

Research Article

A Modality-Adaptive Multimodal Deep Learning Framework With Uncertainty-Aware Attention for Robust Heart Disease Prediction

Sonali¹, Vinay Chopra²

¹Student, ²Assistant Professor, Department of Computer Science & Engineering, AI&ML, DAV Institute of Engineering and Technology, Jalandhar, Punjab, India

DOI: <https://doi.org/10.24321/2394.6539.202603>

I N F O

Corresponding Author:

Sonali, Department of Computer Science & Engineering, AI&ML, DAV Institute of Engineering and Technology, Jalandhar, Punjab, India

E-mail Id:

sonalisona2028@gmail.com

Orcid Id:

<https://orcid.org/0009-0009-7408-954X>

How to cite this article:

Sonali, Chopra V. A Modality-Adaptive Multimodal Deep Learning Framework With Uncertainty-Aware Attention for Robust Heart Disease Prediction. *J Adv Res Med Sci Technol* 2026;13(1&2):16-23.

Date of Submission: 2026-05-10

Date of Acceptance: 2026-06-12

A B S T R A C T

Introduction: Heart disease is still a major cause of death around the globe, so spotting it early can really help save lives. Today's healthcare systems churn out tons of varied medical data like ECG reports, clinical records, and lab tests. But a lot of the traditional machine learning models just use one type of data, which means they might miss out on understanding the complex nature of heart problems. In real-life clinical settings, data isn't always perfect - ECG signals might have glitches, and clinical records could be missing stuff.

Objective: To tackle these issues, this study presents a new multimodal deep learning framework that's good at predicting heart disease, even when things are uncertain.

Findings: This system mixes ECG images with well-organized medical information, letting it use different data types together. A Convolutional Neural Network (CNN) finds key spatial patterns in the ECG pictures, while a Feed-Forward Neural Network (FFNN) spots hidden connections in the medical details. The special part is the Uncertainty-Aware Modality-Adaptive Attention Fusion (UA-MAAF). Unlike regular methods, this one pays attention differently depending on how good and useful the data is. If some data is messy or missing, it can still rely on the solid bits, making predictions more stable.

Conclusion: Tests on common datasets showed this system is more accurate than other methods using one or more data types, according to accuracy, F1-score, and AUC-ROC.

Keywords: Multimodal Deep Learning, Heart Disease Prediction, ECG Image Analysis, Clinical Data Modeling, Attention Fusion, Uncertainty-Aware Learning, Deep Neural Networks

Introduction

Cardiovascular diseases, heart disease are among the top causes of death worldwide killing millions of people every

year.^{8,9} Getting an accurate diagnosis is crucial because it allows doctors to start the right treatment improve patient survival rates and reduce the overall healthcare

Journal of Advanced Research in Medical Science and Technology (E-ISSN: 2394-6539)

Copyright (c) 2025: Author(s). Published by Advanced Research Publications



burden.^{12,18} Over the past decade artificial intelligence and deep learning have changed the way medical diagnosis systems are developed making it possible to automatically analyze clinical data on a large scale.^{10,21}

Hospitals generate an amount of different medical data every day including electrocardiogram signals, diagnostic images, patient medical histories, blood pressure readings and cholesterol values.¹¹ Despite the progress in technology creating heart disease prediction systems is still a significant challenge because medical data is complex and comes in many different forms.^{1,13}

Information about heart disease can be found across sources, including ECG signals, medical images and clinical attributes like age, blood pressure, cholesterol level and medical history.^{9,15} Traditional machine learning approaches have mostly relied on a source of data, which limits their ability to capture relationships between different medical data sources.⁶ Deep learning models that focus on a modality have shown good performance in many studies.^{16,17,20} but they fail to use the complementary information available across different modalities resulting in suboptimal prediction accuracy in real clinical environments.^{3,4}

To overcome these limitations multimodal deep learning has emerged as an approach for combining different types of medical information.^{2,13} By processing modalities together these frameworks can learn complementary representations and provide better predictive performance.^{1,14} Several studies have confirmed that combining imaging data with clinical attributes yields disease prediction than unimodal systems.^{5,18,19} However most existing multimodal frameworks use fixed fusion strategies that treat all modalities equally regardless of data quality or informativeness.^{7,13} In world clinical settings data quality can vary significantly: ECG signals may contain errors patient records can be incomplete and certain clinical measurements may be inconsistent or incorrect. As a result fixed fusion strategies often struggle in conditions.^{8,11}

