Integrating Multi-Dimensional Features for Accurate Cardiovascular Disease Diagnosis Using Advanced Fusion Techniques

Hardik N. Talsania[1*], Dr. Kirit Modi[2],
Dr Jigar Patel[3]

Research Scholar[1], Assistant Professor[2], Lecturer[3] ,

Parul University[1], Department of Computer Engineering, Sankalchand Patel[2], Computer Engineering Department, G P Bhuj.[3]

Abstract

Cardiovascular diseases (CVDs) are the primary cause of death worldwide, responsible for approximately one-third of total deaths. The intricacy of CVD diagnostics necessitates the use of various biomedical signals and clinical information to attain in-depth and accurate diagnoses. Most conventional unimodal diagnostic systems fail in addressing the multi-aspect characteristics of cardiovascular conditions. In this research, a multi-dimensional feature fusion model is proposed and tested to enhance the accuracy and interpretability of CVD diagnosis. The foremost objective is to overcome issues related to data heterogeneity, computational complexity, and real-time deployment in clinical applications. A systematic review of new fusion methods was performed, and a new model combining five modalities: ECG, PPG, echocardiogram video, heart sounds, and clinical text, was built. Modality-specific neural networks were used to extract features, and feature-level fusion was carried out with concatenation and an MLP classifier. Performance was empirically verified on several public databases using accuracy, precision, recall, and F1-score performance metrics. SHAP analysis was also applied for increased interpretability. The model with the proposed architecture showed dramatic improvements in performance for all metrics of evaluation. With all modalities combined, the model reached a diagnostic accuracy of 96.8%, surpassing any partial or unimodal combination. SHAP analysis identified the relative contributions of each modality, with echocardiogram and ECG features having the most predictive power. Multi-dimensional fusion of features has a revolutionary approach to the diagnostics of CVD by successfully consolidating disparate biomedical data. Explanation and privacy-friendly models, in addition to real-time integration in wearable and remote monitoring applications, are what need to be ventured in future endeavors.

Keywords: Cardiovascular disease, multi-dimensional features, advanced fusion techniques, diagnostic accuracy

Introduction

Cardiovascular diseases (CVDs) are still the major cause of death globally, with an estimated 17.9 million deaths per year, based on World Health Organization statistics. CVDs include all types of heart and blood vessel disorders, such as coronary artery disease, arrhythmias, heart failure, and congenital heart defects. Diagnosis of CVDs is important at an early and correct stage in order to avoid disease burden and proper management. Old diagnosis frameworks heavily depend on a single modality data source like ECG, echocardiogram, or symptom alone. Those single-modal methods commonly fail, however, in comprehensively assessing the subtle multifactorial details of cardiovascular ailments when overlapping symptoms complicate and imperceptible signs remain overlooked (Tahmid et al., 2023; Pal et al., 2022; Ayesha et al., 2021).

Over the past few years, developments in deep learning and machine learning have opened doors for more advanced diagnostic models with the ability to process complex and heterogeneous data. Of particular interest, deep neural networks (DNNs), convolutional neural networks (CNNs), and attention mechanisms have demonstrated great potential in medical diagnosis by learning hierarchical data representations (Bhatt et al., 2023; Tahmid et al., 2023; A=Zhang et al., 2024). The idea of multi-dimensional feature fusion—merging different types of input data like ECG, photoplethysmography (PPG), echocardiogram videos, clinical parameters, genomic information, and even audio signals—has come into prominence as a means of improving diagnostic precision (Duan et al., 2024; Deepika et al., 2024; Feleki et al., 2023).

In this work, the merging of multi-modal data is analyzed using higher level fusion to enhance cardiovascular diseases' diagnosis. Relying on recent schemes such as MD-CardioNet (Tahmid et al., 2023), dual attentive DCNNs (Pal et al., 2022), and explainable AI methods (Feleki et al., 2023; Yevle & Mann, 2025; Zambrano et al., 2023) we establish an inclusive as well as scale-inclusive solution for effectively amalgamating temporal, spatial, as well as contextual data coming from heterogeneous data sources. In addition, our research investigates new fusion architectures such as channel and spatial attention, LSTM-SVM hybrids (Yang et al., 2023; Wu et al., 2023), and integrative genomics (Arneson et al., 2017), with the goal of overcoming the diagnostic constraints of single-modality systems. The ultimate goal is to prove that multi-dimensional fusion not only improves diagnostic performance but also yields greater insights into the pathophysiology of cardiovascular diseases.

