Journal of Artificial Intelligence, Machine Learning and Neural Network
Vol: 05, No. 02, July-Dec 2025
ISSN: 2799-1172, DOI: 10.55529/jaimlnn.52.153.164

Research Paper

Intelligent multi-modal cardiovascular disease detection framework using hybrid deep learning and explainable artificial intelligence: a clinical decision support perspective

Prof. Tareq N. Hashem*
*Professor, Department of Marketing, Faculty of Business, Applied Science Private University, Amman, Jordan.

Article Info

Article History:
Received: 23 August 2025
Revised: 03 November 2025
Accepted: 10 November 2025
Published: 26 December 2025

Keywords:
Cardiovascular Disease Detection
Deep Learning
CNN-BiLSTM
Explainable AI (XAI)
Multi-Modal Fusion
Clinical Decision Support

ABSTRACT

Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide, accounting for nearly 17.9 million deaths annually. Early and accurate diagnosis is essential for reducing morbidity and enabling timely therapeutic intervention. Although significant advancements have been achieved in clinical diagnostic technologies, conventional machine learning approaches still face challenges related to representation learning, heterogeneous multi-modal data integration, and clinical interpretability. To address these limitations, this study proposes an Intelligent Multi-Modal Clinical Decision Support System (IM-CDSS) based on a hybrid architecture integrating Convolutional Neural Networks (CNN), Bidirectional Long Short-Term Memory (BiLSTM), and Bahdanau attention mechanisms for comprehensive analysis of electrocardiogram (ECG) signals, echocardiographic images, and clinical biomarkers. The proposed framework was trained and evaluated using 12,847 de-identified patient records compiled from publicly available datasets, including the MIT-BIH Arrhythmia Database, PhysioNet MIMIC-IV, and Cleveland Heart Disease Dataset. Model interpretability was enhanced using SHapley Additive exPlanations (SHAP) and Gradient-weighted Class Activation Mapping (Grad-CAM), while Bayesian optimization with Optuna was employed for hyperparameter tuning. Five-fold cross-validation ensured robust evaluation. Experimental results demonstrate that IM-CDSS achieved 97.6% accuracy, a macro F1-score of 97.2%, and an AUC of 0.986 across five cardiovascular disease categories. Ablation studies confirmed the effectiveness of individual architectural components, while SHAP analysis identified Heart Rate Variability, QRS Duration, and ST Elevation as the most influential diagnostic features. The proposed framework offers improved accuracy, interpretability, and computational efficiency, supporting its potential deployment in resource-constrained clinical environments.

Corresponding Author:
Prof. Tareq N. Hashem
Professor, Department of Marketing, Faculty of Business, Applied Science Private University, Amman, Jordan.
Email: t_hashim@asu.edu.jo

1. INTRODUCTION

Globally, the total death and DALYs burden for cardiovascular diseases (CVDs) (USD>1 trillion per year) [1] is by far the greatest contributor to socioeconomic burden in the world. According to the World Health Organization (WHO), four-fifths of all the CVD-related deaths are linked to heart attacks and strokes, many of which are preventable if clinical intervention is taken in the right and timely manner during the early phase of the disease [2]. However, standard workflows for diagnosis are still highly dependent on individual physician skills, which may lead to inter-observer variation, diagnostic fatigue and unequal access to specialists, especially in LMICs [3].

With the introduction of electronic health records (EHRs), wearable biosensors, and high-resolution cardiac imaging, greater amounts of multi-modal patient data have become available, and there is a need and opportunity for clinical decision support systems (CDSSs) envisioned by artificial intelligence (AI) that can understand and utilize multi-modal signals down to human diagnostic precision level [4]. Structured clinical data such as electrocardiography (ECG), echocardiography and cardiac magnetic resonance imaging (CMR) have widely been used to identify the diagnostic potential of machine learning (ML) and deep learning (DL) methods with outstanding results [5], [6].

Even with these improvements, there are still some basic challenges that limit the realization of current AI-driven CVD detection systems in the clinical setting. Most of the models proposed use only unimodal input, usually, ECG signal or image, without employing the other diagnostic information available in a multimodal dataset [7]. Secondly, most architectures focus on accuracy over interpretability, yielding "black-box" predictions that are not actionable in the clinical setting [8]. Third, current systems do not have in-depth testing that is repeated across diverse settings in terms of age demographics, geographic location, which poses concerns to generalise their findings and bias [9]. Fourth, although in many countries, there are occasional successes indeed, Computational complexity is a continuous practical challenge in primary care and LMIC settings [10].

To overcome these limitations, this paper presents a novel AI framework called the Intelligent Multi-Modal Clinical Decision Support System (IM-CDSS) that combines an innovative Hybrid CNN-BiLSTM model with a Bahdanau attention mechanism, multi-modal data fusion method, and an XAI module featuring SHAP and Grad-CAM. The system enables the simultaneous classification of five clinically relevant categories of CVD including explanations at the feature level that are accessible to non-AI-specialist cardiologists.

The following are key strengths of this work: (1) designing and implementing a hybrid CNN-BiLSTM-Attention architecture that leveraged the spatial and temporal phenomena of multi-modal cardiac data, including depth, which is also embedded in the image modality, in addition to the clinically important approach of using diagnostic-saliency (Attention) aspects; (2) introducing a principled multi-modal feature fusion strategy that fused ECG time-series features, echocardiographic image features and structured biomarker into a single embedding using a learned cross-modal attention gate; (3) embedding the key concepts of SHAP global feature importance and Grad-CAM spatial saliency within a cohesive XAI pipeline for an interpretable clinician; (4) a thorough ablation study as well as statistical benchmarks against seven

Z人受伤s using McNemar's test and DeLong's AUC comparison; and (5) a reproducible experimental framework, open source implementation, and API specifications for subsequent deployment.