Recently attention-based fusion mechanisms have been explored to introduce modality weighting into multimodal learning.^{7,18} While these approaches improve adaptability many still assume that all modalities are equally reliable and seldom account for uncertainty explicitly.¹⁹ This represents a limitation in healthcare applications, where noisy or low-quality data can substantially degrade prediction performance. Therefore there is a need for multimodal frameworks that can jointly model both modality relevance and data reliability.^{12,13}

To address these gaps this work proposes a novel Uncertainty-Aware Adaptive Attention Multimodal

Framework for heart disease prediction. Unlike attention-based methods the proposed system explicitly estimates data-level uncertainty and dynamically adjusts each modality's contribution based on both its informativeness and its trustworthiness. When a modality is degraded by noise or incompleteness the framework reduces its influence. Redistributes emphasis toward more reliable sources thereby enhancing robustness in realistic clinical scenarios where data imperfection is inevitable. The main contributions of this work are as follows:

- We propose a multimodal deep learning framework that jointly models modality importance and data reliability through an uncertainty-aware adaptive fusion mechanism.^{1,7}
- We introduce an attention mechanism that explicitly incorporates uncertainty estimation, significantly improving robustness under noisy and incomplete data conditions.^{13,18}
- We conduct experimental evaluations on benchmark datasets demonstrating superior performance, over both unimodal baselines and existing multimodal approaches.^{3,14,19}
- We provide interpretability through modality-level attention analysis supporting the framework's applicability within clinical decision support systems.^{12,21}

Proposed Methodology

This study is about a way to predict heart disease. It uses a kind of deep learning that can handle uncertain things. The framework is called the uncertainty-modality-adaptive multimodal deep learning framework, or UA-MAAF for short. The UA-MAAF is made to work with kinds of medical data like electrocardiogram images and clinical data from hospitals.

The UA-MAAF has to deal with some problems like data that is all over the place, noise that can be confusing, and missing information that is important. The whole system has three parts:

- Getting features from electrocardiogram images using a convolutional neural network, or CNN.
- Making sense of data using a feed-forward neural network or FFNN.
- Using Uncertainty-Aware Modality-Adaptive Attention Fusion, or UA-MAAF, to combine everything.

You can see what the whole system looks like in Figure 3. The UA-MAAF is a part of this system, and it helps with heart disease prediction by using the uncertainty-aware modality-adaptive multimodal deep learning framework.

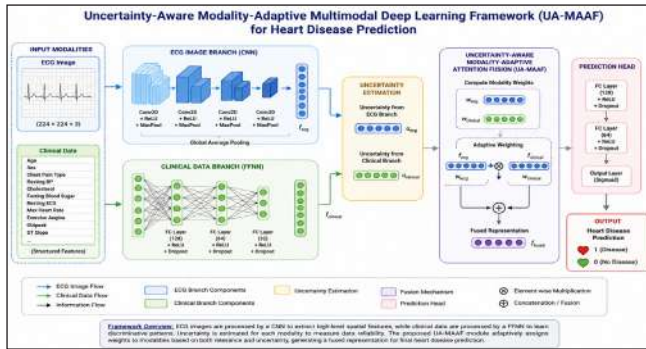


Figure 1. UA-MAAF Framework architecture showing CNN for ECG processing, FFNN for clinical data, and Uncertainty-Aware Attention Fusion for heart disease prediction

ECG Image Feature Extraction

Electrocardiogram (ECG) signals are first preprocessed and transformed into image representations to enable spatial feature extraction. A Convolutional Neural Network (CNN) is employed to automatically learn hierarchical feature representations from ECG images.

$$F_{i,j,k}^l = \sigma(\sum_{m=1}^M \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} W_{m,k,q,p}^l \times X_{i+p,j+q,m}^{l-1} + b_k^l) \quad 1$$

where $F_{i,j,k}^l$ represents the feature map at position (i,j) in channel k of layer l , $W_{m,k,q,p}^l$ is the convolutional kernel, $X_{i+p,j+q,m}^{l-1}$ is the input from the previous layer, b_k^l is the bias term, M is the number of input channels, $P \times Q$ is the kernel size, and σ is the activation function (e.g., ReLU).

