Multi-Label Cardiovascular Disease Classification from 12-Lead ECG Using Enhanced DenseNet-121 with CBAM and Grad-CAM Explainability

Munji Gayathri[1*], Suresh Chittineni[2]
[1,2]Department of Computer Science and Engineering, GITAM School of Technology, GITAM (Deemed to be University), Visakhapatnam, AP, India
E-mail: gmunji@gitam.in, schittin@gitam.edu

Keywords: 12-Lead ECG, cardiovascular disease classification, enhanced DenseNet-121, explainable AI, multi-label classification

Received: July 3, 2025

Early and accurate diagnosis of cardiovascular diseases (CVDs) is critical, as they are one of the most prevalent causes of death across the globe. We propose a multi-label classification model consisting of an Enhanced DenseNet-121 (EDN-121) deep learning framework for 12-lead ECG signals obtained initially from the CPSC 2018 dataset (in total 6,877 samples) for eight cardiovascular conditions. Data Preprocessing: A Butterworth high-pass filter was employed to remove noise from the ECG signals, followed by normalization to ensure the data was within a consistent range. Then, the Synthetic Minority Oversampling Technique (SMOTE) was applied to fix the class imbalance problem. EDN-121 architecture incorporates a Convolutional Block Attention Module (CBAM), which contributes to superior spatial and channel-wise representation features. Additionally, Gradient-weighted Class Activation Mapping (Grad-CAM) ensures the explainability of the prediction by identifying the clinically relevant region on the ECG. Experimental results based on five-fold cross-validation show that it outperforms the baseline VGG-16 with 97.90% accuracy, 97.75% F1-score, and 97.34% recall. Conclusion: This study demonstrates the performance and interpretability of a machine learning framework that has the potential to be implemented in clinical settings for the real-time, automated ECG-based CVD screening.

Povzetek: Predstavljen je razložljiv model EDN-121, ki z visoko natančnostjo omogoča avtomatizirano presejanje srčno-žilnih bolezni iz EKG.

1 Introduction

One of the top causes of death worldwide, cardiovascular diseases, led to 17.9 million deaths in 2019 32 % of all deaths. Furthermore, the World Health Organisation reports that 75% of these deaths occur in low- and middle-income nations. Thus, it's essential to identify these cardiac issues as soon as possible so that medication and counselling can start as part of the treatment. An electrocardiogram, or ECG, is a waveform representation of the heart's electrical activity acquired by applying electrodes to the skin.

The heart and its blood arteries are vulnerable to a group of disorders collectively known as CVD, which is a leading cause of mortality globally. The accumulation of fatty deposits within the arteries is the most common cause of this condition [1]. From 12.3 million (25.8% of the total) in 1990 to 17.9 million (32.1%) in 2015, the number of fatalities worldwide attributable to CVD has risen significantly [2-3]. Therefore, to help clinicians accurately diagnose different CVDs, an automated technique is required for heart disease diagnosis. When it comes to diagnosing heart conditions, electrocardiograms (ECGs) are among the most widely used diagnostic tools for this purpose.

The typical ECG signal structure, as shown in Figure 1, consists of three primary components: the P wave, which denotes atrial depolarization, the QRS complex, which represents ventricular depolarization, and the T wave, which denotes ventricular repolarization. The signal also consists of PR, ST segments, and QT intervals. This electrical signal is a popular non-invasive method for detecting irregularities in a patient's cardiovascular system. Nevertheless, ECG analysis is a specialized and time-consuming procedure that requires highly skilled cardiologists to inspect and identify abnormal ECG recording patterns closely. This difficulty, together with the exponential growth in Computer-aided, automatic ECG analysis, makes ECG data increasingly important, particularly in low- and middle-income nations where

there is a severe shortage of skilled and knowledgeable cardiologists.

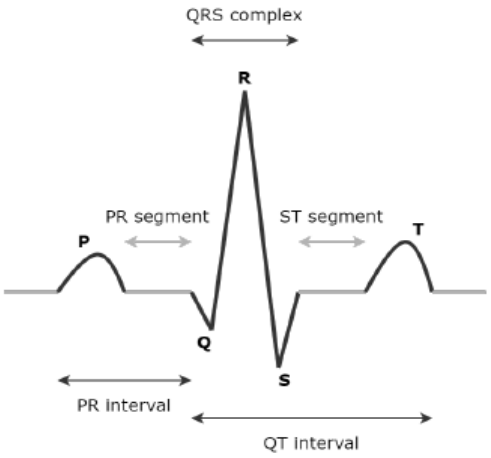


Figure 1: Structure of ECG signal

Therefore, there is a growing interest in the precise and automated diagnosis of ECG signals. The MIT-BIH Arrhythmia database as well as wide open-source ECG datasets are also accessible [4], the 2017 Physionet Challenge/CinC dataset [5], the dataset from the 2018 China Physiological Signal Challenge (CPSC2018) [6], and the PTB-XL database [7], automatic ECG diagnosis has been extensively studied in recent decades.

The study developed a DNN for detecting cardiovascular diseases and performing automatic multilabel classification of different cardiac diseases using 12-lead ECG signals. The simple work process schematic is shown in Figure 2. The evaluation of work is also performed on single-lead ECG signals. An inclusive and explainable DNN is performed for the classification of ECG signals by utilizing the technique of gradient-weighted class activation mapping.