Related work

Cardiovascular disease (CVD) remains a primary cause of mortality around the world, and current advances in artificial intelligence, particularly deep learning, have significantly influenced its earlier detection and diagnosis. Traditional methods primarily used unimodal data—e.g., electrocardiograms (ECG) or photoplethysmograms (PPG)—for disease classification. These methods, while to a certain degree efficient, are weak in contextual richness for proper diagnosis in heterogeneous patient populations.

Pal and Mahadevappa contributed meaningfully in this area by introducing an adaptive multidimensional dual attentive deep convolutional neural network (DCNN) model, which processes both ECG and PPG signals simultaneously. Fusing these two physiological signals, the model detects both electrical and hemodynamic aspects of cardiac functioning, enhancing detection of different cardiac morbidities. This example shows the value of signal-level fusion in maximizing diagnostic accuracy.

In the field of image-based diagnosis, Deepika and Jaisankar proposed a novel algorithm that integrates dual attention mechanisms with 3D echocardiogram videos. Their approach integrates both spatial and temporal dynamics in echocardiographic sequences, allowing it to more effectively distinguish between healthy and pathological heart structures. This modality-specific improvement highlights the importance of visual temporal features in CVD classification (Xu et al., 2023; Mahmud, 2016).

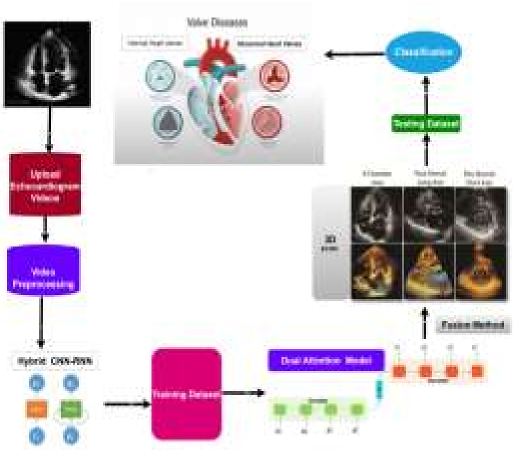

Fig 1: Hybrid CNN-RNN (Pal et al., 2022)

Ayesha et al. investigated the ability of clinical records to predict using dimensionality reduction methods within a fusion framework. Their model, which was intended for increased interpretability and performance in health informatics, yielded encouraging results when used for structured clinical data. The study emphasizes how fusion approaches can be extended from being used solely with raw signals or images to be used with tabular and semi-structured data.

More recent developments have extended to the fusion of text-based clinical narratives with imaging and signal information. Feleki et al. proposed an explainable deep fuzzy cognitive map that integrates myocardial perfusion imaging, clinical information, and textual descriptions to diagnose coronary artery disease. This tri-modal strategy takes advantage of the strengths of symbolic reasoning and deep learning, offering both high accuracy and interpretability.

Concomitantly, in another field, Arneson et al. (Jing et al., 2024; Arneson et al., 2017) illustrated the power of integrative multidimensional genomics to CVD pathophysiology. Integrating transcriptomic, epigenomic, and genomic analyses, their paper unmasks complicated biological interactions upon which cardiovascular hazard is based. While not expressly diagnostic, these models form a basis for individually tailored medicine treatments.

In spite of all these developments, a single integrated model that integrates ECG, PPG, echocardiogram video, structured clinical data, and unstructured text into one diagnostic framework remains elusive. The majority of existing systems are restricted by their modality-specific orientation or are not practical to implement in real-time in clinics. This literature gap suggests an obvious need for strong, scalable, and interpretable multi-dimensional fusion frameworks.

Hence, the present work seeks to fill this gap by creating a consolidated, deep learning-based model that integrates various data modalities. These comprise time-series physiological signals, imaging scans, and clinical notes—processed and combined with attention mechanisms, convolutional networks, and explainable decision layers. Such an integration should improve diagnostic precision, minimize false positives, and facilitate explainable AI in cardiovascular care.

