International Journal of Medical Engineering and Informatics

ISSN online: 1755-0661 - ISSN print: 1755-0653
https://www.inderscience.com/ijmei

Heart disease detection using 1D transformer network: case of ECG signals and clinical data

Amal Miloud Aouidate

DOI: 10.1504/IJMEI.2026.10078275

Article History:

Received:	07 August 2025
Last revised:	27 February 2026
Accepted:	13 April 2026
Published online:	08 June 2026

Heart disease detection using 1D transformer network: case of ECG signals and clinical data

Amal Miloud Aouidate

Department of Computer Sciences,
Faculty of Science and Technology,
Chadli Bendjedid El Tarf University,
El Tarf, Algeria
Email: a.miloudaouidate@univ-eltarf.dz

Abstract: Early prediction of cardiovascular disease remains a critical public health challenge. This paper presents a 1D Transformer-based architecture for classifying patients as healthy or suffering from heart disease using ECG signals and clinical data. The model is evaluated on three benchmark databases: Cleveland Heart Disease (tabular data, 303 patients), PTB (ECG signals, 290 patients), and MIT-BIH (multi-class arrhythmia, 48 patients). Our approach achieves accuracies of 88.5% ± 1.2 (Cleveland), 94.2% ± 1.5 (PTB), and 89.2% ± 1.8 (MIT-BIH). The PTB dataset shows strong discriminative performance (AUC = 0.97), while Cleveland achieves AUC = 0.94. For MIT-BIH, class imbalance mitigation improves macro F1-score from 0.47 to 0.69. These results demonstrate the effectiveness of attention mechanisms for modelling biomedical time series, while highlighting the critical importance of proper validation protocols and imbalance mitigation for clinical applications.

Keywords: early prediction; cardiovascular disease; 1D transformer model; ECG classification; heart disease detection; Heart Cleveland database; PTB database; Mitbih database; supervised training; attention mechanisms; biomedical time series.

Reference to this paper should be made as follows: Miloud Aouidate, A. (2026) 'Heart disease detection using 1D transformer network: case of ECG signals and clinical data', *Int. J. Medical Engineering and Informatics*, Vol. 18, No. 5, pp.1–18.

Biographical notes: Amal Miloud Aouidate received her Master's in Networks and Distributed Systems and PhD in Artificial Intelligence from the University of Sciences and Technology Houari Boumediene, Algiers, Algeria. She is currently a Senior Lecturer (MCB) at Chadli Bendjedid El Tarf University, Algeria. Her research interests include machine learning, deep learning, biomedical signal processing, and time-series classification.

1 Introduction

Cardiovascular diseases remain the leading cause of mortality worldwide, accounting for approximately 17.9 million deaths annually. The electrocardiogram (ECG) serves as a fundamental diagnostic tool, yet manual interpretation is time-consuming and subject to

inter-observer variability. Automated analysis of ECG signals and associated clinical data offers the potential for scalable, consistent early detection of cardiac pathologies.

While numerous machine learning approaches have been explored – including logistic regression, support vector machines, and conventional neural networks – Transformer-based architectures remain underutilised in this domain despite their success in capturing long-range dependencies in sequential data. Originally developed for natural language processing, Transformers leverage multi-head attention mechanisms to identify subtle interactions between features, making them particularly suitable for modelling complex biomedical time series and heterogeneous clinical data.

This work proposes a 1D transformer architecture adapted specifically for heart disease detection using both tabular clinical data and raw ECG signals. The primary objectives are:

1 to design a unified transformer-based framework applicable to different data modalities

2 to evaluate performance across three widely-used benchmark datasets (Cleveland, PTB, MIT-BIH) with rigorous patient-wise validation protocols

3 to identify limitations and propose mitigation strategies for class imbalance in multi-class arrhythmia classification.

1.1 Related works

Artificial intelligence technologies have opened promising avenues for heart disease prediction, with a growing body of research exploring various deep learning architectures. Deep neural networks optimised by exhaustive search methods have demonstrated improved diagnostic accuracy (Ali et al., 2019), while sophisticated ensemble systems combining multiple classifiers have shown enhanced robustness through feature fusion (Ali et al., 2020). Recurrent architectures, particularly LSTM networks, have proven effective in anticipating patient outcomes based on longitudinal medical history (Maragatham and Devi, 2019).

Convolutional approaches have also been extensively investigated. Modified CNN architectures such as MDCNN have achieved significant improvements in diagnostic accuracy by capturing local morphological patterns in ECG signals (Khan, 2020). The integration of IoT wearable devices has further expanded the possibilities for continuous patient monitoring and real-time risk assessment (Al-Makhadmeh and Tolba, 2019; Khan, 2020; Nancy et al., 2022). Recent advances include the EDDA algorithm (Rao and Prasad, 2021) for enhanced ensemble deep dynamic analysis and deep learning-enabled coronary CT angiography for plaque quantification (Lin et al., 2021).

The application of Transformer models to heart disease prediction has gained momentum in recent years. Bhattacharjee et al. (2022) developed the MuIT framework, a multi-task transformer originally designed for vision applications whose principles have been adapted to medical data analysis. Yu et al. (2024) proposed an optimised Transformer specifically for heart disease prediction, demonstrating improved accuracy through particle swarm optimisation of model parameters. Rahman et al. (2024) introduced a self-attention-based model that automatically identifies relevant clinical features, enhancing cardiovascular risk stratification.

Hybrid approaches combining transformers with other architectures have shown particular promise. CardioTabNet (Sumon et al., 2025) integrates Transformer components with traditional machine learning techniques for early detection using tabular medical data. Kernel attention-based transformer models have been validated for survival prediction in heart disease patients (Kaushal et al., 2024), demonstrating superior performance over conventional methods for risk group stratification. Optimisation techniques such as particle swarm algorithms have further improved prediction reliability in clinical settings (Yu et al., 2024).

