Multi-Disease Cardiovascular Detection from ECG Signals Using an Attention-Driven Deep Network

Atitaya Phoemsuk
School of Computer Science and Electronic Engineering
University of Essex
Colchester, United Kingdom
ap19698@essex.ac.uk

Vahid Abolghasemi
School of Computer Science and Electronic Engineering
University of Essex
Colchester, United Kingdom
v.abolghasemi@essex.ac.uk

***Abstract*—Electrocardiography (ECG) is widely used to diagnose cardiovascular diseases (CVDs), particularly during prescreening. We propose a novel deep learning architecture for classifying multiple CVDs by integrating convolutional layers, residual networks, and attention mechanisms into a unified model. The motivation is to enhance the performance of traditional diagnostic methodologies by employing convolutional neural networks (CNNs) with residual connections to mitigate the vanishing gradient problem, allowing the model to learn complex patterns within the ECG signals while reducing the risk of overfitting. The proposed model integrates attention mechanisms to identify the most relevant features within ECG signals for classification. This model effectively captures both local and global features within ECG data, facilitating a comprehensive analysis of intricate cardiac patterns. Our extensive experimental results demonstrate that the proposed model effectively achieves an average classification accuracy of 99.54%, which is superior to existing deep learning-based models, and enables the detection of multiple heart conditions from a single ECG reading.**

***Index Terms*—Electrocardiogram, Coronary Artery Disease, Arrhythmia, Atrial Fibrillation, Convolutional Neural Networks, Residual Networks, Attention Mechanisms.**

I. Introduction

Cardiovascular diseases (CVDs) continue to be the leading cause of mortality worldwide. The global burden of CVDs has increased substantially, with the number of deaths rising to nearly 19.9 million in 2021, representing a 72% increase compared to 1990, when approximately 12.3 million deaths were recorded [1]. This substantial rise underscores the critical need for enhanced preventive strategies and timely therapeutic interventions in healthcare systems worldwide. Atrial fibrillation (AF) is the most common sustained cardiac arrhythmia and is associated with a significantly increased risk of thromboembolic events. Similarly, coronary artery disease (CAD), characterised by the narrowing of coronary arteries due to the formation of atherosclerotic plaques, is a major contributor to ischaemic heart disease and myocardial infarction. Arrhythmias can vary from benign to life-threatening, disrupting normal heart rhythms and significantly impairing cardiac function. The early detection of these conditions is crucial for preventing sudden cardiac death and often necessitates specific treatments.

Electrocardiography (ECG) is routinely utilised as the primary tool for the initial screening of CVDs in general practice. It serves as a non-invasive method for monitoring heart function and identifying abnormalities in cardiac activity. Numerous studies have demonstrated its effectiveness in diagnosing conditions such as arrhythmias [2], [3], CAD [4], [5], and AF [6]–[8]. Given the increasing prevalence of these conditions, enhancing ECG classification methods is crucial for improving patient outcomes and optimising healthcare resources. ECG signals provide fine-grained information in both time and amplitude domains, which are essential for accurate classification. Developing deep learning models that can effectively capture and interpret these relevant features across various cardiac conditions presents inherent complexities. Additionally, achieving model universality over a range of diseases using a single model is challenging due to significant variations in ECG features among patients.

Traditional machine learning methods have been widely employed in automatic ECG interpretation, typically involving multi-step processes such as signal preprocessing, feature extraction, and classification using algorithms such as Support Vector Machines (SVM) and Random Forest [9]. However, these approaches often rely on manual feature engineering, which may not effectively capture the intricate relationships within ECG data [10]. Therefore, deep learning (DL) techniques are introduced as these methods can address some of the limitations of traditional methods by automatically learning intricate features directly from the raw ECG data. Convolutional neural networks (CNNs) have demonstrated the ability to learn spatial hierarchies of features and are particularly effective at capturing local patterns within the ECG signal, making CNN well-suited for the detection of CVDs [3], [5]. In addition to the traditional CNN, Residual Networks (ResNets) have been introduced for the classification of various CVDs. ResNet, a specific type of Convolutional Neural Network (CNN), addresses several limitations associated with traditional CNN architectures. By incorporating shortcut connections, ResNets facilitate the construction of deeper network architectures, effectively mitigating the vanishing gradient problem often encountered in very deep networks [3], [6], [11]. In [12], Long Short-Term Memory (LSTM) is employed for the early detection of heart diseases. LSTM is a specialised type of recurrent neural network (RNN) designed to address the vanishing gradient problem, which can impede RNNs from learning long-term dependencies. Consequently, numerous research papers combine CNN and Long LSTM due to their complementary strengths [4], [13]. CNN is utilised to extract meaningful features from ECG data, such as local patterns, while LSTM is employed to model the temporal relationships and dependencies over time. Not only has ResNet been introduced to address gradient problems, but attention mechanisms have also been recently integrated into DL-based models [7], [14], [15]. In [8], an attention mechanism can