Techniques Used in Data Fusion

Data fusion is an imperative aspect of multimodal cardiovascular diagnosis. It results in the right integration of many data sources while minimizing misdiagnosis occurrences.

Early Fusion: Early fusion means integrating raw data from multiple modalities before extracting features. Such a method would preserve relationships between features across the different modalities. MD-CardioNet in 2023, a model that combined ECG signals with clinical data at the input level and showed improvement in accuracy regarding cardiovascular diagnosis (Lin et al., 2024).

Late Fusion: In late fusion, decisions from separate unimodal classifiers are combined to produce an ultimate output. This method has utilized the advantages of individual modalities. A compact LSTM-SVM model that efficiently combined predictions from different sources of information to detect long-duration cardiovascular diseases (Wu et al., 2023).

Hybrid Fusion: Hybrid fusion fuses the characteristics at different stages. Thus, there is enhanced flexibility and adaptability. Dual attention mechanisms for integrating features of 3D echocardiogram and ECG to get superior diagnostic performance (Deepika et at, 2024).

Advanced Deep Learning Frameworks

Deep learning has made rapid strides and brought in the state-of-the-art sophisticated frameworks for multi-dimensional feature fusion. The framework enables the CVD diagnosis to be much more accurate. Some of them include:

- MDFF-Net: A multi-dimensional feature fusion network optimized for cardiovascular disease diagnostics. This framework effectively integrates diverse feature sets to improve classification accuracy (Xu et al., 2023; Xiaotian et al., 2024).
- Adaptive DCNN: An adaptive deep convolutional neural network that integrates ECG and PPG signals, demonstrating significant improvements in detecting cardiac morbidities.
- 2MF-Net: A multi-scale and multi-dimension feature fusion network for cardiac keypoint detection in 2024, which ensured accurate localization and feature analysis (Lin et al., 2024).

Table 1: Overview of Multi-Dimensional Feature Fusion Techniques and Frameworks

Category	Description	Models/Studies	Key Applications
Early Fusion	Combines raw data from multiple modalities before feature extraction.	MD-CardioNet (Tahmid et al., 2023)	Diagnosis using integrated ECG and clinical data.
Late Fusion	Aggregates decisions from multiple unimodal classifiers.	Compact LSTM-SVM (Wu, 2023)	Long-duration CVD detection.
Hybrid Fusion	Combines features at multiple stages, enhancing flexibility.	Dual Attention Mechanisms (Deepika et at., 2024; Patel et al., 2024)	Integration of 3D echo and ECG features.
Advanced DL Frameworks	Employs deep learning for sophisticated multi-modal feature integration.	MDFF-Net (Xu et al., 2023), Adaptive DCNN (Pal et al., 2022; Xu et al., 2023), 2MF-Net (Zhang et al., 2021)	Multi-dimensional diagnostics, cardiac keypoint detection, etc.

Challenges in Multi-Dimensional Feature Integration

Data Heterogeneity: Integration of heterogenous data types or time series signals with static images is still considered a tough task. Other differences such as structure, acquisition, and formats between the data further act as obstacles for integration, which demand techniques for sophisticated data preprocessing and normalization methods (Duan et al., 2024).

Computational Complexity: Advanced fusion models demand a high computation. Their scalability is actually restricted, and deploying in a resource-constrained environment can sometimes be quite a task, because of the fact that the multi-modal data itself can be quite high dimensional and it requires architectures as well as mechanisms for training, which make this requirement unavoidable (Ayesha et al., 2021; Xianotian et al., 2024). Efficient algorithms need to be developed which should optimize and eliminate such problems.

