JOURNAL OF CLINICAL AND BIOMEDICAL SCIENCES

ORIGINAL ARTICLE

* **Corresponding author**.

febigenius@gmail.com

Advanced Arrhythmia Classification Using Transformer-Based CNN

Febeena K R[1]*, **Kurian Cini**[2]

1 Research Scholar, School of Computer Sciences, Mahatma Gandhi University, Kottayam, Kerala, India
2 Principal, Al-Ameen College, Edathala, Aluva, Kerala, India

Abstract

Background: The Electrocardiogram (ECG) is a vital clinical signal for recognizing cardiovascular ailments (CVDs) such as Arrhythmia. However, manual assessment of ECG signals is challenging due to subtle physiological variations in both regular and irregular cases, mainly when dealing with a large volume of cardiac patients. From this perspective, automated sorting of ECG signals can offer substantial relief to healthcare experts, facilitating precise analysis. **Objective**: This study aims to develop an automated system for sorting ECG signals to ease the workload of healthcare experts and enhance the precision of cardiac condition analysis. The ultimate goal is to provide healthcare professionals with a reliable tool that streamlines the interpretation process, enabling timely and accurate diagnoses, thereby improving patient outcomes and reducing healthcare burdens. **Method & Material**: Current approaches predominantly rely on convolutional neural networks (CNNs) to extract ECG signal features. However, these may fail to capture nuanced differences in pathological features across different diseases. Transformer networks, known for their prowess in handling sequence data, offer advantages in feature extraction but often rely on extensive datasets, making the complete network intricate. This proposed model utilizes CNN and Transformers for arrhythmia classification. This study was conducted on the MIT-BIH Arrhythmia database (MIT-ArrhyDB), classifying five distinct classes of arrhythmias based on their morphological features. **Result**: The proposed model exhibits an impressive F1 Score of 98.52% and classification accuracy of 98.95%. Conclusion: Comparative analysis with standard CNN exposes the superior performance of our proposed model. This highlights its outstanding overall performance and potential utility in clinical applications.

Keywords: Electrocardiogram (ECG); Arrhythmia; Convolutional Neural Network (CNN); Transformer

1 Introduction

An irregular heartbeat, known as arrhythmia, occurs when the heart's rhythm is disrupted, causing it to beat erratically, excessively quickly, or unusually slowly, distracting its ability to pump blood efficiently. It can be analysed with an Electrocardiogram (ECG), which records the heart's electrical activity without invasive procedures. Traditionally, diagnosing arrhythmia relied on human observation, which was time-consuming. However, automatic detection and classification now save time and improve accuracy. Over the past few decades, cardiovascular disease diagnosis has increasingly relied on machine learning (ML) and deep learning (DL) techniques. Outmoded ML requires extensive feature extraction from ECG signals, while DL models simplify this process by automatically identifying relevant features for prediction and classification[1]. Yanfang Dong et al.[2] proposed CNN-DVIT, the new architecture for deep learning classification of multi-label arrhythmias using 12-lead ECG signals containing different-length recordings. This model combines CNNs with depth wise separable convolutions and incorporates a ViT with a deformable attention mechanism to extract spatial and temporal characteristics from ECG data efficiently. Sattar et al.[3] developed a deep learning-based approach for ECG classification by digitising ECG images to time-series data and applying models such as CNN, LSTM, and SSL-based autoencoders. The CNN model achieves the highest accuracy for classifying cardiac arrhythmias. The results have proven that the digitised ECG signals allow such accurate real-time monitoring for cardiologists, offering an essential instrument for the early and efficient diagnosis of cardiac diseases. Answer: Ansari et al.[4] review deep learning (DL) architectures for ECG arrhythmia detection, comparing models like CNNs, MLPs, Transformers, and RNNs used from 2017 to 2023. It offers a roadmap for researchers entering the field, providing insights into current trends and practical models for detecting ECG anomalies. The survey also highlights areas for future research, aiming to inspire further advancements in ECG arrhythmia detection and classification.

ECG is the most widely employed and pertinent method for evaluating a patient's cardiac activity. Each cardiac phase embraces successive atrial and ventricular depolarisation[5], emanating from the atrial sinoatrial node and spreading across the heart. This process produces electrical currents on the body's surface, inducing skin surface electrical potential variations. Surface electrodes capture these signals, which are visually depicted in the ECG. A standard ECG cycle's key attributes include amplitudes, morphologies, and durations of waves such as P, QRS, and T waves[3], as illustrated in Figure 1. A simple ECG wave starts with a P wave. The QRS wave, notably, is the most significant waveform in the ECG signal and represents the depolarisation of the ventricles. T waves represent ventricular repolarisation.