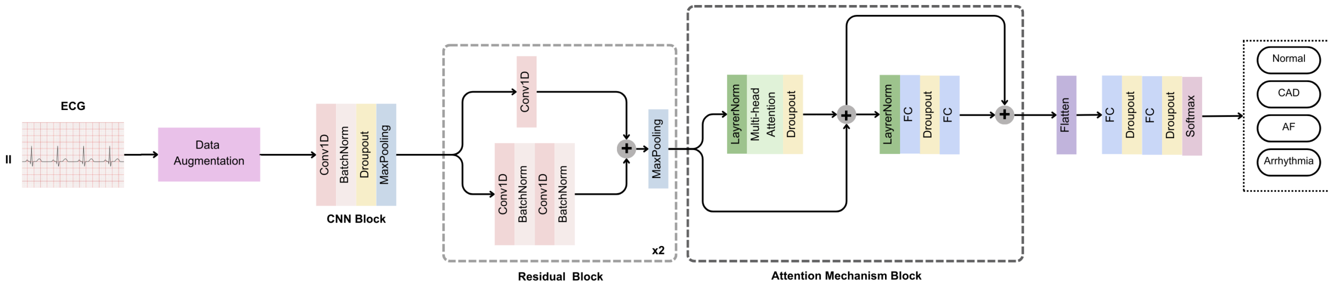


Fig. 1: The proposed model architecture.

assign greater importance to leads that provide the most useful features to detect a particular heart condition, while giving less weight to leads that are less informative. This helps the network learn more meaningful representations by highlighting the most important features, potentially leading to better accuracy and diagnostic performance.

According to the literature, current models face significant challenges in multi-class disease classification from ECGs due to limitations in data availability, class imbalance, and inadequate feature extraction, which adversely affect their performance across various cardiovascular conditions [10]. These challenges underscore the need for a more effective and practical solution, prompting the development of our proposed model, which integrates advanced DL-based techniques to enhance both accuracy and robustness in the diagnosis of multiple cardiovascular diseases.

In this study, we propose a novel model for the classification of multiple CVDs, integrating convolutional layers, residual networks, and attention mechanisms. We propose to apply a data augmentation process to serve two purposes: 1) enriching the dataset to enhance its robustness, generalisability, and applicability to real-world scenarios, etc. and 2) examining the impact of noise power on ECG signal classification to determine the optimal settings for augmentation. Our multi-label classification enables the detection of multiple conditions from a single ECG reading, potentially improving treatment decisions and revealing diseases that may have previously gone undetected.

II. Methodology

Figure 1 illustrates an overview of our proposed approach. As seen from this figure, the process begins with raw ECG signals obtained from four classes (Normal, CAD, AF, and Arrhythmia) and only one ECG channel (lead II), which will be explained in detail in Section II-A. Subsequently, the raw data will be randomly selected and used for data augmentation (see Section II-B). The proposed model (with its structure described in Sections II-C, II-D, and II-E) is then employed to classify the ECG signals. In our study, we propose a novel model that integrates CNN with residual blocks and multi-head attention mechanisms to enhance the processing and classification of ECG signals. This hybrid architecture effectively captures both local patterns and long-range dependencies, thereby improving feature extraction and overall model performance in identifying cardiac diseases.

A. Data preprocessing

The dataset used in this study consists of four distinct categories of ECG data: atrial AF, arrhythmia, CAD, and healthy patients. Arrhythmia data are obtained from a large-scale 12-lead electrocardiogram database [16]. AF data are sourced from the 4th China Physiological Signal Challenge 2020 [17], while CAD data are retrieved from the MIMIC III database [18]. Data from healthy patients are acquired from the Fantasia database [19]. All of these databases are publicly accessible via PhysioNet. Twenty patients were carefully selected from each respective database, and their ECG data were segmented into 1-second intervals. This segmentation allows for a detailed analysis of variations and trends in the signals over short time intervals, a methodology extensively used in previous research to ensure consistency in ECG signal evaluation [4]. Each segment comprises 250 samples, representing approximately one full cycle of the ECG signal. Class labels were assigned as follows: 0 for normal, 1 for CAD, 2 for arrhythmia, and 3 for AF. Subsequently, 200 segments from each class were randomly selected for data augmentation and further used in the experimental process.

