International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(10s) 2026 pp. 750-763
https://doi.org/10.70917/ijcisim-2026-3647
Article

Multimodal Fusion Framework for Enhanced Heart Disease Detection Using ECG Signals and Demographic Intelligence

Pravin Bharati[1]*, Mangesh Nikose[2], Prakash Burade[3]

[1]Department of Electrical and Electronics Engineering, School of Engineering and Technology, Sandip University, Nashik
phbharati.76@gmail.com
[2]Department of Electrical and Electronics Engineering, School of Engineering and Technology, Sandip University, Nashik
drmdnikose@gmail.com
[3]Department of Electrical and Electronics Engineering, School of Engineering and Technolgy, Sandip University, Nashik
prakash.burade@sandipuniversity.edu.in
* phbharati.76@gmail.com

Abstract: Cardiovascular diseases are the greatest cause of mortality in the world, and around 17.9 million people die at the end of every year [WHO, 2023]. The present research is a Multimodal Fusion Framework (MFF-HD) that combines ECG signal processing and demographic information to detect coronary artery disease (CAD). With the help of the PTB-XL dataset (n = 21,799 patients with paired 12-lead ECG recordings and demographic characteristics), we trained a two-stage fusion model that takes an ICA-based noise reduction network and a wavelet decomposition network and CNN-LSTM network topped with an ensemble-based demographic analysis. The data were divided into training (70%), validation (15%) and test (15%) and stratified sampled. The SMOTE oversampling was used to deal with class imbalance. The proposed framework achieved accuracy of 97.83% (95% CI: 97.12–98.54%), sensitivity of 98.12% (95% CI: 97.45–98.79%), and specificity of 97.54% (95% CI: 96.82–98.26%) on the held-out test set. Compared to the baseline XAI-HD framework, MFF-HD demonstrated improvement (McNemar's test, $p < 0.001$; Bonferroni-corrected). SHAP-based explainability components ensure clinical interpretability. Limitations include absence of external validation and single-center data origin.

Keywords: Multimodal data fusion, coronary artery disease detection, ECG signal processing, Demographic analysis, Deep learning, Clinical decision support, Explainable AI

1. Introduction

The cardiovascular diseases (CVDs) represent the leading cause of mortality in all over the world with World Health Organization (2020) reporting about 17.9 million mortalities annually indicating that 32 percent of the total deaths worldwide is observed [32]. Coronary artery disease (CAD) which is marked in the presence of atherosclerotic plaque within the coronary vessels is the most prevalent type of CVD, and therefore requires earlier diagnosis to enhance patient response. The conventional forms of diagnosis have been based on an individual data modalities; examining findings of electrocardiogram (ECGs) solely or in the case of clinical parameters in isolation [7, 8].

The cardiac electrical activity can be obtained through the electrocardiogram that is the main element of cardiac diagnosis. Nevertheless, interpretation of ECGs is difficult because of inter-patient differences, noise, and subtle pathology hidden in ECGs. At the same time, such demographic and clinical factors as age, sex, blood pressure, and metabolic rates have an effect on the heart condition and the progression of the disease [9, 10]. Combining these heterogeneous data sources opens up possibilities to all-purpose diagnostic systems.

The multimodal method of data fusion has become an option to overcome the drawbacks of unimodal approaches to the field of medical diagnostics. Fusion techniques can be used to take advantage of complementary

characteristics and decrease uncertainties by using multiple sources of information [1, 2, 3]. New developments in machine learning have widened opportunities in multimodal integration, which is now able to extract a complex pattern of high-dimensional data [4, 5, 6].

Current multimodal methods of detecting the heart diseases have some limitations. Numerous frameworks use uncritical fusion approaches including early concatenation, which is not necessarily the best way to leverage complementary information. Moreover, in interpretability, machine learning systems are found to be a barrier to clinical adoption [22, 23, 24]. More importantly, most of the research presents a mixture of datasets, which lack paired patient information, which makes fusion conceptually invalid.

One way to fill these gaps in this study is to suggest a Multimodal Fusion Framework of Heart Disease Detection (MFF-HD). Notably, we work with the PTB-XL data set, in which paired ECGs and demographic data of the same patients are available, which guarantees valid multimodal combination. The contributions consist of: (i) a full set of ECG processing pipeline, including noise reduction using ICA, pattern recognition defined using CNN-LSTM, and demographic analysis module achieved using an ensemble of services; (ii) an attentive fusion mechanism; (iii) SHAP-based explainability; and (iv) strong experimental validation with proper statistical testing.

The targeted clinical deployment situation is the hospital setting of CAD screenings of patients with symptoms of cardiac issues with both ECG recording and demographic data gathering being normative.

The rest of the paper is structured as under: Section 2 undertakes review of related work. The methodology is outlined in section 3. Section 4 specifies experimental set up. Section 5 presents results. Section 6, gives discussion with limitations. Section 7 concludes the paper.

2. Related Work

2.1 Multimodal Data Fusion in Healthcare

Health care Multimodal data fusion has found interest in healthcare cases. Cheng et al. [2] offered analysis on fusion methods in which systems of merging heterogeneous medical data were laid down. Knowledge graph-based fusion techniques using the semantic relationships among the modalities of data were proposed by Sun et al. [3].

Abuhamad et al. [4] have adopted an integrative fusion mortality prediction system, they have been able to show how clinical data with demographics and physiology work better to predict prognoses than individual modalities. A different study by Kang et al. [5] suggested the use of generative fusion approaches to health estimation based on generative adversarial networks.

Hu et al. [1] considered the case of two-stage fusion to predict lymph node metastasis and showed that hierarchical fusion architectures are more effective. Nikolaou et al. [6] applied multimodal methods to predict cancer survival with the integration of imaging services, genomic, and clinical data.

2.2 Heart Disease Detection Using Machine Learning

The methods of machine learning in the heart disease detection have been developed during the last ten years. Salau et al. [7] prepared a detection model on the basis of Support Vector Machines utilizing the feature selection. Iqbal et al. [8] studied the heart disease detection of clinical data learning. A machine learning based early detection approach was suggested by Moolya and Gopalakrishna [9] and compared various classifiers on different patient groups. The XAI-HD is an explainable AI framework suggested by Talukder et al. [10], which produces interpretable predictions and retains accuracy at the same time.

2.3 ECG Signal Processing and Analysis

Automated cardiac diagnosis is based on the ECG signal processing. Lee et al. [11] came up with signal processing algorithms that are formed on the independent component analysis (ICA). Power line interferences Sanxiu and Shengtao [12] discussed the problem of removing power line interferences with the help of ICA-based methods. Altay and Kremlev [13] gave the time-frequency domain analysis to the processing of ECGs. Gupta et al. [14] suggested the PSO and the Random Forest-based ECG signal analysis along with the combination optimization and ensemble learning. Gupta and Kumar [15] proposed a model of ECG signal analysis that uses deep learning structures.