Baseline wander, Power Line Interference, contact noise, and motion artefacts represent diverse categories of noise found in ECG signals[6]. M Wu et al.[7,8] proposed a resilient and effective deep 1D CNN with 12 layers to categorise five arrhythmia classes. X Hua et al.[9] presented a comprehensive ECG signal classification technique that leverages a unique 1D CNN segmentation strategy. MMR Khan et al.[10] presented a 1-D CNN based on DL for the automated categorisation of ECG heartbeats, aiming to classify five distinct types of abnormal rhythms. BM Maweu et al.[11] presented the CNN Explanations Proposed model for ECG Signals (CEFEs), designed to provide interpretable explanations. Y Zhao et al.[12] introduced an automated technique for ECG signal classification exploiting a deep CNN, complemented by wavelet transform for data filtering. Savalia S and Emamian V[13] introduced a computerised classification system employing the Multi-Layer Perceptron (MLP) network and the CNN.
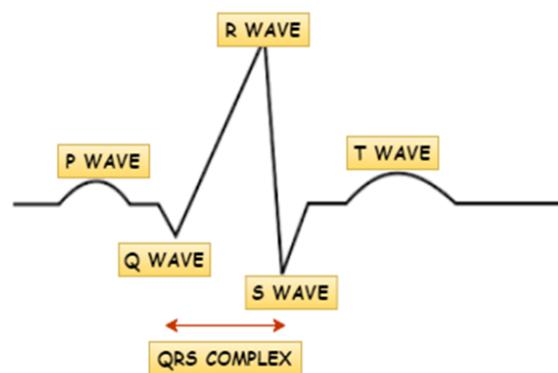

Fig 1. Typical diagram of ECG Waveform

Despite their effectiveness, DL-based methods face specific challenges[14]. They often require intricate convolution and recursive structures, producing a series of concealed states with limited parallelisation due to dependencies on previous states. While CNNs have become famous for pattern classification, they may not capture pathological variations in ECG signals, such as irregularities in shape, duration, or timing[15]. Transformer networks present a promising alternative, offering advantages in handling sequential data and extracting relevant features to classify cardiac abnormalities better.

Key contributions of the paper include:

(a) In pre-processing, denoising, signal segmentation, and data resampling are meticulously executed.

(b) The proposed model employs transformer network (TransNet) based CNN for feature mining.

(c) A comparative study is performed against the standard CNN architecture to evaluate the model's effectiveness.

ECG monitoring generates a continuous flow of data, yielding large amounts of complicated waveforms that cannot be analysed manually, especially in real-time. The vol-

ume becomes even larger in large-scale clinical environments where numerous patients are monitored simultaneously. Arrhythmia can take such diverse forms and severities. Some are subtle and not easily detectable and require specialists' expert analyses for pattern recognition. Manual interpretation is prone to error, particularly regarding infrequent or unusual arrhythmias. This analysis takes more time and requires great attention to detail, which would be impractical for a clinical setting that needs prompt diagnoses to intervene effectively. It is hard to find many professionals in clinical settings whose qualifications warrant extensive training and experience for proper ECG interpretation. In high-volume settings, the absence of available experts might reduce the capability to keep pace with demand. Automated techniques, like Transformer-based Convolutional Neural Networks, help address these challenges through efficient, scalable, and accurate tools for classifying arrhythmias. Specifically, Transformer-based CNNs apply deep learning techniques to capture temporal and spatial features in ECG data, allowing for improved accuracy and speed up, which is critical in vast, real-time applications.

The proposed methodology integrates TransNet with the CNN network to address its limitations, particularly its suboptimal performance in handling temporal features. The paper's structure is outlined as follows: The proposed model is detailed in Section 2; preliminary findings are presented in Section 3; and a concise overview and concluding remarks are provided in Section 4.

2 Materials and Methodology

The proposed model is represented in Figure 2. In Figure 2 ECG signals from MIT-ArrhyDB are given as input to the model. Due to undesired random instabilities, the signal is exposed to eradicate noise as part of the pre-processing. After the signal has been de-noised, R-peak information is accessible from the database, which separates the individual heartbeats. Next, resampling has been used to enrich the data, and lastly, the classifier architecture receives the processed and expanded data. The following subsections include a full description of the complete procedure.
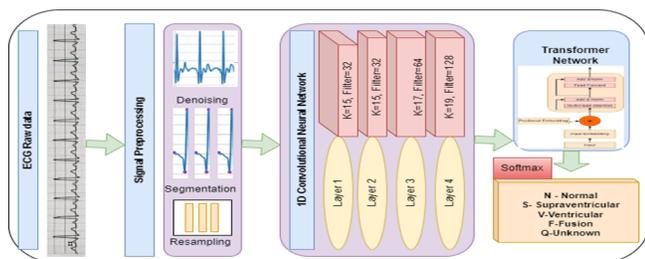