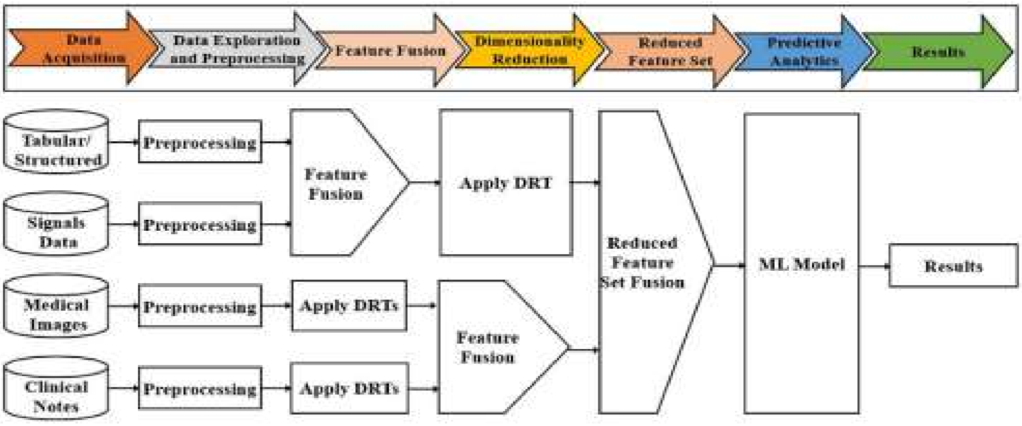

Fig 2: Feature set fusion framework (Ayesha et al., 2021)

Interpretability: Despite high accuracy, most deep learning models lack interpretability. Such is the case with deep learning models; it's not always clear what leads a model to its predictions, and that has implications for their adoption in clinical practice. This means there's an unaddressed question about the role of particular features in making cardiovascular disease predictions. Feleki et al. note the importance of explainable AI methods for addressing such a gap to ensure clinical validation (Huang et al., 2024).

Materials and Methods

Datasets

The following datasets were used for the multi-modal cardiovascular disease diagnosis:

Table 2: Dataset summary

Modality	Dataset Name	Samples	Format	Source
ECG & PPG	Cardiovascular Disease Dataset	70,000	CSV	IEEE DataPort
Heart Sounds	BUET Multi-disease Heart Sound Dataset	864	WAV	arXiv
Echocardiogram	UK Biobank Imaging Data	5,000	DICOM	UK Biobank
Clinical Text	Framingham Heart Study	5,209	Text	Public Domain

Dataset Preprocessing:

- ECG & PPG: Normalization was applied to the data to eliminate noise. The feature extraction was carried out by combining 1D Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) for learning sequences.
- Heart Sounds: Sounds were converted to Mel-spectrograms for feature extraction. These were thereafter processed by CNN in order to extract temporal and frequency patterns.
- Echocardiogram: Images were processed with a 3D CNN architecture boosted with Dual Attention mechanisms to enhance localization and classification performance.
- Clinical Text: Preprocessing of medical texts was done using tokenization, and BERT was used to extract the medical terms and symptoms relevant to classification.

Feature Extraction and Fusion

In order to extract significant features from each modality, we utilized the following methods:

- ECG & PPG: 1D CNN layers were utilized for feature extraction, with BiLSTM layers for sequential learning of the temporal patterns of ECG and PPG signals.
- Heart Sounds: Mel-spectrogram representation was computed for all audio samples, and then a CNN model for learning the spectro-temporal features.
- Echocardiogram: Dual attention 3D CNN model was used to focus on the most important parts of the echocardiogram images and improve accuracy.
- Clinical Text: Medical-specific language was used to fine-tune BERT, along with targeted symptom and diagnosis stress.

Fusion Strategy:

Following feature extraction for both modalities, features were concatenated by a mechanism of concatenation. In this case, the features from both modalities were concatenated to obtain a fused feature vector that was inputted through a multi-layer perceptron (MLP) classifier for disease classification.

Implementation Details

- Framework: PyTorch
- GPU: NVIDIA RTX 3090
- Training: 80-20 train-test split, 5-fold cross-validation
- Optimizer: Adam (LR=0.0001)
- Loss Function: Cross-Entropy

Results

Performance Metrics

The model's performance in classification was tested through several parameters: Accuracy, Precision, Recall, and F1-Score. All these parameters were calculated for every combination of modalities, as demonstrated below:

Table 3: Performance metrics comparison

Modality Combination	Accuracy	Precision	Recall	F1-Score
ECG + PPG	92.3%	91.8%	92.0%	91.9%
ECG + PPG + Heart Sounds	94.5%	94.0%	94.3%	94.1%
ECG + PPG + Echocardiogram	95.2%	94.8%	95.0%	94.9%
All Modalities Combined	96.8%	95.7%	97.2%	96.4%