B. Data augmentation

Data augmentation is employed by introducing random noise to ECG segments, significantly enhancing model generalisation and helping to prevent overfitting. Commonly applied in ECG signal analysis [10], this method effectively replicates the noisy conditions typically encountered in real-world scenarios. Gaussian noise is added to generate synthetic examples, to improve the model's generalisation. A zero-mean white Gaussian noise, denoted as $\mathcal{N}(0, \sigma^2)$, with a standard deviation σ, is added to each sample. In this study, Gaussian noise is applied to each original ECG segment to generate two augmented segments as shown in Fig. 2. The white Gaussian noise is introduced as follows:

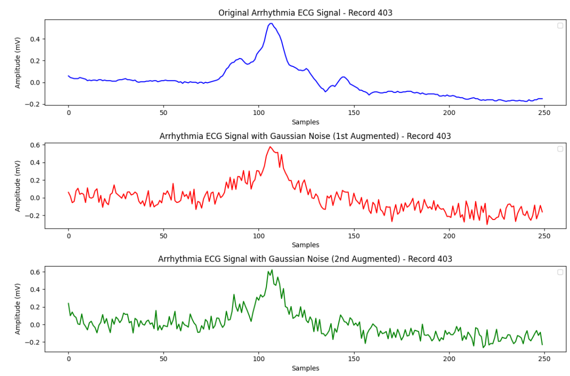
$$X_{\text{augmented}} = X + \mathcal{N}(0, \sigma^2), \tag{1}$$

where $X = \{x_1, x_2, x_3, \ldots, x_n\}$ represents the original ECG segment composed of n samples, $\mathcal{N}(0, \sigma^2)$ represents Gaussian noise with mean 0 and variance σ^2.

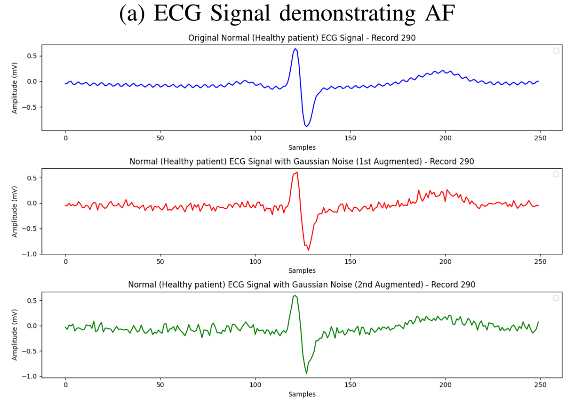
C. CNN block

A CNN block utilises 50 convolutional filters, each with a kernel size of 28, to process the input ECG data. The convolutional operation involves sliding these filters across the temporal dimension of the input sequence to capture local patterns and features. The activation function applied is the Rectified Linear Unit (ReLU), which introduces non-linearity into the model, enabling it to learn more complex patterns from the ECG signals. Furthermore, L2-norm regularisation is applied to the convolutional weights with a penalty term

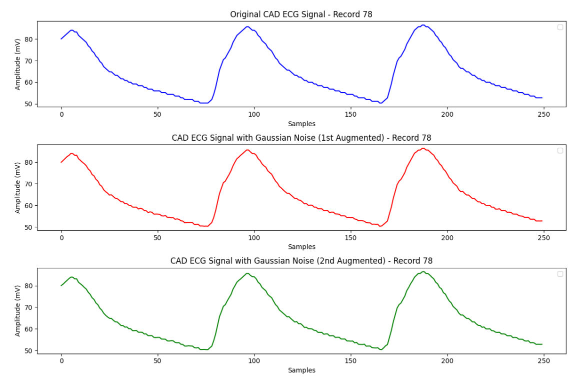
(a) ECG Signal demonstrating AF



(b) ECG Signal demonstrating arrhythmia



(c) ECG signal demonstrating normal sinus rhythm in a healthy patient



(d) ECG Signal demonstrating CAD

Fig. 2: Example of ECG signals from four different classes (Normal, CAD, Arrhythmia, and AF) with applied data augmentation.

of 0.01. This regularisation technique helps to prevent overfitting by penalising large weights, thereby promoting model generalisation when interpreting ECG signals.

Batch normalisation is applied to adjust the activations from the previous layer by normalising them across the mini-batch. This process standardises the mean and variance of the inputs to each subsequent layer, which helps to reduce internal covariate shifts and stabilise the distribution of ECG signal features. To further address overfitting, the block incorporates a dropout layer with a dropout rate of 0.1, which randomly disables a fraction of neurons during training. This technique reduces the likelihood of overfitting by preventing the network from becoming excessively dependent on any individual neuron, thereby enhancing its ability to generalise to unseen ECG data. A max pooling layer with a pool size of 3 is used to decrease feature map dimensionality by selecting the maximum values from non-overlapping segments. This pooling operation reduces computational load and introduces translational invariance by emphasising the most significant features while discarding less relevant information from the ECG signal.