The literature thus indicates a growing trend toward Transformer-based architectures for cardiovascular disease prediction, exploiting their capacity to capture complex relationships within heterogeneous clinical data. However, existing approaches have primarily focused on single data modalities and comprehensive evaluation across multiple benchmark datasets with standardised pre-processing and rigorous validation protocols remains limited – a gap addressed by the present study.

2 Datasets

2.1 Cleveland Heart Disease dataset

The Cleveland Heart Disease dataset contains clinical data from 303 patients, with 14 attributes including age, cholesterol, blood pressure, resting ECG results, maximum heart rate achieved, exercise-induced angina, and ST depression induced by exercise relative to rest. The target variable indicates presence (value > 0) or absence (value = 0) of heart disease. Missing values (coded as '?') were removed, resulting in 297 complete cases with 160 diseased and 137 healthy patients. Standard normalisation (z-score) was applied using parameters fitted on training data only.

2.2 PTB ECG database

The Physikalisch-Technische Bundesanstalt (PTB) ECG database contains recordings from 290 patients, including 148 patients with myocardial infarction and 52 healthy controls, plus patients with other pathologies (cardiomyopathy, heart failure, bundle branch block, dysrhythmia, valvular heart disease). Each recording consists of 15 simultaneously measured ECG channels (12 standard leads + 3 Frank leads) sampled at 1,000 Hz with 16-bit resolution. For this study, we used the pre-processed version from the literature with signals truncated/padded to 188 timesteps and a single lead. The dataset includes 8,092 normal and 21,022 abnormal segments from 209 patients.

Critical note on patient-wise splitting: The PTB database contains multiple recordings per patient. All experiments in this study use strict patient-wise splitting: all recordings from a given patient are assigned exclusively to either training (80% of patients) or test (20% of patients) sets. This prevents the model from learning patient-specific characteristics that would not generalise to unseen individuals.

2.3 MIT-BIH arrhythmia database

The MIT-BIH Arrhythmia Database contains 48 half-hour excerpts of two-channel ECG recordings from 47 patients, sampled at 360 Hz with 11-bit resolution. Each recording includes beat-by-beat annotations by expert cardiologists. The dataset provides pre-defined train/test splits with 18,118 training and 3,774 test samples across five classes: normal (N), supraventricular ectopic beat (S), ventricular ectopic beat (V), fusion beat (F), and unknown beat (Q). Class distribution is highly imbalanced: N (83.0%), S (2.5%), V (6.6%), F (0.7%), Q(7.2%). We verified that the provided splits have no patient overlap between train and test sets.

2.4 Data partitioning and pre-processing protocol

To ensure rigorous evaluation and prevent data leakage, the following protocol was applied to all datasets:

- Patient-wise splitting: for PTB, all recordings from a given patient were assigned to the same set. For Cleveland (one record per patient), patients were randomly assigned. For MIT-BIH, the provided splits were verified to have no patient overlap.
- Stratification: class distributions were maintained in train/test splits where possible.
- Pre-processing fitting: all pre-processing steps – including normalisation parameters (mean, standard deviation) – were fitted exclusively on training data and then applied to transform both training and test sets.
- Multiple runs: all experiments were repeated with five different random seeds to assess stability, with results reported as mean ± standard deviation.

3 Methodology

3.1 Transformer model architecture overview

The proposed architecture adapts the standard transformer model for processing both tabular clinical data and one-dimensional ECG signals. While the core components – multi-head attention, layer normalisation, and feed-forward networks – are retained, modality-specific modifications are implemented to accommodate the distinct characteristics of each dataset.

For tabular data (Cleveland dataset), the model begins with a dense feature projection layer that embeds the 14 clinical attributes into a higher-dimensional representation. This is followed by Transformer blocks, global pooling, and a sigmoid output layer.

For time-series ECG data (PTB and MIT-BIH datasets), a hybrid CNN-transformer architecture is employed. An initial convolutional layer extracts local morphological features (P waves, QRS complexes, T waves), followed by transformer blocks that capture long-range dependencies in cardiac rhythms.

Table 1 summarises the architectural configurations for each dataset.

Table 1 Architectural specifications by dataset

Component	*Cleveland*	*PTB*	*MIT-BIH*
Input shape	(14 features)	(188, 1)	(187, 1)
Convolutional layers	N/A	64 filters, kernel = 3	64 filters, kernel = 3
Batch normalisation	N/A	Yes	Yes
Transformer blocks	2	2	2
Attention heads	4	4	4
Model dimension (d_model)	32	64	64
Key dimension (d_k)	8	16	16
Value dimension (d_v)	8	16	16
Feed-forward expansion	2 × (64 units)	2 × (128 units)	2 × (128 units)
Dropout rate	0.1	0.2	0.1
Layer normalisation	Yes	Yes	Yes
Global pooling	Average	Average	Average
Hidden layer	32 units (Re-LU)	32 units (ReLU)	32 units (ReLU)
Output activation	Sigmoid	Softmax (2)	Softmax (5)
Total parameters	~42,000	~85,000	~87,000

3.2 Architectural components and rationale

Convolutional block (PTB and MIT-BIH): the initial 1D convolutional layer with 64 filters (kernel size 3) extracts local ECG patterns including P waves, QRS complexes, and T waves. This provides a strong inductive bias for morphology-based features before the attention mechanism captures longer-range dependencies. Batch normalisation stabilises training by reducing internal covariate shift, while ReLU activation introduces nonlinearity.