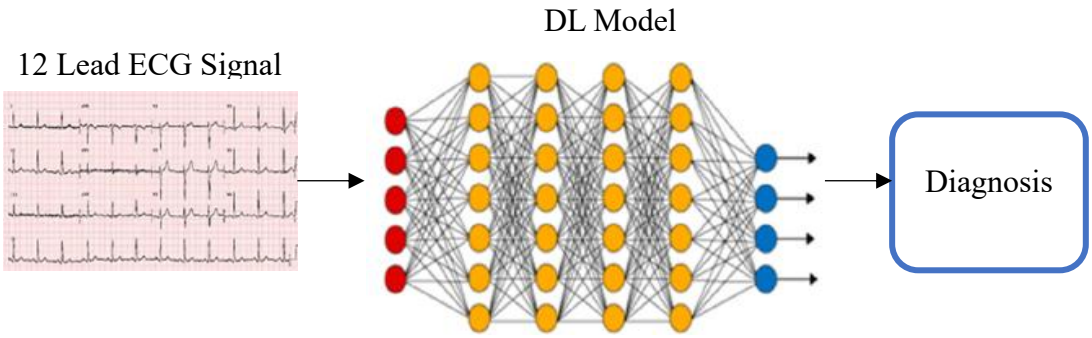


Figure 2: Diagnosis model

The work performed in this paper is summarized as follows:

a. A DNN concept is developed for the automatic identification of cardiovascular diseases (CVD).
b. The identification of CVD is performed using a 12-lead ECG dataset.
c. The results obtained using the proposed DCNN model are compared with other existing classifiers.
d. The experiment is conducted on 12 lead signals, in which leads I, aVR, and V5 perform well in detecting CVD.
e. Finally, for improved results, an explainable DCNN is used to establish proper segments of the ECG input and achieve perfect classification depending on the corresponding heart disease.

This is how the remainder of the paper is structured. We provide a summary of relevant studies on the automated diagnosis of ECG in Section II. Section III explains the deep neural network architecture and the 12-lead ECG dataset. In Section IV, we present the experimental outcomes and discuss the model's interpretability. In Section V, we finally draw the paper's conclusion.

How this research translates into a method is hard to establish and to identify it, the latter section outlines the problems that need to be solved, and the objectives of this study.

1.1 Problem definition

ECG signals can have numerous intermingling abnormalities — which makes them complex to handle — in a single ECG recording so the task of diagnosing heart disease using an ECG signal is a rather complicated one. This inherently creates a multi-label classification problem, as a given sample may often have more than one condition in common simultaneously. This study develops an automated deep learning framework that can simultaneously identify 8 cardiovascular conditions from

12-lead ECG data. We aim to develop a tool for clinicians to aid in clear and efficient diagnoses without the burden of manually interpreting ECG recordings.

The objective of this research is to design a model that learns to correlate raw 12-lead ECG input signals to their respective diagnostic outputs. We evaluate the performance of the proposed framework by three central evaluation criteria, accuracy, F1-score and Hamming Loss. Accuracy indicates the general correctness of the predictions, and the F1-score tells us more about the balance between precision and recall, which is of much importance with imbalanced datasets for cardiac conditions, where some exist much less frequently than others. Another perspective on how well the model is dealing with multi-label classification is Hamming Loss, which evaluates the fraction of wrong predictions by counting the total number of labels.

This paper seeks to elaborate on some key research questions: It first explores whether applying to DenseNet-121 a Convolutional Block Attention Module (CBAM) can help the model learn basic spatial and channel-level representations of the subtly varying features in ECG signals. Second, it investigates the performance of Synthetic Minority Oversampling Technique (SMOTE) on class imbalance, a typical problem that arises in real-life clinical datasets. Third, it assesses whether the application of Grad-CAM on the model improves the interpretability of the model outputs by providing insight into ECG regions that may contribute to the decision-making process of the model that can be understood by clinicians. Finally, a comparison of the proposed Enhanced DenseNet-121 against a baseline CNN architecture, VGG-16, is performed to assess improvements in performance and clinical applicability.

Specifically, answering these questions would allow us to build an automatic model with high accuracy for ECG signal-based cardiovascular disease detection while allowing us to better interpret the results of such a model which the proposed research aims to do. It can close the gap between sophisticated deep learning methods and practical clinical needs to improve accessibility, efficiency, and reliability of ECG analysis.

2 Related work

To identify cardiovascular diseases (CVDs), one typical medical method is the categorization of cardiac arrhythmias, or abnormal heartbeats. This is the subject of our review. The author in [8] suggests a machine learning model, specifically SVM, for detecting heart disease. The KNN algorithm, a machine learning model, is employed by the author in [9] for predicting heart disease, achieving an accuracy of 92.46%. In recent years, DL has been used to classify cardiac abnormalities from ECG readings automatically. DL has demonstrated impressive performance in medical diagnosis. The learned mapping of relevant medical categories from ECG data can be observed across multiple sensory neural layers in deep learning (DL) models. The DL model's inference power is enhanced by tuning the neuron weights using training datasets to narrow the gap between the inferred categories and the ground-truth categories of the training data [10]. Due to its strong multi-level abstraction capability in feature extraction, DL-based ECG classification can more accurately map the features of ECG signals to their specific categories [11]. In comparison to more conventional ML-based classification techniques, such as clustering and SVM, DL-based ECG classification can yield more accurate results.

The potential usefulness of AI-ECG has been established through clinical studies of AI-ECG diagnostics conducted at the Mayo Clinic for the identification of various cardiovascular diseases [12]. They do, however, conclude that the use of AI-ECG diagnosis is in stages. As a result, although DL approaches are successful for classifying ECGs in the research community, obstacles related to both DL techniques and ECG data have hindered their use in actual clinical settings.