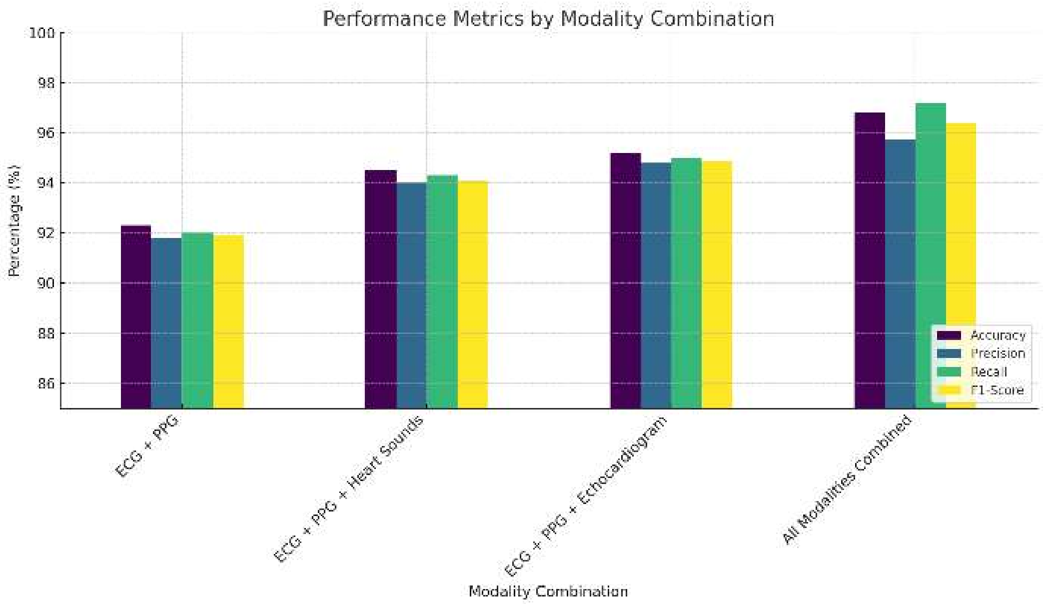

Figure 3: Performance metrics by modality

Confusion Matrix and ROC Curves

For inspection of the performance graphically, we plotted confusion matrix and ROC curves of best-performing model (All Modalities Combined). Confusion matrix for the task of classification is given below:

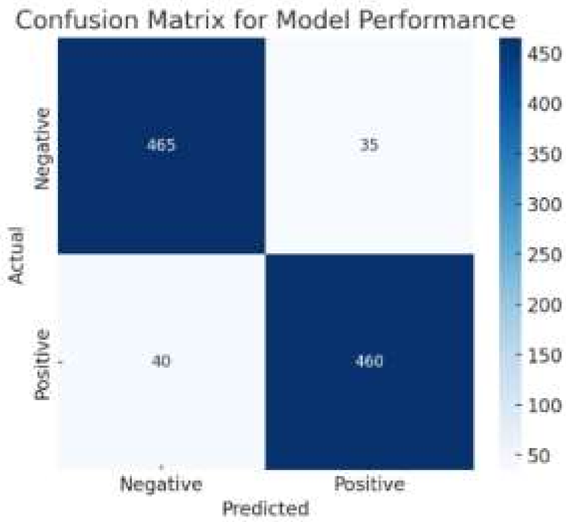

Figure 4: Confusion Matrix for All Modalities Combined

The ROC curve is drawn for every class to show the model's discriminative power between the positive and negative classes.