Fig 2. Architecture of proposed TransNet-based CNN Model

In Figure 2, the raw data is pre-processed to eliminate unwanted noise or artefacts of the ECG signal, improving the signal quality. Then, we extract all the relevant segments that might correspond to individual heartbeats or specific intervals, after which we bring the data into a uniform sampling rate for consistency, which is valuable for further processing. Filter count and kernel dimension with values of 32, 64, and 128 and dimensions of 1.5, 1.7, and 1.9, respectively, are used within each layer for identifying the different patterns appearing in ECG signals. After extracting the features by CNN, these are fed into a Transformer Network, which, using attention mechanisms, helps to understand complex relationships between data points. This approach also accounts for sequential relationships among heartbeats in ECGs, making it suitable for analysing ECG data, which is inherently time-dependent. This output from the Transformer Network is given a softmax function to provide different classes' probabilities. The pipeline implements CNN for feature extraction and Transformer networks for sequential learning of ECG data to classify it into different heartbeat categories. The approach efficiently suits the appropriately complex temporal patterns of ECG signals and will potentially contribute to automated arrhythmia detection systems.

2.1 Dataset

The data hired in this research were traced from the MIT-ArrhyDB[16,17]. By leveraging the R-peak data from MIT-ArrhyDB, we fragmented the signal into individual beats grounded around each R-peak and documented their respective types. Each beat fragment contained 300 data points, with 150 preceding the R-peak and 149 following the peak. Adhering to AAMI recommendations, we elicited 109,446 fragments categorised into five categories: 90,592 Non-ectopic beats (N), 2,781 Supraventricular ectopic beats (S), 7,232 Ventricular ectopic beats (V), 802 Fusion beats (F), and 8,039 Unknown beats (Q). The dataset is divided into 80:20 ratios for training and evaluate the proposed model.

2.2 Pre-Processing of ECG Input Signal

Denoising, signal segmentation and resampling are done as part of pre-processing. The prime noise in most electrophysiological signals is power line interference (PLI)[18]. The difficulties of transient interferences associated with digital notch filters are addressed by the proposal of an adaptive notch filter (ANF)[19]. In this method, apply fast Fourier transform (FFT) on the input data, $e(t)$ with the sampling frequency of f_s and length N. Utilizing a pair of spectral bins positioned in the primary lobe of the Fourier spectrum permits the assessment of PLI parameters, including amplitude A_{PLI}, frequency f_{PLI}, and initial phase $\emptyset_{PLI}$ with the help of ratio based spectrum correction methods (RBSC). This estimation facilitates the creation of a compensation signal, $C_{PLI}(t)$. The

subsequent subtraction of this recompense signal from the original signal effectively suppresses the occurrence of PLI. The mathematical representation of the above steps is shown in Equations (1), (2), (3) and (4), respectively, *et* represents the ECG signal, *ef* represents the ECG signal after the application of FFT, and *CPLIt* is the compensation signal.

$$e(t) \xrightarrow{FFT} \hat{e}(f) \tag{1}$$

$$\hat{e}(f) \xrightarrow{RBSC} A_{PLI}, f_{PLI}, \emptyset_{PLI} \tag{2}$$

$$C_{PLI}(t) = A_{PLI}.\cos(2\pi f_{PLI}t + \emptyset_{PLI}) \tag{3}$$

$$\bar{e}(t) = e(t) - C_{PLI}(t) \tag{4}$$

The signal is further treated by segmenting it into individual heartbeats following the denoising step. A significant imbalance in the dataset prompted the implementation of resampling techniques to achieve data balance.

2.3 Proposed model

The proposed architecture embraces three primary segments: (a) an embedded network based on one-dimensional convolution layers for extracting unprocessed data from divided ECG waves, (b) a transformer network, and (c) a classification segment.

(a) CNN segment
CNN operates feed-forwardly and is extensively employed for extracting features[20]. The CNN segment consists of four convolutional layers with the following parameters: Layer 1 has 32 filters and features a kernel size (k_s) of 15. Layer 2 has the same parameter as Layer 1. Layer 3 is composed of 64 filters and utilises a k_s of 17, while Layer 4 comprises 128 filters with a k_s of 19. The heartbeats undergo processing using four 1D convolutional layers. Following representation learning, the characteristic vector $x_{f'} = x_{f1}, x_{f2}, \ldots\ldots, x_{fn}$ is generated, and this vector serves as the input to a TransNet. Padding and stride are kept identical in all layers. To give the network non-linearity and to address the vanishing gradient problem, the non-linear function is configured to a rectified linear unit (ReLU).