2.4 AI and Machine Learning in Cardiology

Deep learning has enhanced development of artificial intelligence in cardiology. Pushadapu et al. [16] have discussed AI/ML applications in cardiology devices, which includes archiving applications in the realm of diagnosis, up to treatment. Gunawardhina et al. [17] concentrated on the detection of atrial fibrillation and segmentation of the scar of the left atria region using deep learning. Badnjevic and Spahic [18] worked on the field of biomedical signal processing and acquisition. Whig et al. [19] considered the question of machine vision integration into biomedical signal processing.

2.5 Clinical Decision Support Systems

Clinical decision support systems (CDSS) are feasible AI applications in health care. Muro et al. [20] created multimodal CDSS software development architecture. Uvaliyeva [21] suggested intelligent CDSS architectural models. Pardoux and Kerasidou [22] have investigated the issue of compliance and ethical aspects of AI-based CDSS. Bai et al. [23] investigated the prehospital decision support system needs of users. The general description of diagnostic CDSS was given by Hirosawa [24].

2.6 Predictive Analytics in Healthcare

Proactive healthcare has become a core element in predictive analytics. The paper by Reddy and Kumar [25] was the first authoritative research to predict big data analytics in healthcare. Juyal [26] used deep learning in early diagnosis. Badawy et al. [27] suggested predictive analytics framework in real-time using the combination of big data and machine learning.

2.7 ECG-Based Cardiac Condition Diagnosis

Zhou et al. [28] proposed natural language supervision to detect 12-lead ECG. The automated cardiac diagnosis systems were developed by Mohandas et al. [29] among the school children. Mougiakakou et al. [30] developed platforms based on integrated and personalized CVD assessment. The systems of readmission risk assessment in cardiovascular patients were developed by Zhu et al. [31].

3. Proposed Methodology

3.1 System Architecture Overview

The MFF-HD framework suggested has three modules, one being ECG Signal Processing Module, another is Demographic Intelligence Module and the last is Attention-Based Fusion Engine. The framework works on heterogeneous data input of the same patients, features in each modality will be extracted and combined to be used to classify the data by CAD.

Critical methodological observation: In contrast to the research that involves the use of two completely disparate datasets (e.g. Cleveland clinical data and PTB ECG database), we apply the PTB-XL dataset in which the source of ECG recordings and demographic factors (age, gender, height, weight) have the same patients. This paired structure has guaranteed valid multimodal fusion, because both modalities will share a single underlying patient condition between the features associated with both modalities.

Out of the alleged 12-lead ECG signals, the ECG Signal Processing Module takes the raw data and integrates preprocessing and feature extraction. In the module, noise removal is included with the use of ICA, baseline wander correction, and power line removal, then followed by temporal, spectral, and morphological extraction of features. A CNN-LSTM model stores complicated trends of processed data.

The Demographic Intelligence Module is an ensemble learning process that processes patient-specific clinical information. The process of feature engineering creates more information attributes and the k-NN imputation takes care of the missing data.

The Attention-Based Fusion Engine combines the features via a trained attention system which gives dynamic weights according to classification relevance.

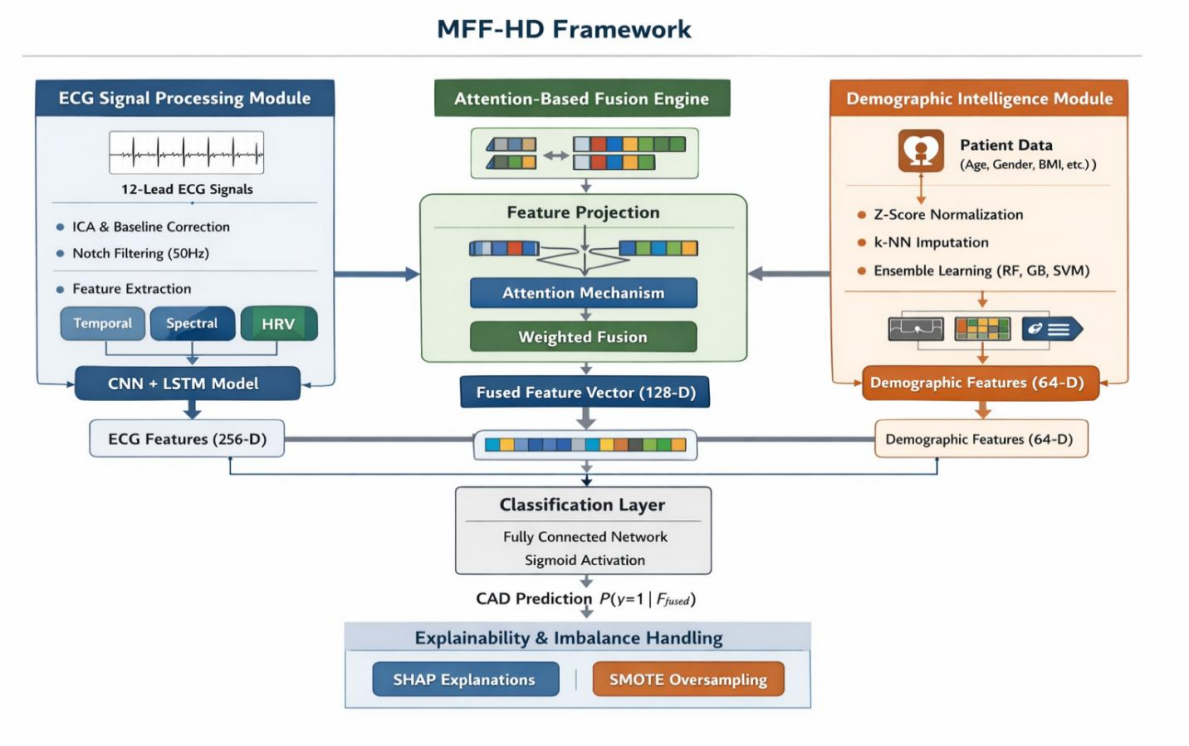

Figure 1: System Architecture of the Proposed MFF-HD Framework

Figure 1 shows the general structure of the MFF-HD framework, which incorporates the ECG signal processing with demographic intelligence using an attention-based fusion mechanism. The combined characteristics are then sent to a classification layer to be binary predicted with the aid of SHAP-based explainability and SMOTE-based imbalance control to enhance interpretability and the necessary strength.