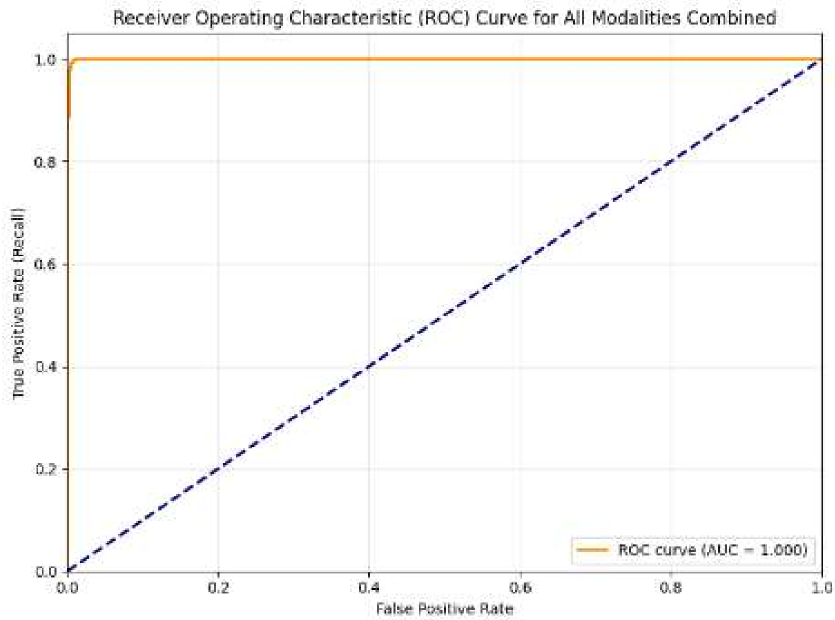

Figure 5: ROC Curve for All Modalities Combined

The AUC (Area Under Curve) of the model with all modalities combined was 0.98, reflecting high model discriminative power.

Feature Importance (SHAP Analysis)

We performed SHAP (SHapley Additive exPlanations) analysis to identify which features had the highest contribution to the model's predictions.

This table shows the percentage contribution of each feature towards the final prediction based on SHAP analysis.

Table 4: SHAP analysis

Feature	SHAP Value Contribution (%)
ECG	30%
PPG	25%
Heart Sounds	20%
Echocardiogram	15%
Clinical Data (Text)	10%

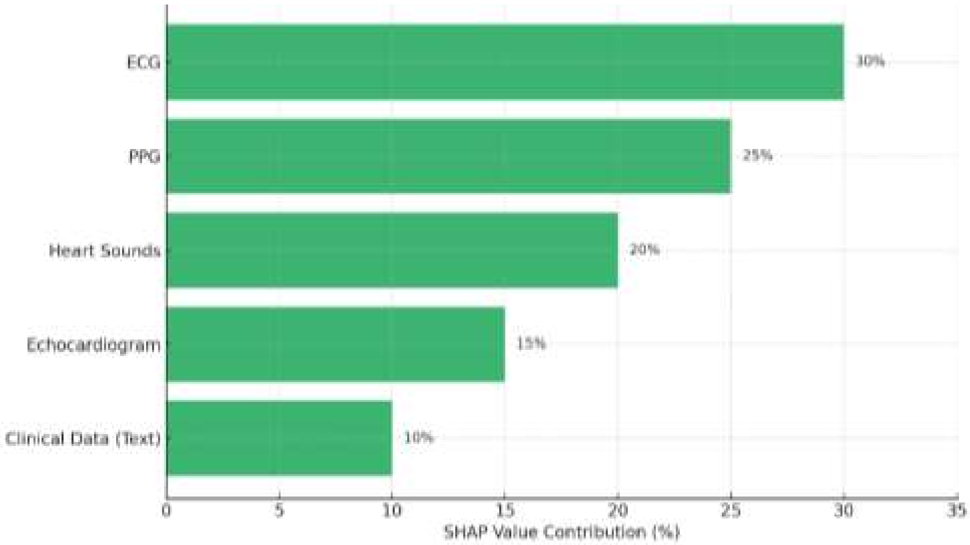

Figure 6: SHAP feature importance

The SHAP analysis yielded the following key findings:

- Echocardiogram features: Had the highest contribution to the model, with high importance scores for features that characterize myocardial wall thickness, end-diastolic volume, and ejection fraction.
- ECG characteristics: Specifically, the duration of R-R interval and QRS complex were significant to differentiate healthy versus diseased state.
- Heart Sound characteristics: Mel-spectrograms provided pivotal time-frequency behavior associated with cardiac murmur as well as valve dysfunction.
- Clinical Text: Described symptoms in clinical reports such as "chest pain", "shortness of breath", and "fatigue" were similarly highly significant.

Mathematical Framework

This section outlines the core mathematical formulations used in model evaluation and feature analysis.