D. Residual block

A residual block is incorporated to enhance the model's ability to learn effectively as the network depth increases. The residual block contains two main convolutional layers; The first layer performs a one-dimensional convolution with 128 filters and 16 kernel size, using ReLU activation function. The filter size dictates the number of features extracted by each convolutional layer, whereas a kernel size of 16 specifies the window across which the convolution is applied, enabling the network to identify significant patterns within the temporal data. Each convolutional layer is followed by a batch normalisation layer, which standardises the outputs and improves the efficiency of convergence during training. A shortcut connection is introduced to allow the output from the CNN block to pass through directly if the dimensions of the input align with the dimensions of the processed output before addition. Mathematically, this can be expressed as:

$$\mathbf{y} = \mathcal{F}(\mathbf{x}, \{W_i\}) + \mathbf{x}, \tag{2}$$

where $\mathbf{x}$ denotes the input signal, $\mathcal{F}(\mathbf{x}, \{W_i\})$ represents the output of the convolutional layers with the learned weights $\{W_i\}$, and $\mathbf{y}$ corresponds to the final output of the residual block. If the dimensions of $\mathbf{x}$ and $\mathcal{F}(\mathbf{x}, \{W_i\})$ do not align, a one-dimensional convolution with a kernel size of 1 is applied to the shortcut connection to match the dimensions, thereby enabling the addition to be performed successfully. Following the two convolutional layers, a max pooling layer is applied with a pool size of 2 to downsample the feature map, reducing its dimensionality and summarising the most prominent features within each pooling window.

E. Attention block

An attention block integrates multi-head self-attention with feed-forward neural networks, using residual connections and layer normalisation to enhance the model's ability to capture complex dependencies in ECG data. This block is designed to maintain stable gradient flow, minimise overfitting, and address vanishing gradient issues, all of which are essential for effectively training deep neural networks. The block begins with layer normalisation, which standardises input activations across feature dimensions for each ECG instance. This is particularly beneficial for ECG data, where signal amplitudes and baselines vary. After normalisation, the block applies multi-head self-attention to capture relationships between different elements of the ECG sequence. The multi-head self-attention mechanism is defined as:

$$\mathbf{Attention}(Q, K, V) = \mathbf{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{3}$$

where $Q = W_Q x$, $K = W_K x$, and $V = W_V x$ are the query, key, and value matrices derived from the ECG input x through linear projections, and W_Q, W_K, and W_V are learnable weight matrices. The term $\sqrt{d_k}$ scales the dot product to maintain gradient stability. The softmax function converts the similarity scores into attention weights, which are used to compute a weighted sum of the value vectors. By employing 256 attention heads, each processing 512 dimensions of the input, the model captures a broader range of dependencies across the ECG sequence, enabling it to focus on different aspects or patterns within the ECG signal. After the multi-head operation, a residual connection is applied, combining the original ECG input x with the output from the multi-head operation. Following the multi-head attention mechanism, a feed-forward neural network is applied. This network comprises two fully connected layers. The first fully connected layer in the feed-forward network expands the input dimensionality to a size of 128, matching the size of the hidden layer. This expansion enables the model to learn richer and more abstract feature representations. By applying a ReLU activation, the layer introduces non-linearity, which enhances the model's ability to capture and process more complex patterns, a particularly important feature for ECG data. The second dense layer reduces the data back to its original input dimensionality. Regularisation is applied through both L1-norm and L2-norm penalties in this layer, while dropout is used between the layers to prevent overfitting. The feed-forward network operates independently on each time step of the ECG sequence, ensuring that the features extracted from each step are processed separately. Finally, the output of the feed-forward network is combined with the residual connection from the multi-head attention layer, preserving the original input information and facilitating effective gradient flow during backpropagation.

III. Experimental results

Four distinct classes were formed in our dataset, with each class containing 200 ECG segments, resulting in a total of 800 ECG segments, as mentioned in Section II-A. After applying data augmentation, the dataset was expanded to 1,600 ECG segments, equivalent to 400,000 ECG samples, which were used for the subsequent experiments. To ensure a robust evaluation of the model's performance, we employed 10-fold cross-validation. In each fold, 90% of the data was used for training, while the remaining 10% was reserved for testing. This partitioning method is widely employed due to its ability to provide substantial training data while simultaneously allocating sufficient resources to assess the model's generalisation capability.

To evaluate the model, performance metrics including accuracy, precision, recall, and F1 score were used to measure the performance of the proposed model:

$$\mathbf{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

$$\mathbf{Precision} = \frac{TP}{TP + FP} \tag{5}$$

$$\mathbf{Recall} = \frac{TP}{TP + FN} \tag{6}$$

$$\mathbf{F1\ score} = 2 \times \frac{\mathbf{Precision} \times \mathbf{Recall}}{\mathbf{Precision} + \mathbf{Recall}} \tag{7}$$

where True Positives (TP) refer to instances where the model correctly identifies a class, while True Negatives (TN) denote cases where the model correctly identifies the absence of a class. False Positives (FP) occur when the model incorrectly assigns a case to a specific class, and False Negatives (FN) arise when the model fails to correctly identify a case belonging to the relevant class. In a multi-class classification, these metrics are computed for each class, with macro averages employed to aggregate performance across all classes.

