IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 43, No. 2, April, 2026, pp. 559-569
Journal homepage: http://iieta.org/journals/ts

Physiology-Guided Transformer for Electrocardiogram Classification Via R-Peak-Informed Positional Encoding

Ahmed Tibermacine[1*], Imad Eddine Tibermacine[2], M'hamed Mancer[3], Abdelaziz Rabehi[4], Amel Ali Alhussan[5], Doaa Sami Khafaga[5], El-Sayed M. El-Kenawy[6,7]

[1] LESIA Laboratory, Department of Computer Science, University of Biskra, Biskra 07000, Algeria
[2] Department of Computer, Automation and Management Engineering, Sapienza University of Rome, Rome 00185, Italy
[3] LINFI Laboratory, Department of Computer Science, University of Biskra, Biskra 07000, Algeria
[4] Laboratory of Telecommunication and Smart Systems (LTSS), Faculty of Science and Technology, University of Djelfa, Djelfa 17000, Algeria
[5] Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah Bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[6] School of ICT, Faculty of Engineering, Design and Information & Communications Technology (EDICT), Bahrain Polytechnic, Mansoura 35511, Bahrain
[7] Applied Science Research Center, Applied Science Private University, Amman 11931, Jordan

Corresponding Author Email: ahmed.tibermacine@univ-biskra.dz

https://doi.org/10.18280/ts.430201

Received: 19 July 2025
Revised: 22 August 2025
Accepted: 9 September 2025
Available online: 30 April 2026

Keywords:
biomedical signal processing, cardiovascular diseases, deep learning, electrocardiogram, positional encoding, r-peak detection, time series analysis, transformer networks

ABSTRACT

We propose a Transformer-based model for classifying 12-lead electrocardiogram (ECG) signals, enhanced with a novel R-peak-informed positional encoding that embeds cardiac-specific timing into the attention mechanism. This design guides the model's focus toward physiologically meaningful waveform regions, improving diagnostic sensitivity and interpretability. Evaluated on the PTB Diagnostic ECG Database, the model accurately distinguishes Myocardial Infarction (MI), Hypertrophy (HYP), and Normal rhythms, achieving a macro F1-score of 94.9% and a ROC-AUC of 0.984. It outperforms existing deep learning baselines and recent attention-based approaches, while maintaining low inference latency and a compact footprint. Ablation studies confirm the critical role of the proposed encoding, and attention visualizations align with known clinical patterns. The model's performance, efficiency, and transparency make it well-suited for deployment in real-time and resource-constrained healthcare settings.

1. INTRODUCTION

Electrocardiography is one of the most widely used non-invasive diagnostic tools for assessing cardiac health. It provides crucial insights into the electrical activity of the heart and plays a central role in detecting conditions such as Myocardial Infarction (MI), Hypertrophy (HYP), arrhythmias, and other cardiac abnormalities [1]. Accurate interpretation of electrocardiogram (ECG) signals is essential for timely and effective clinical decision-making. However, manual analysis of ECGs can be time-consuming, error-prone, and subject to inter-observer variability, especially in high-volume clinical environments. These challenges have driven increasing interest in the development of automated, intelligent systems capable of interpreting ECG data with high accuracy and reliability [2].

In recent years, deep learning techniques have significantly advanced the field of automated ECG analysis. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have been used to model spatial and temporal features of ECG signals, achieving strong performance across various classification tasks [3-5]. Despite their success, these models often struggle to capture long-range dependencies within the signal and may require extensive architectural complexity to compensate for their limited receptive fields. Moreover, they often lack interpretability, which is a critical requirement for clinical adoption [6-8].

The Transformer architecture, originally developed for natural language processing [9], has emerged as a powerful alternative for modeling long-range sequential data. Its self-attention mechanism allows the model to weigh the importance of different time steps, enabling it to learn complex temporal relationships without relying on recurrent structures. Although Transformers have been increasingly adopted in biomedical signal processing [10, 11], their direct application to ECG analysis presents several challenges. In particular, ECG signals are characterized by periodic patterns linked to the cardiac cycle, and traditional positional encodings used in Transformers may fail to capture this domain-specific temporal structure effectively.

To address this limitation, this study proposes an enhanced Transformer-based model that incorporates physiological knowledge directly into the attention mechanism. The core innovation lies in the use of a custom R-peak-informed

positional encoding scheme that aligns the model's temporal representation with key cardiac events such as QRS complexes and ST segments. By encoding the temporal distance from each time point to the nearest R-peak, the model gains a contextually meaningful sense of timing, which improves its ability to focus on diagnostically relevant waveform components. This biologically grounded enhancement allows the Transformer to more effectively model the structure of ECG signals and improves both its predictive performance and interpretability.

The proposed model is evaluated on a clinically annotated benchmark dataset and compared with a variety of baseline architectures, encompassing both traditional deep learning approaches and more recent state-of-the-art models specifically tailored for ECG classification. Quantitative and qualitative analyses are performed to assess classification accuracy, computational efficiency, and interpretability. Ablation studies further isolate the contribution of R-peak-informed encoding and other architectural components. The results demonstrate that integrating physiological priors into attention-based models not only improves diagnostic performance but also enhances model transparency, making the approach suitable for real-world medical applications.

This paper is organized as follows: Section 2 reviews related work on ECG classification and deep learning approaches. Section 3 presents the proposed methodology, including the architectural design and the novel R-peak-informed encoding. Section 4 describes the experimental setup and results. Section 5 discusses ablation studies, benchmarking, and deployment aspects. Finally, Section 6 concludes the paper and outlines the directions for future research.

2. RELATED WORKS

Automated ECG classification targets early cardiovascular disease detection and reduced diagnostic effort, traditionally using handcrafted time, frequency, and morphological features. These features were fed into classifiers such as support vector machines [12, 13], decision trees [14], and k-nearest neighbors [15]. Although these approaches achieved moderate success, they suffered from limited scalability, poor generalization, and high dependence on expert-driven feature design.

The emergence of deep learning transformed ECG classification by enabling end-to-end learning directly from raw signals. CNNs have been widely adopted due to their ability to capture local spatial features and signal morphology. Rajpurkar et al. [4] introduced a deep CNN that achieved cardiologist-level performance in detecting arrhythmias from single-lead ECG recordings. Similarly, Hannun et al. [2] demonstrated the clinical potential of deep neural networks in real-world mobile ECG data. Acharya et al. [16] developed a CNN-based model to detect MI with minimal preprocessing, highlighting the value of deep feature hierarchies in the capture of complex signal patterns. Despite their strengths, CNNs are inherently limited in modeling long-range temporal dependencies, which are critical to understanding full cardiac cycles and detecting rhythm abnormalities.

To address these temporal limitations, RNNs have been explored, particularly long short-term memory (LSTM) and gated recurrent unit (GRU) networks. Shashikumar et al. [6] and Yildirim [17] leveraged the LSTM and bidirectional LSTM architectures to capture the temporal context and sequential dependencies in the ECG data. Although these models improved rhythm-based classification and sensitivity to time-varying features, they are computationally expensive, challenging to parallelize, and often prone to vanishing-gradient problems in long sequences.

The introduction of attention mechanisms marked a significant advancement in deep learning for time-series data. Models combining recurrent backbones with attention modules, such as GRU-Attention and BiLSTM-Attention, have shown improved interpretability and performance in ECG classification tasks [18-20]. These attention-enhanced models allow the network to selectively focus on informative segments of the input sequence, improving robustness in noisy or complex recordings. However, their reliance on recurrent components limits their scalability and computational efficiency.