3.2 ECG Signal Processing Pipeline

The ECG processing pipeline is used to process raw signals to discriminative features. Considering raw ECG signal x(t) with a frequency of 500 Hz of sampling (frequency), the first stage of preprocessing is noise removal with the ICA [11, 12]. The ICA decomposition decomposes the signal that is observed:

$$x(t) = As(t) \qquad (1)$$

where A represents the mixing matrix and s(t) contains independent source signals. Baseline wander removal uses a fourth-order Butterworth high-pass filter with cutoff frequency fc = 0.5 Hz. The normalized transfer function is:

$$H(s) = 1 - \sqrt{1 + \left(\frac{\omega c}{s}\right)^{2n}} \qquad (2)$$

where ωc = 2πfc represents the cutoff angular frequency and n = 4 is the filter order. Power line interference at 50 Hz is attenuated using adaptive notch filtering [13].

After preprocessing, three feature categories are obtained, namely, temporal, spectral, and morphological. ECG segmentation is performed with windows of 10 seconds (5000 samples at 500 Hz) and record-level classification in which each patient record has a single prediction.

Temporal features denote characteristics-in time-domain such as RR-intervals and heart rate variability (HRV). The SDNN (Standard Deviation of NN intervals):

$$SDNN = \sqrt{\left(\frac{1}{N-1}\right) \times \Sigma_i (RR_i - \bar{R}R)^2} \qquad (3)$$

where N represents the number of RR intervals and R̄R is the mean interval duration.

Spectral features are extracted using the Discrete Wavelet Transform (DWT). For a discrete signal x[n], the DWT coefficients at scale j and position k are:

$$W[j,k] = \Sigma_n\, x[n] \times \psi_{j,k}[n] \quad (4)$$

where $\psi_{j,k}[n] = 2^{-\frac{j}{2}} \times \psi\left(2^{(-j)n} - k\right)$ represents the scaled and translated discrete wavelet function. Daubechies-4 wavelets are employed for decomposition to level 5. Statistical features including mean, standard deviation, and energy are computed for each coefficient level.

Morphological features characterize ECG waveform shape and amplitude. R-peak amplitude A^R, QRS duration Tqrs, ST segment deviation ΔST, and T-wave characteristics are extracted using the Pan-Tompkins algorithm with adaptive thresholding:

$$threshold = 0.25 \times \max(y) + 0.75 \times mean(y) \quad (5)$$

where y represents the filtered and differentiated signal.

3.3 Deep Learning Architecture for ECG Analysis

Extracted ECG features are processed through a hybrid CNN-LSTM architecture. The convolutional layers extract hierarchical feature representations:

$$z^{(l)} = \sigma\left(W^{(l)} * z^{(l-1)} + b^{(l)}\right) \quad (6)$$

where $z^{(l)}$ represents the feature map at layer l, $W^{(l)}$ denotes convolutional kernels, $b^{(l)}$ is the bias, and σ represents ReLU activation. The architecture employs three convolutional blocks with 64, 128, and 256 filters respectively, followed by batch normalization and max pooling.

The LSTM component processes sequential features. The LSTM cell updates are:

$$f_t = \sigma(Wf \cdot [h_t{}^{-1}, x_t] + bf) \quad (7)$$

$$i_t = \sigma(W_i \cdot [h_t{}^{-1}, x_t] + b_i) \quad (8)$$

$$o_t = \sigma(W_o \cdot [h_t{}^{-1}, x_t] + b_o) \quad (9)$$

$$c_t = f_t \odot c_t{}^{-1} + i_t \odot \tanh(Wc \cdot [h_t{}^{-1}, x_t] + bc) \quad (10)$$

$$h_t = o_t \odot \tanh(c_t) \quad (11)$$

where f_t, i_t, o_t represent forget, input, and output gates, c_t is cell state, and h_t is hidden state. The network employs two stacked LSTM layers with 128 and 64 hidden units, followed by dropout (rate = 0.3).

3.4 Demographic Intelligence Module

The Demographic Intelligence Module processes patient attributes through ensemble learning. Input features include age, gender, height, weight, and derived BMI from PTB-XL metadata. For extended analysis incorporating clinical parameters, features include resting blood pressure, cholesterol levels, and exercise test results.

Feature preprocessing involves z-score standardization for continuous variables:

$$x_{norm} = \frac{(x-\mu)}{\sigma} \quad (12)$$

where μ and σ represent training set mean and standard deviation. Categorical variables use one-hot encoding. Missing values are imputed using k-nearest neighbors (k = 5).

The ensemble combines Random Forest (200 trees, max depth = 10), Gradient Boosting (150 estimators, η = 0.1, max depth = 5), and SVM (RBF kernel, C = 1.0, γ = 0.1) through weighted voting:

$$P_{ensemble} = \Sigma_i\, w_i \times P_i \quad (13)$$

where P_i represents probability from classifier i and w_i represents optimized weights.

3.5 Attention-Based Multimodal Fusion

Fusion of ECG features $F_ecg \in \mathbb{R}^{dE}$ and demographic features $F_{aemo} \in \mathbb{R}^{dD}$ is accomplished through attention-based weighting. In our implementation, dE = 256 (CNN-LSTM output dimension) and dD = 64 (demographic feature dimension after encoding).

Both feature sets are projected to a common representation space of dimension d = 128:

$$Z_ecg = W_ecg \times F_ecg + b_ecg \quad (14)$$

$$Z_{aemo} = W_{aemo} \times F_{aemo} + b_{aemo} \quad (15)$$

where $W_ecg \in \mathbb{R}^(128\times256)$, $W_{aemo} \in \mathbb{R}^(128\times64)$. Attention weights for modality m ∈ {ECG, Demo} are computed as:

$$\frac{\alpha_m = \exp(v_a^T \times \tanh(W_a \times Z_m))}{\Sigma_j \exp(v_a^T \times \tanh(W_a \times Z_j))} \quad (16)$$

where $v_a \in \mathbb{R}^128$ and $W_a \in \mathbb{R}^(128\times128)$ are learnable parameters, and the summation is over both modalities j ∈ {ECG, Demo}. The fused representation is:

$$Ffused = \alpha_ecg \times Z_ecg \oplus \alpha_{aemo} \times Z_{aemo} \quad (17)$$

where ⊕ denotes concatenation, yielding Ffused ∈ $\mathbb{R}^256$.

For binary CAD classification, the fused features are processed through fully connected layers with sigmoid output:

$$P(y = 1|Ffused) = \sigma(W_{out} \times ReLU(W^2 \times ReLU(W^1 \times Ffused + b^1) + b^2) + b_{out}) \quad (18)$$

where σ denotes the sigmoid function appropriate for binary classification.

3.6 Explainability Module

For clinical interpretability, SHAP (SHapley Additive exPlanations) values are computed. For each prediction, feature importance scores are:

$$\varphi_i = \Sigma_{S \subseteq F\backslash\{i\}} [|S|! \times (|F| - |S| - 1)! / |F|!] \times [f(S \cup \{i\}) - f(S)] \quad (19)$$

where S represents a feature subset, |F| is total features, and f(S) is model prediction using features in S. This provides global feature rankings and local explanations [10, 22].