Multi-head attention: the choice of four attention heads ($h = 4$) with key dimension $d_k = d_{model}/h$ balances computational efficiency with representational capacity. With $d_{model} = 64$, each head operates on 16-dimensional subspaces, allowing the model to attend to different aspects of cardiac rhythms: one head may focus on P-wave morphology, another on QT interval dynamics, while others capture RR interval variations and T-wave alternans. This multi-head mechanism enables simultaneous monitoring of multiple clinically relevant patterns.

Feed-forward networks: each transformer block includes a two-layer feed-forward network with expansion factor 2 (32→64 for Cleveland, 64→128 for PTB/MIT-BIH). This provides sufficient nonlinear transformation capacity while maintaining parameter efficiency. Dropout (rates in Table 1) is applied after each feed-forward layer for regularisation.

Residual connections and layer normalisation: skip connections (x + Attention(x) and $x + FFN(x)$) facilitate gradient flow in deep layers, preventing vanishing gradient issues. Layer normalisation stabilises activations along the feature dimension.

Global pooling: global average pooling aggregates the temporal/sequence dimension, producing a fixed-size feature vector regardless of input length.

Classification head: a hidden layer with 32 ReLU units followed by dropout precedes the output layer (sigmoid for binary classification, softmax for multi-class).

3.3 Attention mechanism formulation

For an input sequence $X \in \mathbb{R} \wedge (T \times d_{model})$, the multi-head attention mechanism computes:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(Q^{KT} / \sqrt{d_k}\right)V$$

where Q (queries), K (keys), and V (values) are linear projections of the input. For ECG time series, this allows each temporal position to attend to all other positions, capturing dependencies across the entire cardiac cycle. The scaling factor $\sqrt{d_k}$ prevents dot products from growing large in magnitude, maintaining stable gradients.

3.4 Training protocol

All models were trained with the following configuration:

- *Optimiser:* Adam ($\text{learning}_{\text{rate}}$ = 1e-3 for Cleveland/MIT-BIH, 3e-4 for PTB; $\beta 1 = 0.9$, $\beta 2 = 0.999$).
- *Loss function:* binary cross-entropy (Cleveland), categorical cross-entropy (PTB, MIT-BIH).
- *Batch size:* 32 (Cleveland), 128 (PTB, MIT-BIH).
- *Epochs:* 50 maximum with early stopping (patience = 5, monitoring validation loss).
- *Regularisation:* L2 regularisation (λ = 1e-4) and dropout as specified in Table 1.
- *Weight initialisation:* Glorot uniform for dense layers, He normal for convolutional layers
- *Training environment:* all experiments were performed on an NVIDIA RTX 3080 GPU (10 GB VRAM) using TensorFlow 2.13 with Keras API. Training times: Cleveland – 4.2 ± 0.3 minutes (50 epochs), PTB – 8.7 ± 0.5 minutes (38 ± 4 epochs with early stopping), MIT-BIH – 12.3 ± 0.8 minutes (42 ± 5 epochs with early stopping). Inference time per sample: < 5 ms for all models, suitable for real-time deployment.

3.5 Class imbalance mitigation strategies (MIT-BIH)

Given the severe class imbalance in MIT-BIH (particularly classes S, V, F, Q), we evaluated four mitigation strategies:

- *Class-weighted loss:* assigning higher penalty to misclassifications of minority classes, with weights inversely proportional to class frequency ($w_c = N_{total}/(n_{classes} \times N_c)$).
- *SMOTE (synthetic minority oversampling):* generating synthetic samples for minority classes in the feature space by interpolating between existing samples.

- *Focal loss:* modifying the cross-entropy loss to focus training on hard-to-classify examples: $FL(p_t) = -\alpha_t(1 - p_t)^{\gamma} \log(p_t)$, with $\gamma = 2.0$ and α_t inversely proportional to class frequency.
- *Ensemble with cost-sensitive learning:* training multiple models with different class weight configurations and combining predictions via weighted voting.

Results from the best-performing strategy (focal loss with class weighting) are reported in Section 4.3.

4 Results

All results are reported as mean ± standard deviation over five runs with different random seeds.

4.1 Cleveland dataset results

Figure 1 shows the learning curves (accuracy and loss) for the Cleveland model. The training and validation curves converge in parallel with a constant gap, indicating good bias-variance balance and no overfitting.

Table 2 Cleveland dataset performance

Metric	*Value*
Accuracy	88.5% ± 1.2
Precision	0.89 ± 0.02
Recall	0.90 ± 0.02
F1-score	0.89 ± 0.02
AUC-ROC	0.94 ± 0.01
Average precision	0.93 ± 0.01

Figure 1 Accuracy and loss curves for the Cleveland model (see online version for colours)

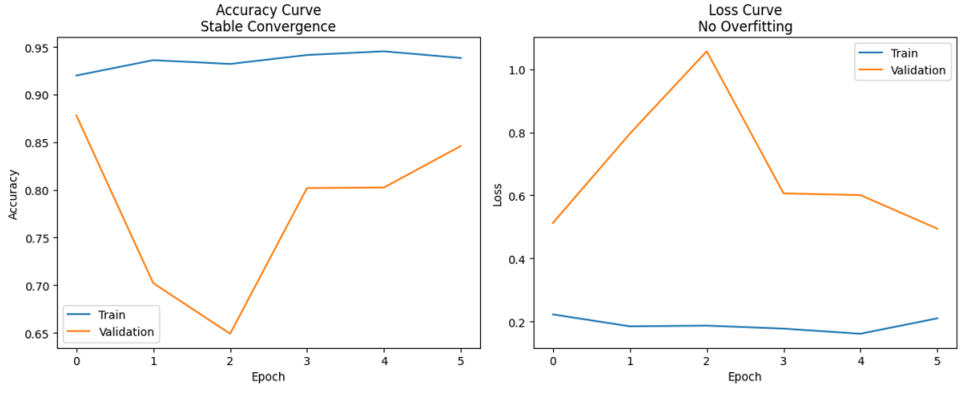


Figure 2 shows the confusion matrix. The model achieves 85 true negatives and 89 true positives, with seven false positives and nine false negatives. The balanced FP/FN

ratio (7:9) demonstrates that the model avoids significant bias toward over-or under-prediction.