The CNN extracts a compact feature vector that captures discriminative patterns from ECG signals relevant to cardiac abnormalities.

Clinical Data Representation

Structured clinical data, including patient attributes such as age, blood pressure, cholesterol levels, and heart rate, are processed using a Feed-Forward Neural Network (FFNN). The FFNN models complex non-linear relationships among clinical features.

The transformation is defined as:

$$F^l = \sigma(W^l \times F^{l-1} + b^l) \quad 2$$

where F^l is the feature vector at layer l , and W^l and b^l are the weight matrix and bias vector for the layer, and σ is a non-linear activation function

The output of this module is a dense representation encoding patient-specific clinical information.

Uncertainty Estimation

To account for variability and noise in medical data, an uncertainty estimation mechanism is incorporated for each modality. The uncertainty score u_m for modality m is computed as:

$$u_m = 1 - \text{confidence}(f_m) \quad 3$$

Where f_m represents the feature vector of modality m . Confidence can be estimated using prediction entropy or variance from model outputs.

This formulation ensures that modalities with higher uncertainty (i.e., lower confidence) contribute less to the final prediction.

Uncertainty-Aware Modality-Adaptive Attention Fusion (UA-MAAF)

To effectively combine multimodal features, we propose a novel uncertainty-aware attention fusion mechanism. Unlike conventional attention models that consider only feature importance, the proposed method integrates both modality relevance and data reliability.

Unlike conventional attention mechanisms, the proposed UA-MAAF explicitly incorporates uncertainty estimation into the fusion process, enabling robustness under noisy and incomplete data conditions.

First, attention weights are computed as:

$$\alpha_m = \text{Softmax}(W[f_{\text{image}}; f_{\text{tabular}}] + b) \quad 4$$

Next, uncertainty-adjusted weighting is applied:

$$\tilde{\alpha}_m = \alpha_m \cdot (1 - u_m) \quad 5$$

Finally, the fused feature representation is computed as:

$$F_{\text{fused}} = \sum_m \tilde{\alpha}_m \cdot f_m \quad 6$$

This formulation ensures that the model dynamically prioritizes reliable modalities while suppressing noisy or less informative inputs.

Classification Layer

The fused feature vector is passed through fully connected layers followed by a sigmoid activation function to predict the probability of heart disease:

$$\gamma = \sigma(w_f F_{\text{fused}} + b_f) \quad 7$$

where γ represents the predicted probability.

Training Strategy

The model is trained using the binary cross-entropy loss function:

$$L = -\frac{1}{N} \sum_{i=1}^N [\gamma_i \log(\gamma_i) + (1 - \gamma_i) \log(1 - \gamma_i)] \quad 8$$

The Adam optimizer is used for parameter optimization. To improve generalization, techniques such as dropout and early stopping are applied.

The proposed model is implemented using TensorFlow and Keras. Training is performed using a custom fusion data generator that synchronizes ECG image batches with corresponding structured clinical features. The model is trained with a learning rate of 0.001, a batch size of 32,

and 50 epochs. The Adam optimizer is used for parameter updates, and validation is performed at each epoch to monitor generalization performance. Dropout and early stopping are applied to prevent overfitting.

Results and Discussion

Overview

This chapter presents the experimental evaluation of the proposed Uncertainty-Aware Modality-Adaptive Attention Fusion (UA-MAAF) framework for heart disease prediction. The performance of the proposed model is compared with unimodal and conventional multimodal baselines using multiple evaluation metrics.

Quantitative Performance Evaluation

The proposed Modality-Adaptive Attention Fusion (MAAF) model achieves superior performance on the disease prediction task.

Table 1. Performance Evaluation of Proposed Model

-	ECG	CSV	MAAF
Metric	Value	Value	Value
Accuracy	0.8135	0.8480	0.9491
Precision	0.7761	0.8446	1.0
Recall	0.8813	0.8529	0.8938
F1-score	0.8253	0.8487	0.9464
AUC Score	0.8888	0.9140	0.9982

Accuracy Analysis

The ECG-based CNN model achieves an accuracy of 81.35%, indicating moderate predictive capability due to variability in ECG signals. The clinical model achieves 84.80% accuracy, demonstrating the importance of structured medical attributes.

The proposed UA-MAAF model significantly improves accuracy to 94.91%, showing that multimodal integration effectively enhances prediction performance.