The author in [13] aims to familiarise the physician with the core principles of DL, the state-of-the-art before it was applied to ECG analysis, and the present usage of DL on ECGs, along with their shortcomings and potential areas of future development. In particular, the 1D-CNNs' structure has made them suitable for handling ECG data. The author in [14] created a deep (CNN) with nine layers to recognise five distinct pulse types in ECG readings automatically. Original and attenuated ECG signal sets from a publicly accessible database were used for the experiment. The number of occurrences of each of the five types of heartbeats was artificially increased, and high-frequency noise was removed by filtering. After being trained on the supplemented data, the CNN was able to diagnose heartbeats with an accuracy of 94.03% in the original ECG and 93.47% in the noise-free one.

Using 12-lead ECG data to automatically classify cardiac arrhythmias, the author in [15] created a deep neural network, i.e., 1D-ResNet34. The average F1 score attained by the suggested model was 81.3%. From 12-lead ECG datasets, the author in [16] developed an algorithm that automatically detects 27 ECG anomalies. The study employed two rule-based models and one SE_ResNet model to enhance the classification performance of various ECG abnormalities. Our suggested method ranked us third out of 40 in the official ranking of the PhysioNet/Computing in Cardiology Challenge (2020), with a challenge validation score of 68.2% and a complete test score of 51.4%.

Researchers in [17] suggested a (TI-CNN). It featured > 90% in real-time processing computation savings and halved parameter quantity. Recurrent cells were used to introduce flexibility in input length for CNN models.

According to the experiment's findings, TI-CNN achieved a total classification accuracy of 77.3%. Srinivas et al. CardiacNet, a CNN-based model with Improved Harris Hawks Optimization (IHHO) for feature selection achieves an accuracy of 97.57% in arrhythmia detection using 12-lead ECG data [18]. While their work exemplifies how optimized deep learning frameworks can be applied to classify cardiovascular diseases accurately, it is limited by an explainability [11] and multi-label classification capability [13], which our study overcomes.

Aswani et al. An explainable clinical decision support framework with the use of Random Forest models has been proposed [19], contributing towards making healthcare AI systems more transparent and interpretable. This highlights the crucial necessity of having explainable models in medical applications. Although quite successful, it does not utilize deep learning or ECG data, thus making our initial integration with Grad-CAM, a significant step toward explainable diagnostics.

Zhang et al. Hexaxial feature mapping was introduced for ECG signal representation in [20], which is effective in extracting spatial and directional features from 12-lead ECG data. The principles behind their work on COVID-19 detection guided our own pre-processing techniques, combining filtering and normalization in order to boost downstream deep learning classification by conveying better information through the representation of signals.

The author of [21] said that a limb 6-lead ECG might also be used to identify myocardial infarction in addition to the standard 12-lead ECG. The research contributes to the growing body of evidence demonstrating that ECG can detect a range of cardiovascular issues, not just a select few. DL models' opaque character may be overlooked in many situations, but it undermines accountability and confidence in judgements made in delicate fields like healthcare and medicine. As a result, scientists have begun incorporating well-known XAI methods for image data into ECG data. Using 12-lead ECG data, the author in [22] trained a CNN model to predict the existence of 38 diagnostic classifications across five categories. The ECG segments that contributed to CNN diagnosis were visualised using an explainability approach termed LIME for CNN.

To diagnose cardiac abnormalities, the author in [23] used a variety of DNNs on a dataset of PTB-XL ECG signals that was made publicly accessible. The developed ST-CNN-GAP-5 model achieved an AUC of 93.41%, outperforming the current state-of-the-art results on this dataset. Lastly, the trained ST-CNN-GAP-5 is subjected to an analysis of the ECG data using SHAP to evaluate the explainability or interpretability of the DCNN model's conclusions.

Explainable artificial intelligence (XAI) has emerged and is rapidly expanding as a solution to the black box problem [24]. This review focuses on perturbation-based XAI techniques, which enable the exploration of DNN models through input perturbations and output modifications. For ECG-based healthcare applications, the author in [25] created a novel (XAI) based DL architecture in a federated environment. The federated option is used to address privacy concerns and data availability. Additionally, the autoencoder and classifier, both of which are based on convolutional neural networks (CNNs), operate effectively under the suggested framework setting to classify arrhythmias. Furthermore, we recommend incorporating an XAI-based module into the suggested classifier to provide explanations of the classification outcomes, thereby assisting medical professionals in making prompt and accurate judgments.

This research [26] presents X-ECGNet, an explainable multi-branch CNN for categorizing multi-label 12-lead ECG data. With this method, characteristics from each lead in the 12-lead ECG can be extracted individually and then concatenated to maximize the amount of data obtained. This model uses this architectural layout to apply the (Grad-CAM) method to produce explanations for every ECG lead. Lastly, we test our model's performance on the PTB-XL dataset. The AUC score achieved is 93.6%.