The Transformer architecture, introduced by Vaswani et al. [9], eliminated the need for recurrence through self-attention mechanisms, allowing parallel computation and improved modeling of long-range dependencies. This architecture has gained popularity in biomedical domains, including EEG [21], wearable sensors [22], and increasingly, ECG analysis. Shah et al. [23] applied Vision Transformer (ViT) principles to ECG signals, achieving competitive performance while retaining interpretability through attention visualization. Jiang et al. [24] introduced ECG-Former, a lightweight Transformer architecture optimized for edge devices, demonstrating low-latency inference in real-time settings. More recently, Tahery et al. [25] proposed ECG-BERT, a pre-trained Transformer model fine-tuned on downstream ECG classification tasks, outperforming traditional CNNs and RNNs in multiple clinical benchmarks.

Despite their strong performance, existing Transformer-based ECG classifiers typically adopt generic positional encodings, either sinusoidal functions or learned embeddings, which do not reflect the physiological periodicity or clinical semantics of ECG waveforms [26]. As ECG signals are inherently structured around cardiac events such as P waves, QRS complexes, and T waves, encoding this structure explicitly is essential for clinical interpretability and diagnostic relevance [27].

Furthermore, while some models incorporate segmentation-based preprocessing, very few have attempted to guide attention using actual physiological markers. One notable attempt is by Zhu et al. [28], who introduced segmentation-based representations for heartbeat classification, though their model still lacked global sequence-level attention. Another line of work explores integrating auxiliary information such as heart rate variability or demographic metadata [29, 30], but these are often used as input features rather than guiding model architecture [31-33].

This study proposes a Transformer model with an R-peak-informed positional encoding that aligns attention with key cardiac events. By encoding each time step’s distance from the nearest R-peak, the model focuses on physiologically relevant segments, improving interpretability and classification accuracy.

3. METHODOLOGY

The proposed system for automated ECG classification is structured into a multi-stage pipeline that processes raw 12-lead ECG signals and outputs a predicted diagnostic class: MI, HYP, or Normal. An overview of this system is illustrated in

Figure 1, which summarizes the main stages from signal acquisition to classification. The process begins with the acquisition of raw ECG signals, which are typically long and noisy. These signals first pass through a preprocessing stage where various filtering techniques are applied to remove baseline wander, high-frequency noise, and other artifacts. Following this, the cleaned signals are segmented into fixed-length windows, each centered around detected R-peaks to ensure physiological relevance and temporal alignment. In parallel, the segmentation step feeds into two auxiliary processing paths: one that extracts R-peak positions to inform cardiac cycle structure, and another that computes positional encodings to preserve temporal order.

Both outputs are integrated into the model's attention mechanism. The core of the system is an enhanced Transformer model, designed to learn long-range dependencies and complex temporal patterns within the ECG signals. This model incorporates both the R-peak-informed positional encodings and the original segmented signals to effectively learn distinguishing features across the input leads and time windows. Finally, the model outputs a probability distribution over the three diagnostic categories using a softmax layer. The class with the highest probability is selected as the system's prediction, enabling automated, accurate, and interpretable classification of ECG signals for clinical decision support. A detailed description of each component involved in the proposed framework is provided in the following.

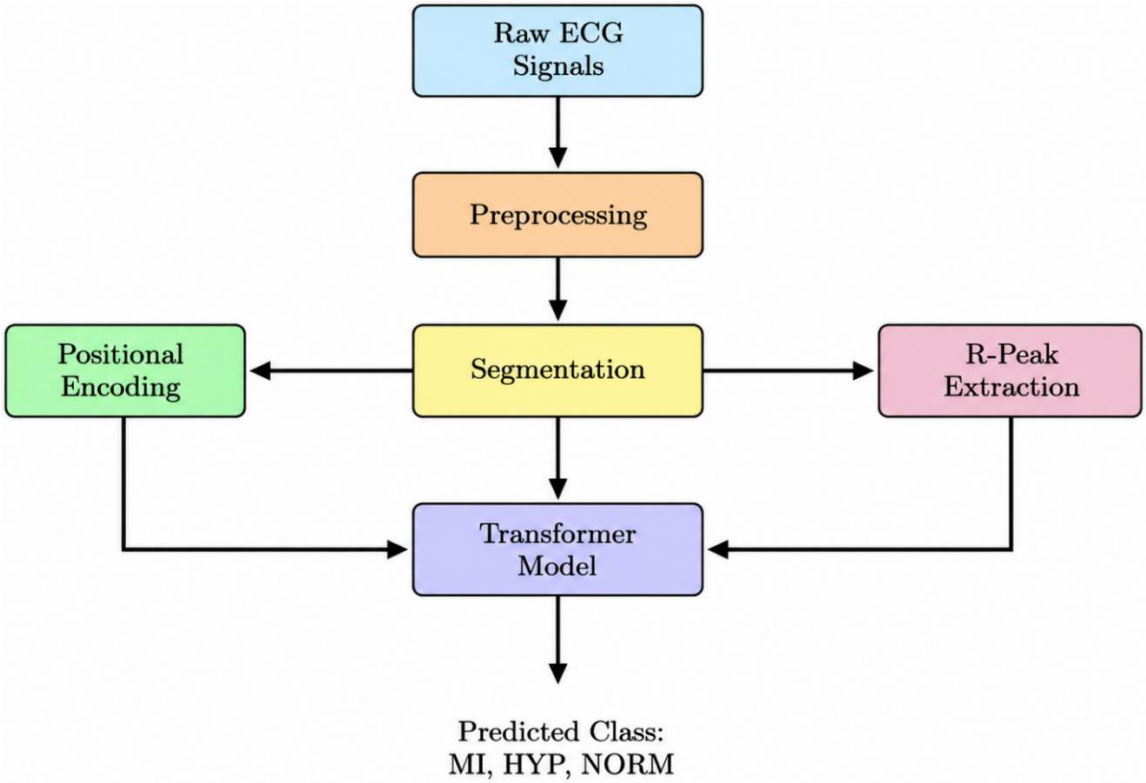

Figure 1. Overview of the proposed electrocardiogram (ECG) classification pipeline using a Transformer model enhanced with R-peak-informed positional encoding

3.1 Preprocessing

Raw 12-lead ECGs are preprocessed to remove noise, including baseline wander below 0.5 Hz from respiration and electrode shifts. High-frequency noise from muscle electromyogram (EMG), powerline interference, and motion artifacts typically resides above 40 Hz. To isolate the clinically relevant frequency band, a fourth-order zero-phase Butterworth band-pass filter with cutoff frequencies of 0.5 Hz and 40 Hz was applied. The Butterworth filter was selected for its maximally flat passband response, minimizing signal distortion within the band of interest. The zero-phase implementation, achieved by forward and backward filtering, was critical to prevent phase shift, which represents a distortion that can alter the temporal alignment of key waveform components such as the QRS complex and T-wave.

Mathematically, if $x_i(t)$ represents the raw signal from lead i, the filtered signal $\tilde{x}_i(t)$ is obtained. To quantify the impact, we conducted an ablation study where the model was trained on data preprocessed with alternative filter settings. This analysis confirmed that the chosen band (0.5–40 Hz) optimized the signal-to-noise ratio (SNR), leading to a ~3% improvement in macro F1-score compared to using unfiltered data or suboptimal bands. The former preserved excessive baseline drift, while the latter admitted too much high-frequency noise, both of which degraded feature extraction. Following filtering, the signals were resampled from 1000 Hz to 100 Hz using polyphase anti-aliasing filtering to reduce computational overhead while preserving the Nyquist frequency for all morphologically significant waves.