Precision, Recall, and F1-score Analysis

The precision of the proposed model reaches 1.0000, indicating that all predicted positive cases are correctly classified. However, this perfect precision may vary when evaluated on larger and more diverse datasets. The perfect precision score indicates no false positives in the validation set; however, this may vary when evaluated on larger and more diverse datasets. The recall value (0.8938) shows strong ability to detect actual heart disease cases.

The F1-score (0.9464) reflects a balanced trade-off between precision and recall, confirming the reliability of the proposed model.

Training and Validation Accuracy Analysis

The training and validation accuracy curves provide insights into the learning behavior of the models.

- The ECG model exhibits fluctuations, indicating sensitivity to noise. Figures 3.1
- The clinical model shows stable convergence. Figures 3.2
- The proposed model demonstrates smooth and stable convergence, indicating effective learning. Figures 3.3

This confirms that uncertainty-aware fusion stabilizes training and improves generalization.

Loss Convergence Analysis

The loss curves show the optimization behavior of the models:

- **ECG model:** Irregular loss pattern Figures 3.4
- **Clinical model:** Steady convergence Figures 3.5
- **UA-MAAF model:** Balanced and smooth convergence Figures 3.6

This indicates that multimodal fusion improves learning efficiency and reduces overfitting.

ROC Curve Analysis

The ROC curves further validate model performance:

- ECG model shows moderate curve Figure 3.7
- Clinical model shows better curve Figure 3.8
- Proposed model shows near-perfect ROC curve Figure 3.9

This confirms the superior classification ability of the UA-MAAF framework.

The high AUC score may be influenced by dataset size and controlled experimental conditions; further validation on larger real-world datasets is required.

Confusion Matrix Analysis

The confusion matrix analysis highlights classification performance:

- ECG model shows higher false negatives Figure 3.10
- Clinical model improves detection Figure 3.11
- UA-MAAF model achieves maximum correct predictions Figure 3.12

The reduction in false negatives is particularly important for medical diagnosis, where missing a disease case can have serious consequences.

Modality Attention Weight Analysis

The attention mechanism provides interpretability: Figure 3.13

- Clinical data receives higher weight in most cases

- ECG data contributes when strong signal patterns exist
- This demonstrates that the model dynamically adapts to input data quality and relevance.

Comparison with Existing Models

The performance of the proposed model is compared with traditional machine learning models:

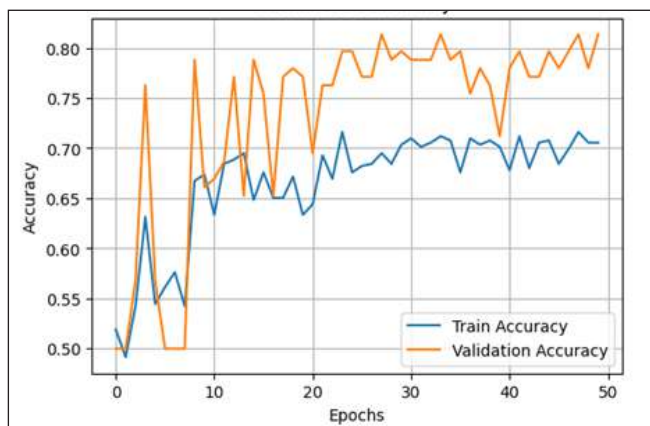


Figure 2.Heart ECG Accuracy

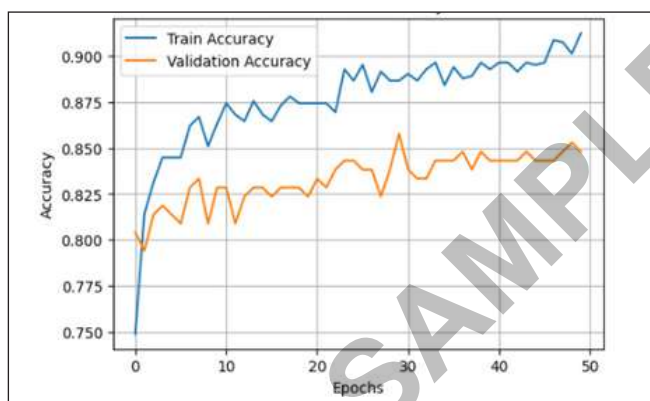


Figure 3.Heart CSV Accuracy

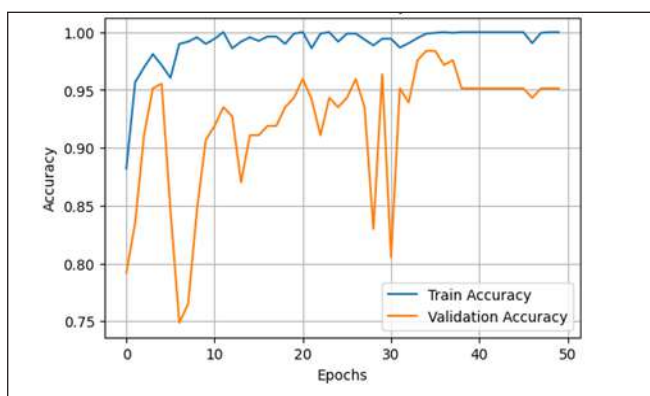


Figure 4.MAAF Accuracy

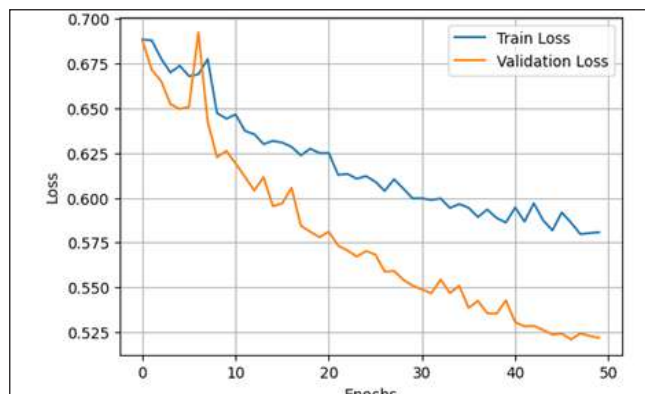


Figure 5.Heart ECG Loss

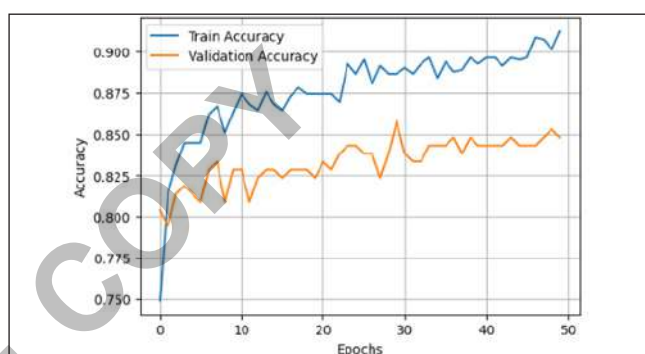


Figure 6.Heart CSV Loss

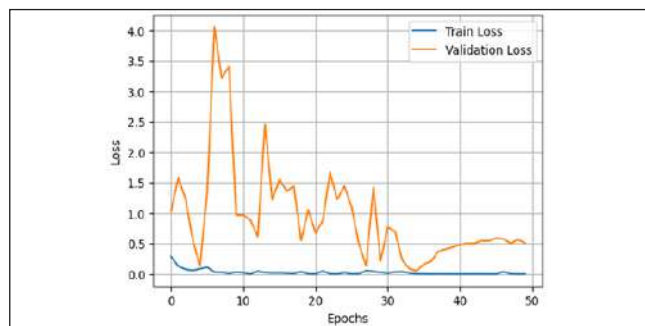


Figure 7.MAAF Loss

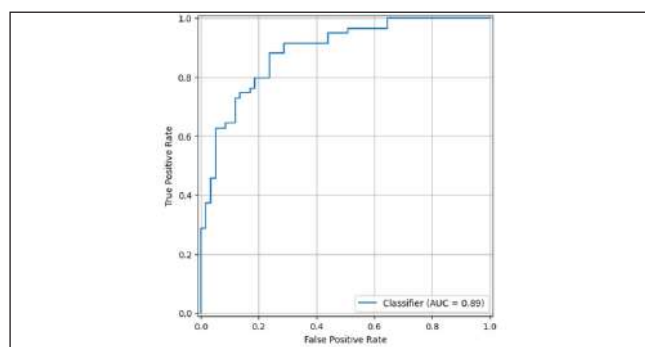


Figure 8.Heart ECG ROC Curve

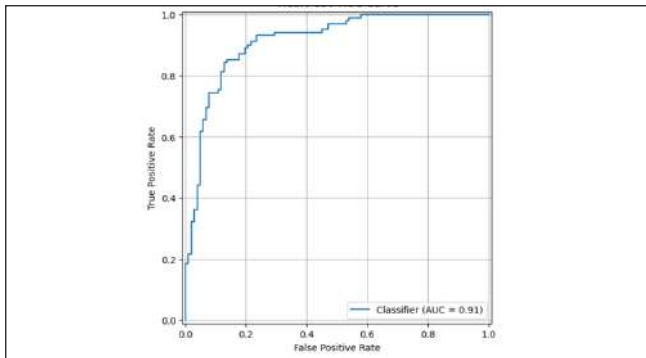


Figure 9.Heart CSV ROC Curve

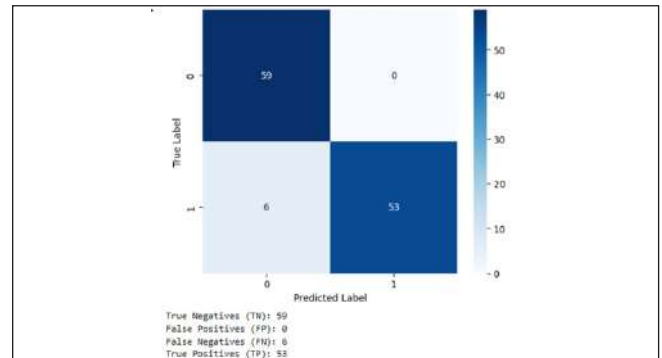