Additionally, t-distributed Stochastic Neighbor Embedding (t-SNE) analysis is conducted to visualise high-dimensional data and to investigate the impact of each important block in the proposed model on classifying between classes. This technique facilitates the observation of the distribution and clustering of data points in a lower-dimensional space [20] . By applying t-SNE, insights can be gained into the relationships between different classes of ECG signals, aiding in the identification of patterns that may indicate various CVDs.

Table I presents a comparative analysis of the existing traditional ML and DL-based model and our proposed model using 10-fold cross-validation. SVM achieves a training accuracy of 80.04%, but its test accuracy drops significantly to 63.25%, suggesting that the model may have overfitted the training data. Although SVM captures patterns in the training set, its generalisation capability is limited, as indicated by moderate precision of 65.52%, recall of 63.10%, and F1 score of 62.08%. KNN exhibits poor performance, with a training accuracy of 61.72% and a test accuracy of 50.50%, implying that it struggled to model the underlying relationships in the data. Its F1 score of 46.79% further reflects its difficulty in effectively handling this multi-class classification task. Logistic Regression shows the lowest overall performance among the traditional ML methods, with a training accuracy of 52.81% and a test accuracy of 48.45%. The result indicates that the linear nature of Logistic Regression is not suitable for the complexity and non-linearity of the data. In contrast, Naive Bayes outperforms both KNN and Logistic Regression, achieving a relatively balanced training accuracy of 67.95%

TABLE I: Comparative performance of existing Traditional ML and DL-based models and our proposed model using 10-fold cross-validation

Model	Train Accuracy	Test Accuracy	Precision*	Recall*	F1 Score*
SVM	0.8004	0.6325	0.6552	0.6310	0.6208
KNN	0.6172	0.5050	0.6314	0.5053	0.4679
Logistic Regression	0.5281	0.4845	0.6085	0.4904	0.4800
Naive Bayes	0.6795	0.6754	0.7907	0.6753	0.6448
RNN-LSTM [12]	0.7977	0.7845	0.8146	0.7845	0.7839
BiLSTM [13]	0.4356	0.4195	0.4885	0.4195	0.3614
VGG19 [21]	0.3860	0.3866	0.4023	0.3866	0.2997
ResNet-16 [6]	0.7127	0.7104	0.7538	0.7104	0.7061
Convolutional Attention-based [15]	0.9811	0.9750	0.9767	0.9736	0.9742
Baseline CNN [5]	0.9449	0.9713	0.9723	0.9715	0.9711
Proposed model	**0.9979**	**0.9954**	**0.9953**	**0.9955**	**0.9953**

* Precision, recall, and F1-scores are reported as macro averages.

and a test accuracy of 67.54%, indicating improved generalisation. Its higher precision of 79.07% suggests greater effectiveness in avoiding false positives; however, the moderate F1 score of 64.48% indicates that, like the other traditional models, it still struggles to fully capture the complexity of the ECG data

Among existing models, the Convolutional Attention-based model [15] exhibits strong performance, achieving a training accuracy of 98.11% and a testing accuracy of 97.50%. Its precision, recall, and F1 score are closely aligned, each around 97%, demonstrating a well-balanced and robust ability to handle both positive predictions and true positives effectively. Similarly, our baseline CNN model [5] also displays commendable results with a test accuracy of 97.13%, precision of 97.23%, and an F1 score of 97.11%, indicating reliable and consistent performance across key metrics. These results place it among the models, slightly behind the attention-based architecture but still demonstrating excellent generalisability.

However, several other architectures struggle to reach comparable levels of performance as depicted in Table I. For example, the BiLSTM model [13] and the VGG19 model [21] exhibit notably lower test accuracies of 41.95% and 38.66%, respectively. The VGG19 model, in particular, suffers from poor precision, with an F1 score as low as 29.97%, highlighting its imbalanced performance and difficulty in accurately predicting correct outputs. Similarly, the BiLSTM model produces only moderate results, with a precision of 48.85% and an F1 score of 36.14%, underscoring the limitations of these architectures for multi-class heart disease classification.

The RNN-LSTM [12] and ResNet-16 [6] models, although performing better than the BiLSTM and VGG19, still demonstrate only moderate performance. The RNN-LSTM achieves a test accuracy of 78.45%, precision of 81.46%, recall of 78.45% and an F1 score of 78.39%, while the ResNet-16 model achieves a test accuracy of 71.04%, precision of 75.38%, recall of 71.04% and an F1 score of 70.61%. In contrast, the proposed model clearly outperforms all existing models across the evaluated metrics. With a training accuracy of 99.79% and a test accuracy of 99.54%, it not only surpasses its competitors but also sets a new benchmark for performance in multiclass heart disease classification. Furthermore, its precision, recall, and F1 score are exceptionally high, reflecting the model's superior ability to balance true positive classifications with false positives, with values of 99.07%, 98.99%, and 98.97%, respectively. The close alignment of precision and recall further indicates that the model achieves a well-balanced trade-off between sensitivity and precision, which is crucial for ensuring reliable and accurate predictions in real-world applications.