(b) Transformer segment
To address the limitation of CNNs in modelling long-range correlations in time series data, a TransNet is integrated into the proposed model to extract extended patterns from the features identified by the CNN. The TransNet[21] was devised based on the attention mechanism, comprising both an encoder and a decoder. In the proposed model the transformer encoder has been used to record interactions and long-range dependencies across time instances. The sole portion employed in the input signal categorisation problem is the encoder section, whose structure is depicted in Figure 3. Before applying the attention module, the output of the convolution segment, xf', is subjected to positional encoding. As shown in Figure 3, the encoder function is explained as follows.

Self-Attention Module: The Inputs Q, K, and V correspond to the query, key, and value, respectively. The attention score is computed based on the resemblance between the Q and K. Subsequently, the attention context is established according to this attention score. The calculation for the scaled dot-product attention employed by the model can be expressed using Equation (5) where $\frac{1}{\sqrt{d_k}}$ is the scaling factor.

$$Atten(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \tag{5}$$

Multi-head attention: The multi-head attention mechanism splits different attention products after projecting Q, K, and V through 'n' distinct linear transformations. The values of Q, K, and V are identical in the self-attention process. The formulas are represented in Equations (6) and (7). The model's input size is 256, and it is executed through four transformer encoder blocks.

$$MultiHead(Q,K,V) = Concat(head_1, head_2, \ldots\ldots, head_n) \tag{6}$$

$$head_i = Atten(QW_i^Q, KW_i^K, VW_i^V \tag{7}$$

Position-wise feed-forward networks: - Every layer of the encoder has a fully linked feed-forward network, a two-level layer linear mapping, and an attention sub-layer. Linear transformations in the network using weight matrices W1, W2, and biases b1, b2, followed by layer normalisation, are represented as $FFN(x)$ and calculated using Equation (8). The resulting output of the transformer network, denoted as $o_{tf} : \{o_1, o_2, \ldots\ldots, o_n\}$ represents a learned vector for each feature.

$$FFN(x) = max(0, xW_1 + b_1)W_2 + b_2 \tag{8}$$

Positional encoding: The sequence's absolute or relative position is appended to the input at the encoder's top to utilise the sequence's order.

Certain conditions in medical datasets, such as certain arrhythmias, are much less frequent than others. Transformers struggle with imbalanced datasets since they can become biased towards more frequently occurring patterns

and, therefore, do not perform optimally in rare but critical conditions. Transformers have high computational and memory demands, especially because the self-attention mechanism scales quadratically with the input sequence length. Long sequences are typical in medical data: continuous ECG signals and extended MRI scans consume a lot of resources, making training and inference slow and resource-hungry. Big transformers need a great deal of labelled data to generalise well. In medical domains, however, labelled data is usually limited due to privacy concerns, costs of annotations and the requirement of expert interpretation. In such cases, it can be challenging to attain optimal model performance. Sometimes Transformers have poor performance in local feature extraction. This may require supplementary convolutional layers, or hybrid approaches to improve sensitivities to local features. These will add to the complexity of the model.

(c) Classification segment

The output of TransNet o_{tf} is attached to the fully connected layer designed for multiclass categorisation. This tier categorises five different classes of arrhythmias.
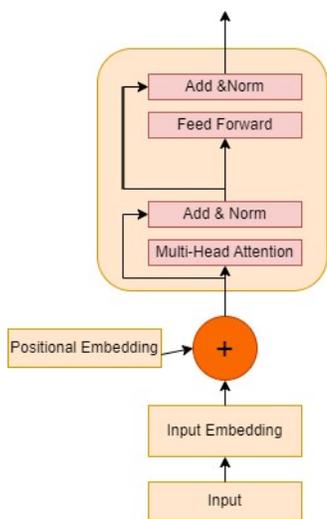

Fig 3. Encoder architecture of TransNet Model

In Figure 3, the architecture of a Transformer Encoder Layer is quite typical in deep learning models to treat sequential data, such as text or time-series data, such as an ECG signal. Raw Input Data: This could be an ECG signal segment, text, or sequential data. Input Embedding: This block converts the input into a dense vector representation with which the model can work. This embedding captures the inductive features of the input data. Since Transformer models are inherently position agnostic, a positional embedding is added to the input embedding. This helps it learn the order of the data point, which is very important for sequences like the signal from an ECG. This layer lets the model focus on differ-

ent parts of the input sequence simultaneously by computing attention scores for multiple "heads." Each head learns a different aspect of the relationships in the data, allowing the model to capture complex dependencies across time steps in the ECG signal. The Add & Norm layer normalises the output of the attention layer. Then it adds it to the input embedding along with a residual connection, which supports the flow of information through the network and stabilises training. This layer applies to each position separately with a fully connected feed-forward network that further transforms the representation and supports the model in catching more complex patterns. Here, another Add & Norm process is introduced so that the standardisation of feed-forward output is combined with the previous output for better memorisation and a stabilisation model. By this Transformer Encoder Layer, the model becomes helpful in understanding complex dependencies and contextual information that remains in such sequential data. The multi-head attention combined with feed-forward layers makes this architecture particularly well suited to learning both short-term and long-term dependencies; this gives its strength to applications such as electrocardiogram signal classification and natural language processing.