Table 1: Summary of existing studies on ECG-based cardiovascular disease classification

Author/Year	Method Type	Dataset Used	Leads	Performance Metrics	Limitations
Al-Jammali [8] (2023)	SVM	Custom Dataset	12	Accuracy: 89%	Traditional ML, not explainable
Joseph & Kartheeban [9] (2025)	KNN	Custom Dataset	12	Accuracy: 92.46%	No deep learning, low scalability
Dongdong Zhang et al. [15] (2021)	1D-ResNet34	12-lead ECG Dataset	12	F1 Score: 81.3%	Limited class coverage, no XAI
Zhu et al. [16] (2021)	SE-ResNet + Rule-Based	PhysioNet Challenge	12	Validation Score: 68.2%	Poor generalization, low AUC
Ganeshkumar et al. [26] (2023)	Multi-Branch CNN + Grad-CAM	PTB-XL	12	AUC: 93.6%	No SMOTE, imbalance not addressed
Cai et al. [44] (2020)	Multi-ECGNet	CPSC	12	F1 Score: 86.3%	Limited explainability features
Proposed Work (2025)	**Enhanced DenseNet-121 + CBAM + Grad-CAM**	**CPSC 2018**	**12**	**Accuracy: 97.90%, F1 Score: 97.75%**	**Addresses class imbalance, explainability, and multi-label classification**

Recent works on ECG-based cardiovascular disease classification can be found summarized in Table 1. While most previous works have focused on classification accuracy with deep learning architectures like ResNet and CNNs, they have been limited to non-explainability, imbalanced datasets, and a limited coverage of diagnostic classes. Some approaches, like [15] and [16], gained moderate F1 scores but did not include any mechanism, like Grad-CAM, to interpret the model. Explainability was also done in [26], although it did not address class imbalance. Our Enhanced DenseNet-121 framework, which incorporates the Combined Attention Module, Synthetic Minority Over-Sampling Technique, and Grad-CAM, can overcome these limitations by extracting clinically relevant and interpretable features from the images.

3 Methodology

The methodology describes the study of a DL model to classify the ECG signals and an explainable deep neural network for obtaining proper segments of the ECG signal. The section also describes the Grad-CAM for DCNN. Using a collection of 12-lead ECG recordings and the associated co-occurring illness labels, the objective is to predict the presence of several co-occurring cardiac conditions in each ECG record. Correct categorization of various conditions is challenging, as co-occurring disorders often produce irregular ECG characteristics across all 12 leads. The proposed methodology is shown in Figure 3.

A. Input dataset

The dataset utilized to process the proposed model is CPSC 2018 [27]. The signals available in the dataset are obtained from 11 hospitals in China with 12-lead ECG data. This dataset consists of ECGs with a size of 6877, and the signal length varies from 6 seconds to 60 seconds. The sampling rate at which the signals are collected is 500Hz. Nine diagnostic classes are identified on these ECGs: ST elevation (STE), premature atrial contraction (PAC), premature ventricular contraction (PVC), left bundle branch block (LBBB), right bundle branch block (RBBB), normal sinus rhythm (NSR), and ST depression (STD).

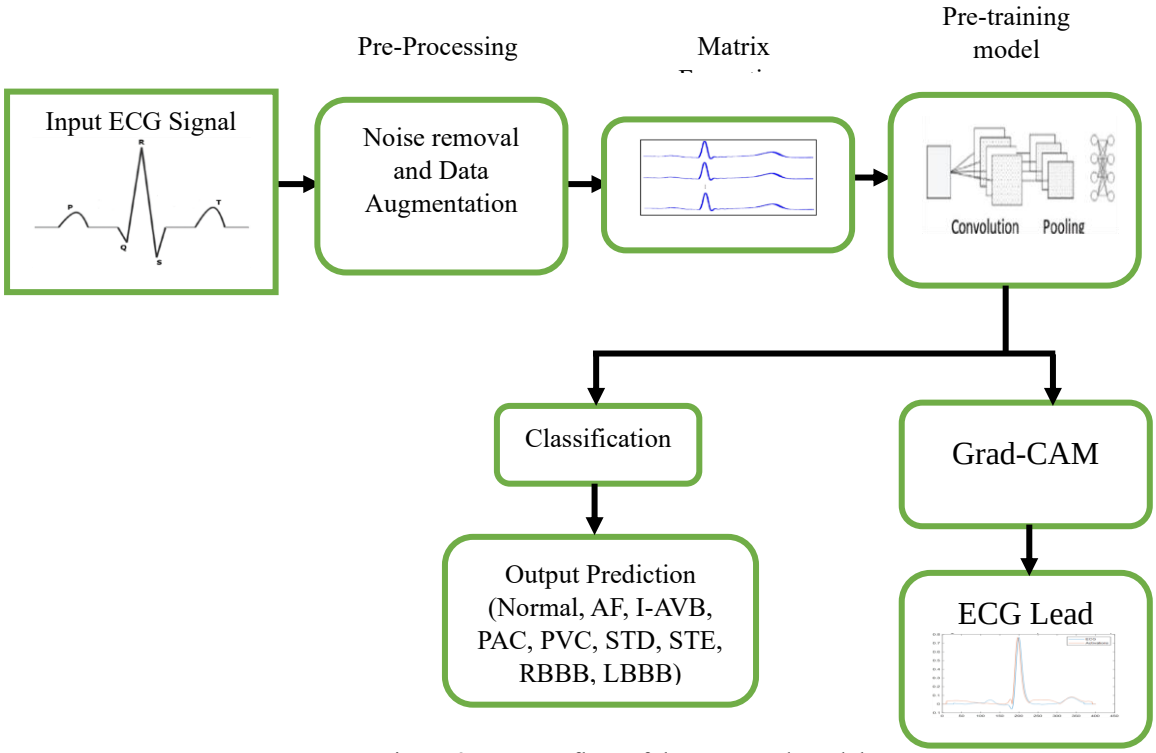


Figure 3: Process flow of the proposed model

B. Pre-processing

Several disturbances, including (BW), (PLI), and muscle activity, can significantly impact the practical ECG signals and alter their morphology [28]. As a result, the ECG-lead signal is a high-pass filter, which has a passband ranging from 0.5 to 45 Hz, to reduce any potential noise effects. Additionally, because the amplitudes of ECG signals vary across participants, the ECG lead is z-score normalized to maintain a mean of zero. For the given input signal as shown in Figure 4, the pre-processed signal after removing the noise is shown in Figure 5.