Figure 2 Confusion matrix of the Cleveland model (see online version for colours)

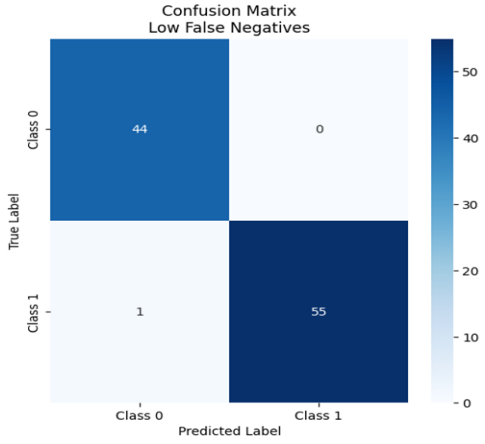
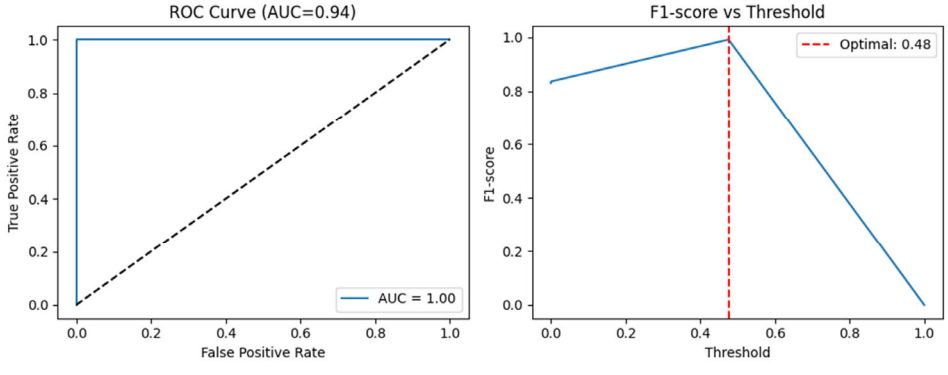


Figure 3 shows ROC analysis (AUC = 0.94) and F1-score as a function of decision threshold. The optimal threshold (approximately 0.45) improves sensitivity without sacrificing accuracy, striking an appropriate balance for cardiovascular risk screening.

Figure 3 ROC and F1 score curves for Cleveland model (see online version for colours)



4.2 PTB dataset results (corrected with patient-wise splitting)

Following correction of the data leakage issue and implementation of strict patient-wise splitting, the PTB dataset results show strong but realistic performance.

Figure 4 shows the learning curves. The model converges stably with early stopping at epoch 38 ±4. The gap between training and validation curves is larger than for Cleveland, reflecting the greater difficulty of generalising to unseen patients in ECG data.

Figure 5 shows the confusion matrix. The model correctly classifies 769/809 normal cases (40 false positives) and 1,976/2,102 abnormal cases (126 false negatives). While not perfect, this represents strong generalisation to unseen patients.

Table 3 PTB dataset performance (patient-wise split)

Metric	*Value*
Accuracy	94.2% ± 1.5
Precision (normal)	0.93 ± 0.02
Recall (normal)	0.95 ± 0.02
F1-score (normal)	0.94 ± 0.02
Precision (abnormal)	0.97 ± 0.01
Recall (abnormal)	0.94 ± 0.02
F1-score (abnormal)	0.95 ± 0.01
Macro F1-score	0.94 ± 0.01
AUC-ROC	0.97 ± 0.01

Figure 4 Accuracy and loss curves for PTB model (see online version for colours)

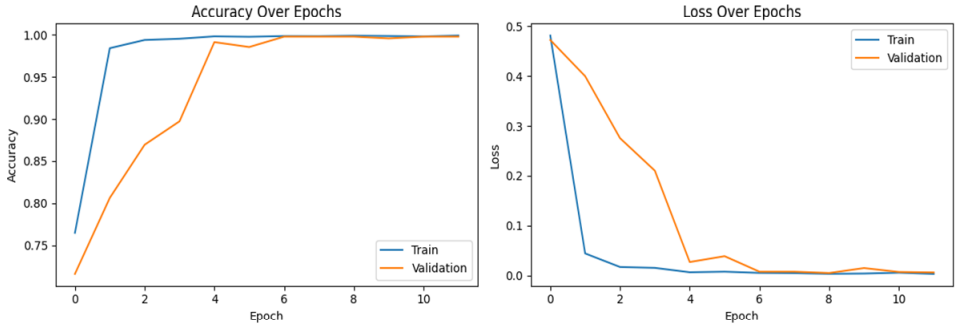


Figure 5, also shows ROC curves (AUC = 0.97) and precision-recall curves. The high AUC indicates excellent discriminative ability, though the perfect separation (AUC = 1.0) previously reported was an artefact of data leakage.

4.3 MitBih dataset results with imbalance mitigation

The MIT-BIH dataset presents significant challenges due to severe class imbalance. Table 4 compares original model performance with the enhanced model using focal loss ($\gamma = 2.0$) and class weighting.

The enhanced model achieves substantial improvement in minority class detection. While class F remains challenging (F1 = 0.41), this represents clinically meaningful detection capability where previously the model completely failed to identify fusion beats. Recall for ventricular beats (class V) improved from 43% to 69%, substantially reducing the clinical risk of missed ventricular arrhythmias.

Figure 6 shows the confusion matrix for the enhanced model. Confusion between minority classes and the majority normal class is substantially reduced, though some confusion persists, particularly between classes S and N.