3 Experimental Results

This architecture uses 109,446 ECG signal segments from the MIT-ArrhyDB to evaluate model performance. The dataset is divided 80:20 for learning and evaluation, with 87,554 beats for training and 21,892 for assessment. The model classifies five categories of CVDs—N, S, V, F, and Q achieving an overall accuracy of 98.95% and an F1-score of 98.52%. The training was conducted for 50 epochs. A gradual increase in training accuracy was observed up to the 40th epoch, after which it reached a saturation point, maintaining stability from the 40th to the 50th epoch. Therefore, training was halted at 50 epochs. The Adam optimiser with a learning rate 0.001 was utilised to avoid overfitting. Despite some misclassifications in the normal class, these are minor relative to the total number of beats tested. Although S and F beats are less frequent, the model achieves 83% and 86% accuracy for these classes, respectively. In contrast, a reference model [20] using a transformer for ECG classification reported an accuracy of 90.52%. The commonly employed metrics, namely Accuracy, Precision, Recall, and F1-score (represented in Equations (9), (10), (11) and (12), respectively), have been utilised to evaluate the proposed classification model quantitatively. True positives and true negatives are denoted as TP and TN, while false positives and false negatives are represented as FP and FN.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (9)$$

$$Precision = \frac{TP}{TP + FP} \qquad (10)$$

$$Recall = \frac{TP}{TP+FN} \qquad (11)$$

$$F_1 \; score \; = 2.\frac{precision.recall}{precision+recall} \qquad (12)$$

Each category is evaluated based on accuracy, precision, recall, and F1-score. The model achieves high accuracy across all categories, with N and Q at 99.00%, V at 95.00%, S at 83.00%, and F at 86.00%. Precision and recall values closely match accuracy, indicating consistent performance in correctly identifying each category. F1-scores, which balance precision and recall, range from 83.00% to 99.00%, highlighting the model's effectiveness in classifying ECG signals across diverse cardiac conditions.

To analyse the performance of the proposed model, it is also trained using a convolutional neural network, which gives an overall accuracy of 96.82%. When the transformer block is incorporated with CNN, the overall accuracy is enhanced to 98.95%. Table 1 shows the comparison of the transformer-based CNN model with previous architectures. Figures 4 and 5 represent training and validation accuracy and training and validation loss, respectively. Figures 6 and 7 illustrate the confusion matrices for the model employing CNN-Transformer and CNN, respectively. The application of CNN alone results in an F1-score of 96.31% in the proposed model. Incorporating the transformer into the CNN leads to a notable improvement, with the F1-score improving by 2.21%.

Figure 4 compares the training and validation accuracy of two models over 50 epochs. Figure 4(a) shows the performance of a CNN, where training accuracy stabilises close to 1.0, but validation accuracy fluctuates and stabilises around 0.975, showing potential overfitting. Figure 4 (b) depicts the performance of a combined CNN and Transformer model, with both training and validation accuracies increasing rapidly and stabilising close to 1.0, representing better generalisation and reliable performance. The enclosure of transformers augments the model's capability to capture complex patterns, leading to superior overall performance compared to CNN alone. Figure 5 (a) shows that the training loss decreases rapidly and stabilises near zero, while the validation loss also decreases but remains higher and fluctuates more, suggesting potential overfitting. Figure 5 (a) shows that the training and validation losses decrease rapidly and closely follow each other, stabilising near zero, indicating that the TransNet-based CNN model achieves better generalisation and more stable performance on both training and validation data.