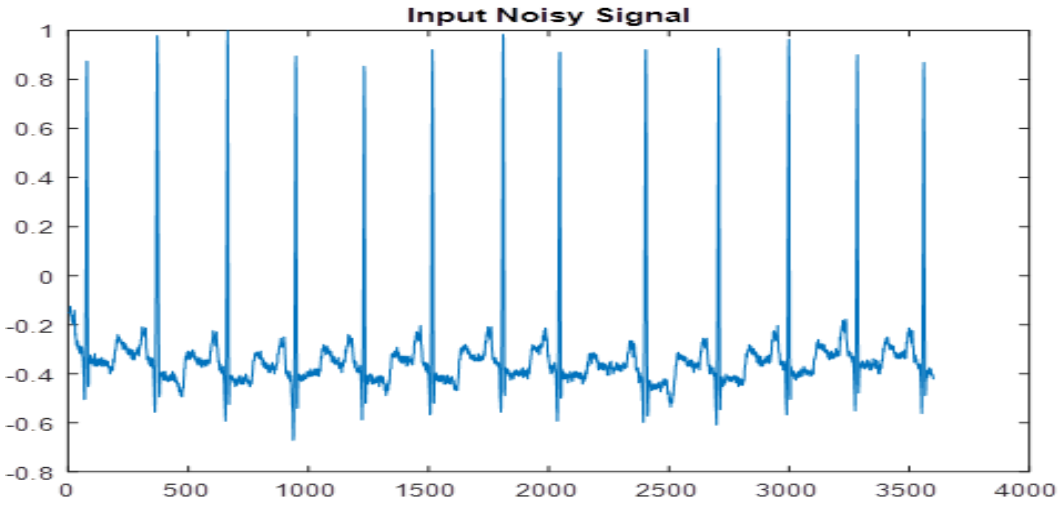


Figure 4: Input noisy signal

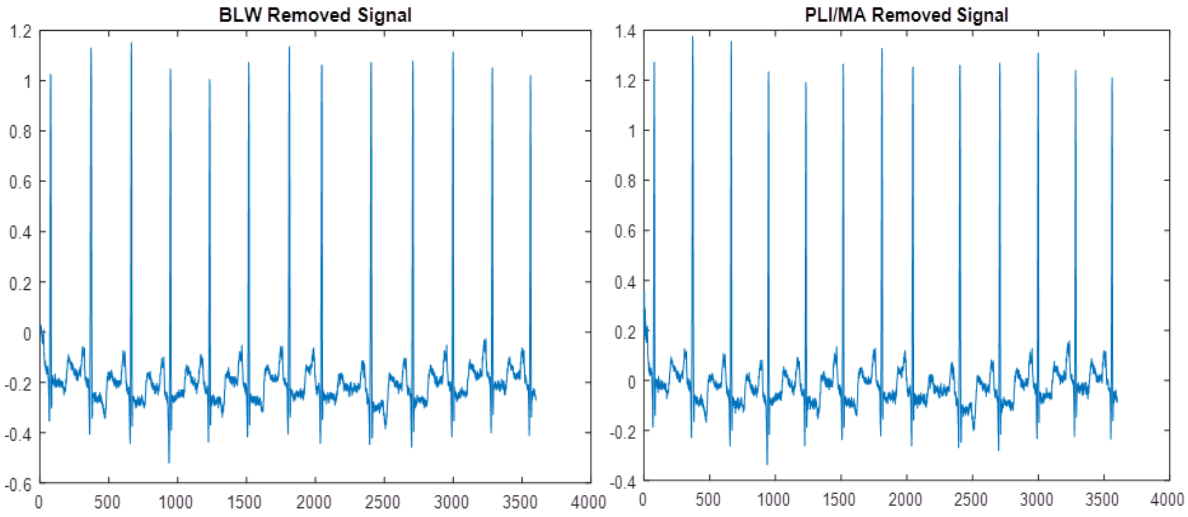


Figure 5: BW and PLI noise removed from the ECG signal

After removing noise, the imbalance of data and insufficiency problem leads to incorrect diagnosis of CVDs. We used scaling and shifting for data augmentation to solve this issue. Data augmentation can promote resilience against hostile cases and lessen overfitting in models. This is why data augmentation is performed to ensure there is no imbalance. Oversampling members of underrepresented groups in training datasets is a part of data augmentation. We employ SMOTE [47], which stands for Synthetic Minority Over-sampling Technique, to increase the sample size of the minority group. To achieve a more even distribution of classes throughout the dataset, it employs an oversampling strategy.

C. Matrix formation

The ECG signal is divided into segments, also known as windows. Every segment is arranged in a matrix form, either as a row or a column. Each row or column indicates the segment of the ECG signal. The matrix formation of the signal is shown in Figure 4. As stated by the author in [29], the leads aVF, V5, and V6 perform well in identifying CVDs. This helps in reducing the complexity of work. Further beat segmentation and normalization are performed on the ECG signal, which is then given as input to the DNN. This will improve disease detection, and heart rate variability will be analyzed.

Beat segmentation

Segmentation of beat entails recognising the start and end points of each heartbeat within the ECG signal. This generally includes detecting essential locations such as the R-peaks (the highest points in the QRS complex) since they are the most noticeable elements in the ECG signal. Beat segmentation uses the Pan-Tompkins method [46].

