM(LRF)	0.92	0.76	0.49	0.82
Single-model(ECG)	0.87	0.62	0.51	0.41
Single-model(ECHO)	0.88	0.70	0.63	0.55

Table 1 The Comparison of the Results of Two Single-modal Methods and Two Multimodal Methods with Different Fusion Methods

To enhance our understanding of how the fusion model differentiates between hypertrophic cardiomyopathy (HCM) and hypertensive (HTN) left ventricular hypertrophy (LVH), we employed the SHAP Gradient Explainer, a game theory-based method designed to interpret the outputs of machine learning models[14]. This method calculates the gradient of the model's prediction for each timestep across all leads in the input signal for the ECG component. Similarly, the gradient is computed for each pixel in the input frames for echocardiogram videos. These gradients are then compared against a background distribution derived from the training data. The resulting values are assigned importance scores, where high-impact gradients (highlighted in red) positively contribute to the model's predictions, while low-impact gradients (shown in blue) negatively affect the predictions. Figures 3 and 4 illustrate these results, showcasing the key features influencing the model's decisions.



Figures 3 SHAP of ECG



Figures 4 SHAP of ECHO

Discussion

In this study, we report the first multimodal (ECG and echocardiogram-based) deep learning model in clinical cardiology and use it to predict the etiology of left ventricular hypertrophy. Combining complementary knowledge from multiple modalities can improve diagnostic performance in clinical practice. The trained model demonstrates a high discriminatory ability in distinguishing HCM. Furthermore, different fusion methods provided independent support for future work. Combining complementary information from multiple modalities is intuitively appealing for improving the performance of learning-based approaches.

Conclusion

In summary, this study demonstrates a deep learning model incorporating ECG and echocardiogram time series data and apply it to help identify HCM patients. We use well known fusion methods of combining data streams from multiple modalities and compare these comprehensively to single-modal models. Finally, we end up with late-average fusion, which is better. Future research should focus on optimizing the model further, expanding the scope of multimodal data integration, and exploring its application to other diseases.

Reference

[1] "Detection of Hypertrophic Cardiomyopathy Using a Convolutional Neural Network-Enabled Electrocardiogram — Mayo Clinic." Accessed: Aug. 29, 2024. [Online]. Available: https://mayoclinic.elsevierpure.com/en/publications/detection-of-hypertrophic-cardiomyopathy-using-a-convolutional-ne

[2] J. Kwon *et al.*, "A deep learning algorithm to detect anaemia with ECGs: a retrospective, multicentre study," *The Lancet Digital Health*, vol. 2, no. 7, pp. e358–e367, Jul. 2020, doi: 10.1016/S2589-7500(20)30108-4.

[3] Z. I. Attia *et al.*, "Age and Sex Estimation Using Artificial Intelligence From Standard 12-Lead ECGs," *Circulation: Arrhythmia and Electrophysiology*, vol. 12, no. 9, p. e007284, Sep. 2019, doi: 10.1161/CIRCEP.119.007284.

[4] "Fully Automated Echocardiogram Interpretation in Clinical Practice - PubMed." Accessed: Aug. 29, 2024. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/30354459/

[5] E. Pennacchini, M. B. Musumeci, S. Fierro, P. Francia, and C. Autore,

"Distinguishing Hypertension From Hypertrophic Cardiomyopathy as a Cause of Left Ventricular Hypertrophy," *J Clin Hypertens (Greenwich)*, vol. 17, no. 3, pp. 239–241, Jan. 2015, doi: 10.1111/jch.12481.

[6] "Echocardiographic differentiation of hypertensive heart disease and hypertrophic cardiomyopathy. | Heart." Accessed: Aug. 29, 2024. [Online]. Available: https://heart.bmj.com/content/44/4/395.short

[7] X. Yao *et al.*, "Artificial intelligence–enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial," *Nat Med*, vol. 27, no. 5, pp. 815–819, May 2021, doi: 10.1038/s41591-021-01335-4.

[8] "Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease | npj Digital Medicine." Accessed: Aug. 29, 2024. [Online]. Available: https://www.nature.com/articles/s41746-018-0065-x

[9] Y. Chou, S. Hong, Y. Zhou, J. Shang, M. Song, and H. Li, "Knowledge-shot learning: An interpretable deep model for classifying imbalanced electrocardiography data," *Neurocomputing*, vol. 417, pp. 64–73, Dec. 2020, doi: 10.1016/j.neucom.2020.07.045.

[10] "Full article: Misclassification of hypertrophic cardiomyopathy: validation of diagnostic codes." Accessed: Aug. 28, 2024. [Online]. Available: https://www.tandfonline.com/doi/full/10.2147/CLEP.S139300

[11] "Genotype and Lifetime Burden of Disease in Hypertrophic Cardiomyopathy | Circulation." Accessed: Aug. 28, 2024. [Online]. Available: https://www.ahajournals.org/doi/full/10.1161/CIRCULATIONAHA.117.033200

[12] "Madani: Deep echocardiography: data-efficient supervised... - Google Scholar." Accessed: Aug. 29, 2024. [Online]. Available: https://scholar.google.com/scholar_lookup?title=Deep%20echocardiography%3A%20data-efficient%20supervised%20and%20semi-supervised%20deep%20learning%20towards%20automated%20diagnosis%20of%20cardiac%20disease&author=A%20Madani&author=JR%20Ong&author=A%20Tibrewal&author=MRK%20Mofrad&publication_year=2018&journal=NPJ%20Digit%20Med&volume=1&pages=59

[13] D. Tran, H. Wang, L. Torresani, J. Ray, Y. LeCun, and M. Paluri, "A Closer

Look at Spatiotemporal Convolutions for Action Recognition," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT: IEEE, Jun. 2018, pp. 6450–6459. doi: 10.1109/CVPR.2018.00675.

[14] X. Lan, D. Ng, S. Hong, and M. Feng, "Intra-Inter Subject Self-Supervised Learning for Multivariate Cardiac Signals," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 4, Art. no. 4, Jun. 2022, doi: 10.1609/aaai.v36i4.20376.