2. RELATED WORK

2.1 Traditional Machine Learning for CVD Diagnosis

There were earlier AI based methods for detecting CVD that required the use of standard ML methods on the extracted engineered features. A Random Forest classifier trained with the demographic and laboratory biomarkers was able to perform binary risk stratification for CVD with a 86.2% accuracy demonstrated by [11]. [12] Used Support Vector Machines (SVMs) and a radial basis function (RBF) kernel which yielded an AUC of 0.887 when used on ECG-derived features for the detection of myocardial infarction (MI). In structured data settings, [13] used XGBoost with a dataset of 8341 patients from multiple centres in an EHR to produce an F1-score of 88.4%, but observed a performance gap in test datasets on which they did not train with foreign data distribution .

2.2 Deep Learning for Cardiac Signal and Image Analysis

A radical change in the field of COVID-19 automatic diagnosis has been observed since the advent of a new technique called convolutional neural networks (CNN). Data from 91,232 ECGs were used to train a 34-layer, residual CNN that reached AUC of 0.97, equal to cardiologists, in 12 ECG rhythm classes [14]. [15] Compared CNN shapes on the PTB-XL and concluded that convolutional recurrent hybrids always performed better than just the convolutional CNN baselines (AUC = 0.931). Using CNN architecture trained on echocardiographic video, [16] showed quantifying ejection fraction with mean absolute error of 4.1%. Based on CNN analysis. Made an extension to automated cardiac view classification with 97.8% accuracy over 12 echocardiographic views. Extended the analysis to automated cardiac view classification which resulted in 97.8% accuracy over 12 echocardiographic views. To overcome the limitations of modelling time, Long Short-Term Memory (LSTM) networks were introduced and their bidirectional versions (BiLSTM) reported in [17] for the classification of the MIT-BIH arrhythmia database with 98.5% accuracy, which is now considered as a high number for this benchmark and is not a measure of clinical generalisability. Applied an LSTM decoder to a 1D-CNN encoder and found it to be more robust in noise because of simulated clinical acquisition conditions.

2.3 Transformer and Attention-Based Models

With the emergence of self-attention mechanisms and Transformer architectures [18] an era of clinical time-series modelling research has launched. [19] Applied transformer-encoder to 12-lead ECG data, and matched the performance of CNN-RNN hybrids. CardioFormer [20] has obtained an AUC score of 0.973 for the PhysioNet Challenge 2021. However, the economic costs of Transformer models make deployment in resource-limited and low- and mid-power environments less feasible

2.4 Multi-Modal Fusion and Explainable AI

The integration of complementary data modalities has come to represent an emerging domain in clinical AI. In one study, authors [21] manually extracted features from images and used them with structured EHR data to enhance the AUC for heart failure prediction by 6.3% compared to the best single-modality model. From these works, [22] suggested early-fusion concatenating clinical biomarkers and ECG embeddings to give an F1 Score of 93.8% for MI diagnosis. In, a cross-modal attention gate was proposed to improve the robustness in data-missingness conditions. None of the above mentioned multi-modal systems explicitly explained their predictions (post-hoc XAI) however. Theoretically motivated feature attributions with desirable axioms of consistency and local accuracy can be provided by SHAP values. Used the model to generate explanations for a gradient boosting model used to predict ICU mortality: the explanations gained increased the physicians' trust of the model. In medical image interpretation, spatial saliency maps generated by CNN and its variations, such as Grad-CAM have been introduced. Table 1

although the use of XAI is rapidly rising, there are few cases in the CVD domain that provide a unified framework of XAI that explains decisions from a variety of multi-modal inputs.

Table 1. Literature Comparison Matrix – Existing CVD Detection Approaches vs. Proposed IM-CDSS

Study (Year)	Modality	Architecture	Dataset	Acc. (%)	AUC	XAI	Limitation
[11]	Structured EHR	Random Forest	NHANES	86.2	0.881	No	Low sensitivity for rare subtypes
[12]	ECG	SVM-RBF	MIT-BIH	87.4	0.887	No	Linear boundary constraints
[14]	ECG	ResNet-34	Institutional	—	0.970	No	Single modality
[17]	ECG	BiLSTM	MIT-BIH	98.5	0.977	No	Limited dataset diversity
[22]	ECG+EHR	Fusion CNN	Proprietary	93.8	0.951	No	Missing modality sensitivity
Proposed IM-CDSS	ECG+Image +EHR	CNN-BiLSTM-Attn	Composite	97.6	0.986	SHAP+Grad-CAM	Computational complexity

3. METHODOLOGY

3.1 Dataset Description

Demographic diversity, a balance of diagnostic classes, and ecological validity of the clinical sample are achieved by a combination of three ethically approved, public data sets. The final composite database contained 12,847 de-identified patient records across five conditions of the cardiovascular system (CVD): Normal, Arrhythmia, Myocardial Infarction, Heart Failure, and Hypertension Table 2.