Period normalization

The files contain 12-lead ECG samples with varying durations, averaging 10 seconds, which is a beat normalization. Deep neural networks require uniformly sized inputs; therefore, all datasets were pre-processed to ensure they were 10 seconds in length. This was achieved by clipping samples that were longer than 10 seconds and padding shorter samples with zeros. As the length of the ECG signal is restricted to 10 seconds, the matrix formation of the signal will be simplified. The sampling rate is 500Hz. The shape of the matrix is 12 × 2500. There is no overlap between the segments, and each row in the matrix is considered as one ECG lead.

Finally, the ECG signals are treated as 2D arrays (12 leads × 2500 samples), which are then fed into a 1-channel pseudo-image format to match the requirements of DenseNet-121.

D. Design of DNN

The dot products serve for the input's convolution of layers of input and weights. The set of filters is then moved up and down in the direction of the input. In the ReLU layer, the threshold process sets values below zero to zero. In the max pooling layer, later downsampling is performed by splitting the input into a rectangular pooling region to find the maximum value of each zone. Before adding the bias vector to the fully connected layer, the input is multiplied by a weight matrix. The main deep CNN architecture is shown in Figure 6.

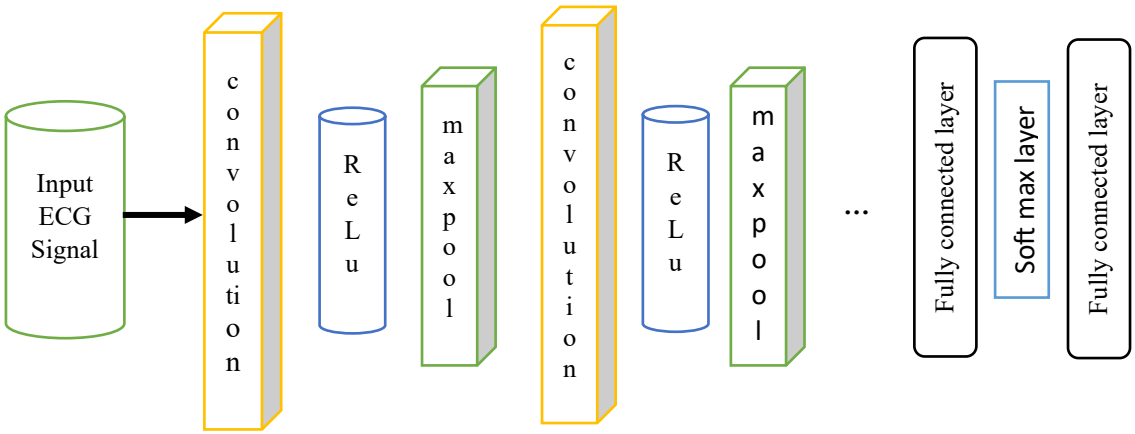


Figure 6: DCNN architecture

The network utilized is the Enhanced DenseNet-121, and the process of training the Enhanced DenseNet-121 model is shown in Figure 7. This network connects one layer to another layer in a feed-forward way. More specifically, the feature maps of all previous layers are fed into the next layer. Dense connectivity refers to this type of connection structure. The key elements of DenseNet-121 are the dense block, transition layer, and bottleneck layer.

We used five-fold cross-validation to determine the initial learning rate of 0.001, and trained the model for a total of 100 epochs. The learning rates tested were 0.01, 0.005, and 0.0005, and 0.001 struck the best balance of speed of convergence to stability. Given that we start from a pre-catapult point (pre-trained), applying early stopping and a learning rate scheduler to prevent overfitting were really only a matter of keeping the two curves smooth from training to validation. The ultimate configuration showed consistent results in all folds which confirmed the robustness of the selected hyperparameters.