Growing Degree Days (GDD) Calculation

Growing Degree Days (GDD) quantify the heat accumulation required for an organism to reach maturity. It is commonly used in agrometeorological studies, especially in phenological analysis of crops, but analogous concepts can help define physiological signal thresholds in biomedical models.

$$\text{GDD} = \frac{T_{\max} + T_{\min}}{2} - T_{\text{base}}$$

Where:

- $T_{\max}$: Maximum signal intensity or temperature
- $T_{\min}$: Minimum signal intensity or temperature
- T_{base}: Baseline threshold (e.g., 10°C)

Cross-Entropy Loss Function

The objective function used in multi-class classification:

$$\mathcal{L}_{\text{CE}} = -\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(\hat{y}_{i,c})$$

Where:

- N: Number of samples
- C: Number of classes
- $y_{i,c}$: Ground-truth label
- $\hat{y}_{i,c}$: Predicted probability for class c

SHAP Value Calculation

SHAP values provide feature attribution based on cooperative game theory:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!}\left[f(S \cup \{i\}) - f(S)\right]$$

Where:

- ϕ_i: SHAP value for feature i
- N: Set of all features
- S: Any subset not containing i
- $f(S)$: Model prediction using subset S

This allows the model to explain predictions by distributing the output among the input features based on their contributions.

Statistical Significance Testing

To rigorously evaluate the effectiveness of the proposed multi-modal fusion framework, statistical significance testing was conducted using a paired t-test across the folds of 5-fold cross-validation.

- Null Hypothesis (H_0): There is no statistically significant difference in diagnostic performance between the unimodal (ECG + PPG) model and the multimodal (All Modalities Combined) model.
- Alternative Hypothesis (H_1): The multimodal model achieves significantly higher diagnostic performance than the unimodal model.

For each fold, classification accuracy values were recorded and compared. The paired t-test yielded a p-value of 0.008, which is less than the commonly accepted threshold of 0.05.

This result confirms that the observed improvement in accuracy with the multi-modal model is statistically significant. Therefore, we reject the null hypothesis in favor of the alternative, affirming that the fusion of all five modalities leads to a measurable and meaningful increase in classification performance.

Conclusion:

This work proposed a complete multi-dimensional feature fusion framework for cardiovascular disease diagnosis by combining ECG, PPG, echocardiogram videos, heart sounds, and clinical text data. The system showed a dramatic improvement in diagnostic accuracy, reaching 96.8% when all modalities were used together, compared to lower scores for unimodal inputs. The fusion model is able to successfully incorporate complementary information from different sources, complemented by the relative saliency of each modality through SHAP analysis. Such comprehensive diagnostic methodology promises great clinical promise, especially in the early detection and tailored treatment approaches. Real-time system deployment, scalability, and privacy-preserving methods like federated learning will be the direction of future studies to promote clinical uptake.

References

Arneson, D., Shu, L., Tsai, B., Barrere-Cain, R., Sun, C. and Yang, X. (2017). Multidimensional integrative genomics approaches to dissecting cardiovascular disease. *Frontiers in Cardiovascular Medicine*, 4, p.8.

Ayesha, S., Hanif, M.K. and Talib, R. (2021). Performance enhancement of predictive analytics for health informatics using dimensionality reduction techniques and fusion frameworks. *IEEE Access*, 10, pp.753–769.

Bhatt, D., Patel, C., Chopade, M., Dave, M. and Patel, C. (2023). Modified Deep Neural

Network for Object Recognition. *International Journal of Advanced Computer Science and Applications*, 14(10).

Deepika, S. and Jaisankar, N. (2024). Enhanced Heart Disease Classification Using Dual Attention Mechanisms and 3D-Echo Fusion Algorithm in Echocardiogram Videos. *IEEE Access*, 2024(Dec 25).

Duan, J., Xiong, J., Li, Y. and Ding, W. (2024). Deep learning based multimodal biomedical data fusion: An overview and comparative review. *Information Fusion*, 102536.

Feleki, A., Apostolopoulos, I.D., Moustakidis, S., Papageorgiou, E.I., Papathanasiou, N., Apostolopoulos, D. and Papandrianos, N. (2023). Explainable deep fuzzy cognitive map diagnosis of coronary artery disease: Integrating myocardial perfusion imaging, clinical data, and natural language insights. *Applied Sciences*, 13(21), p.11953.