Figure3 visualises the impact of each block on feature extraction using t-SNE. In Figure 3(a), the CNN block shows moderate class separation with some overlap. After the first residual block (Figure 3(b)), clearer clusters emerge, further refined after the second residual block (Figure 3(c)). Eventually, the attention block (Figure 3(d)) yields distinct class separation, with minimal overlap between the normal and arrhythmia classes, highlighting the role of residual and attention blocks in improving feature extraction.

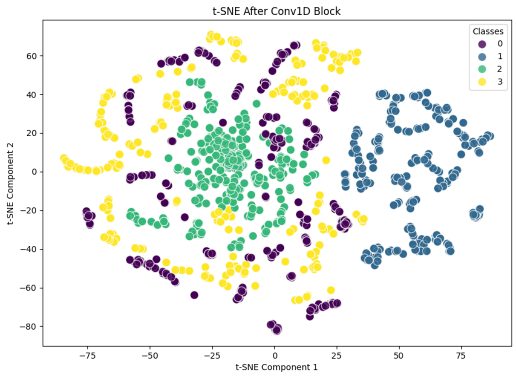
Figure 4 illustrates the relationship between noise power and classification accuracy on the testing set for our proposed model. The noise power, a parameter in data augmentation that employs white Gaussian noise with a standard deviation (σ), ranges from 0.01 to 0.1. The figure shows that moderate noise augmentation (i.e., noise power around 0.06) can enhance the model's robustness by mitigating overfitting and promoting improved generalisation. Therefore, a noise power of 0.06 was selected as the optimal level, as it provided the highest accuracy while also improving the model's generalisation and robustness.

TABLE II: Comparison of Model Block Configurations in the Proposed Model

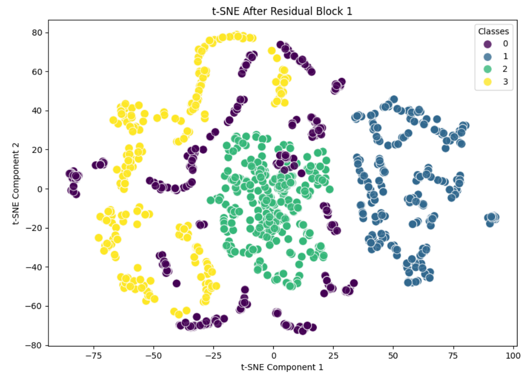
Block			Accuracy	Precision	Recall	F1 Score
CNN	Residual	Attention				
✓	X	X	0.9904	0.9912	0.9902	0.9905
✓	✓	X	0.9920	0.9925	0.9916	0.9919
✓	✓✓	X	0.9900	0.9908	0.9899	0.9898
✓	✓✓	✓	**0.9954**	**0.9953**	**0.9955**	**0.9953**

X: no block used, ✓: one block used, ✓✓: two blocks used.

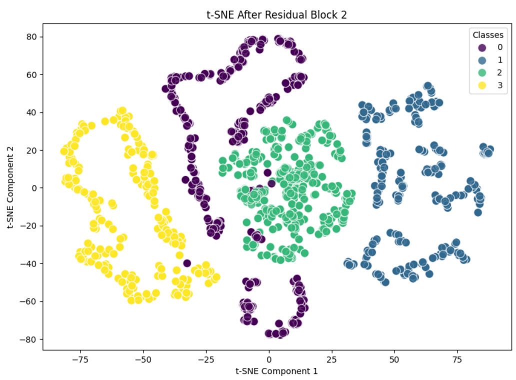
An ablation study was performed to assess different configurations of the convolutional, residual, and attention blocks (see Table II). The baseline CNN achieved an accuracy



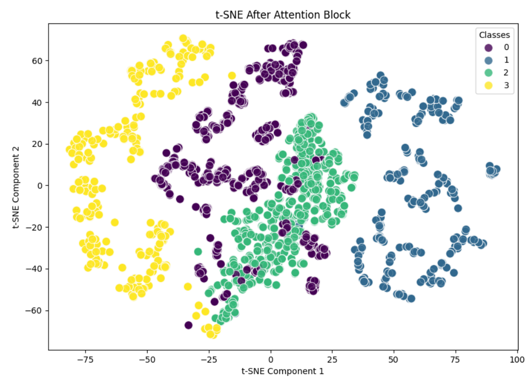
(a) t-SNE visualisation after CNN block



(b) t-SNE visualisation after Residual block 1



(c) t-SNE visualisation after Residual block 2



(d) t-SNE visualisation after attention block

Fig. 3: t-SNE visualisation of the feature space, with class labels: 0 (Normal), 1 (CAD), 2 (Arrhythmia), and 3 (AF).