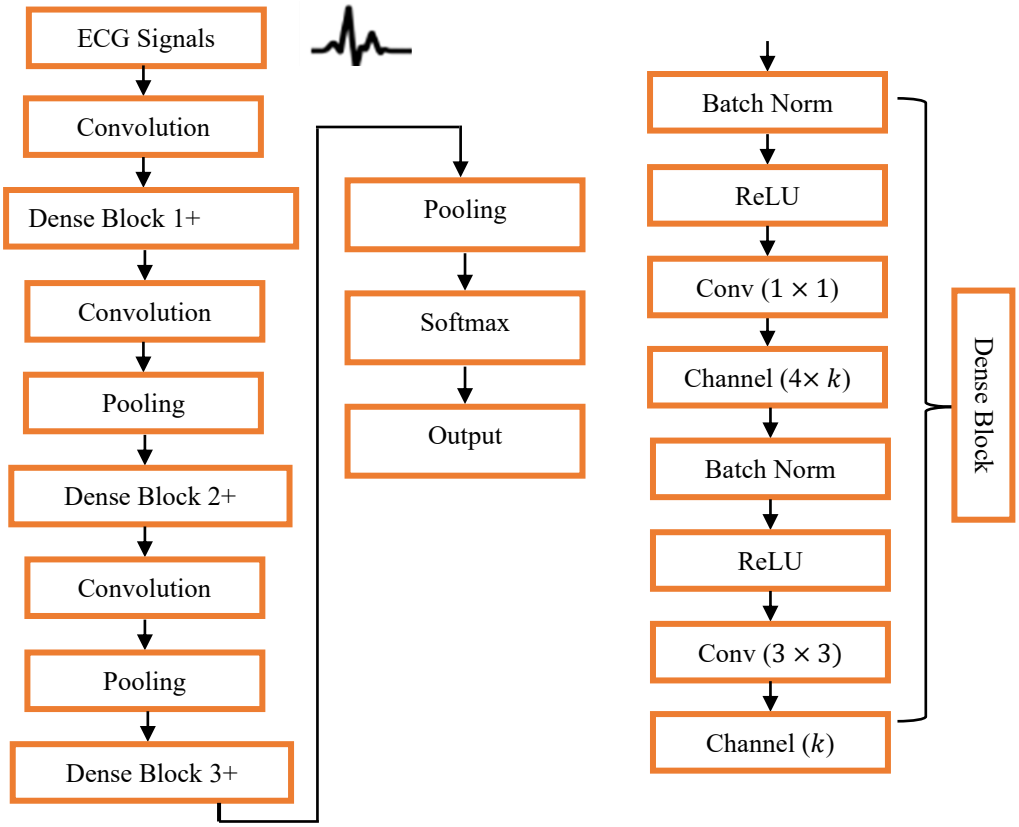


Figure 7: Proposed enhanced densenet-121 Model

The architecture features a 1x1 convolution layer and a 2x2 average pooling layer, which follow the dense blocks. The transition layers are simple to concatenate since the feature map sizes are the same. The feature size map is reduced with the aid of the pooling layer. Finally, a global average pooling is carried out after the dense block and is connected to a softmax classifier. Each class is given a probability distribution function, which is how the softmax layer operates. Softmax assigns a decimal probability to each class in a multi-class problem—their decimal odds sum to 1.0. Training converges more quickly with this additional restriction than it would have otherwise. To get the desired results, this layer is positioned before the output layer. The model is enhanced by combining a dense block with CBAM, i.e., a CAM and a SAM are added after performing the dense block operation to improve the DenseNet 121 model. The CBAM is shown in Figure 8.

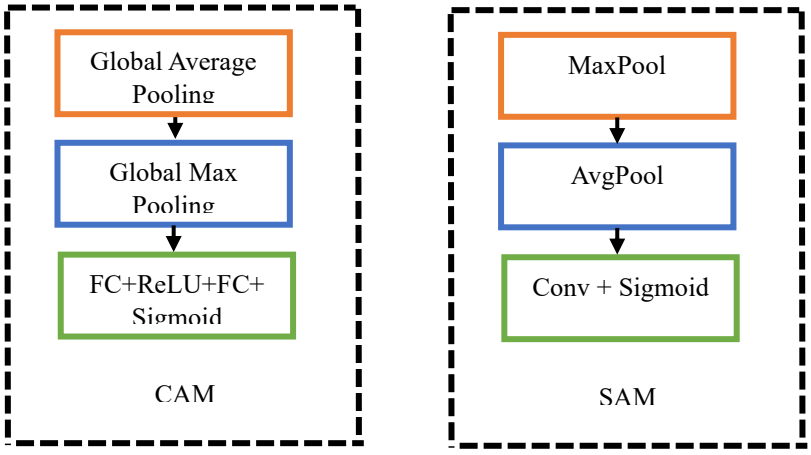


Figure 8: CBAM block

E. Explainable DNN

Explainable methods for machine learning have been developed in the past few years to provide consumers with more information, improve data insights, and give algorithms a more descriptive approach. We propose a new method to enhance the interpretability of DNNs by changing encoded input into helpful information. An in-depth understanding of how a DNN modifies input for classification will be provided by examining the data structure in each hidden layer representation. Here, the identified problem is solved using the Grad-CAM approach. Adding CBAM after DenseBlocks improves feature refinement.

Gradient-weighted class activation mapping (Grad-CAM)

One of the most popular XAI methods for improving our understanding of CNN-based models is Grad-CAM. Grad-CAM enables the identification of the most significant input region for predictions. The weighted class will be assigned a score and is termed as O^c, in which 'c' is said to be the class. Now the gradient of the CNN layer i.e, the last convolution layer need to evaluated and is termed as $\frac{\partial o^c}{\partial R^n}$, R^n are said to be the features maps. Next, to estimate the neuron significance weights $\hat{w}_n^c$, which are specified as:

$$\hat{w}_n^c = \frac{1}{Z}\sum_i \sum_j \frac{\partial o^c}{\partial R_{i,j}^n} \quad (1)$$

where (i, j) the pixels and Z tells the number of pixels exist in a feature map. At last, a weighted mix of R^n is computed:

$$\hat{w}_n^c = ReLU \sum_n \hat{w}_n^c R^n \quad (2)$$

where $ReLU$ is the rectified linear unit transfer function.

A standard Grad-CAM output is a significance heatmap that highlights the input regions contributing most strongly to the classification decision. The phrasing seeks to clearly summarise this output. In our setup, the heatmap represents the 2D input structure (Leads × Time) and graphically represents low to high relevance using a blue-to-red colour gradient. The text will be revised to clearly state the spatial dimensions (Leads × Time) and emphasize that the colour scale indicates the relative relevance scores from the Grad-CAM activation map.

A weighted map generated via Grad-CAM highlights the key areas of the input that the CNN uses to determine the class label. The class-activation maps of the sample ECG matrices that our CNN was able to classify correctly were obtained using Grad-CAM. The activation of the ECG matrices along the beat space under study (Leads × Time) is calculated by averaging the Grad-CAM activations across all detected beats within a sample, as the class activation maps are found. This process is explained in Figure 9.

The Grad-CAM heatmaps were visually inspected to ensure that identified regions of the ECG that were indicated to be important for diagnosis corresponded to expected regions consistent with prior knowledge, including P-waves, QRS complexes, and ST-segments. While we did not validate the clinical interpretability of the model outputs in this study through a domain expert, we will provide cardiologist annotations in future work, along with quantitative metrics (e.g., IoU), to formally assess the clinical interpretability of the model outputs.