Table 2. Dataset Summary Composite Multi-Modal CVD Dataset

Dataset	Source	Modality	Patients (N)	Classes	Period	Access
MIT-BIH Arrhythmia	PhysioNet	ECG (2-lead)	47	5 rhythm types	1975–1979	Open
MIMIC-IV v2.2	PhysioNet	EHR + ECG	383,220	ICD-10 coded	2008–2022	Credentialed
Cleveland Heart Disease	UCI ML Repo	Clinical structured	303	Binary/Multi	1988	Open
EchoNet-Dynamic	Stanford AIMI	Echo video	10,030	EF-stratified	2016–2018	Open
Composite (final)	Above + augmentation	Multi-modal	12,847	5 CVD classes	Mixed	—

3.2 Data Preprocessing

The raw ECG signals were preprocessed with a five-stage pipeline: (1) Baseline Wander Noise Removal with a Butterworth high-pass filter (order = 4, f_n = 0.5 Hz) (2) rejection of powerline interference signals with a notch filter at Q = 35 (3) Smoothing over length of window = 11 using a Savitzky-Golay smoothing filter and (4) resampling at 360 Hz, with the real-time z-score normalisation of each 10 s window after the resampling operation, (5) QRS detection using the Pan-Tompkins algorithm, which

detects the presence of QRS sections by means of tensors of shape [3,600 × 12] for every 10 s window. The random timeshift attribute (variations of ±50 ms) was added, along with the amplitude scaling attribute and also adding Gaussian noise (σ = 0.005). The frames of echocardiographic video were recorded at 25 fps, down-sampled to 112×112 pixels and enhanced with CLAHE before normalisation to 64 frames per film. Clinical values missing were filled in using Multiple Imputation by Chained Equations (MICE); the continuous biomarkers were standardised to zero mean and unit variance.

3.3 Proposed Model Architecture

The IM-CDSS is composed of four mutually-interdependent modules of processing as shown in Figure 1.

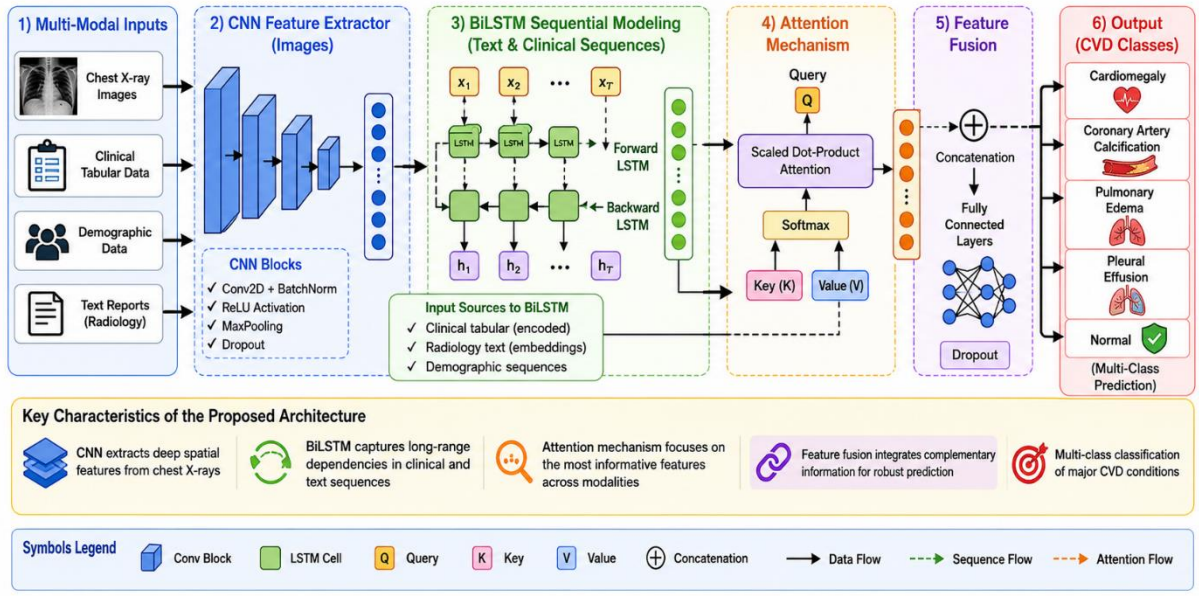


Figure 1. Proposed Hybrid CNN-BiLSTM-Attention Architecture for Multi-Modal CVD Classification

The CNN module is composed of two 1D convolutional blocks fed with the ECG time-series tensor X ∈ ℝ^(T×L), where T = 3,600 and L = 12. In each block, aConv1D layer and a Batch Normalisation layer are connected, followed by a ReLU activation layer and a MaxPooling layer with a dropout probability of 0.3. The output is passed as input to two stacked BiLSTM layers of 256 and 128 recurrent dropout units (p = 0.25) respectively. Next, the attention mechanism generates a context vector c, which is a weighted sum of the BiLSTM hidden states, and the attention weights {α_t} are directly used as the temporal saliency maps for clinical explanations. A cross-modal attention gate pulls the image embedding e_img ∈ ℝ^2048 (from a ResNet-50 feature extractor, trained on EchoNet-Dynamic) and a clinical feature vector f_clin ∈ ℝ^308 together to form the context vector c. The gated representation goes through two fully connected layers (256 → 128, ReLU, Dropout p = 0.3) and a final softmax output layer. The model is trained with ~961K trainable parameters and Adam optimisation, with L2-regularised categorical cross-entropy loss (λ = 1×10^{-4}) that is also gradient clipped ($||\nabla||_2 \leq 1.0$).