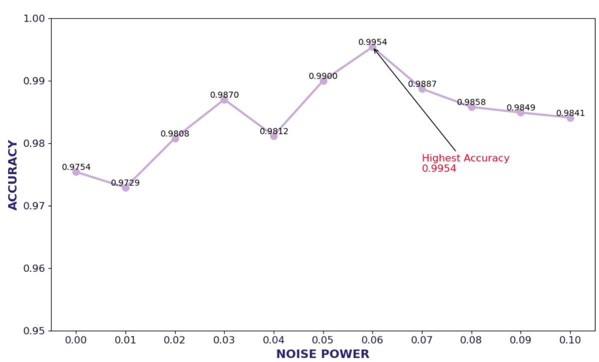


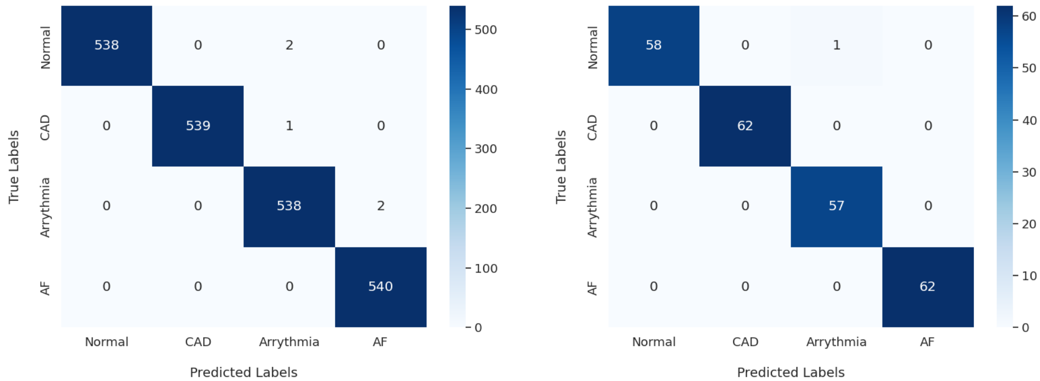
Fig. 4: Impact of noise power variation on the accuracy of multi-class classification.

of 0.9904, demonstrating strong performance with simple architecture. Adding one residual block improved accuracy to 0.9920, showing the benefit of residual connections in enhancing gradient flow. However, using two residual blocks slightly reduced accuracy to 0.9900, suggesting that increasing complexity doesn't always boost performance. The best result, 0.9954, was obtained by integrating CNN, two residual blocks, and the attention mechanism, highlighting the importance of attention in focusing on key features for improved classification.

Figure 5 illustrates the corresponding confusion matrix for the proposed multi-class classification model in both the training and testing phases. Figure 5(a) demonstrates a high level of training accuracy across all categories, correctly classifying 538 instances of Normal, 539 of CAD, 538 of Arrhythmia, and 540 of AF. Notably, only two instances from the Normal class were misclassified as Arrhythmia, and one instance of CAD was similarly misclassified as Arrhythmia. Furthermore, Figure 5(b) shows the correct classification of 58 instances of Normal, 62 of CAD, 57 of Arrhythmia, and 62 of AF in the testing phase. Notably, only one instance from the Normal class was misclassified as Arrhythmia, while no misclassifications were observed for CAD, Arrhythmia, or AF. These results highlight the robustness of the model in accurately differentiating between the classes, even when applied to unseen data.

IV. Conclusion

In this paper, we propose a novel model for classifying multiple CVDs that integrates convolutional layers, residual networks, and attention mechanisms. This model effectively



(a) On training set

(b) On testing set

Fig. 5: Confusion Matrix for the proposed model.

addresses noise in real-world ECG datasets through robust data augmentation, ensuring reliable performance across diverse cardiac conditions. Integrating noise into ECG signals before inputting the ECG data into the model is essential for simulating real-world clinical scenarios. This approach not only enhances the robustness of the model but also improves its accuracy in detecting cardiovascular conditions in the presence of typical environmental noise. The proposed model can serve as a pre-screening tool, as ECG is non-invasive compared to other medical examination methods in hospitals. Future research may focus on refining the model for real-time monitoring and wearable devices, enhancing its clinical utility, and supporting proactive patient management.