Figure 13.MAAF Confusion Matrix

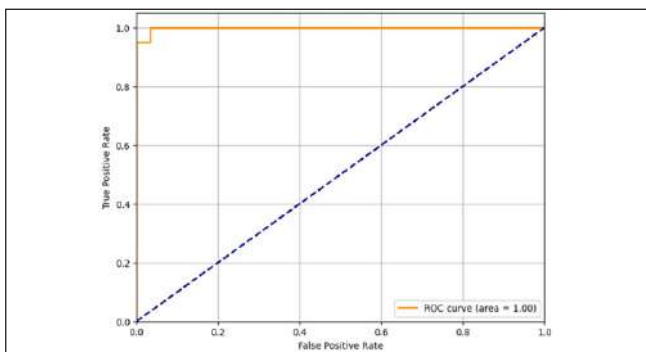


Figure 10.ROC Curve - MAAF Model



Figure 14.Average Modality Attention Weights

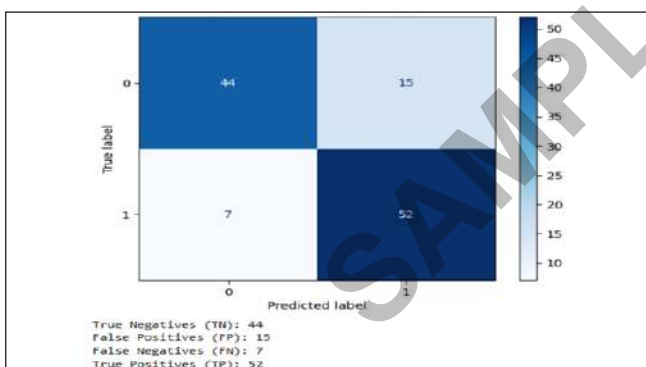


Figure 11.Hear ECG Confusion Matrix

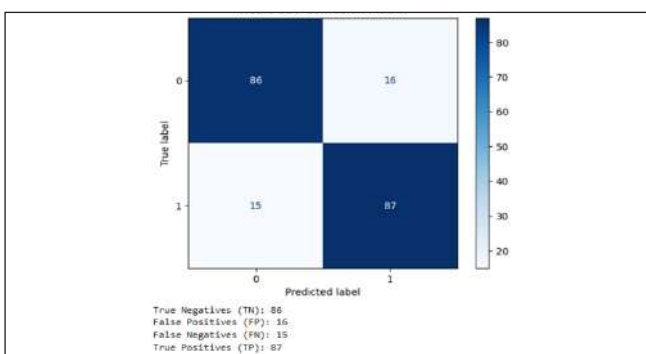


Figure 12.Hear CSV Confusion Matrix

Figure 2.Comparison of Different Algorithms

Model	Data Used	Fusion Type	Accuracy
MN-STDT. ⁸	X-ray + EHR (Temporal)	Multimodal Fusion	0.8186
Multimodal Fusion. ⁵	Retinal + Blood + Notes	Late Fusion	0.859
DeAF. ¹	Image + Clinical	Static Fusion	0.8135
UN-MAAF	Image + Clinical	Adaptive Attention Fusion	0.9491