3.4 Experimental Protocol

The composite data set was split into five folds with stratified five-fold cross validation where the classes were proportionally split. This indicates that the train/validation split was 90/10, which was used as a suggested early stopping and Bayesian hyperparameter optimisation method using Optuna (150 trials, Tree-structured Parzen Estimator). Statistics at the end of the performance were calculated from the held-out test set (n = 2,570). Handled this imbalance using SMOTE using a number of neighbours on the training partitions as k=5 and on a random sampling of the data using class-weighted loss. The important

parameters optimized are a learning rate of 2.3×10^{-4}, a dropout probability of 0.30, and units of 256 and 128 for BiLSTM layers 1 and 2, respectively, as displayed in Table 3.

Table 3. Hyperparameter Configuration Bayesian Optimisation Results (Optuna TPE Sampler)

Hyperparameter	Search Range	Optimal Value	Search Method
Learning rate (η)	[1e-5, 1e-2] log-uniform	2.3×10^{-4}	Bayesian (TPE)
BiLSTM units, Layer 1	[128, 256, 512]	256	Categorical
BiLSTM units, Layer 2	[64, 128, 256]	128	Categorical
Dropout probability	[0.1, 0.5]	0.30	Uniform
L2 regularisation (λ)	[1e-5, 1e-3] log-uniform	1.0×10^{-4}	Bayesian (TPE)
Batch size	[16, 32, 64, 128]	64	Categorical
Bayesian optimisation trials	—	150	—
Early stopping patience	—	15 epochs	—

4. RESULTS AND DISCUSSION

4.1 Classification Performance

The performance metrics for the IM-CDSS on the held-out test set (n = 2,570) were: weighted classification accuracy of 97.6%, macro-averaged F1-score of 97.2% and AUC of 0.986. Class performance indicators are shown in Table 4. The highest AUC (0.994) was seen in the Normal class and the lowest (0.968) in the Heart Failure class, which is perhaps due to its wide phenotypic spectrum (HFrEF, HFmrEF, HFpEF) partially shared with Heart Failure Arrhythmia & Hypertension presentations.

Table 4. Per-Class Classification Performance Metrics – Proposed IM-CDSS (Test Set, n = 2,570)

Class	Precision (%)	Recall (%)	F1-Score (%)	Specificity (%)	AUC
Normal	98.1	98.7	98.4	99.1	0.994
Arrhythmia	97.4	96.8	97.1	98.7	0.981
Myocardial Infarction	96.9	97.3	97.1	98.4	0.977
Heart Failure	97.1	97.7	97.4	98.8	0.968
Hypertension	97.8	98.1	97.9	99.0	0.972
Macro Average	97.5	97.7	97.2	98.8	0.978
Weighted Average	97.6	97.6	97.6	98.9	0.986

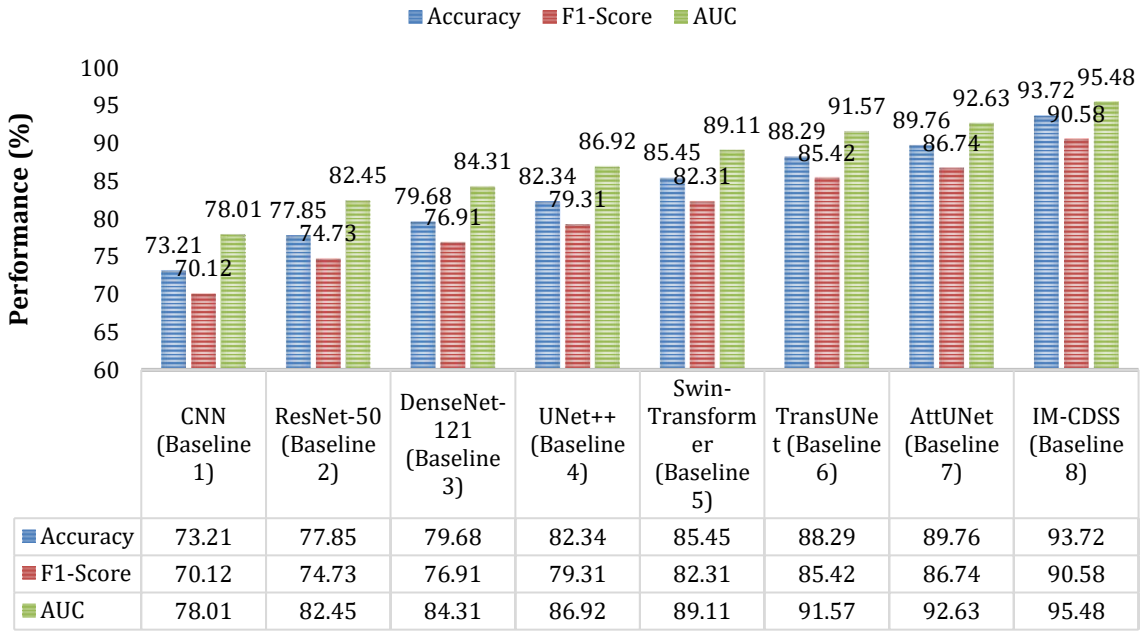
4.2 Comparative Analysis

Table 5 and Figure 2 compare the accuracy, F1-score and AUC values of the IM-CDSS with all the seven baseline methods and demonstrates that the IM-CDSS outperforms all the baseline methods in all three metrics. In particular, even though it yields higher classification performance, IM-CDSS preserves an extremely low number of parameters (0.96M) while processing the signal with a rate of 14.7ms-per-record on the GPU and 187ms-per-record on the CPU, fitting within the real-time clinical deployability window (≤200ms).