References

[1] S. S. Martin, A. W. Aday, Z. I. Almarzooq, C. A. Anderson, P. Arora, C. L. Avery, C. M. Baker-Smith, B. B. Gibbs, A. Z. Beaton, A. K. Boehme, Y. Commodore-Mensah, M. E. Currie, M. S. Elkind, K. R. Evenson, G. Generoso, D. G. Heard, S. Hiremath, M. C. Johansen, R. Kalani, D. S. Kazi, D. Ko, J. Liu, J. W. Magnani, E. D. Michos, M. E. Mussolino, S. D. Navaneethan, N. I. Parikh, S. M. Perman, R. Poudel, M. Rezk-Hanna, G. A. Roth, N. S. Shah, M.-P. St-Onge, E. L. Thacker, C. W. Tsao, S. M. Urbut, H. G. V. Spall, J. H. Voeks, N.-Y. Wang, N. D. Wong, S. S. Wong, K. Yaffe, L. P. Palaniappan, on behalf of the American Heart Association Council on Epidemiology, P. S. Committee, and S. S. Subcommittee, "2024 heart disease and stroke statistics: A report of us and global data from the american heart association," *Circulation*, vol. 149, no. 8, pp. e347–e913, 2024. [Online]. Available: https://www.ahajournals.org/doi/abs/10.1161/CIR.0000000000001209

[2] S. Nahak and G. Saha, "Ensembled feature based multi-label ecg arrhythmia classification," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2023, pp. 1–4.

[3] E. Merdjanovska and A. Rashkovska, "Patient-specific heartbeat classification in single-lead ecg using convolutional neural network," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2021, pp. 932–936.

[4] J. H. Tan, Y. Hagiwara, W. Pang, I. Lim, S. L. Oh, M. Adam, R. S. Tan, M. Chen, and U. R. Acharya, "Application of stacked convolutional and long short-term memory network for accurate identification of cad ecg signals," *Computers in Biology and Medicine*, vol. 94, pp. 19–26, 2018.

[5] A. Phoemsuk and V. Abolghasemi, "Coronary artery disease classification using one-dimensional convolutional neural network," in *2024 IEEE Conference on Artificial Intelligence (CAI)*. IEEE, 2024, pp. 389–394.

[6] Z. Xiong, M. K. Stiles, and J. Zhao, "Robust ecg signal classification for detection of atrial fibrillation using a novel neural network," in *2017 Computing in Cardiology (CinC)*, 2017, pp. 1–4.

[7] W. Xie, C. Chen, R. Zhao, and Y. Lu, "Detection of atrial fibrillation based on feature fusion using attention-based bilstm," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2023, pp. 1–4.

[8] Y. Liu, H. Li, J. Lin, H. Li, H. Lei, C. Xia, C. Xiao, and B. Lei, "Gated cnn-transformer network for automatic cardiovascular diagnosis using 12-lead electrocardiogram," in *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2023, pp. 1–4.

[9] T. Subba and T. Chingtham, "Comparative analysis of machine learning algorithms with advanced feature extraction for ecg signal classification," *IEEE Access*, vol. 12, pp. 57 727–57 740, 2024.

[10] M. M. Rahman, M. W. Rivolta, F. Badilini, and R. Sassi, "A systematic survey of data augmentation of ecg signals for ai applications," *Sensors*, vol. 23, no. 11, p. 5237, 2023.

[11] E. Prabhakararao and S. Dandapat, "Congestive heart failure detection from ecg signals using deep residual neural network," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 53, no. 5, pp. 3008–3018, 2023.

[12] F. Corradi, J. Buil, H. D. Cannière, W. Groenendaal, and P. Vandervoort, "Real time electrocardiogram annotation with a long short term memory neural network," in *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2019, pp. 1–4.

[13] J. Cheng, Q. Zou, and Y. Zhao, "Ecg signal classification based on deep cnn and bilstm," *BMC medical informatics and decision making*, vol. 21, pp. 1–12, 2021.

[14] P. Xu, H. Liu, X. Xie, S. Zhou, M. Shu, and Y. Wang, "Interpatient ecg arrhythmia detection by residual attention cnn," *Computational and Mathematical Methods in Medicine*, vol. 2022, no. 1, p. 2323625, 2022.

[15] T. Zhu, W. Luo, and F. Yu, "Convolution-and attention-based neural network for automated sleep stage classification," *International Journal of Environmental Research and Public Health*, vol. 17, no. 11, p. 4152, 2020.

[16] J. Zheng, H. Guo, and H. Chu, "A large scale 12-lead electrocardiogram database for arrhythmia study (version 1.0.0)," PhysioNet, 2022. [Online]. Available: https://doi.org/10.13026/wgex-er52

[17] X. Wang, C. Ma, X. Zhang, H. Gao, G. D. Clifford, and C. Liu, "Paroxysmal atrial fibrillation events detection from dynamic ecg recordings: The 4th china physiological signal challenge 2021 (version 1.0.0)," PhysioNet, 2021. [Online]. Available: https://physionet.org/content/challenge-2021/1.0.0/

[18] A. E. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark, "Mimic-iii, a freely accessible critical care database," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[19] N. Iyengar, C. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz, "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 271, no. 4, pp. R1078–R1084, 1996.

[20] C. Xiao and J. Sun, *Introduction to deep learning for healthcare*. Springer Nature, 2021.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.