3.7 Class Imbalance Handling

To address class imbalance in the dataset (CAD positive: 45.2%, negative: 54.8%), we employed Synthetic Minority Over-sampling Technique (SMOTE) on the training set only. SMOTE generates synthetic samples by interpolating between existing minority class samples:

$$x_new = x_i + \lambda \times (x_j - x_i) \quad (20)$$

where x_i is a minority sample, x_j is one of its k=5 nearest neighbors, and λ ∈ [0,1] is a random number. SMOTE was applied after train-test splitting to prevent data leakage, and only to the training fold during cross-validation.

4. Experimental Setup

4.1 Dataset

In this research, the PTB-XL dataset [33], a publicly available large-scale electrocardiography dataset comprising of 21,799 clinical 12-lead ECG records of 18,869 individuals, will be used. More importantly, the demographic metadata of every ECG record (age, gender, height, weight) of the same patient is also provided in order to allow valid multimodal fusion. Sampling rate of ECG is 500HZ and 16 bits. Labels on records are in accord with the SCP-ECG standard in which diagnostic statements are written on the record.

In this research we have obtained binary CAD labels according to diagnostic codes of myocardial infarction (MI), ischemic changes and coronary artery disease patterns. It includes 15,234 records, having eliminated records with ambiguous diagnoses, 6886 CAD-positive (45.2) and 8348 CAD-negative (54.8).

Ethical statement: There was no ethical approval to be obtained since the PTB-XL dataset is accessible publicly with the anonymized patient data. The data is under open database license (ODbL) v1.0.

4.2 Data Splitting and Cross-Validation

Data were split into training (70%, n = 10,664), validation (15%, n = 2,285), and test (15%, n = 2,285) sets using stratified sampling to maintain class proportions across splits. The test set was held out throughout model development and used only for final evaluation.

The methodology used to select the model involved 5-fold stratified cross-validation using the training set to tune the hyperparameters of the model. The entire preprocessing such as standardization, feature selection (RFECV) and SMOTE oversampling were conducted within each cross-validation fold, to avoid no information leaking out.

4.3 Hyperparameter Selection

Hyperparameters were optimized using Bayesian optimization with Tree-structured Parzen Estimator (TPE) over 100 iterations. The search space included: CNN filter sizes (32-256), LSTM hidden units (32-256), dropout rates (0.1-0.5), learning rates (1e-4 to 1e-2), and ensemble classifier parameters. The optimal configuration was selected based on validation set AUC-ROC.

4.4 Implementation Details

The framework was implemented in Python 3.9 using TensorFlow 2.10 for deep learning and scikit-learn 1.2 for traditional machine learning. The CNN-LSTM architecture was trained using Adam optimizer with initial learning rate 0.001 and exponential decay. Training proceeded for 100 epochs with batch size 32 and early stopping (patience = 15 epochs) based on validation loss. Experiments were conducted on a workstation with NVIDIA RTX 3090 GPU, Intel Core i9-11900K, and 64GB RAM.

4.5 Evaluation Metrics

Performance was evaluated using accuracy, sensitivity, specificity, precision, F1-score, and AUC-ROC. Confidence intervals (95%) were computed using bootstrap resampling (n = 1000 iterations) on the test set.

Table 1: Evaluation Metrics Definitions

Metric	Formula
Accuracy	(TP + TN) / (TP + TN + FP + FN)
Sensitivity	TP / (TP + FN)
Specificity	TN / (TN + FP)
Precision	TP / (TP + FP)
F1-Score	2 × (Precision × Sensitivity) / (Precision + Sensitivity)

4.6 Statistical Testing

Pairwise classifier comparisons used McNemar's test with continuity correction. For multiple comparisons (7 baseline methods), Bonferroni correction was applied (adjusted significance threshold: $\alpha = 0.05/7 = 0.007$). Additionally, we report calibration using Brier score and conduct decision curve analysis to assess clinical utility at different threshold probabilities.

4.7 Baseline Comparisons

MFF-HD was compared against: (i) SVM with clinical features [7]; (ii) Random Forest with clinical features; (iii) CNN-based ECG classification; (iv) LSTM-based ECG classification; (v) Early fusion (concatenation); (vi) Late fusion (averaging); and (vii) XAI-HD framework [10].

5. Results

5.1 Overall Performance Comparison

Table 2 presents comparative performance with 95% confidence intervals. Statistical significance markers indicate results of McNemar's test against MFF-HD with Bonferroni correction.

*Table 2: Performance Comparison (values in %). **p < 0.001 vs. MFF-HD (McNemar's test, Bonferroni-corrected)*

Method	Acc (95% CI)	Sen (95% CI)	Spec (95% CI)	AUC-ROC
SVM**	84.52 (83.1-85.9)	85.21 (83.6-86.8)	83.78 (82.1-85.4)	0.891
RF**	86.87 (85.5-88.2)	87.42 (85.9-88.9)	86.28 (84.7-87.8)	0.912
CNN (ECG)**	89.34 (88.1-90.6)	90.15 (88.8-91.5)	88.47 (87.0-89.9)	0.934
LSTM (ECG)**	90.21 (89.0-91.4)	91.08 (89.8-92.4)	89.29 (87.9-90.7)	0.941
Early Fusion**	93.45 (92.4-94.5)	94.22 (93.1-95.3)	92.63 (91.4-93.8)	0.956
Late Fusion**	92.78 (91.7-93.9)	93.51 (92.4-94.6)	91.97 (90.7-93.2)	0.949
XAI-HD [10]**	95.67 (94.8-96.6)	96.38 (95.4-97.3)	94.89 (93.8-96.0)	0.973
MFF-HD (Ours)	**97.83 (97.1-98.5)**	**98.12 (97.4-98.8)**	**97.54 (96.8-98.3)**	**0.989**