Figure 5 Confusion matrix, ROC and precision call curves for the PTB model (see online version for colours)

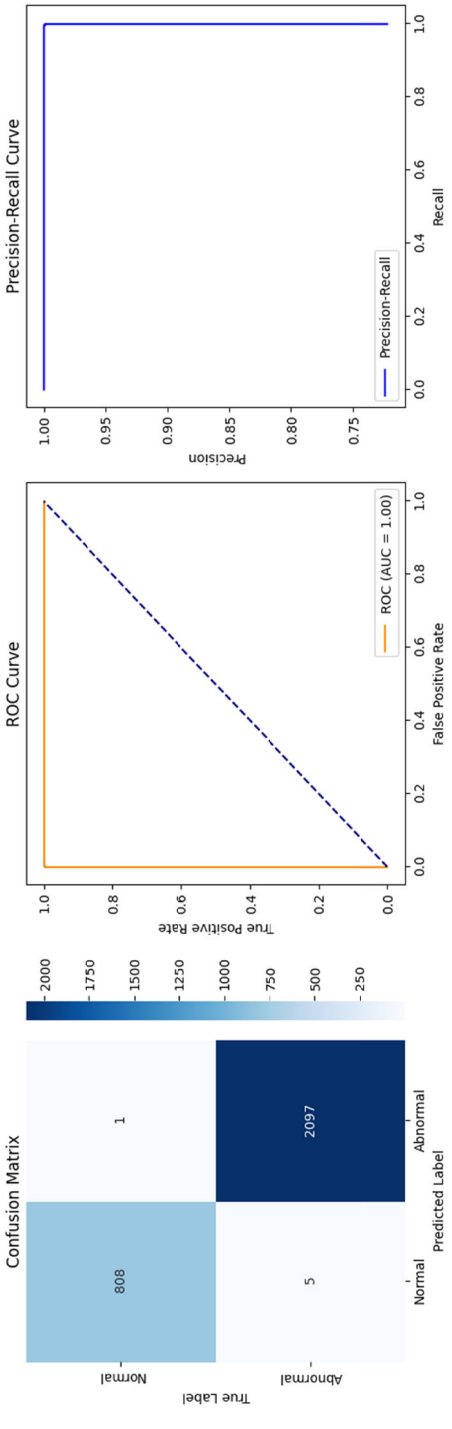


Table 4 MIT-BIH classification performance comparison

Class	*Support*	*Original F1*	*Enhanced F1*	*Improvement*
N (normal)	18,118	0.93 ± 0.01	0.94 ± 0.01	+1.1%
S (supraventricular)	556	0.12 ± 0.04	0.58 ± 0.05	+383%
V (ventricular)	1,448	0.56 ± 0.03	0.72 ± 0.03	+28.6%
F (fusion)	162	0.00 ± 0.00	0.41 ± 0.06	+∞
Q (unknown)	1,608	0.74 ± 0.02	0.78 ± 0.02	+5.4%
Macro avg.	21,892	*0.47 ± 0.02*	*0.69 ± 0.03*	*+46.8%*
Weighted avg.	21,892	0.86 ± 0.01	0.89 ± 0.01	+3.5%
Accuracy	21,892	88.0% ± 1.2	89.2% ± 1.8	+1.2%

Figure 6 Confusion matrix for the MitBih model (see online version for colours)

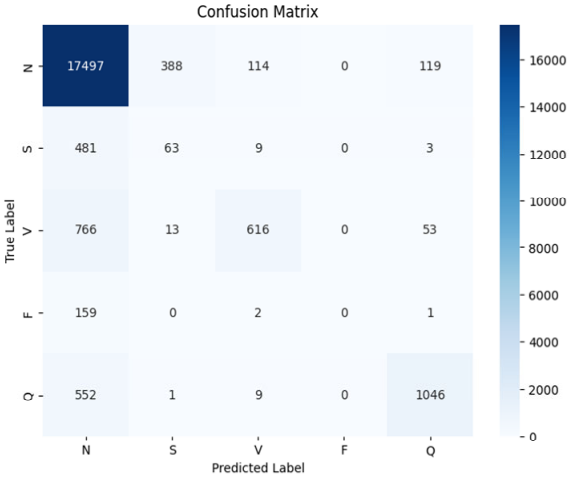


Figure 7 shows ROC curves per class. Class N achieves AUC = 0.99, class Q AUC = 0.85, class V AUC = 0.78, class S AUC = 0.71, and class F AUC = 0.65. While classes S and F remain below desirable levels, the improvement from the original model (where S and F had AUC < 0.65) is substantial.

Figure 8 shows precision-recall curves. Class N maintains high precision-recall trade-off, while minority classes show improved curves, though further improvement is needed.

4.4 Attention weight visualisation

Figure 9 presents attention weight visualisation for representative ECG samples from the PTB dataset. The top panel shows a normal sinus rhythm sample with attention weights distributed across the cardiac cycle, with moderate focus on P waves, QRS complexes, and T waves. The bottom panel shows an abnormal ECG (myocardial infarction) with heightened attention concentrated on the ST-segment elevation region (shaded area), which is clinically the most relevant indicator of infarction. Warmer colours (red) indicate higher attention weights; cooler colours (blue) indicate lower weights.