Discussion

The experimental results show that using multimodal learning makes a difference in predicting heart disease. Multimodal learning does this by combining ECG images and clinical data in a way that's more effective than using just one type of data. The new way of combining ECG images and clinical data, which is aware of uncertainty, makes the system work better. This happens because the system gives weight to the types of data that are most relevant and reliable. This makes the multimodal learning system more robust and able to handle data.

The multimodal learning model is also better at generalizing and learning because the data is balanced and the way it combines ECG images and clinical data is adaptive. However, this study only used data that's publicly available and did not test it in real clinical settings. This study also did not compare the multimodal learning model to recent models that use transformers for multimodal learning.

Overall, the proposed multimodal learning framework has a lot of potential for helping doctors make decisions in real-world clinics. The multimodal learning framework does this by providing interpretable and reliable results. Multimodal learning is really good at predicting heart disease, and the uncertainty-aware fusion mechanism makes the multimodal learning system work better.

Abbreviations

Abbreviation	Full Form
AI	Artificial Intelligence
CVD	Cardiovascular Disease
ECG	Electrocardiogram
CNN	Convolutional Neural Network
FFNN	Feed-Forward Neural Network
UA-MAAF	Uncertainty-Aware Modality-Adaptive Attention Fusion
AUC-ROC	Area Under the Receiver Operating Characteristic Curve
ReLU	Rectified Linear Unit
BCE	Binary Cross-Entropy
ML	Machine Learning

Declarations

Availability of Data and Material

The datasets used and analyzed during the current study are publicly available on the Kaggle repository, including ECG image datasets and heart disease clinical datasets.

Authors' Contributions

Sonali: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, Writing – Original Draft.

Dr. Vinay Chopra: Supervision, Validation, Writing – Review & Editing.

Acknowledgements

The author would like to thank DAV Institute of Engineering and Technology (DAVIET), Jalandhar, for providing academic support and research resources.

Conflict of Interest: None

Source of Funding: None

REFERENCES

1. Li K, Chen C, Cao W, Wang H, Han S, Wang R, Ye Z, Wu Z, Wang W, Cai L, Ding D. DeAF: a multimodal deep learning framework for disease prediction. *Computers in Biology and Medicine*. 2023 Apr 1;156:106715. [Google Scholar]
2. Yao Z, Lin F, Chai S, He W, Dai L, Fei X. Integrating medical imaging and clinical reports using multimodal deep learning for advanced disease analysis. In 2024 IEEE 2nd International Conference on Sensors, Electronics and Computer Engineering (ICSECE) 2024 Aug 29 (pp. 1217-1223). IEEE. [Google Scholar]
3. Kumar M, Shelke NA, Singh J, Sharma KN, Sharma R, Kumar R. A Multimodal Deep Learning Approach for Advancing Liver Disease Diagnosis and Prognosis Prediction. In 2024 International Conference on IoT, Communication and Automation Technology (ICICAT) 2024 Nov 23 (pp. 838-843). IEEE. [Google Scholar]
4. Kumar S, Ivanova O, Melyokhin A, Tiwari P. Deep-learning-enabled multimodal data fusion for lung disease classification. *Informatics in Medicine Unlocked*. 2023 Jan 1;42:101367. [Google Scholar]
5. Mr Prasad. N. Maldhure, "Multimodal Deep Learning Framework for Enhanced Diabetes Prediction Using Retinal Images, Blood Test Data, and Clinical Notes," *International Journal of Creative Research Thoughts*, vol. 13, issue 8 August 2025. <https://www.ijcrt.org/papers/IJCRT2508271.pdf>
6. Akil Arsath J., Suganthi S, Rakeshwaran S., and Karthiga S., "Multimodal Disease Prediction using Machine Learning and Deep Learning Techniques," *International Research Journal on Advanced Science Hub*, vol. 05, issue 05 May. https://www.researchgate.net/publication/372126155_Multimodal_Disease_Prediction_using_Machine_Learning_and_Deep_Learning_Techniques
7. Ashwin M. Nayak and Dr Vibha M. B. Vibha, "Graph-Enhanced Multimodal Deep Learning for Disease Prediction: A Hybrid Approach," *International Journal of Engineering Research & Technology* vol. 11, issue 06. <https://www.ijert.org/graph-enhanced-multimodal-deep-learning-for-disease-prediction-a-hybrid-approach> Li D, Xing W, Zhao J, Shi C, Wang F. Multimodal deep learning for predicting in-hospital mortality in heart failure patients using longitudinal chest X-rays and electronic health records. *The International Journal of Cardiovascular Imaging*. 2025 Mar;41(3):427-40. [Google Scholar]