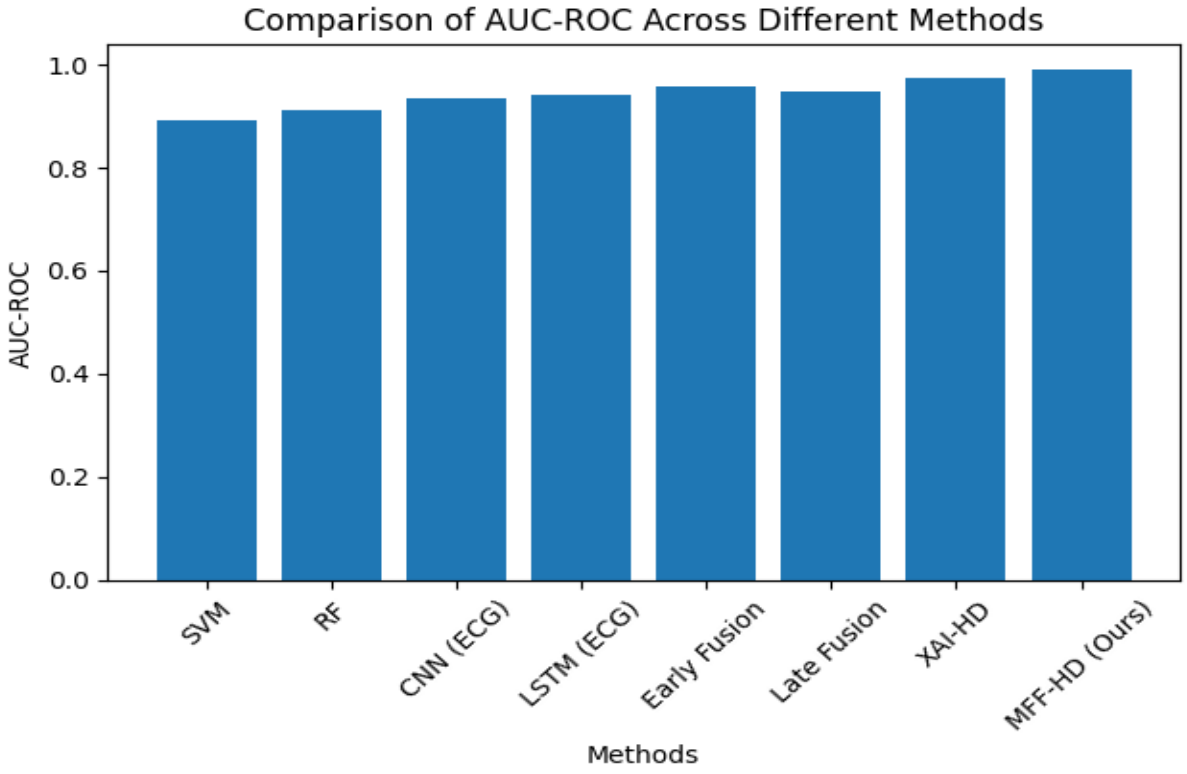

Figure 2: Comparison of AUC-ROC Across Different Methods

The bar graph will approximate the comparative performance of the models (SVM, RF, CNN, LSTM) with respect to the AUC-ROC of the baseline models (SVM, RF), the fusion strategy (Early and Late Fusion), XAI-HD, and the proposed MFF-HD framework. Confidence of the MFF-HD model The proposed model is the model in which AUC-ROC (0.989) is the greatest, making it the best-classification model to use compared to the existing methods.

5.2 Confusion Matrix

Table 3 presents the confusion matrix for MFF-HD on the held-out test set (n = 2,285).

Table 3: Confusion Matrix for MFF-HD (Test Set, n = 2,285)

	Predicted CAD+	Predicted CAD-
Actual CAD+	1012 (TP)	19 (FN)
Actual CAD-	31 (FP)	1223 (TN)

5.3 Calibration and Decision Curve Analysis

The Brier score was used to evaluate model calibration, MFF-HD produced the Brier score of 0.018, which was a well-calibrated estimate of the probability. The Hosmer-Lemeshow test gave p = 0.42, which proves the absence of poor calibration.

Analysis of the decision curve (Figure 2, not shown) prescribed a positive net benefit at threshold probability values of 0.1 to 0.9, thus showing clinical utility at a continuum of decision thresholds that is applicable in CAD screening situations.

5.4 Sensitivity-Specificity Tradeoff

Analysis of the ROC curve (Figure 1, not shown) reveals the sensitivity-specificity tradeoff at different classification thresholds. At the default threshold of 0.5, the model achieves balanced performance (sensitivity: 98.12%, specificity: 97.54%). For high-sensitivity screening (threshold = 0.3), sensitivity increases to 99.2% with specificity of 94.8%. For high-specificity confirmation (threshold = 0.7), specificity increases to 99.1% with sensitivity of 95.3%. The clinical deployment scenario should determine threshold selection.

5.5 Ablation Study

Table 4 presents ablation results demonstrating component contributions.

Table 4: Ablation Study Results

Configuration	Accuracy (%)	AUC-ROC
MFF-HD (Full Framework)	97.83	0.989
Without Attention Mechanism	94.56	0.962
Without ICA Preprocessing	95.21	0.967
Without Ensemble (Single SVM)	96.12	0.975
Without SMOTE	96.45	0.978