As shown in Figures 6 and 7, the CNN with TransNet performs better than the basic CNN. The proposed model demonstrates an accuracy of 86% for class S and 83% for class F, despite these classes having a smaller quantity of data.
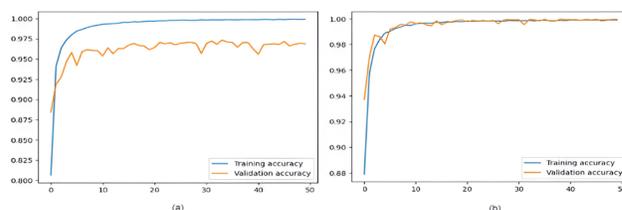

Fig 4. (a) Training and validation accuracy of CNN (b) Training and validation accuracy of TransNet-based CNN
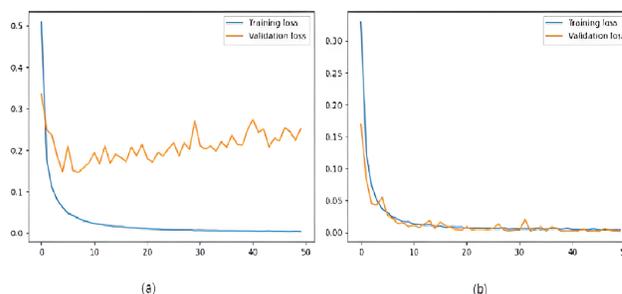

Fig 5. (a) Training and validation loss of CNN (b) Training and validation loss of TransNet-based CNN
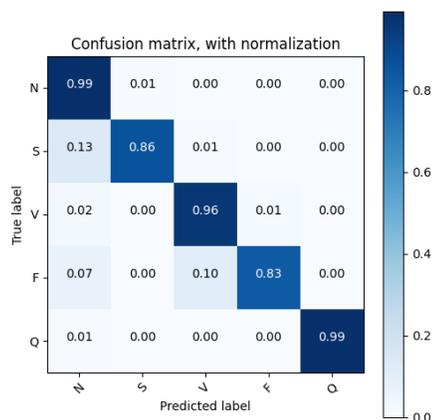

Fig 6. Confusion matrix of TransNet-based CNN (in probability)

Table 1 showcases the superior performance of the proposed TransNet-based CNN model, which achieves an impressive accuracy of 98.95% and an F1 score of 98.52%. The table demonstrates that the proposed architecture significantly outperforms other compared models, particularly excelling in accurately classifying low-quantity signals such as Supraventricular and Fusion beats.

Implementing an automated Transformer-based CNN system for classifying arrhythmia in a real-world clinical environment presents several potential challenges. Hospitals and clinics use many legacy systems that cannot support advanced deep-learning models; thus, much modification and integration may be necessary for seamless data transfer
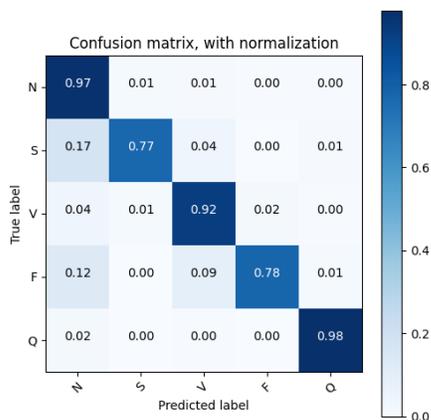
Fig 7. Confusion matrix of CNN (in probability)

Table 1. Performance comparisons of TransNet-based CNN with other models

Model	Year of publication	Accuracy	Precision	Recall	F1-score
Transformer[22]	2022	90.52%	88.50%	86.46%	87.47%
CNN-LSTM[23]	2021	95.81%	74.94%	69.20%	71.06%
CNN[20]	2021	94.70%	93.70%	89.00%	88.90%
Ensemble Multilabel Classification[24]	2020	75.20%	80.80%	71.60%	75.20%
1D CNN[9]	2020	97.45%	—	97.00%	97.00%
CNN-BiLSTM[25]	2020	96.77%	81.24%	74.89%	77.84%
CNN[26]	2019	93.71%	88.30%	91.25%	89.75%
1D CNN[27]	2018	95.20%	92.52%	93.52%	92.45%
Proposed-CNN	—-	96.82%	96.03%	96.26%	96.31%
Proposed-TransNet based CNN	—-	98.95%	98.35%	98.24%	98.52%

and processing. Integrations with electronic health records and other medical databases are technically complicated and resource-demanding processes. Healthcare data is most sensitive, and implementing automated systems requires strict adherence to privacy regulations in the healthcare sector, such as HIPAA in the U.S. or GDPR in Europe. Secure data handling, encryption, and anonymisation protocols will protect privacy but add some layers of complexity to the deployment of such a system. Clinical environments would particularly demand near real-time if not real-time, processing of ECG data for expedited patient diagnosis and treatment. Transformer-based models are powerful but could also be computationally expensive, posing significant challenges in achieving adequate speed and efficiency without high-performance hardware. ECG data differ considerably from patient to patient due to age, physical condition, and sensor quality. Noises and artefacts from patient movement, electrode placement, or other environmental reasons can affect signal quality and make accurate classification difficult. The system built should ensure robustness against such variability to limit misdiagnosis. Advanced deep learning models integrated and maintained within healthcare require special hardware, maintenance, and technical support for proper functionality. These costs can be highly prohibitive for smaller clinics or facilities with limited budgets, slowing adoption across diverse clinical environments.