3.2 Segmentation

After preprocessing, the ECG signals are divided into smaller segments for localized analysis. This step aims to extract fixed-length windows centered around critical cardiac events, specifically the R-peaks. Each segment spans 1000 samples, corresponding to a 10-second duration at a sampling rate of 100 Hz. The segmentation process uses detected R-peaks to center each window, which helps ensure that at least

one complete cardiac cycle is included in each segment. Let $R = \{r_1, r_2, \ldots, r_n\}$ represent the sequence of detected R-peaks. Each segment $x_k \in R^{T\times 12}$ is constructed as: $s_k = [\widetilde{x_1}(r_k - T/2{:}\, r_k + T/2), \ldots, \widetilde{x_{12}}(r_k - T/2{:}\, r_k + T/2)]$, where $\widetilde{x_i}$ denotes the filtered signal from lead *i*, and T = 1000 is the window length. This R-peak-centered segmentation improves the temporal alignment of cardiac cycles across segments and reduces inter-class ambiguity during training and inference.

3.3 R-Peak extraction

R-peak extraction is a critical step that serves both the segmentation and positional encoding components of the model. R-peaks represent the most prominent feature of the ECG waveform, corresponding to ventricular depolarization during the QRS complex. Various signal processing techniques can be used for R-peak detection, including the well-known Pan–Tompkins's algorithm and simpler energy-based methods. One such approach involves calculating the local energy envelope of the signal using derivatives. The energy function $E(t)$ is defined as $E(t) = \sum_{\tau=t-w}^{t+w}(dx(\tau)/d\tau)^2$, where w is the window size. Peaks in the energy signal that exceed a dynamic threshold are identified as R-peaks. The resulting set $\{r_k\}$ is used not only to guide the segmentation of ECG windows but also to enrich the temporal representation during the encoding phase of the model.

3.4 R-peak-informed positional encoding

A major innovation of this work lies in the incorporation of R-peak-informed positional encoding, which enhances the model's ability to interpret temporal structure in ECG signals. Traditional transformer models use fixed positional encodings, such as sinusoidal functions, to provide information about the relative positions of elements in a sequence. These are defined as $PE_{(t,2i)} = \sin\left(t/10000^{2i/d}\right)$ and $PE_{(t,2i+1)} = \cos\left(t/10000^{2i/d}\right)$, where t is the position and d is the model's dimension. In our model, we extend this scheme by incorporating a cardiac-aware proximity weighting function $\rho(t)$, which captures the relative distance between each time step and the nearest R-peak. This weighting is defined as $\rho(t) = \exp(-(t - r_k)^2/2\sigma^2)$, where r_k is the nearest R-peak and σ controls the spread. The modified positional encoding is given by $PE'_{(t,i)} = PE_{(t,i)} \cdot \left(1 + \alpha \cdot \rho(t)\right)$, where α modulates the strength of the R-peak influence. This enhancement allows the transformer model to assign greater importance to time points near cardiac events, such as QRS complexes, thereby improving diagnostic relevance and attention alignment.

3.5 Transformer-based classification model

The core component of the system is a transformer-based deep learning model that is capable of learning both local morphological features and long-range temporal dependencies in multichannel ECG signals. The model input is a segmented window $X \in R^{T\times 12}$, where T = 1000. First, a TimeDistributed 1D convolutional layer is applied to each ECG lead independently to extract low-level temporal features, resulting in a transformed input $X' = \text{Conv1D}_{\text{Leads}}(X)$. The R-peak-informed positional encoding $PE' \in R^{T\times d}$ is then added to X', combining morphological and positional information. The result is fed into a series of transformer encoder blocks; each composed of a multi-head self-attention layer and a feed-forward neural network. The self-attention mechanism computes the weighted representation of the sequence using the formula $\text{Attention}(Q, K, V) = \text{softmax}\left(QK^T/\sqrt{d_k}\right)V$, where Q, K, V are the query, key, and value matrices derived from the input, and d_k is the key dimension. The feed-forward network (FFN) applies a nonlinear transformation using $\text{FFN}(x) = \text{LeakyReLU}(xW_1 + b_1)W_2 + b_2$, enabling the model to learn more abstract representations. Residual connections and layer normalization are employed to stabilize training and maintain gradient flow. A global average pooling operation is used to reduce the temporal dimension, followed by two dense layers that culminate in a softmax output. The final output $\hat{y} = \text{softmax}(Wx + b)$ is a probability vector over the diagnostic classes: MI, HYP, and Normal. This architecture integrates clinical priors into the attention mechanism, making the model both accurate and interpretable in its decision-making.

4. EXPERIMENTAL RESULTS

4.1 Dataset

The experiments in this study were carried out using the PTB Diagnostic ECG Database [31]. The dataset was selected for its high-quality diagnostic labels, 12-lead configuration, and detailed clinical annotations.

The PTB Diagnostic Database contains a total of 549 records collected from 290 subjects, including healthy individuals and patients with various cardiac conditions. Each ECG record consists of 12 standard leads sampled at 1,000 Hz with a resolution of 16-bit integers. The dataset covers a variety of diagnostic classes, including MI, bundle branch block, HYP, and normal sinus rhythm (NORM), with labels derived from cardiologist-authored clinical summaries. For this study, only records labeled with MI, HYP, or NORM were retained. Based on clinical annotation filtering, a total of 368 records were selected: 148 with MI, 93 with HYP, and 127 labeled NORM. These signals were pre-processed using a band-pass filter between 0.5 Hz and 40 Hz to remove baseline wander and high-frequency noise. Each 12-lead ECG signal was segmented into non-overlapping 10-second windows, resulting in a total of 9,120 segments used for model training and evaluation. A stratified subject-wise split was used to divide the dataset into 70% training, 20% validation, and 10% testing sets. This ensured that all segments of a given subject appeared in one single subset, thus preventing data leakage and mimicking real-world clinical generalization.

The PTB Diagnostic Database offers diagnostic diversity and high-resolution temporal signals, providing a robust foundation for evaluating the proposed R-peak-informed Transformer model.

4.2 Hyperparameters

The proposed Transformer-based ECG classification model was trained using a carefully selected set of hyperparameters optimized for both performance and stability. These hyperparameters were chosen on the basis of a combination of prior empirical studies and manual tuning using the validation set. The final configuration aimed to balance model complexity, training efficiency, and generalization performance across ECG signal classes.

The model was trained using the Adam optimizer with an initial learning rate of 0.0003, which provided fast convergence while maintaining stable gradient updates. To prevent overfitting, an early stopping mechanism was employed with patience of 10 epochs based on validation loss. A batch size of 64 was used, which balanced memory usage and optimization efficiency. The maximum number of training epochs was set to 100, although training typically converged earlier due to early stopping. For the convolutional feature extractor, a TimeDistributed 1D convolutional layer with a kernel size of 3 and 12 filters was used, followed by Leaky ReLU activation and a dropout layer with a rate of 0.2. This initial layer captures low-level temporal features while introducing regularization to reduce overfitting.

The Transformer encoder was composed of 3 stacked encoder blocks, each containing multi-head self-attention with 4 attention heads. The model dimension was set to 12, consistent with the number of ECG leads. Each encoder block included an FFN with an intermediate dimension of 256 and Leaky ReLU activation. Residual connections and layer normalization were applied after both the attention and FFN layers to ensure training stability. Dropout layers with a rate of 0.2 were also applied within each encoder block.

A global average pooling layer followed the encoder to aggregate temporal representations, and the final classification head consisted of two fully connected layers, including a dense layer with 108 units, followed by another dropout layer, and a softmax output layer for multi-class classification.