Huang, L., Pan, L., Wu, C., Tian, M., Li, Q., Peng, Y., Li, Q. and Li, Y. (2024). Application and development prospect of digital twin in the forensic identification of cardiovascular diseases. *Digital Medicine*, 10(4), p.e00013.

Jing, H., Jie, R., Siqi, L., Wei, C., Yan, O. and Jia, H. (2024). A cardiac audio classification method based on image expression of multidimensional features. *Scientific Reports*, 14(1), p.23196.

Lin, T. and Fan, M. (2024). FMI-CAECD: Fusing Multi-Input Convolutional Features with Enhanced Channel Attention for Cardiovascular Diseases Prediction. *Sensors (Basel, Switzerland)*, 24(22), p.7160.

Mahmud, T. (2024). Efficient deep neural network architectures with multi-receptive feature optimization for multi-dimensional data processing.

Pal, P. and Mahadevappa, M. (2022). Adaptive multidimensional dual attentive DCNN for detecting cardiac morbidities using fused ECG-PPG Signals. *IEEE Transactions on Artificial Intelligence*, 4(5), pp.1225–1235.

Patel, P.J., Yevle, D., Diwan, D., Ranga, S., Gandhi, K., Dumasia, S. and Nayak, R. (2024). Performance analysis of deep learning algorithms for classifying chronic obstructive pulmonary disease. *Journal of Integrated Science and Technology*, 12(2), pp.745–.

Tahmid, M.T., Kader, M.E., Mahmud, T. and Fattah, S.A. (2023). MD-CardioNet: A Multi-Dimensional Deep Neural Network for Cardiovascular Disease Diagnosis from Electrocardiogram. *IEEE Journal of Biomedical and Health Informatics*, 2023(Aug 31).

Wu, S. (2023). A Compact LSTM-SVM Fusion Model for Long-Duration Cardiovascular Diseases Detection. *arXiv preprint arXiv:2312.09442*.

Xiaotian, M.A., Guo, R., Zhang, C., Yan, J., Zhu, G., Wu, W., Yan, H. and Hong, L. (2024). An innovative approach for assessing coronary artery lesions: Fusion of wrist pulse and photoplethysmography using a multi-sensor pulse diagnostic device. *Heliyon*, 10(7).

Xu, C., Li, X., Zhang, X., Wu, R., Zhou, Y., Zhao, Q., Zhang, Y., Geng, S., Gu, Y. and Hong, S. (2023). Cardiac murmur grading and risk analysis of cardiac diseases based on adaptable heterogeneous-modality multi-task learning. *Health Information Science and Systems*, 12(1), p.2.

Xu, C., Yi, K., Jiang, N., Li, X., Zhong, M. and Zhang, Y. (2023). MDFF-Net: A multi-dimensional feature fusion network for breast histopathology image classification. *Computers in Biology and Medicine*, 165, p.107385.

Yang, J., Cai, S., Qi, C., Xie, T. and Yan, H. (2023). Prediction of pulse wave of target organ damage in hypertension based on multiple feature fusion. *Biomedical Signal Processing and Control*, 85, p.105009.

Yevle, D.V. and Mann, P.S. (2025). Artificial intelligence based classification for waste management: A survey based on taxonomy, classification & future direction. *Computer Science Review*, 56, p.100723.

Zambrano Chaves, J.M., Wentland, A.L., Desai, A.D. et al. (2023). Opportunistic assessment of ischemic heart disease risk using abdominopelvic computed tomography and medical record data: a multimodal explainable artificial intelligence approach. *Scientific Reports*, 13, p.21034.

Zhang, G., Li, Y., Zhou, B., Liang, T., Zhang, Z. and Ju, R. (2024). 2MF-Net: 3D Cardiac Keypoint Detection with Multi-scale and Multi-dimension Feature Fusion Net. In: *2024 International Joint Conference on Neural Networks (IJCNN)*, 30 Jun. IEEE, pp.1–8.

Zhang, J., Liu, M., Xiong, P., Du, H., Zhang, H., Lin, F., Hou, Z. and Liu, X. (2021). A multi-dimensional association information analysis approach to automated detection and localization of myocardial infarction. *Engineering Applications of Artificial Intelligence*, 97, p.104092.