**Abstract**

In contemporary medical research, multimodal data analysis has become essential for enhancing the accuracy and reliability of disease diagnosis. This study develops a multimodal disease diagnosis system using electrocardiogram (ECG) and echocardiogram (ECHO) data from the MIMIC-III and MIMIC-IV databases, which contain comprehensive clinical and physiological information from ICU patients at Beth Israel Deaconess Medical Center.

The research begins with the preprocessing and feature extraction of ECG and ECHO data to ensure high quality and consistency. Convolutional neural networks (CNNs) are employed to process ECG signals, while recurrent neural networks (RNNs) with attention mechanisms are used to analyze ECHO data. Advanced multimodal fusion techniques integrate these data types into a unified analytical framework.

The study is validated through multiple cohorts, including a development cohort for model training, an internal validation cohort to assess performance within the same dataset, a prospective validation cohort to test generalization on new data.

This multimodal system is applied to the automated diagnosis of common cardiac diseases, such as arrhythmias, cardiomyopathies, and heart failure, and is compared with traditional single-modality diagnostic methods to demonstrate its advantages. This research highlights the importance of multimodal data fusion in clinical diagnosis and proposes new methodologies for future medical AI applications.

**Introduction**

Multimodal data analysis has gained increasing importance in medical research for improving the precision and reliability of disease diagnosis. With the rapid expansion of medical data and the advancement of artificial intelligence, approaches based on multimodal data are receiving growing attention. Traditional single-modality diagnostic methods often fail to capture the complexity of certain medical conditions, prompting a shift towards combining multiple data types, such as ECG and ECHO, for more accurate cardiac disease diagnosis.

However, effectively integrating diverse medical data types remains a significant challenge. This study aims to address this by developing a robust multimodal disease diagnosis system utilizing ECG and ECHO data from the MIMIC-III and MIMIC-IV databases. Unlike previous research, which often focuses on single-modality analysis, this study explores the deep integration of multimodal data using convolutional neural networks (CNNs) and recurrent neural networks (RNNs) with attention mechanisms. The model's generalizability and robustness are ensured through comprehensive validation across multiple cohorts.

Recent advancements in multimodal data analysis have demonstrated its potential, particularly in fields like medical imaging and natural language processing. Multimodal approaches have shown superior performance compared to single-modality methods, especially in cardiac disease diagnosis, where integrating various physiological signals offers a more complete view of a patient's health. Despite the progress, studies that deeply integrate ECG and ECHO data using large-scale datasets are still scarce. This research aims to fill this gap by developing a comprehensive multimodal diagnostic system.

**Method**

The study utilizes the MIMIC-III and MIMIC-IV databases, extracting ECG and ECHO data that meet specific criteria. This data undergoes rigorous preprocessing and feature extraction to ensure consistency and quality. The ECG data is processed using CNNs, while the ECHO data is analyzed by RNNs equipped with attention mechanisms. The multimodal data fusion is achieved through these attention mechanisms, effectively integrating the different data types into a cohesive model.

The experiments are structured across three stages: a development cohort for model training, an internal validation cohort to assess performance within the same dataset, a prospective validation cohort to test the model's generalization capabilities on new data.

**Results**

The performance of the multimodal fusion model is evaluated across the various validation cohorts, with a particular focus on its generalizability and robustness, especially in the external validation cohort. The model's results are compared with traditional single-modality diagnostic methods, demonstrating the clear advantages of multimodal fusion in improving diagnostic accuracy.

**Discussion**

An in-depth analysis of the experimental results is provided, discussing variations in model performance across the different validation cohorts and evaluating the strengths and limitations of the multimodal approach. The potential clinical applications of the multimodal diagnostic system are explored, along with suggestions for future research directions.

**Conclusion**

This study demonstrates the effectiveness of multimodal data fusion for cardiac disease diagnosis and highlights its potential applications in clinical settings. Future research should focus on optimizing the model further, expanding the scope of multimodal data integration, and exploring its application to other diseases.