4 Conclusion

This research proposed an automated categorisation model that combines a CNN and TransNet to categorise ECG signals. Before being fed into the CNN, the ECG signal undergoes pre-processing procedures, including denoising, segmentation, and resampling. The parameter information extracted by the CNN retains temporal characteristics. Through the synergistic combination of the CNN and an enhanced transformer, the proposed model achieves a categorisation accuracy of 98.95%. The suggested model may be used because of its performance and the fact that transformers were included to address the complexity issues with sequential models. This model can be used with wearable technologies in future research to help save more lives in emergencies by providing ongoing monitoring. Future scopes in this direction are promisingly developing Transformer-based CNNs for arrhythmia classification in clinical settings, which extend to all constituencies of healthcare domains. Further work could incorporate these models with wearable devices for health monitoring, such as smartwatches and mobile ECG monitors. This would enable seamless, in situ, immediate arrhythmia detection outside clinics, allowing for remote monitoring and early interventions, especially among high-risk or elderly patients. Automated systems can permit clinicians to monitor and manage patients with arrhythmias who reside in areas inaccessible by specialists. To combine it with telemedicine, future systems can be designed to integrate with existing platforms for reviewing and diagnosing patients by remote specialists, which is highly valuable in managing chronic cardiac conditions.

References

1) Gao J, Zhang H, Lu P, Wang Z. An Effective LSTM Recurrent Network to Detect Arrhythmia on Imbalanced ECG Dataset. *Journal of Healthcare Engineering*. 2019;2019:1–10. Available from: https://doi.org/10.1155/2019/6320651.

2) Dong Y, Zhang M, Qiu L, Wang L, Yu Y. An Arrhythmia Classification Model Based on Vision Transformer with Deformable Attention. *Micromachines*. 2023;14(6):1–12. Available from: https://dx.doi.org/10.3390/mi14061155.

3) Sattar S, Mumtaz R, Qadir M, Mumtaz S, Khan MA, De Waele T, et al. Cardiac Arrhythmia Classification Using Advanced Deep Learning Techniques on Digitized ECG Datasets. *Sensors*. 2024;24(8):1–23. Available from: https://dx.doi.org/10.3390/s24082484.

4) Ansari Y, Mourad O, Qaraqe K, Serpedin E. Deep learning for ECG Arrhythmia detection and classification: an overview of progress for period 2017–2023. *Frontiers in Physiology*. 2023;14:1–20. Available from: https://doi.org/10.3389/fphys.2023.1246746.

5) McSharry PE, Clifford GD, Tarassenko L, Smith LA. A dynamical model for generating synthetic electrocardiogram signals. *IEEE Transactions on Biomedical Engineering*. 2003;50(3):289–294. Available from: https://dx.doi.org/10.1109/tbme.2003.808805.

6) Rashmi N, Begum G, Singh V. ECG denoising using wavelet transform and filters. In: 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). IEEE. 2018;p. 2395–2400. Available from: https://doi.org/10.1109/WiSPNET.2017.8300189.

7) Guan J, Wang W, Feng P, Wang X, Wang W. Low-Dimensional Denoising Embedding Transformer for ECG Classification. In: ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE. 2021;p. 1285–1289. Available from: https://doi.org/10.1109/ICASSP39728.2021.9413766.

8) Wu M, Lu Y, Yang W, Wong SY. A Study on Arrhythmia via ECG Signal Classification Using the Convolutional Neural Network. *Frontiers in Computational Neuroscience*. 2021;14:1–10. Available from: https://dx.doi.org/10.3389/fncom.2020.564015.

9) Hua X, Han J, Zhao C, Tang H, He Z, Chen Q, et al. A novel method for ECG signal classification via one-dimensional convolutional neural network. *Multimedia Systems*. 2022;28(4):1387–1399. Available from: https://doi.org/10.1007/s00530-020-00713-1.

10) Khan MMR, Siddique MAB, Sakib S, Aziz A, Tanzeem AK, Hossain Z. Electrocardiogram heartbeat classification using convolutional neural networks for the detection of cardiac Arrhythmia. In: 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). IEEE. 2020. Available from: https://doi.org/10.1109/I-SMAC49090.2020.9243474.

11) Maweu BM, Dakshit S, Shamsuddin R, Prabhakaran B. CEFEs: A CNN Explainable Framework for ECG Signals. *Artificial Intelligence in Medicine*. 2021;115:102059. Available from: https://dx.doi.org/10.1016/j.artmed.2021.102059.