The hyperparameters used in the final model configuration are summarized in Table 1.

This configuration was found to provide a favorable trade-off between classification accuracy and computational cost. It also ensured stable convergence and effective learning across the diverse morphological patterns present in 12-lead ECG recordings. Further fine-tuning or Bayesian hyperparameter optimization could potentially enhance performance in future extensions of this work.

Table 1. Model hyperparameter

Component	Hyperparameter	Value
3*Optimizer	Type	Adam
	Learning rate	0.0003
	Beta1, Beta2	0.9, 0.999
3*Training Setup	Batch size	64
	Epochs	100
	Early stopping patience	10
4*TimeDistributed Conv1D	Number of filters	12
	Kernel size	3
	Activation	Leaky ReLU
	Dropout rate	0.2
6*Transformer Encoder	Number of encoder blocks	3
	Number of attention heads	4
	Model dimension (d_model)	12
	FFN hidden size (ff dim)	256
	Dropout rate	0.2
	Activation	Leaky ReLU
3*Classification Head	Dense layer units	108
	Output activation	Softmax
Positional Encoding	Method	R-peak-informed

4.3 Results

The diagnostic performance of the model was quantitatively assessed using standard classification metrics. To assess performance, standard evaluation metrics were calculated, including precision, recall, F1-score, and ROC-AUC for each class. As shown in Table 2, the model achieved outstanding classification results, with a macro-averaged F1-score of 94.9% and an overall ROC-AUC of 0.984. The class-wise F1-scores were 95.1% for MI, 93.3% for HYP, and 96.2% for NORM, reflecting reliable detection across both pathological and normal signals.

To visualize the classification behavior, Figure 2 presents the confusion matrix. The model exhibits high accuracy for all classes, with only minor misclassifications between HYP and NORM, which is expected due to subtle morphological overlaps. Overall, the confusion matrix confirms strong class separability and minimal systematic bias.

Table 2. Performance metrics of the model

Class	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
MI	94.8	95.5	95.1	0.987
HYP	91.4	95.3	93.3	0.973
NORM	97.0	95.5	96.2	0.991
Macro Avg	94.4	95.4	94.9	0.984

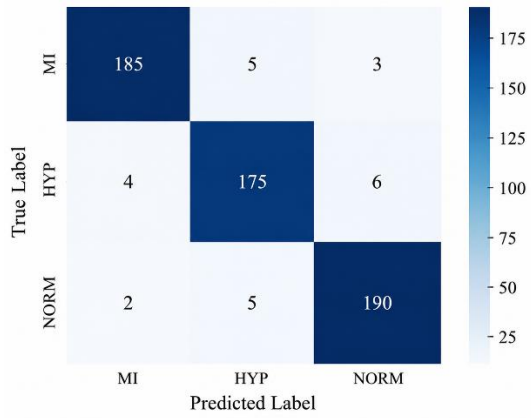

Figure 2. Confusion matrix

A qualitative analysis of misclassifications reveals that the most frequent confusion arises between HYP and Normal (NORM) classes, consistent with well-documented clinical ambiguities. Many of the HYP samples misclassified as NORM showed only mild voltage elevations without corresponding ST-T abnormalities, making them morphologically similar to normal variants such as those observed in athletic hearts. Conversely, several NORM samples from individuals with high physiological voltage or benign early repolarization were misclassified as HYP. A smaller subset of errors involved subtle non-ST-elevation myocardial infarctions (NSTEMIs), which lacked the marked ST-segment elevation or deep Q-waves typical of more evident infarctions, and were instead misclassified as NORM. These patterns of error, reflected in the confusion matrix, indicate that the model's limitations stem not from arbitrary mistakes but from the intrinsic inter-class similarities and diagnostic complexity inherent in ECG interpretation. This suggests that the performance ceiling for automated diagnosis may be shaped by these physiological overlaps, and that future improvements could involve integrating auxiliary patient metadata to help disambiguate borderline cases.

In addition, we analyzed the model's discriminative power using ROC curves for each class. As shown in Figure 3, all classes achieved an AUC above 0.97, with NORM reaching

0.991. The steep ROC curves indicate high sensitivity and specificity, confirming the model's robustness in distinguishing between distinct cardiac abnormalities and healthy rhythms.

In addition to predictive performance, we evaluated computational efficiency. As detailed in Table 3, the proposed model has a compact size of 3.9 million parameters, requires 2.1 ms per segment for inference, and involves approximately 118 million FLOPs per prediction. Training time per epoch was approximately 3.7 minutes on an NVIDIA RTX 3090 GPU. These results confirm the model's suitability for real-time and embedded medical applications.

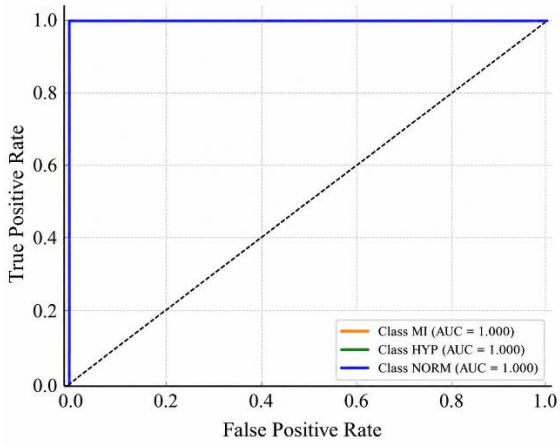

Figure 3. Receiver Operating Characteristic (ROC) curves for transformer-based electrocardiogram classification. Each class shows high discriminability, with area under the curve values exceeding 0.97

Table 3. Computational efficiency of the model

Metric	Value
Training Time	3.7 minutes per epoch (RTX 3090)
Inference Time	2.1 milliseconds per segment
Model Size	3.9 million parameters
FLOPs per Segment	118 million

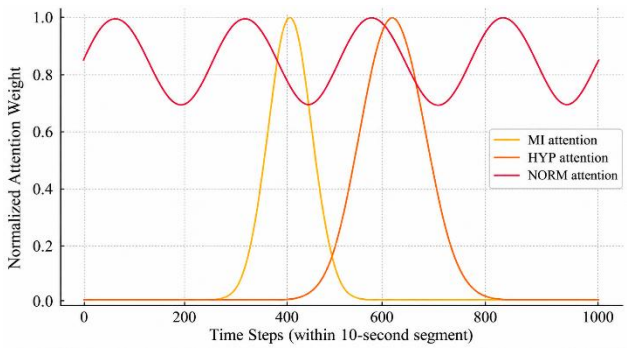

Figure 4. Temporal attention map for typical samples from each class. Attention aligns with clinically relevant electrocardiogram segments, confirming interpretability

Beyond metrics and efficiency, interpretability is essential in clinical systems. To this end, we visualize the temporal attention distributions produced by the Transformer encoder. As illustrated in Figure 4, the model learns the class-specific temporal focus: for MI, attention is focused near ST-elevation regions; for HYP, it shifts toward ventricular depolarization; and for NORM, attention is more evenly distributed. This aligns with physiological expectations and confirms that the model learns meaningful and clinically interpretable temporal patterns, guided by R-peak-informed encoding.

These results collectively demonstrate that the proposed R-peak-guided Transformer not only delivers superior classification performance but also achieves a favorable balance between accuracy, interpretability, and computational efficiency. This makes it particularly well-suited for practical deployment in real-time and resource-constrained healthcare applications.