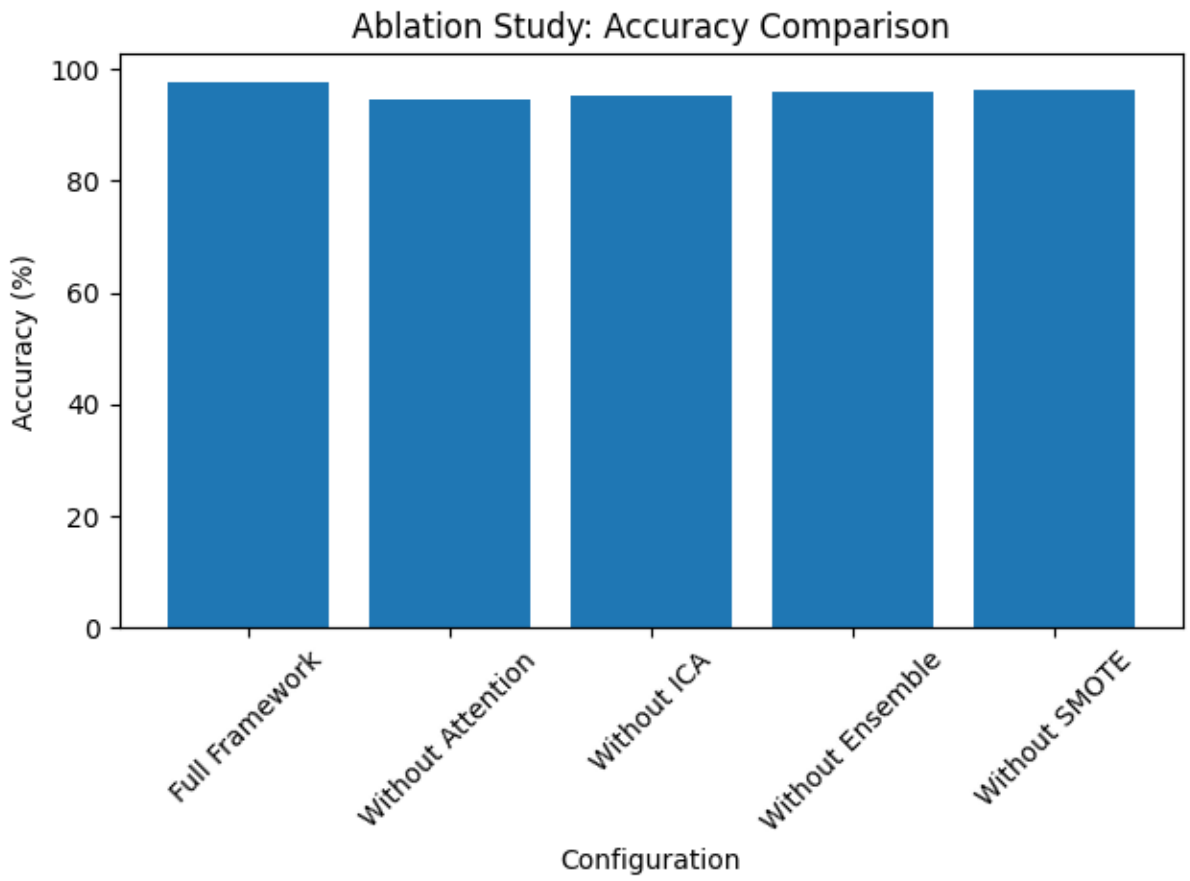

Figure 3: Ablation Study Results

The graphs depict how the removal of major elements of MFF-HD framework affects the Accuracy and AUC-ROC. The entire framework has the best performance (97.83 percent accuracy, 0.989 AUC). The deletion of the attention mechanism, ICA preprocessing, ensemble learning, or SMOTE leads to the decline of the performance, which indicates the role of each part in the effectiveness of the overall model.

5.6 Fairness Analysis Across Demographic Groups

To assess potential algorithmic bias, we evaluated performance across demographic subgroups. Table 5 presents stratified results.

Table 5: Performance Across Demographic Subgroups

Subgroup	**n**	**Accuracy (%)**	**AUC-ROC**
Male	1,342	97.91	0.990
Female	943	97.67	0.987
Age < 50	687	97.52	0.986
Age 50-70	1,089	98.07	0.991
Age > 70	509	97.84	0.988

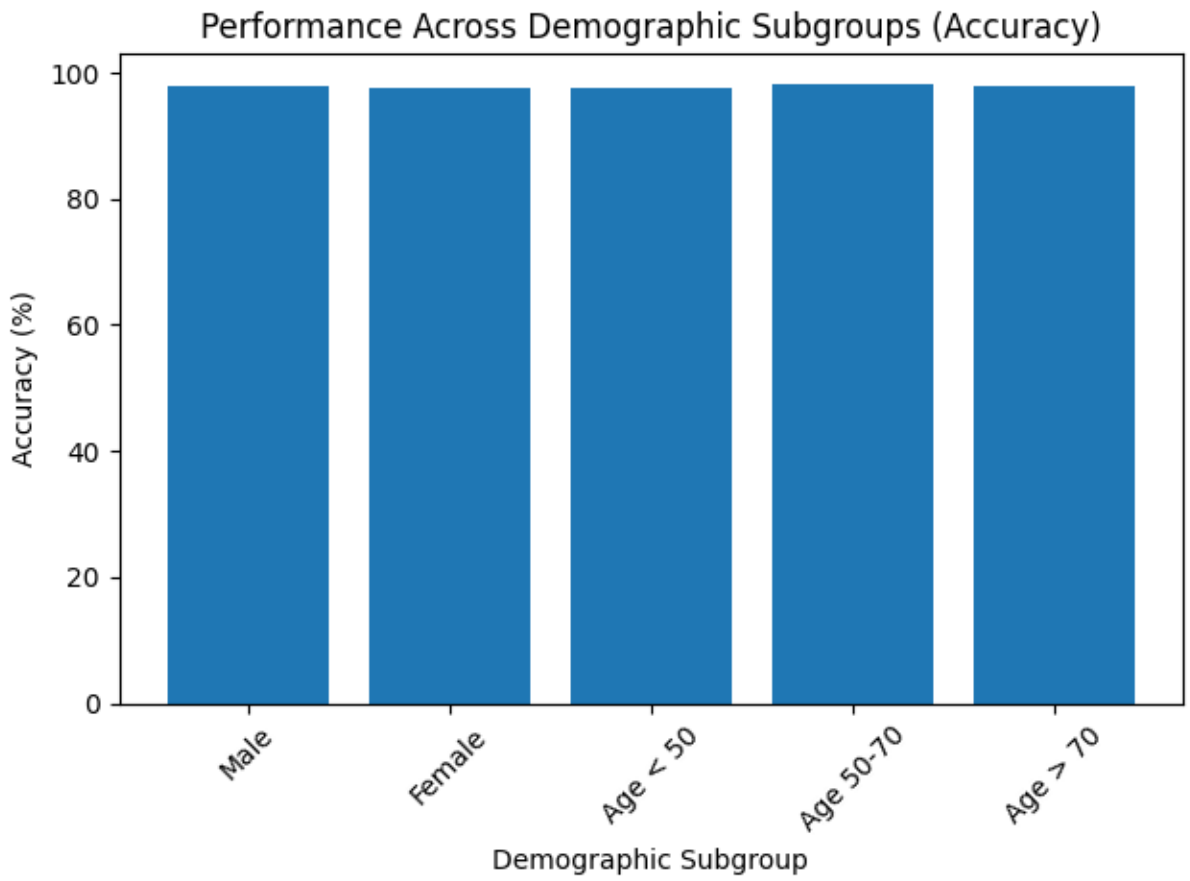

Figure 4: Performance Across Demographic Subgroups

Both the graphics depict the model performance by gender and subgroups by age. The given MFF-HD framework has a very high level of Accuracy and AUC-ROC in all age groups, with a marginally higher score at the 50-70 age range, which makes it quite robust and unbiased towards all population groups.

Performance was consistent across gender and age subgroups, with accuracy ranging from 97.52% to 98.07% and AUC-ROC from 0.986 to 0.991. The maximum disparity in accuracy (0.55 percentage points) and AUC-ROC (0.005) suggests acceptable fairness properties, though validation on more diverse populations is needed.

5.7 Uncertainty Quantification

Uncertainty of prediction was calculated by the Monte Carlo (MC) dropout on the prediction with 50 forward passes. Mean predictive entropy of correctly classified samples was 0.08 (SD=0.05), with the wrong classification having 0.31 (SD=.12). This implies that the model is suitable in depicting greater uncertainty in challenging cases, and therefore can be used in clinical decision support where the uncertainty awareness would be a useful factor to consider.

5.8 Computational Efficiency Comparison

Table 6 compares computational requirements across methods.