8. Sivasubramaniam S, Balamurugan SP. Early detection and prediction of Heart Disease using Wearable devices and Deep Learning algorithms. *Multimedia Tools and Applications*. 2025 Mar;84(9):6187-201. [Google Scholar] D. R. Sarvamangala and R. V. Kulkarni, "Convolutional neural networks in medical image understanding: a survey," *Evolutionary Intelligence* (Springer Nature), vol. 15, pp. 1–22, 2022. DOI: 10.1007/s12065-020-00540-3 [Google Scholar]
9. Liu S, Wang X, Hou Y, Li G, Wang H, Xu H, Xiang Y, Tang B. Multimodal data matters: Language model pre-training over structured and unstructured electronic health records. *IEEE Journal of Biomedical and Health Informatics*. 2022 Oct 28;27(1):504-14. [Google Scholar]
10. Almazroi AA, Aldhahri EA, Bashir S, Ashfaq S. A clinical decision support system for heart disease prediction using deep learning. *IEEE Access*. 2023 Jun 12;11:61646-59. [Google Scholar]
11. Zhao F, Zhang C, Geng B. Deep multimodal data fusion. *ACM computing surveys*. 2024 Apr 24;56(9):1-36. [Google Scholar]
12. Kumar S, Ivanova O, Melyokhin A, Tiwari P. Deep-learning-enabled multimodal data fusion for lung disease classification. *Informatics in Medicine Unlocked*. 2023 Jan 1;42:101367. [Google Scholar]
13. Rai HM, Chatterjee K. Hybrid CNN-LSTM deep learning model and ensemble technique for automatic detection of myocardial infarction using big ECG data: Hybrid CNN-LSTM deep learning model and ensemble technique for automatic detection of myocardial infarction using big ECG data. *Applied Intelligence*. 2022 Mar;52(5):5366-84. [Google Scholar]
14. Bhavakar GS, Goswami AD. A hybrid model for heart disease prediction using recurrent neural network and long short term memory. *International Journal of Information Technology*. 2022 Jun;14(4):1781-9. [Google Scholar]
15. Rath A, Mishra D, Panda G, Satapathy SC, Xia K. Improved heart disease detection from ECG signal using deep learning based ensemble model. *Sustainable Computing: Informatics and Systems*. 2022 Sep 1;35:100732. [Google Scholar]
16. Lee YC, Cha J, Shim I, Park WY, Kang SW, Lim DH, Won HH. Multimodal deep learning of fundus abnormalities and traditional risk factors for cardiovascular risk prediction. *NPJ digital medicine*. 2023 Feb 2;6(1):14. [Google Scholar]
17. Soto JT, Weston Hughes J, Sanchez PA, Perez M, Ouyang D, Ashley EA. Multimodal deep learning enhances diagnostic precision in left ventricular hypertrophy. *European Heart Journal-Digital Health*. 2022 Sep 28;3(3):380-9. [Google Scholar]
18. Islam MS, Islam MN, Hashim N, Rashid M, Bari BS, Al Farid F. New hybrid deep learning approach using BiGRU-BiLSTM and multilayered dilated CNN to detect arrhythmia. *IEEE Access*. 2022 May 30;10:58081-96. [Google Scholar]
19. Zhuang H, Lin Z, Yang Y, Toh KA. An analytic formulation of convolutional neural network learning for pattern recognition. *Information Sciences*. 2025 Jan 1;686:121317. [Google Scholar]