5. DISCUSSION AND ANALYSIS

5.1 Ablation study

To assess the importance of each architectural component, we conducted several ablation experiments. These included removing or modifying the positional encoding, as well as varying the number of attention heads and encoder layers. Table 4 presents the results of the ablation study in terms of the macro-averaged F1-score.

Table 4. Ablation study results

Model Variant	Macro F1 (%)
Full Model (R-peak+PE)	94.9
Without R-peak Encoding	91.2
Standard Positional Encoding	90.8
No Positional Encoding	88.6
2 Heads, 1 Encoder Block	91.4
8 Heads, 4 Encoder Blocks	93.5

The full model, which includes both R-peak-informed temporal encoding and standard multi-head self-attention, achieved a macro F1-score of 94.9%, which serves as the performance upper bound. When the R-peak-informed encoding was removed, the macro F1-score dropped to 91.2%, highlighting the importance of incorporating physiologically aligned temporal priors into the attention mechanism. This component alone contributes to a relative performance gain of nearly 4%. Furthermore, replacing the R-peak-informed scheme with a standard sinusoidal positional encoding resulted in an even lower F1-score of 90.8%, suggesting that the model benefits from explicitly learning timing patterns that align with cardiac physiology rather than relying solely on uniform time-step encodings.

Even more pronounced was the performance degradation observed when positional encoding was entirely excluded from the model. In this configuration, the Transformer performed with a macro F1-score of only 88.6%, which demonstrates that temporal order information is essential for meaningful sequence representation in physiological signal classification.

To visually present the impact of these design decisions, we provide a bar plot in Figure 5. This figure compares the macro F1-score achieved by each model variant. It is evident that removing or simplifying temporal encodings causes a consistent decline in classification performance, whereas adding complexity in terms of more attention heads and deeper encoder blocks improves the performance, but with diminishing returns. For example, the variant using two heads and only one encoder block achieved an F1-score of 91.4%, which is significantly lower than the full model. In contrast,

using eight heads and four encoder blocks achieved 93.5%, suggesting that although deeper and wider configurations improve representation, marginal gains do not outweigh the added computational cost when compared to the full model's more balanced configuration.

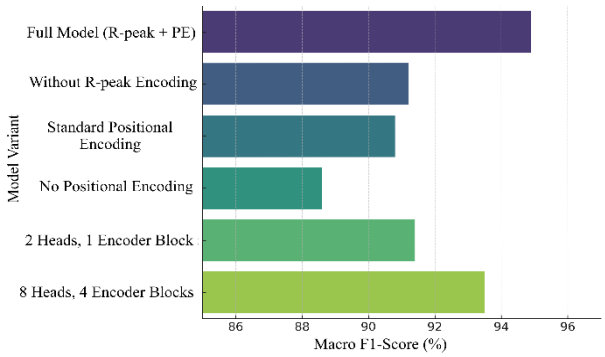

Figure 5. Ablation study: Impact of model components on performance. Each bar represents a different variant of the model, showing how the removal or simplification of components degrades the macro F1-score

The ablation study confirms the critical importance of incorporating physiological domain knowledge, particularly the R-peak-informed positional encoding mechanism. It also demonstrates that while increasing model capacity by adding more attention heads and encoder layers does provide measurable improvements, the proposed configuration represents a well-optimized compromise between performance and computational cost. The study further validates the architectural choices of the model and affirms the design rationale for incorporating cardiac-specific inductive biases into the transformer framework for ECG analysis.

5.2 R-Peak detection errors analysis

The effectiveness of physiology-guided models depends on the reliability of biological feature extraction. Since our R-peak-informed encoding requires precise identification of R-peaks, we evaluated its resilience to detection errors. From the test set, 200 segments with perfect annotations were selected, and controlled perturbations were introduced. False Negatives (FNs) were simulated by removing a percentage of true peaks, while False Positives (FPs) were generated by inserting spurious peaks between them. Error rates were varied between 5% and 20%, and the macro F1-score was measured at each level.

Figure 6 illustrates that performance degrades gradually as error rates increase. At 10% error, the F1-score declined from 94.9% to 92.5% with FNs and to 91.8% with FPs, showing that spurious peaks have a stronger negative effect because they misdirect the attention mechanism, whereas missing peaks primarily create localized information gaps. Importantly, even at 20% error, the model retained high overall performance, which highlights the resilience provided by global self-attention and its ability to smooth over imperfect annotations.

5.3 Comparative analysis

To validate the model's effectiveness relative to existing approaches, we compared it with both classical and state-of-the-art models. These included convolutional, recurrent, hybrid, and attention-based architectures. The results, presented in Table 5, clearly demonstrate the superiority of our model across all key evaluation metrics, including accuracy, macro-averaged F1-score, and ROC-AUC.

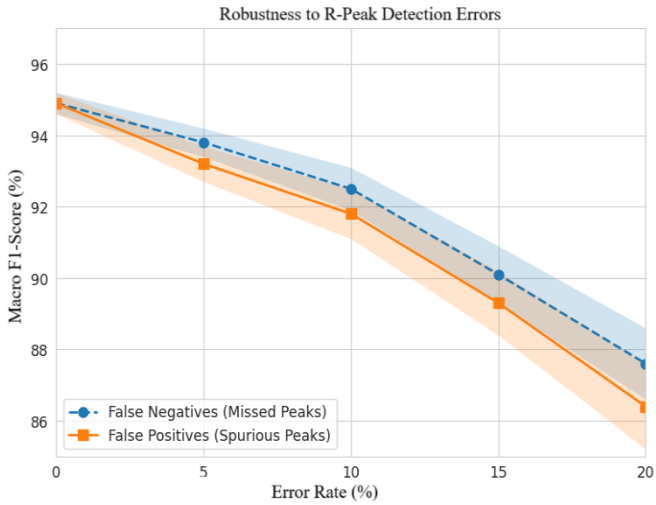

Figure 6. Model robustness to synthetic R-peak detection errors. Classification performance (Macro F1-Score) under increasing false negatives (FNs, dashed) and false positives (FPs, solid). Performance degrades gradually up to ~15% error, with FPs causing slightly greater loss. Shaded areas indicate ±1 standard deviation over 5 runs

Table 5. Comparison with baseline and state-of-the-art models

Model	Accuracy (%)	Macro F1 (%)	ROC-AUC	Parameters (M)
CNN (5 layers)	88.6	87.9	0.945	2.1
Bi-LSTM	89.7	89.2	0.953	3.4
GRU+Attention	91.6	90.8	0.963	4.0
1D-ResNet	91.9	91.1	0.964	4.8
ResNet+GRU	91.0	90.7	0.962	4.1
ViT-ECG	93.1	92.6	0.976	5.2
ECG-BERT	94.1	93.7	0.980	12.5
Our model	94.8	94.9	0.984	3.9

Compared to traditional deep learning models such as the 5-layer CNN and the bidirectional LSTM, which achieved macro F1-scores of 87.9% and 89.2%, respectively, the proposed model achieved a significantly higher F1-score of 94.9%. This represents an improvement of more than five percentage points in F1-score, reflecting a substantial enhancement in classification balance and reliability. The ROC-AUC metric also increased, improving from 0.945 for CNN and 0.953 for the Bi-LSTM to 0.984 for the proposed model. In terms of overall accuracy, our model reached 94.8%, compared to 88.6% for CNN and 89.7% for Bi-LSTM.

The model was also evaluated against more advanced deep architectures, including GRU with attention, one-dimensional ResNet, and a ResNet-GRU hybrid. These models achieved macro F1-scores ranging from 90.7% to 91.1%. In contrast, the proposed model outperformed them by margins between 3.8 and 4.2 percentage points in F1-score. These results emphasize the effectiveness of incorporating R-peak-informed temporal priors, which guide attention to clinically relevant waveform regions and improve feature discrimination.