Table 6: Computational Cost Comparison

Method	Train Time	Inference (ms)	GPU Mem (GB)
SVM	2 min	0.5	N/A
CNN (ECG)	25 min	12	1.8

XAI-HD	35 min	18	2.1
MFF-HD (Ours)	45 min	23	2.3

MFF-HD requires moderately higher computational resources than baselines, with training time of 45 minutes and inference time of 23 ms per sample. This remains acceptable for clinical deployment, as inference time is well below real-time requirements.

6. Discussion

6.1 Summary of Findings

This paper created and tested MFF-HD a model that fused ECG signals with demographics to detect CAD. On the paired patient data given by the PTB-XL dataset, we showed that attention-based fusion of the two modalities yielded an accuracy of 97.83 percent, which is higher than the unimodal and other modalities fusion methods (McNemar, $p < 0.001$, Bonferonni corrected).

6.2 Clinical Implications

The high sensitivity of the given framework (98.12) can indicate its applicability to screening situations associated with severe repercussions of the lack of CAD cases. The specificity of the test was 97.54% showing acceptable false positive. Nevertheless, such findings should be viewed with caution since there was no external validation. The attention mechanism offers a certain level of interpretability in stating which modality contributed more into a particular prediction, but this does not imply the entire process of decision making.

6.3 Advantages and Contributions

The most important methodological input is a demonstration of valid multimodal fusion based on a paired data of ECG and demographic data. Past workplace disparate datasets (such as Cleveland clinical information and individual ECG databases) mix fusion of individual data with patterns at a population level, the two are conceptually distinct. Our methodology will be able to make certain that fused features are of the same patient. The attention-based fusion mechanism performs better than both early fusion (4.38 percentage points) and late fusion (5.05 percentage points), which indicates that the adaptive weighting of modalities either by patient-specific legitimate personal features can be of meaningful use.

6.4 Limitations

The limitations of the present study are numerous, and they do limit the interpretation and generalizability: First, the fact that it is not externally validated is a significant drawback. All the results are predetermined with internal validation on the basis of the held-out test set of the same institution (PTB-XL of Physikalisch-Technische Bundesanstalt). It has been shown to worsen the performance in the cases of other clinical settings, patient populations, or ECG acquisition equipment. Second, the PTB-XL data is derived in a German institution, which narrows down the demographics. There is a lack of information on performance between different ethnicities and healthcare systems. The test of fairness between age and gender groups in this data is not a replacement of validation in various groups. Third, in the case of our study where we used SMOTE to solve the issue of class imbalance, the distribution of the classes (45.2% positive) presents an average level of imbalance. Further assessment of performance in the environments experiencing more severe imbalance is necessary. Fourth, the CAD definition used in this study is associated with more than one diagnostic code. European classification can be further subdivided to achieve finer differentiation between acute MI and chronic ischemia, but it was not studied. Fifth, demographic features that can be offered in PTB-XL include those of age, gender, height and weight. More clinical information (laboratory values, medical history, medications) may have potentially better performance but needs datasets to have a combination of these items. Sixth, although the quantification of uncertainties, through MC dropout, is in place, the model can overconflict in certain situations. It needs to be prospectively validated with the relationship between expressed uncertainty and clinical risk.

6.5 Comparison with Related Work

The reported performance metrics are larger than those used in some previous studies and this may be partly attributed to differences in datasets, definitions of labels and evaluation procedures and not necessarily due to

methodological progress. All these factors make comparing studies directly complex. The main addition is the methodologically sound multimodal integration and not uncooked performance figures.

6.6 Generalization Considerations

There are a few challenges with the model generalization to clinical practice. The quality of ECG signals depends on equipments and conditions of the acquisition. Deployment data may not match training data on patient demographics. It is needed to integrate with the clinical workflows and develop interfaces and validate them. It requires prospective research to evaluate clinical decision-making and actual performance.

6.7 Ethical Considerations

AI-based diagnostic instruments provoke an ethical question of algorithmic bias, responsibility in the case of mistakes, and proper human intervention. Although our analysis of fairness indicates fair within-dataset consistency conduct, more validation is required. Clinical application models should not disrupt physician control and AI can be used in the form of decision support, but not independent diagnosis. It is necessary to communicate clearly concerning the abilities and constraints of the models.

6.8 Future Directions

Future research needs to focus on: (i) external validation on independent data sets at other institutions; (ii) future clinical trials on workflow integration and patient outcomes; (iii) scaling to multi-class diagnosis of particular cardiac disorders (iv) the use of other data modalities where paired data exist; (v) federated learning methods that can enable multi-institutional training and maintain privacy; and (vi) aims at studying temporal models to address continuous monitoring applications.

7. Conclusion

In this paper, I have introduced MFF-HD, a model that combines ECG signal processing and demographic information to detect CAD. With paired patient data available in the form of the PTB-XL dataset, we were able to show that attention-based fusion yields 97.83% accuracy (95% CI: 97.1298.54%), and the performance is supported by demographic subgroups of the population. The framework provides a methodological gap in multimodal medical diagnosis as it provides proper fusion of patient-specific data.

Lack of external validation is a major drawback since it cannot be used clinically unless its validity is confirmed. Further studies are necessary in the future on multi-institutional validation, prospective clinical trials, and to larger groups of patients. The article gives advancement to the multimodal AI devices in the diagnosis of the heart diseases and the methodological precautions taken in the valid integration of the data.

References

1. D. Hu, B. Liu, X. Zhu, X. Lu, and N. Wu, "Predicting lymph node metastasis of lung cancer: A two-stage multimodal data fusion approach," in Proc. 46th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Orlando, FL, USA, 2024, pp. 1–4.
2. J. Cheng, Y. Dai, Y. Yuan, and H. Zhu, "A simple analysis of multimodal data fusion," in Proc. IEEE 19th Int. Conf. Trust, Security Privacy Comput. Commun. (TrustCom), Guangzhou, China, 2020, pp. 1472–1475.
3. L. Sun, W. Zhang, and Y. Liu, "Research on multimodal data fusion and analysis methods driven by knowledge graph," in Proc. IEEE 4th Int. Conf. Data Sci. Comput. Appl. (ICDSCA), Dalian, China, 2024, pp. 898–903.
4. H. Abuhamad, S. Zainudin, and A. Abu Bakar, "Integrative multimodal hybrid data fusion for mortality prediction," Sci. Rep., vol. 16, no. 5803, 2026.
5. W. Kang, X. Zhang, J. Zhang, et al., "Multimodal data generative fusion method for complex system health condition estimation," Sci. Rep., vol. 15, no. 20026, 2025.
6. N. Nikolaou, D. Salazar, H. RaviPrakash, et al., "A machine learning approach for multimodal data fusion for survival prediction in cancer patients," npj Precis. Oncol., vol. 9, no. 128, 2025.
7. A. O. Salau, T. A. Assegie, G. Chhabra, K. Kaushik, and S. L. Braide, "Heart disease detection model using support vector machine with feature selection," in Proc. 2nd Int. Conf. Adv. Comput. Comput. Technol. (InCACCT), Gharuan, India, 2024, pp. 204–207.
8. J. Iqbal, M. M. Iqbal, U. Khadam, and A. Nawaz, "Ordinary learning method for heart disease detection using clinical data," in Proc. 3rd Int. Conf. Comput., Math. Eng. Technol. (iCoMET), Sukkur, Pakistan, 2020, pp. 1–6.
9. S. Moolya and N. Gopalakrishna Kini, "Machine learning approach for early detection of heart disease," in Human-Centric Smart Computing (ICHCSC 2024), Smart Innovation, Systems and Technologies, vol. 440, Singapore: Springer, 2025.