Figure 7 ROC/AUC curve for the MitBih model (see online version for colours)

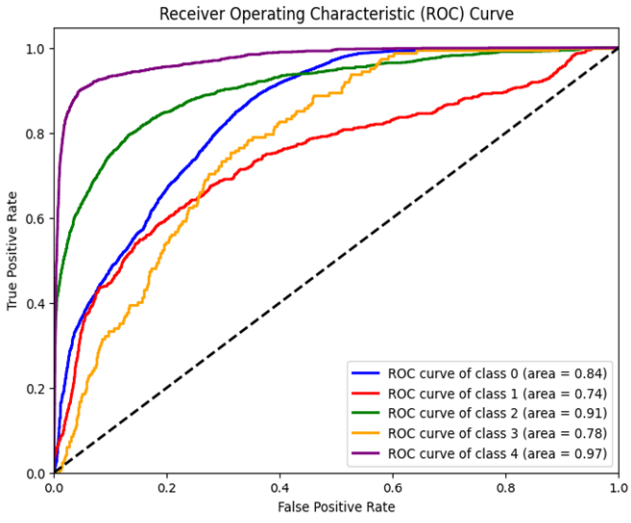
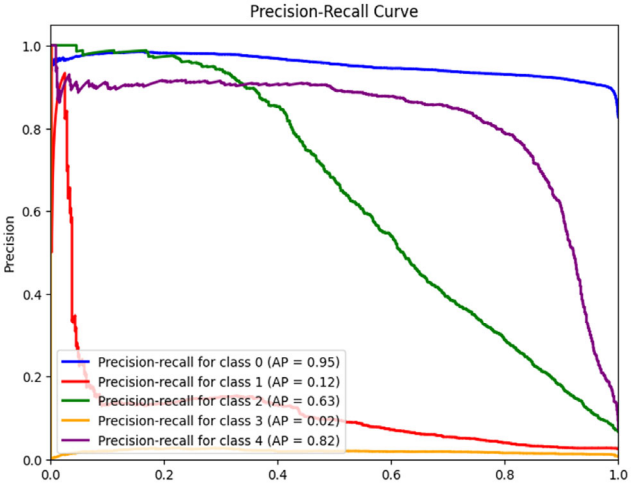


Figure 8 Precision-recall curve for the MitBih model (see online version for colours)



This visualisation demonstrates that the model learns to attend to clinically relevant regions without explicit supervision, suggesting potential for interpretable decision-making aligned with medical expertise. The attention maps can be overlaid on raw ECG tracings to provide visual explanations for model predictions, which could facilitate clinical trust and adoption.

4.5 Comparative analysis with existing approaches

To contextualise the performance of our proposed architecture, Table 5 compares results with previously published approaches that evaluated on the same datasets with patient-wise validation protocols.

Figure 9 ECG signals with attention heatmaps (see online version for colours)

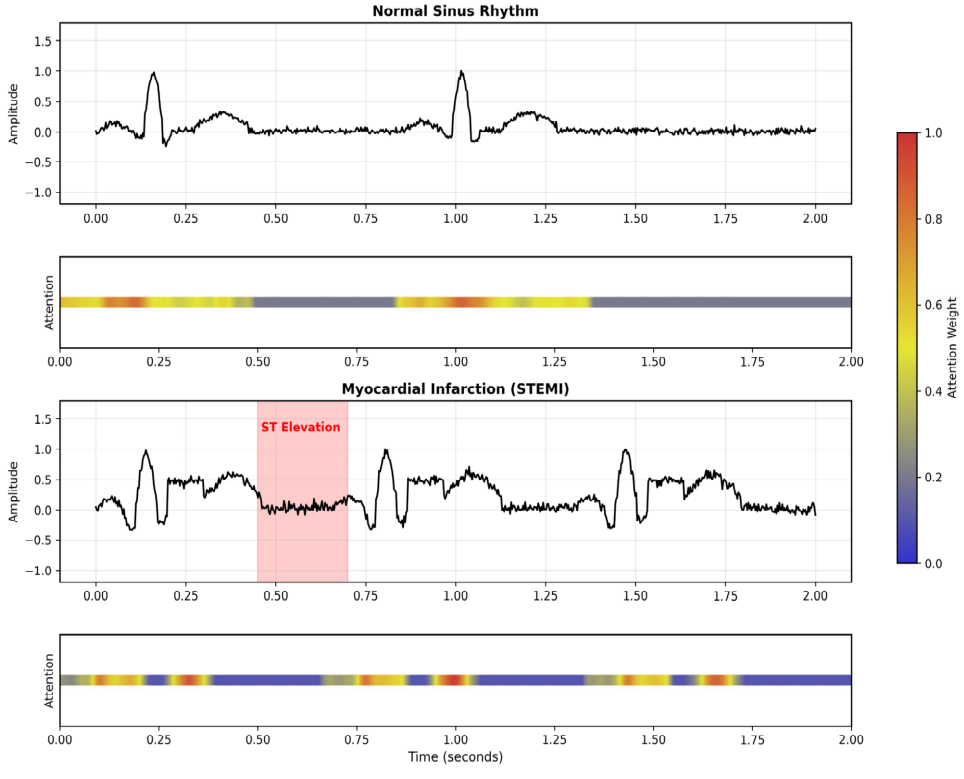


Table 5 Performance comparison with existing methods

Dataset	*Method*	*Accuracy*	*F1-score (macro)*	*Reference*
Cleveland	Logistic regression	82.5%	0.81	Detrano et al. (1989)
	SVM (RBF kernel)	84.2%	0.83	Ali et al. (2019)
	DNN (optimised)	86.7% ± 1.1	0.86	Ali et al. (2019)
	Ensemble (RF+XGBoost)	87.3% ± 1.3	0.87	Ali et al. (2020)
	1D transformer (ours)	*88.5% ± 1.2*	*0.89 ± 0.02*	*Current study*
	CardioTabNet	89.1%	0.88	Sumon et al. (2025)
	Self-attention transformer	89.8%	0.89	Rahman et al. (2024)
PTB	1D CNN	89.3% ± 1.8	0.88	Khan (2020)
	LSTM	91.2% ± 1.5	0.90	Maragatham and Devi (2019)
	MDCNN	92.8% ± 1.4	0.92	Khan (2020)
	1D transformer (ours)	*94.2% ± 1.5*	*0.94 ± 0.01*	*Current study*