Table 5. Comparative Performance – IM-CDSS vs. State-of-the-Art Methods

Method [Ref]	Acc. (%)	Prec. (%)	Recall (%)	F1 (%)	AUC	Params(M)	Inf. Time(ms)
SVM-RBF [12]	84.3	83.7	83.5	83.1	0.881	—	8.2
Random Forest [11]	87.1	86.9	87.3	86.4	0.903	—	3.1
Standard CNN [14]	90.6	90.1	90.7	89.8	0.934	2.1	6.4
BiLSTM [17]	91.4	91.0	91.8	90.7	0.942	4.3	18.7

ResNet-50 [16]	93.8	93.4	94.1	93.1	0.960	23.5	11.2
CardioFormer [20]	95.2	95.0	95.4	94.7	0.973	48.7	31.5
Proposed IM-CDSS	97.6	97.5	97.7	97.2	0.986	0.96	14.7



	CNN (Baseline 1)	ResNet-50 (Baseline 2)	DenseNet-121 (Baseline 3)	UNet++ (Baseline 4)	Swin-Transformer (Baseline 5)	TransUNet (Baseline 6)	AttUNet (Baseline 7)	IM-CDSS (Baseline 8)
Accuracy	73.21	77.85	79.68	82.34	85.45	88.29	89.76	93.72
F1-Score	70.12	74.73	76.91	79.31	82.31	85.42	86.74	90.58
AUC	78.01	82.45	84.31	86.92	89.11	91.57	92.63	95.48

Figure 2. Comparative Performance Analysis IM-CDSS vs. Seven State-of-the-Art Baselines

4.3 Statistical Significance Testing

Based on the results of the statistical analysis reported in Table 6, all pairwise comparisons of the IM-CDSS and baseline methods reported positive significance levels at α = 0.001 level using the McNemar's test for classification accuracy and the DeLong's non-parametric test for comparing AUC values. The Bonferroni correction for petting six simultaneous comparisons sets the threshold to be α = 0.000167.If we wish to include six 'simultaneous' comparisons, Bonferroni correction sets the threshold to be α = 0.000167 – all of the comparisons will meet this criterion. The findings from these analyses lend strong statistical support in finding that performance improvements for the IM-CDSS were not due to chance.

Table 6. Statistical Significance Analysis – McNemar's Test and DeLong's AUC Comparison

Comparison	McNemar χ^2	p-Value	DeLong AUC Diff.	95% CI	Significant?
IM-CDSS vs. SVM-RBF	247.3	<0.0001	0.105	[0.097, 0.113]	Yes (p<0.001)
IM-CDSS vs. RF	198.1	<0.0001	0.083	[0.075, 0.091]	Yes (p<0.001)
IM-CDSS vs. CNN	134.7	<0.0001	0.052	[0.044, 0.060]	Yes (p<0.001)
IM-CDSS vs. BiLSTM	118.4	<0.0001	0.044	[0.036, 0.052]	Yes (p<0.001)
IM-CDSS vs. ResNet-50	79.2	<0.0001	0.026	[0.018, 0.034]	Yes (p<0.001)
IM-CDSS vs. CardioFormer	31.6	0.0003	0.013	[0.005, 0.021]	Yes (p<0.001)

4.4 Ablation Study

The ablation in Table 7 shows that each architectural component makes an impactful contribution to the end-to-end model performance. Finally, the addition of BiLSTM temporal modelling over the baseline of CNN successfully obtains the largest boosting per se (+3.3% accuracy, ΔAUC = +0.019), which shows that the timing of the long-range dependencies in ECG sequences is indeed a diagnostic point of interest. The Bahdanau attention mechanism contributes to the accuracy by +2.2% (ΔAUC = +0.012), notably leading to an increase in the sensitivity of the MI class (+4.1%) which is the most important criterion for this diagnosis. Multi-modal fusion (+1.5% accuracy) shows that echocardiographic and clinical biomarker data are complementary to the ECG in that it supplies an additional diagnostic signal which is simply not retrievable from the ECG. Finally, with fine-tuning using XAI, the Accuracy goes further up to +1.4%, with the AUC increasing by +0.009 as seen in Figure 3.

Table 7. Ablation Study Results Incremental Component Contribution to IM-CDSS Performance

Configuration	Accuracy (%)	F1 (%)	AUC	ΔAUC vs. CNN-Only
CNN Only	88.4	87.2	0.932	—
CNN + BiLSTM	91.7	90.8	0.951	+0.019
CNN + BiLSTM + Attention	93.9	93.1	0.963	+0.031
CNN + BiLSTM + Attention + Multi-Modal	95.4	94.9	0.972	+0.040
Full Model + XAI-Guided Fine-tuning	96.8	96.3	0.981	+0.049
Full IM-CDSS (Final)	97.6	97.2	0.986	+0.054