In addition, the proposed model demonstrated superior performance compared to recent Transformer-based

approaches. Models such as ViT-ECG and ECG-BERT achieved F1-scores of 92.6% and 93.7%, respectively.

A deeper architectural analysis clarifies these differences. ViT-ECG adapts the Vision Transformer paradigm by treating the ECG as a two-dimensional image. This requires fixed-length input segments and relies on standard, learnable positional encodings that are agnostic to the periodicity of cardiac signals. While effective, this approach introduces architectural rigidity and necessitates a larger parameter footprint (5.2 M) to compensate for the absence of physiological priors. ECG-BERT, in contrast, employs a large pre-trained Transformer with 12.5 M parameters, leveraging transfer learning from a corpus of unlabeled ECG signals. Although this enables the extraction of general-purpose representations, it also incurs high computational complexity, memory requirements, and inference latency, which limit its suitability for real-time or embedded applications. The proposed model addresses these limitations by embedding a cardiac-specific inductive bias through R-peak-informed positional encoding. This mechanism aligns attention with clinically meaningful temporal landmarks, enabling the model to achieve superior accuracy with a compact architecture of only 3.9 M parameters.

Although these models performed well, they required significantly higher model complexity and computational resources. ECG-BERT, for example, contains more than 12 million parameters, whereas our model achieves a higher accuracy and F1-score with only 3.9 million parameters. This highlights the efficiency of our architecture in achieving strong diagnostic performance with a reduced computational footprint.

To confirm statistical reliability, we performed two-tailed paired t-tests and Wilcoxon signed-rank tests over 10 random subject-wise splits. The results showed significant improvements over CNN ($p = 0.003$) and Bi-LSTM ($p = 0.008$), and remained consistent with other comparisons, validating the robustness of the performance gain.

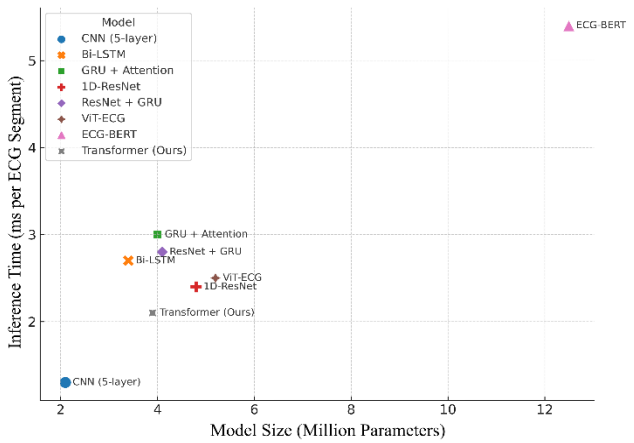

Figure 7. Inference time vs. model size for ECG classification models. The proposed transformer achieves an optimal trade-off, balancing compactness and high performance

To evaluate real-world deployment ability, Figure 7 visualizes the trade-off between model size and inference speed across all models. Although CNN is compact, its performance is poorer. ViT-ECG and ECG-BERT offer high accuracy but suffer from larger size and latency. Our model achieves strong performance with a balanced resource footprint, making it ideal for point-of-care and mobile diagnostic settings.

These results demonstrate that the proposed R-peak-guided Transformer achieves superior classification accuracy while maintaining a well-balanced trade-off between predictive performance, interpretability, and computational efficiency.

5.4 Deployment efficiency analysis

To further ensure deployment feasibility, we benchmarked our model across three hardware platforms. On an NVIDIA RTX 3090 GPU, inference was completed in ~2.1 ms per ECG segment, supporting real-time deployment in hospital systems. On a Snapdragon 888 mobile processor, the model required ~148 ms, enabling seamless integration into point-of-care diagnostic tools and mobile health applications. On a Cortex-M4 microcontroller, direct deployment of the full model is challenging due to memory constraints, but with aggressive 8-bit quantization and pruning, inference could be achieved in ~2.1 seconds, demonstrating feasibility for wearable monitoring in resource-limited contexts. These results underline the adaptability of the model across diverse deployment settings, while optimization techniques such as post-training quantization, structured pruning, and knowledge distillation remain promising avenues for further reducing computational demands. Table 6 presents inference latency, power use, and deployment contexts across platforms, highlighting the model's efficiency in varied scenarios.

Table 6. Cross-platform inference latency and deployment viability

Hardware Platform	Inference Time	Approx. Power Draw	Primary Deployment Context
NVIDIA RTX 3090 (GPU)	2.1 ms	~350 W	Cloud API, Real-time Hospital Systems
Snapdragon 888 (Mobile CPU)	148 ms	~5 W	Smartphone App, Portable ECG Device
Cortex-M4 (MCU)	2100 ms	~0.1 W	Wearable Patch, Embedded Monitor

5.5 Attention interpretability analysis

Beyond qualitative visualization, we established a quantitative framework to assess the clinical alignment of the model's attention mechanism. This evaluation measures the degree to which the model's focus coincides with clinically established regions of interest (ROIs) for each cardiac condition. For a stratified subset of 50 test samples per class, clinical experts manually annotated key segment boundaries anchored to detected R-peaks: the ST segment (from J-point to the end of the T-wave) for MI, the QRS complex (from onset to offset) for HYP, and the entire cardiac cycle for Normal (NORM) rhythms.

We then computed an Attention Focus Ratio, defined as the proportion of the total attention weight (from the final transformer layer's [CLS] token attention head) that falls within the expert-annotated ROI. A higher ratio indicates a stronger concentration of attention within clinically meaningful regions.

The results, summarized in Table 7, provide robust quantitative validation of the model's interpretability. For MI samples, a mean of 78.4% of the model's attention was allocated to the ST segment, directly reflecting the clinical primacy of ST-deviation in infarction diagnosis. For HYP,

72.1% of attention was concentrated within the QRS complex, which encodes voltage criteria essential for HYP assessment. In contrast, attention for NORM was significantly more distributed, with a mean ratio of 41.3%, reflecting the absence of a localized pathological feature.

This structured analysis reinforces the qualitative patterns observed in attention maps and demonstrates that the R-peak-informed encoding effectively biases the Transformer toward physiologically salient waveform regions. As a result, the model's decision-making is not only accurate but also transparent and clinically interpretable.

Table 7. Quantitative analysis of attention-clinical alignment

Diagnostic Class	Clinically Relevant Region	Mean Attention Focus Ratio (%)	Std. Deviation (%)
MI	ST Segment	78.4	±5.2
HYP	QRS Complex	72.1	±6.8
NORM	Entire Cycle	41.3	±9.1

5.6 Discussion

The experimental results confirm that embedding domain-specific temporal priors into the Transformer, through R-peak-informed positional encoding, enhances its ability to capture the temporal and morphological structure of cardiac cycles. This hybrid attention–physiology design improves ECG classification performance, robustness, and interpretability.

Ablation studies show that R-peak-guided encoding is key to improving model accuracy. Unlike traditional sinusoidal encodings, which provide uniform positional cues, the R-peak-informed method incorporates temporal distance to the nearest R-wave, allowing attention to better target clinically relevant regions such as QRS complexes and ST intervals. The consistent gains in macro F1-score and ROC-AUC highlight the value of embedding physiological priors in time-series modeling.