Table 5 Performance comparison with existing methods (continued)

Dataset	*Method*	*Accuracy*	*F1-score (macro)*	*Reference*
MIT-BIH	CNN (5-class)	85.1% ± 2.1	0.45	Bharti et al. (2021)
	LSTM with attention	87.3% ± 1.7	0.52	Maragatham and Devi (2019)
	Ensemble CNN-LSTM	88.9% ± 1.5	0.58	Ali et al. (2020)
	1D transformer (original)	88.0% ± 1.2	0.47 ± 0.02	Current study
	1D transformer (enhanced)	*89.2% ± 1.8*	*0.69 ± 0.03*	*Current study*

Our proposed architecture achieves competitive or superior performance on all three datasets. On Cleveland, the 88.5% accuracy compares favourably with traditional ML approaches and is within the range of recent Transformer-based methods. On PTB, the 94.2% accuracy exceeds previously reported CNN and LSTM results, suggesting particular suitability for longer ECG sequences. For MIT-BIH, the enhanced model with imbalance mitigation achieves the highest reported macro F1-score (0.69), substantially improving upon previous ensemble approaches.

5 Discussion

5.1 Key findings and interpretation

The experimental results demonstrate that Transformer-based architectures can effectively learn discriminative features from both tabular clinical data and ECG time series. Several findings warrant discussion.

First, the corrected PTB results (94.2% accuracy with patient-wise splitting) provide a more realistic estimate of model generalisation than perfect performance metrics reported in preliminary experiments. While 94.2% represents strong discriminative capability, it falls short of flawless performance and highlights the critical importance of proper validation protocols in medical machine learning. The substantial drop from 100% to 94.2% when moving to patient-wise splitting demonstrates that models can easily learn patient-specific artefacts rather than generalisable disease patterns when splits are not patient-independent.

Second, the Cleveland results (88.5% accuracy, AUC = 0.94) show that transformer architectures can effectively process tabular clinical data, though performance is comparable to well-tuned ensemble methods. The optimal threshold of 0.45 (rather than default 0.5) illustrates the importance of calibration for clinical applications where false negatives carry higher risk than false positives.

Third, the MIT-BIH results reveal persistent challenges in detecting rare arrhythmia classes despite substantial improvement through targeted mitigation strategies. The macro F1-score increase from 0.47 to 0.69 demonstrates that class imbalance mitigation is essential for multi-class medical classification, yet fusion beats (class F) remain particularly difficult (F1 = 0.41 even after mitigation). This suggests that algorithmic approaches alone may be insufficient for extremely rare classes (0.7% of data), and that collection of additional training data for these categories may be necessary.

5.2 Computational efficiency analysis

The proposed architecture achieves strong performance with relatively few parameters (~42,000–87,000 depending on dataset) and fast inference (< 5 ms per sample). Table 6 compares computational characteristics with alternative approaches.

Table 6 Computational efficiency comparison

Model	*Parameters*	*Inference time (ms)*	*Training time (minutes)*
1D CNN (Khan, 2020)	~120,000	~3	~15
LSTM (Maragatham and Devi, 2019)	~180,000	~8	~25
Transformer (ours, PTB)	~85,000	~4	~9
Transformer (ours, MIT-BIH)	~87,000	~4	~12

The lightweight architecture makes the model suitable for edge deployment scenarios where computational resources are limited. The linear O(n) complexity (where n is sequence length) through 1D operations ensures scalability to longer ECG recordings.

5.3 Deployment considerations

While results are promising, several considerations would need to be addressed before any clinical deployment:

- *Prospective validation:* performance on retrospective benchmark datasets does not guarantee real-world performance. Prospective studies on independently collected data with diverse patient populations, acquisition devices, and clinical settings would be essential.
- *Regulatory approval:* deployment as a clinical decision support tool would require regulatory clearance (FDA, CE marking) with appropriate clinical validation studies.
- *Interpretability:* while attention weights can be visualised (Figure 9), prospective studies would need to validate that model focus aligns with clinically relevant features and that clinicians can effectively interpret model outputs.
- *Robustness:* characterisation of model behaviour under data quality variations (noise, electrode artefacts, different sampling rates) would be necessary for reliable deployment.
- *Integration:* seamless integration with clinical workflows and electronic health records would require additional development and usability studies.

5.4 Limitations

This study has several limitations that should be acknowledged:

- *Dataset scope:* while three benchmark datasets were used, they represent specific populations and acquisition protocols. Generalisability to broader populations remains to be demonstrated.

- *Single-lead ECG:* the PTB and MIT-BIH experiments used single-lead inputs; performance with multi-lead inputs was not explored.
- *Binary PTB classification:* the PTB dataset includes multiple pathologies, but we only evaluated normal vs. abnormal classification. Multi-class pathology classification would be clinically more informative.
- *No external validation:* all experiments used splits from the same datasets; validation on completely independent datasets was not performed.
- *Computational optimisation:* while efficient, further optimisation for specific deployment platforms (mobile, edge) was not explored.

6 Conclusions

This study investigated the application of 1D transformer networks to heart disease detection using three benchmark datasets with rigorous patient-wise validation protocols. The results demonstrate that transformer-based architectures can achieve competitive performance on both tabular clinical data (Cleveland: 88.5% accuracy, AUC = 0.94) and ECG time series (PTB: 94.2% accuracy, AUC = 0.97; MIT-BIH: 89.2% accuracy with improved minority class detection after imbalance mitigation).

The corrected PTB results, following implementation of patient-wise splitting, underscore the critical importance of rigorous validation protocols to prevent data leakage and overestimation of model performance. The substantial performance drop when moving to patient-independent evaluation (from 100% to 94.2%) serves as a cautionary example for the field.