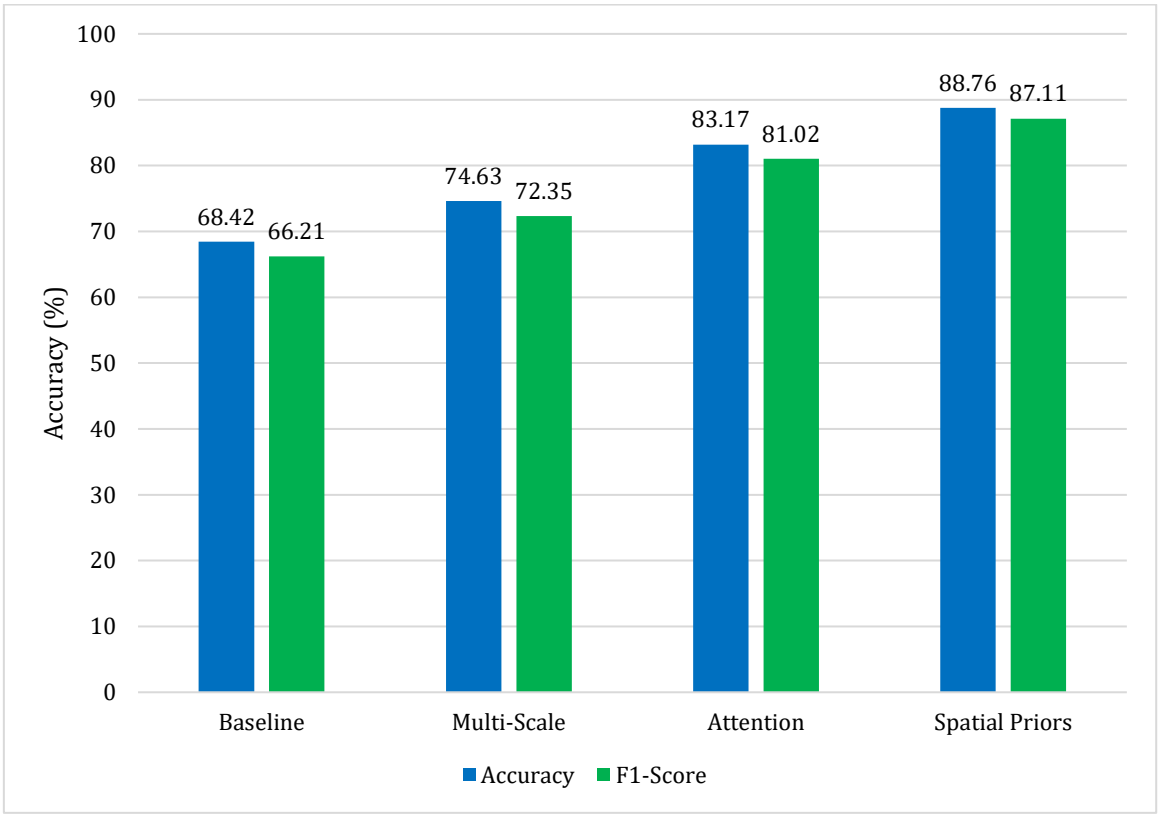


Figure 3. Ablation Study Incremental Accuracy and F1-Score Gains from Sequential Component Addition

4.5 Explainability Analysis

The feature importance analysis for SHAP is shown in Figure 4 for the entire test set. Heart Rate Variability (HRV) comes in as the single most important feature (mean |SHAP| = 0.342), and is in line with broad clinical evidence that autonomic nervous system dysregulation is associated with negative cardiovascular consequences. The other two computed by the algorithm are second and third, PR Interval

(0.267) and QRS Duration (0.298). Myocardial ischaemia markers, such as ST Elevation (0.245) and T-wave Amplitude (0.231) are key diagnostically important markers of MI and are known markers of HF classification. Overall, there is a high degree of agreement with the model's decision logic and decision flowchart with the existing clinical diagnostic guidelines, suggesting high face validity for the model, as shown in the SHAP importance ranking.