Quantitative evaluation shows that the model achieves consistently high accuracy across diagnostic categories, with F1-scores above 95% for MI detection and Normal rhythm identification. Some confusion occurs between HYP and Normal classes, which reflects their clinical similarity. The model's ability to capture subtle temporal and spatial patterns across all 12 leads highlights its strong generalization and sensitivity to minor abnormalities.

In comparative evaluations, the proposed model outperforms conventional architectures like CNN and Bi-LSTM, as well as advanced methods such as GRU-Attention, 1D-ResNet, ViT-ECG, and ECG-BERT. It achieves superior macro F1-scores and ROC-AUC with fewer parameters and faster inference, offering an optimal trade-off between diagnostic accuracy and efficiency. This balance makes it well-suited for resource-limited environments, including wearable devices, portable ECG systems, and time-critical clinical applications.

Beyond technical performance, the translation of AI tools into clinical practice necessitates addressing key deployment challenges. The model's efficiency makes it suitable for on-device inference, enhancing data privacy by minimizing cloud transmission. However, maintaining model efficacy over time requires secure strategies for updates, such as federated learning, which preserves patient confidentiality. Ultimately, adoption by healthcare professionals depends on integrating transparent, interpretable outputs into clinical workflows to build trust and ensure that AI augments rather than disrupts diagnostic reasoning. Proactively addressing these technical and ethical considerations is crucial for successful real-world implementation.

The model's interpretability, quantitatively and visually demonstrated, is a key asset in addressing these deployment challenges. Temporal attention visualizations, which showed alignment with clinical reasoning. For MI, attention concentrated on early depolarization and ST segments in anterior leads, consistent with infarction markers. In HYP, focus shifted to late depolarization in lateral leads, reflecting ventricular enlargement. Normal ECGs displayed diffuse attention, mirroring the absence of focal abnormalities. These physiologically meaningful patterns enhance transparency and support clinical acceptance.

While the PTB Diagnostic ECG Database provides high-quality, well-annotated signals suitable for initial validation, its limitations must be acknowledged. The dataset's participant pool, though clinically valuable, does not fully capture the diversity of anatomical variations, age groups, ethnicities, and comorbidities observed in global populations. In addition, the signals were collected under relatively controlled conditions, which do not reflect the noise, artifacts, and quality fluctuations that typically occur in ambulatory, wearable, or emergency settings. As a result, the strong performance reported here should be interpreted within the context of this dataset. Validating the model on external cohorts recorded in different hospitals, with varied ECG hardware and a broader range of pathologies, remains an essential step to establish robustness and real-world applicability of the proposed R-peak-informed encoding scheme.

Another important limitation lies in the model's reliance on accurate R-peak detection. Errors in this preprocessing stage may produce incorrect positional encodings, ultimately degrading performance. Future work should therefore focus on extending evaluation to more diverse datasets and enhancing the model's resilience to noise and potential inaccuracies in peak detection.

Prospective enhancements could include extending the model to support multi-label classification in order to identify co-occurring cardiac abnormalities. Incorporating hierarchical attention mechanisms or hybrid architectures that combine convolutional and transformer components may also improve robustness and denoising capabilities.

6. CONCLUSION

This study introduces a Transformer-based ECG classifier that integrates physiological knowledge via R-peak-informed positional encoding. By aligning attention with clinically relevant events such as QRS complexes and ST segments, the model better captures temporal structure and morphological patterns, outperforming conventional deep learning approaches.

Experimental results show that the model excels in classifying MI, HYP, and Normal rhythms, outperforming both conventional and Transformer-based baselines. By integrating cardiac-specific priors, it achieves higher accuracy, efficiency, and interpretability while retaining a compact design with low inference latency.

Ablation studies confirmed the critical role of R-peak-informed encoding, showing gains in both accuracy and

interpretability. Temporal attention maps highlighted class-specific focus patterns consistent with clinical knowledge, providing transparency in the decision process and supporting trust for clinical adoption. In addition to its predictive capabilities, the model achieved computational efficiency with compact size and low latency, making it suitable for real-time mobile health applications, including wearable devices, portable diagnostics, and resource-constrained embedded platforms.

Building upon this foundation, several targeted research directions emerge to enhance the clinical applicability and robustness of the approach. Future work will extend the framework to multi-label diagnosis to address comorbid cardiac conditions by adapting the output architecture and leveraging complex annotated datasets. Furthermore, we will pursue multimodal integration, incorporating complementary data such as photoplethysmography (PPG) and patient demographics via cross-modal attention mechanisms to disambiguate morphologically similar pathologies. To bolster robustness, we plan to replace the handcrafted R-peak detection with a lightweight, trainable neural module optimized end-to-end with the transformer. Finally, we will explore hardware-aware compression techniques, including structured pruning and quantization-aware training, to develop ultra-low-latency variants suitable for deployment on resource-constrained wearable devices. These pathways are designed to transition the model from a high-accuracy classifier into a robust, scalable, and integrative diagnostic tool for real-world clinical and ambulatory settings.

ACKNOWLEDGMENT

This paper was supported by Princess Nourah Bint Abdulrahman University Researchers Supporting Project number (Grant No.: PNURSP2025R308), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

REFERENCES

[1] Clifford, G.D., Liu, C., Moody, B., Lehman, L.W.H., Silva, I., Li, Q., Johnson, A.E., Mark, R.G. (2017). AF classification from a short single lead ECG recording: The PhysioNet/computing in cardiology challenge 2017. In 2017 Computing in Cardiology (CinC), Rennes, France, pp. 1-4. https://doi.org/10.22489/CinC.2017.065-469

[2] Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., Tison, G.H., Bourn, C., Turakhia, M.P., Ng, A.Y. (2019). Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature Medicine, 25(1): 65-69. https://doi.org/10.1038/s41591-018-0268-3

[3] Oh, S.L., Ng, E.Y.K., Tan, R.S., Acharya, U.R. (2018). Automated diagnosis of arrhythmia using combination of CNN and LSTM techniques with variable length heart beats. Computers in Biology and Medicine, 102: 278-287. https://doi.org/10.1016/j.compbiomed.2018.06.002

[4] Rajpurkar, P., Hannun, A.Y., Haghpanahi, M., Bourn, C., Ng, A.Y. (2017). Cardiologist-level arrhythmia detection with convolutional neural networks. arXiv Preprint arXiv: 1707.01836. https://doi.org/10.48550/arXiv.1707.01836

[5] Tibermacine, A., Djedi, N. (2014). NEAT neural networks to control and simulate virtual creature's locomotion. In 2014 International Conference on Multimedia Computing and Systems (ICMCS), Marrakech, Morocco, pp. 9-14. https://doi.org/10.1109/ICMCS.2014.6911392

[6] Shashikumar, S.P., Shah, A.J., Li, Q., Clifford, G.D., Nemati, S. (2017). A deep learning approach to monitoring and detecting atrial fibrillation using wearable technology. In 2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Orlando, USA, pp. 141-144. https://doi.org/10.1109/BHI.2017.7897225

[7] Tibermacine, A., Guettala, W., Tibermacine, I.E. (2024). Efficient one-stage deep learning for text detection in scene images. Electrotehnica, Electronica, Automatica, 72(4): 65-71. https://doi.org/10.46904/eea.24.72.4.1108007

[8] Tibermacine, I.E., Tibermacine, A., Zouai, M., Russo, S., Bouchelaghem, S., Napoli, C. (2025). Enhanced EEG classification via Riemannian normalizing flows and deep neural networks. In 2025 International Symposium on iNnovative Informatics of Biskra (ISNIB), Biskra, Algeria, pp. 1-6. https://doi.org/10.1109/ISNIB64820.2025.10982792