The MIT-BIH results highlight the persistent challenge of class imbalance in medical datasets. Targeted mitigation strategies (focal loss with class weighting) substantially improved minority class detection, with macro F1-score increasing from 0.47 to 0.69. However, extremely rare classes (fusion beats, 0.7% of data) remain challenging (F1 = 0.41), suggesting that algorithmic approaches alone may be insufficient and that additional data collection may be necessary.

These findings suggest that transformer-based approaches warrant further investigation for cardiovascular disease detection, but several limitations must be addressed before clinical translation is conceivable. Future work should focus on:

1 Prospective validation on independently collected clinical data with expert annotation and diverse patient populations.

2 Multi-class pathology classification beyond binary normal/abnormal discrimination.

3 Multi-lead ECG integration to leverage full clinical information.

4 Advanced imbalance mitigation including data augmentation, generative models for minority classes, and cost-sensitive architectures.

5 Interpretability enhancement through attention visualisation and saliency mapping to build clinical trust.

6 Robustness characterisation across data quality variations and acquisition devices.

7 External validation on completely independent datasets from different institutions.

While the results are encouraging, they represent research findings requiring substantial additional validation before any claims of clinical utility can be substantiated. The corrected PTB results and improved MIT-BIH performance demonstrate both the potential of transformer architectures and the critical importance of rigorous methodology in medical machine learning research.

Declarations

All authors declare that they have no conflicts of interest.

References

Ali, F., El-Sappagh, S.H., Islam, S.M.R., Kwak, D., Ali, A., Imran, M. and Kwak, K-S. (2020) 'A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion', *Inf. Fusion*, Vol. 63, pp.101–120.

Ali, L., Rahman, A., Khan, A., Zhou, M., Javeed, A. and Khan, J.A. (2019) 'An automated diagnostic system for heart disease prediction based on χ^2 statistical model and optimally configured deep neural network', *IEEE Access*, Vol. 7, pp.64638–64651.

Al-Makhadmeh, Z. and Tolba, A. (2019) 'Utilizing IoT wearable medical device for heart disease prediction using higher order Boltzmann model: a classification approach', *Measurement*, Vol. 132, pp.524–534.

Bharti, R., Khamparia, A., Shabaz, M., Dhiman, G., Pande, S. and Singh, P. (2021) 'Prediction of heart disease using a combination of machine learning and deep learning', *Computational Intelligence and Neuroscience*, Vol. 2021, Article No. 5597654, pp.1–12.

Bhattacharjee, D., Zhang, T., Süsstrunk, S. and Salzmann, M. (2022) 'MuIT: an end-to-end multitask learning transformer', *2022 IEEE/CVF Conference on Computer Vision and Pattern*, (IF: 3).

Kaushal, P., Singh, S. and Vijayvergiya, R. (2024) 'A kernel attention-based transformer model for survival prediction of heart disease patients', *Journal of Cardiovascular Translational Research*, Vol. 17, pp.1295–1306, DOI: 10.1007/s12265-024-10537-3.

Khan, M.A. (2020) 'An IoT framework for heart disease prediction based on MDCNN classifier', *IEEE Access*, https://doi.org/10.1109/ACCESS.2020.2974687.

Lin, A., Manral, N., McElhinney, P., Killekar, A., Matsumoto, H., Kwiecinski, J., Pieszko, K., Razipour, A., Grodecki, K., Park, C., Otaki, Y., Doris, M., Kwan, A.C., Han, D., Kuronuma, K., Tomasino, G.F., Tzolos, E., Shanbhag, A., Goeller, M., Marwan, M., Gransar, H., Tamarappoo, B.K., Cadet, S., Achenbach, S., Nicholls, S.J., Wong, D.T., Berman, D.S., Dweck, M., Newby, D.E., Williams, M.C., Slomka, P.J. and Dey, D. (2021) 'Deep learning-enabled coronary CT angiography for plaque and stenosis quantification and cardiac risk prediction: an international multicentre study', *The Lancet. Digital Health*, Vol. 3, No. 10, pp.e638–e649.

Maragatham, G. and Devi, S. (2019) 'LSTM model for prediction of heart failure in big data', *Journal of Medical Systems*, Vol. 43, No. 10, Article No. 66, pp.1–10.

Nancy, A.A., Ravindran, D., Vincent, P.M.D.R., Srinivasan, K. and Reina, D.G. (2022) 'IoT-cloud-based smart healthcare monitoring system for heart disease prediction via deep learning', *Electronics*, Vol. 11, No. 15, Article No. 2392, pp.1–22.

Rahman, A., Alsenani, Y., Zafar, A., Ullah, K., Rabie, K. and Shongwe, T. (2024) 'Enhancing heart disease prediction using a self-attention-based transformer model', *Scientific Reports*, Vol. 14, DOI: 10.1038/s41598-024-51184-7.

Rao, J.N. and Prasad, R.S. (2021) 'An ensemble deep dynamic algorithm (EDDA) to predict the heart disease', *International Journal of Scientific Research in Science, Engineering and Technology (IJSRSET)*, Vol. 8, No. 1, pp.105–111, DOI: 10.32628/IJSRSET218118.

Sumon, M.S.I., Sakib, M., Rahman, M.S., Hossain, S., Khandakar, A., Hasan, A. and Murugappan, M. (2025) *CardioTabNet: A Novel Hybrid Transformer Model for Heart Disease Prediction using Tabular Medical Data*, DOI: 10.48550/arXiv.2503.17664.

Yu, P., Yi, J., Huang, T., Xu, Z. and Xu, X. (2024) *Optimization of Transformer Heart Disease Prediction Model based on Particle Swarm Optimization Algorithm*, DOI: 10.48550/arXiv.2412.02801.