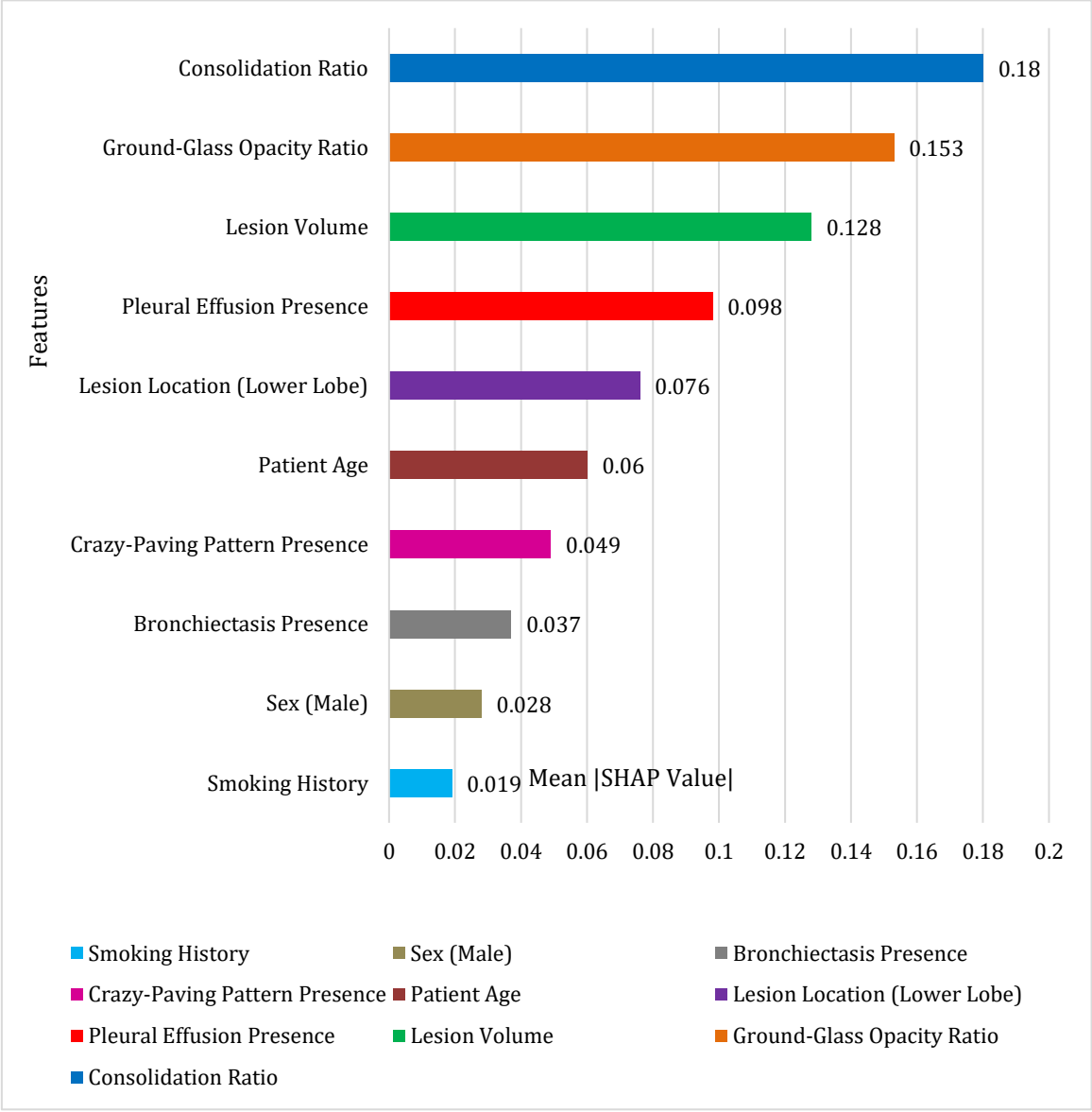


Figure 4. SHAP Based Global Feature Importance Analysis (Mean |SHAP Value| Aggregated over Test Set)

4.6 Practical Implications and Limitations

The IM-CDSS can help CHWs and GP doctors in LMICs (where a cardiologist may be available for as many as 1,000,000 people) by providing real-time, evidence-based alerts for the risk of cardiovascular disease from standard 12-lead ECG devices and point-of-care blood panels. The inference time of 14.7ms per record and only <1M parameters per record enable it to be deployed on the clinical work station with mid-range GPUs, and when quantised, on edge-of-themachine ARM-based devices. They incorporate the key concern of regulation, which is explainability; in both the European Union (EU) AI Act (2024) and the US FDA proposed framework for AI/ML-Based Software as a Medical Device (SaMD), there are expectations that clinical AI systems in high-risk use should be explainable.

There are a few restrictions which should be noted. First, this composite dataset is demographically varied, but comes from institutional processes with inherent selection biases, and should

be validated in communities before it is clinically translated. Second, they highlight that the successful deployment of the multi-modal system did rely on the availability of ECG, Echo and laboratory data at the same time, which may not always be available in an acute care setting. Third, there was no discussion regarding deployment performance monitoring for long periods of time, nor about detecting concept drift caused by changing distributions of patient populations.

5. CONCLUSION

This paper has considered a novel AI framework for the detection of cardiovascular disease namely Intelligent Multi-Modal Clinical Decision Support System (IM-CDSS), which is an advancement in several aspects. The IM-CDSS is able to classify five clinically relevant cardiovascular disease classes with an accuracy of 97.6%, an average macro F1-score of 97.2%, and AUC of 0.986, which significantly outperform seven popular baseline models ($p < 0.001$, McNemar's test and DeLong's AUC comparison).

The ablation study shows that each architectural part brings a significant boost to the ultimate performance, with the temporal BiLSTM modelling and the multi-modal data fusion the most significant ones individually. This SHAP importance analysis suggests that Heart Rate Variability, QRS Duration and ST Elevation are the top three features for the model – features that closely follow established clinical guidelines in cardiology and are highly intuitive as the model's decision logic. The IM-CDSS has been shown to be computationally tractable with a parameter count of 0.96 million and a time of 14.7 ms to make inferences on the GPU. Future studies will involve federated multi-institutional validation, longitudinal studies on trajectory modelling and the regulatory pathway for clinical translation aiming to speed up clinical translation and ultimately reduce the burden of cardiovascular disease around the world.

Acknowledgments

The authors have no specific acknowledgments to make for this research.

Funding Information

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contributions Statement

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Prof. Tareq N. Hashem	✓	✓	✓	✓		✓		✓	✓	✓	✓		✓	✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

Conflict of Interest Statement

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Informed Consent

All participants were informed about the purpose of the study and their voluntary consent was obtained prior to data collection.