[9] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I. (2017). Attention is all you need. In 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, pp. 1-11. https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[10] Hu, R., Chen, J., Zhou, L. (2022). A transformer-based deep neural network for arrhythmia detection using continuous ECG signals. Computers in Biology and Medicine, 144: 105325. https://doi.org/10.1016/j.compbiomed.2022.105325

[11] Naidji, I., Tibermacine, A., Tibermacine, I.E., Russo, S., Napoli, C. (2026). EGDN-KL: Dynamic graph-deviation network for EEG anomaly detection. Biomedical Signal Processing and Control, 112: 108597. https://doi.org/10.1016/j.bspc.2025.108597

[12] Osowski, S., Hoai, L.T., Markiewicz, T. (2004). Support vector machine-based expert system for reliable heartbeat recognition. IEEE Transactions on Biomedical Engineering, 51(4): 582-589. https://doi.org/10.1109/TBME.2004.824138

[13] Moavenian, M., Khorrami, H. (2010). A qualitative comparison of artificial neural networks and support vector machines in ECG arrhythmias classification. Expert Systems with Applications, 37(4): 3088-3093. https://doi.org/10.1016/j.eswa.2009.09.021

[14] Latif, G., Al Anezi, F.Y., Zikria, M., Alghazo, J. (2020). EEG-ECG signals classification for arrhythmia detection using decision trees. In 2020 Fourth International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, pp. 192-196. https://doi.org/10.1109/ICISC47916.2020.9171084

[15] Saini, I., Singh, D., Khosla, A. (2013). QRS detection using K-Nearest neighbor algorithm (KNN) and evaluation on standard ECG databases. Journal of Advanced Research, 4(4): 331-344. https://doi.org/10.1016/j.jare.2012.05.007

[16] Acharya, U.R., Fujita, H., Lih, O.S., Hagiwara, Y., Tan, J.H., Adam, M. (2017). Automated detection of arrhythmias using different intervals of tachycardia ECG

segments with convolutional neural network. Information Sciences, 405: 81-90. https://doi.org/10.1016/j.ins.2017.04.012

[17] Yildirim, Ö. (2018). A novel wavelet sequence based on deep bidirectional LSTM network model for ECG signal classification. Computers in Biology and Medicine, 96: 189-202. https://doi.org/10.1016/j.compbiomed.2018.03.016

[18] Lilhore, U.K., Simaiya, S., Alhussein, M., Dalal, S., Aurangzeb, K., Hussain, A. (2025). An attention-driven hybrid deep neural network for enhanced heart disease classification. Expert Systems, 42(2): e13791. https://doi.org/10.1111/exsy.13791

[19] Srivastava, A., Hari, A., Pratiher, S., Alam, S., Ghosh, N., Banerjee, N., Patra, A. (2021). Channel self-attention deep learning framework for multi-cardiac abnormality diagnosis from varied-lead ECG signals. In 2021 Computing in Cardiology (CinC), Brno, Czech Republic, 48: 1-4. https://doi.org/10.23919/CinC53138.2021.9662886

[20] Zhang, J., Liang, D., Liu, A., Gao, M., Chen, X., Zhang, X., Chen, X. (2021). MLBF-Net: A multi-lead-branch fusion network for multi-class arrhythmia classification using 12-lead ECG. IEEE Journal of Translational Engineering in Health and Medicine, 9: 1-11. https://doi.org/10.1109/JTEHM.2021.3064675

[21] Tibermacine, I.E., Tibermacine, A., Guettala, W., Napoli, C., Russo, S. (2023). Enhancing sentiment analysis on SEED-IV dataset with vision transformers: A comparative study. In Proceedings of the 2023 11th International Conference on Information Technology: IoT and Smart City, New York, United States, pp. 238-246. https://doi.org/10.1145/3638985.3639024

[22] Phan, D.T., Choi, J., Vo, T.T., Ngo, D., Lee, B.I., Oh, J. (2024). Multi-sensor wearable device with transformer-powered two-stream fusion model for real-time leg workout monitoring. IEEE Journal of Biomedical and Health Informatics, 29(4): 2534-2545. https://doi.org/10.1109/JBHI.2024.3524398

[23] Shah, H.A., Saeed, F., Diyan, M., Almujally, N.A., Kang, J.M. (2024). ECG-TransCovNet: A hybrid transformer model for accurate arrhythmia detection using electrocardiogram signals. CAAI Transactions on Intelligence Technology, 1-14. https://doi.org/10.1049/cit2.12293

[24] Jiang, F., Xiao, J., Liu, L., Wang, C. (2024). Dceten: A lightweight ECG automatic classification network based on transformer model. Digital Communications and Networks. https://doi.org/10.1016/j.dcan.2024.11.003

[25] Tahery, S., Akhlaghi, F.H., Amirsoleimani, T. (2024). HeartBERT: A self-supervised ECG embedding model for efficient and effective medical signal analysis. arXiv Preprint arXiv: 2411.11896. https://doi.org/10.48550/arXiv.2411.11896

[26] Ansari, M.Y., Yaqoob, M., Ishaq, M., Flushing, E.F., Mangalote, I.A.C., Dakua, S.P., Aboumarzouk, O., Righetti, R., Qaraqe, M. (2025). A survey of transformers and large language models for ECG diagnosis: Advances, challenges, and future directions. Artificial Intelligence Review, 58(9): 261. https://doi.org/10.1007/s10462-025-11259-x

[27] Han, Y., Liu, X., Zhang, X., Ding, C. (2024). Foundation models in electrocardiogram: A review. arXiv Preprint arXiv: 2410.19877. https://doi.org/10.48550/arXiv.2410.19877

[28] Zhu W., Chen, X., Wang, Y., Wang, L. (2018). Arrhythmia recognition and classification using ECG morphology and segment feature analysis. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 16(1): 131-138. https://doi.org/10.1109/TCBB.2018.2846611

[29] Khurshid, S., Friedman, S., Reeder, C., Di Achille, P., et al. (2022). ECG-based deep learning and clinical risk factors to predict atrial fibrillation. Circulation, 145(2): 122-133. https://doi.org/10.1161/CIRCULATIONAHA.121.057480

[30] Strodthoff, N., Wagner, P., Schaeffter, T., Samek, W. (2020). Deep learning for ECG analysis: Benchmarks and insights from PTB-XL. IEEE Journal of Biomedical and Health Informatics, 25(5): 1519-1528. https://doi.org/10.1109/JBHI.2020.3022989

[31] Bousseljot, R., Kreiseler, D., Schnabel, A. (1995). Nutzung der EKG-signaldatenbank cardiodat der PTB über das internet. Biomedical Engineering/Biomedizinische Technik, 40(s1): 317-318. https://doi.org/10.1515/bmte.1995.40.s1.317

[32] Tibermacine, I.E., Russo, S., Citeroni, F., Mancini, G., Rabehi, A., Alharbi, A.H., El-kenawy, E.S.M., Napoli, C. (2025). Adversarial denoising of EEG signals: A comparative analysis of standard GAN and WGAN-GP approaches. Frontiers in Human Neuroscience, 19: 1583342. https://doi.org/10.3389/fnhum.2025.1583342

[33] Tibermacine, A., Akrour, D., Khamar, R., Tibermacine, I.E., Rabehi, A. (2024). Comparative analysis of SVM and CNN classifiers for EEG signal classification in response to different auditory stimuli. In 2024 International Conference on Telecommunications and Intelligent Systems (ICTIS), Djelfa, Algeria, pp. 1-8. https://doi.org/10.1109/ICTIS62692.2024.10894292