12) Zhao Y, Cheng J, Zhan P, Peng X. ECG Classification Using Deep CNN Improved by Wavelet Transform. *Computers, Materials & Continua*. 2020;64(3):1615–1628. Available from: https://dx.doi.org/10.32604/cmc.2020.09938.

13) Golany T, Lavee G, Yarden ST, Radinsky K. Improving ECG Classification Using Generative Adversarial Networks. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2020;34(08):13280–13285. Available from: https://dx.doi.org/10.1609/aaai.v34i08.7037.

14) Savalia S, Emamian V. Cardiac Arrhythmia Classification by Multi-Layer Perceptron and Convolution Neural Networks. *Bioengineering*. 2018;5(2):1–12. Available from: https://dx.doi.org/10.3390/bioengineering5020035.

15) Chatfield K, Simonyan K, Vedaldi A, Zisserman A. Return of the Devil in the Details: Delving Deep into Convolutional Nets. *Proceedings of the British Machine Vision Conference 2014*. 2014;p. 1–11. Available from: https://doi.org/10.48550/arXiv.1405.3531.

16) Mark R, Moody G. MIT-BIH Arrhythmia Database Directory. Cambridge, MA, USA. Massachusetts Institute of Technology. 1988. Available from: https://physionet.org/physiobank/database/html/mitdbdir/foreword.htm.

17) Acharya UR, Suri JS, Spaan JAE, Krishnan SM. Advances in Cardiac Signal Processing. 1st ed. Berlin, Heidelberg. Springer. 2007. Available from: https://doi.org/10.1007/978-3-540-36675-1.

18) Chandrakar B, Yadav OP, Chandra VK. A survey of noise removal techniques for ECG signals. *International Journal of Advanced Research in Computer and Communication Engineering*. 2013;2(3):1354–1357. Available from: https://www.researchgate.net/publication/303155606_A_survey_of_noise_removal_techniques_for_ecg_signals.

19) Chen B, Li Y, Cao X, Sun W, He W. Removal of Power Line Interference From ECG Signals Using Adaptive Notch Filters of Sharp Resolution. *IEEE Access*. 2019;7:150667–150676. Available from: https://dx.doi.org/10.1109/access.2019.2944027.

20) Wang T, Lu C, Sun Y, Yang M, Liu C, Ou C. Automatic ECG Classification Using Continuous Wavelet Transform and Convolutional Neural Network. *Entropy*. 2021;23(1):1–13. Available from: https://dx.doi.org/10.3390/e23010119.

21) Li J, Si Y, Xu T, Jiang S. Deep Convolutional Neural Network Based ECG Classification System Using Information Fusion and One-Hot Encoding Techniques. *Mathematical Problems in Engineering*. 2018;2018:1–10. Available from: https://dx.doi.org/10.1155/2018/7354081.

22) Hu S, Cai W, Gao T, Wang M. A Hybrid Transformer Model for Obstructive Sleep Apnea Detection Based on Self-Attention Mechanism Using Single-Lead ECG. *IEEE Transactions on Instrumentation and Measurement*. 2022;71:1–11. Available from: https://dx.doi.org/10.1109/tim.2022.3193169.

23) Essa E, Xie X. An Ensemble of Deep Learning-Based Multi-Model for ECG Heartbeats Arrhythmia Classification. *IEEE Access*. 2021;9:103452–103464. Available from: https://dx.doi.org/10.1109/access.2021.3098986.

24) Sun Z, Wang C, Zhao Y, Yan C. Multi-Label ECG Signal Classification Based on Ensemble Classifier. *IEEE Access*. 2020;8:117986–117996. Available from: https://doi.org/10.1109/ACCESS.2020.3004908.

25) Chen A, Wang F, Liu W, Chang S, Wang H, He J, et al. Multi-information fusion neural networks for arrhythmia automatic detection. *Computer Methods and Programs in Biomedicine*. 2020;193:105479. Available from: https://dx.doi.org/10.1016/j.cmpb.2020.105479.

26) Guo L, Sim G, Matuszewski B. Inter-patient ECG classification with convolutional and recurrent neural networks. *Biocybernetics and Biomedical Engineering*. 2019;39(3):868–879. Available from: https://doi.org/10.1016/j.bbe.2019.06.001.

27) Yıldırım Ö, Pławiak P, Tan RS, Acharya UR. Arrhythmia detection using deep convolutional neural network with long duration ECG signals. *Computers in Biology and Medicine*. 2018;102:411–420. Available from: https://dx.doi.org/10.1016/j.compbiomed.2018.09.009.