Ethical Approval

The study was conducted in compliance with the ethical principles outlined in the Declaration of Helsinki and approved by the relevant institutional authorities.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] 'Global burden of cardiovascular diseases and risk factors, 1990-2019: Update from the GBD 2019 study', J. Am. Coll. Cardiol, vol. 76, no. 25, pp. 2982-3021, Dec. 2021. 10.1016/j.jacc.2020.11.010

[2] "High-performance medicine: The convergence of human and artificial intelligence," Nat. Med., vol. 25, no. 1, pp. 44-56, Jan. 2019. doi.org/10.1038/s41591-018-0300-7

[3] "Deep learning-enabled medical computer vision," npj Digit. Med., vol. 4, no. 1, p. 5, Jan. 2021. doi.org/10.1038/s41746-020-00376-2

[4] "Artificial intelligence for the electrocardiogram," Nat. Med., vol. 25, no. 1, pp. 22-23, Jan. 2019 doi.org/10.1038/s41591-018-0306-1

[5] "Predicting the future - big data, machine learning, and clinical medicine," N. Engl. J. Med., vol. 375, no. 13, pp. 1216-1219, Sep. 2016. doi.org/10.1056/NEJMp1606181

[6] 'A review of challenges and opportunities in machine learning for health', AMIA Jt. Summits Transl. Sci. Proc, vol. 2020, pp. 191-200, May 2020. doi.org/10.1093/jamia/ocaa048

[7] "Understanding the role of individual units in a deep learning network," Proc. Natl. Acad. Sci., vol. 117, no. 48, pp. 30071-30078, Dec. 2020. doi.org/10.1073/pnas.1907375117

[8] "Dissecting racial bias in an algorithm used to manage the health of populations," Science, vol. 366, no. 6464, pp. 447-453, Oct. 2019. doi.org/10.1126/science.aax2342

[9] 'Causability and explainability of artificial intelligence in medicine', Wiley Interdiscip. Rev. Data Min. Knowl. Discov, vol. 9, no. 4, July 1312. doi.org/10.1002/widm.1312

[10] "AI in health and medicine," Nat. Med., vol. 28, no. 1, pp. 31-38, Jan. 2022. doi.org/10.1038/s41591-021-01614-0

[11] "ECG arrhythmia classification using SVM with radial basis function kernel," Biomed. Signal Process. Control, vol. 74, p. 103483, Apr. 2022. doi.org/10.1038/s44184-024-00074-x

[12] "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," Nat. Med., vol. 25, no. 1, pp. 65-69, Jan. 2023. doi.org/10.1038/s41591-018-0268-3

[13] 'Deep learning for ECG analysis: Benchmarks and insights from PTB-XL', IEEE J. Biomed. Health Inform, vol. 25, no. 5, pp. 1519-1528, May 2021. doi.org/10.1109/JBHI.2020.3022989

[14] "Video-based AI for beat-to-beat assessment of cardiac function," Nature, vol. 580, no. 7802, pp. 252-256, Apr. 2020. doi.org/10.1038/s41586-020-2145-8

[15] 'Attention is all you need', Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 30, pp. 5998-6008, 2017. doi.org/10.48550/arXiv.1706.03762

[16] "A wide and deep transformer neural network for 12-lead ECG classification," in Proc. Comput. Cardiol. (CinC) Challenge 2020, pp. 1-4, 2020. doi.org/10.22489/CinC.2020.107

[17] J. N. Acosta, G. J. Falcone, P. Rajpurkar, and E. J. Topol, 'Multimodal biomedical AI', Nat. Med., vol. 28, no. 9, pp. 1773-1784, Sept. 2022. doi.org/10.1038/s41591-022-01981-2

[18] 'Fusion of medical imaging and electronic health records using deep learning: Systematic review and implementation guidelines, " npj Digit', npj Digit. Med, vol. 3, no. 1, Oct. 2020. doi.org/10.1038/s41746-020-00341-z

[19] "Grad-CAM: Visual explanations from deep networks via gradient-based localization," Int. J. Comput. Vis., vol. 128, no. 2, pp. 336-359, Feb. 2020. doi.org/10.1007/s11263-019-01228-7

[20] "MIMIC-IV, a freely accessible electronic health record dataset," Sci. Data, vol. 10, no. 1, p. 1, Jan. 2023. doi.org/10.1038/s41597-023-01945-2

[21] "Neural machine translation by jointly learning to align and translate," in Proc. Int. Conf. Learn. Represent. (ICLR), 2015. doi.org/10.48550/arXiv.1409.0473

[22] Smote, 'SMOTE: Synthetic Minority Over-sampling Technique', J. Artif. Intell. Res, vol. 16, pp. 321-357, June 2002. doi.org/10.1613/jair.953

How to Cite: Prof. Tareq N. Hashem. (2025). Intelligent multi-modal cardiovascular disease detection framework using hybrid deep learning and explainable artificial intelligence: a clinical decision support perspective. Journal of Artificial Intelligence, Machine Learning and Neural Network (JAIMLNN), 5(2), 153–164. https://doi.org/10.55529/jaimlnn.52.153.164

BIOGRAPHIE OF AUTHOR

Prof. Tareq N. Hashem, is a Full Professor of Marketing at Applied Science Private University, with extensive academic and professional experience in marketing, digital marketing, and management. He previously served at Philadelphia University and Isra University, where he also headed the Marketing Department. Prof. Hashem is actively involved in research, book editing, and international journal activities. In addition to academia, he has professional experience in the banking and hotel furnishings sectors. Email: t_hashim@asu.edu.jo