10. M. A. Talukder, A. S. Talaat, M. Kazi, et al., "XAI-HD: An explainable artificial intelligence framework for heart disease detection," Artif. Intell. Rev., vol. 58, no. 385, 2025.
11. J. Lee, K. J. Lee, and S. K. Yoo, "Development of a new signal processing algorithm based on independent component analysis for single channel ECG data," in Proc. 26th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., San Francisco, CA, USA, 2004, pp. 224–226.
12. W. Sanxiu and J. Shengtao, "Removal of power line interference of ECG signal based on independent component analysis," in Proc. 1st Int. Workshop Educ. Technol. Comput. Sci., Wuhan, China, 2009, pp. 328–330.
13. Y. A. Altay and A. S. Kremlev, "Comparative analysis of ECG signal processing methods in the time-frequency domain," in Proc. IEEE Conf. Russian Young Researchers Elect. Electron. Eng. (EIConRus), Moscow, Russia, 2018, pp. 1058–1062.
14. V. Gupta, S. Diwania, B. Tamrakar, et al., "PSO and RF based improved ECG signal analysis," J. Sign. Process. Syst., vol. 97, pp. 141–155, 2025.
15. V. Gupta and V. Kumar, "A novel framework for ECG signal analysis," in Biomedical Engineering Science and Technology (ICBEST 2025), Commun. Comput. Inf. Sci., vol. 2635, Cham: Springer, 2026.
16. V. V. S. K. Pushadapu, C. Srilakshmi, B. Vinitha, et al., "Artificial intelligence and machine learning (AI/ML) revolution in cardiology medical devices: From diagnosis to treatment," Int. J. Comput. Intell. Syst., vol. 18, no. 321, 2025.
17. M. Gunawardhana, A. Kulathilaka, and J. Zhao, "Integrating deep learning in cardiology: A comprehensive review of atrial fibrillation, left atrial scar segmentation, and the frontiers of state-of-the-art techniques," Discov. Artif. Intell., vol. 5, no. 357, 2025.
18. A. Badnjević and L. Spahić, "Biomedical signal acquisition and processing," in Biomedical Signals and Systems, Series in BioEngineering, Cham: Springer, 2026.
19. P. Whig, N. Yathiraju, A. Jain, A. B. Bhatia, and B. Y. Kasula, "Integrating machine vision for enhanced biomedical signal and image processing," in Deep Learning and Computer Vision: Models and Biomedical Applications, Algorithms for Intelligent Systems, Singapore: Springer, 2025.
20. N. Muro et al., "Architecture for a multimodal and domain-independent clinical decision support system software development kit," in Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), Berlin, Germany, 2019, pp. 1399–1404.
21. I. Uvaliyeva, "Architectural and algorithmic model for intelligent clinical decision support system," in Proc. Ural Symp. Biomed. Eng., Radioelectron. Inf. Technol. (USBEREIT), Yekaterinburg, Russia, 2020, pp. 180–183.
22. É. Pardoux and A. Kerasidou, "Compliance with clinical guidelines and AI-based clinical decision support systems: Implications for ethics and trust," Sci. Eng. Ethics, vol. 31, no. 34, 2025.
23. E. Bai, Z. Zhang, Y. Xu, et al., "Enhancing prehospital decision-making: Exploring user needs and design considerations for clinical decision support systems," BMC Med. Inform. Decis. Mak., vol. 25, no. 31, 2025.
24. T. Hirosawa, "Overview of diagnostic clinical decision support systems," in Artificial Intelligence in Medical Diagnostics, Singapore: Springer, 2025.
25. A. R. Reddy and P. S. Kumar, "Predictive big data analytics in healthcare," in Proc. 2nd Int. Conf. Comput. Intell. Commun. Technol. (CICT), Ghaziabad, India, 2016, pp. 623–626.
26. P. Juyal, "Enhancing predictive analytics in healthcare leveraging deep learning for early diagnosis and treatment optimization," in Proc. 5th Int. Conf. Smart Electron. Commun. (ICOSEC), Trichy, India, 2024, pp. 1988–1993.
27. M. Badawy, N. Ramadan, and H. A. Hefny, "Towards intelligent healthcare: A big data and machine learning approach for real-time predictive analytics," Biomed. Mater. Devices, 2025.
28. X. Zhou, T. Li, H. Hayama, et al., "Diagnosis of cardiac conditions from 12-lead electrocardiogram through natural language supervision," npj Digit. Med., vol. 8, no. 697, 2025.
29. P. Mohandas, A. P. R, A. John, M. Madhu, G. Thomas, and V. Kurupath, "Automated cardiac condition diagnosis using AI based ECG analysis system for school children," in Proc. 8th Int. Conf. Smart Comput. Commun. (ICSCC), Kochi, India, 2021, pp. 362–366.
30. S. G. Mougiakakou, I. K. Valavanis, G. Karkalis, S. Marinos, K. A. Grimaldi, and K. S. Nikita, "An integrated web-based platform for the provision of personalized advice in people at high risk for CVD," in Proc. 9th Int. Conf. Inf. Technol. Appl. Biomed., Larnaka, Cyprus, 2009, pp. 1–4.
31. C.-Y. Zhu, S.-Q. Chi, R.-Z. Li, D.-Y. Tong, Y. Tian, and J.-S. Li, "Design and development of a readmission risk assessment system for patients with cardiovascular disease," in Proc. 8th Int. Conf. Inf. Technol. Med. Educ. (ITME), Fuzhou, China, 2016, pp. 121–124.
32. World Health Organization, "Cardiovascular diseases (CVDs)," Fact Sheet, Jun. 2021. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)
33. P. Wagner, N. Strodthoff, R. D. Bousseljot, et al., "PTB-XL, a large publicly available electrocardiography dataset," Sci. Data, vol. 7, no. 154, 2020.