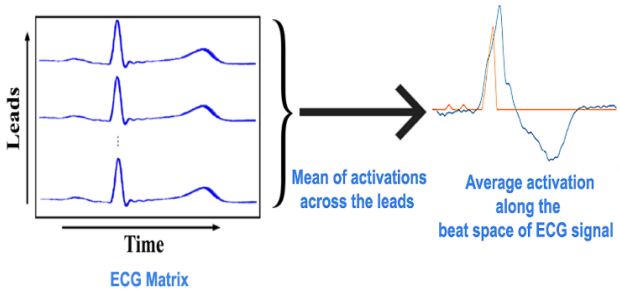


Figure 9: Mean Activation of ECG Matrix [43]

4 Results and discussion

Using electrodes applied to the patient's limbs and chest, a standard 12-lead ECG is recorded. Clinics and hospitals are the settings for these recordings. This means that the 12 ECG leads can be divided into two main groups: six leads related to the limbs (I, II, III, aVR, aVL, and aVF) and six leads pertaining to the chest (V1, V2, V3, V4, V5, and V6).

The conventional 12-lead ECG has been demonstrated in previous research to be useful in various ECG analysis applications. In contrast, obtaining a 12-lead ECG is not essential and largely depends on specialized equipment that is not readily accessible. Advances in ECG technology have led to the development of more user-friendly, compact, and reasonably priced ECG-enabled devices. By removing noise from various ECG signal types, the effectiveness of the suggested framework is assessed. A 5-fold stratified cross-validation with an 80-20 train-test split in a DenseNet-based ECG signal classification task, while ensuring balanced label representation.

4.1 System environment

The model is designed and implemented using the MATLAB software environment with the 2021a version. The Deep Neural Network Toolbox was utilized to evaluate the model. The system configuration considered for training the network and analyzing with GPU- GeForce 940MX; NVIDIA, and computing using Windows 10 64-bit OS- 16 GB RAM.

4.2 Design Implementation

Since all inputs to a deep learning system must have the same duration, all ECG records in both datasets were set to 10 seconds. The number of epochs considered is 100, the learning rate is 0.001, and the training time is approximately 9 hours. For longer recordings, this was accomplished by trimming the portion that lasted longer than the initial 10 seconds and padding the shorter ones with zeros. In this paper, the considered signal is represented in matrix form, and particular leads, such as aVF, V5, and V6, from each ECG recording are given more importance when fed into our algorithm for processing. Various combinations of leads are considered and evaluated to assess the model's performance. In every combination of leads, we utilized aVF, V5, and V6**.** The results obtained using the proposed model are shown in Figures 10 to 18.

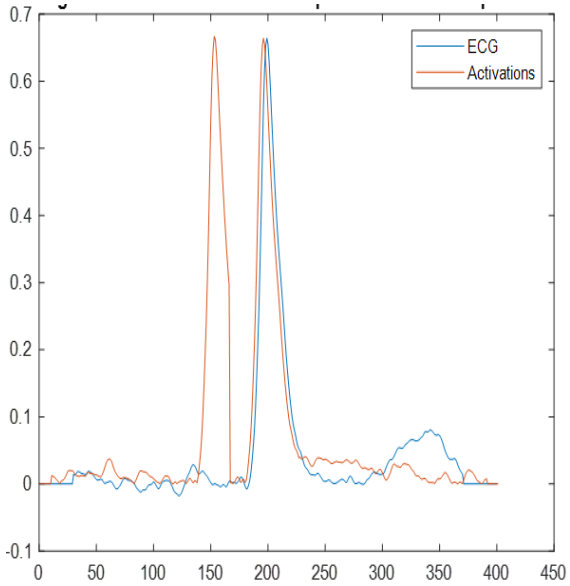


Figure 10: AF based ECG sample vs CNN-activated sample

In Figure 10, the ECG signal with atrial fibrillation (AF) is shown. The duration of the heartbeat is shortened in the case of AF. Zero-padding is applied to standardize signal length for model input. The process is performed during the signal pre-processing. The DCNN helps identify ECG points with AF.

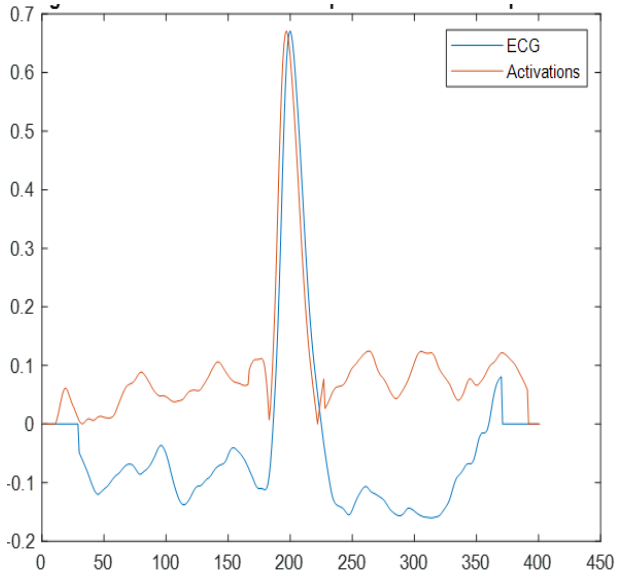


Figure 11: I-AVB-based ECG Sample vs CNN-activated sample

In Figure 11, the first-order degree atrioventricular block (I-AVB) disease activation level of the ECG sample is shown with respect to the plain ECG sample. In this case, the P-wave and T-wave are diminished or altered. In Figure 11, the CNN correctly identifies diminished P-wave activations indicative of I-AVB.

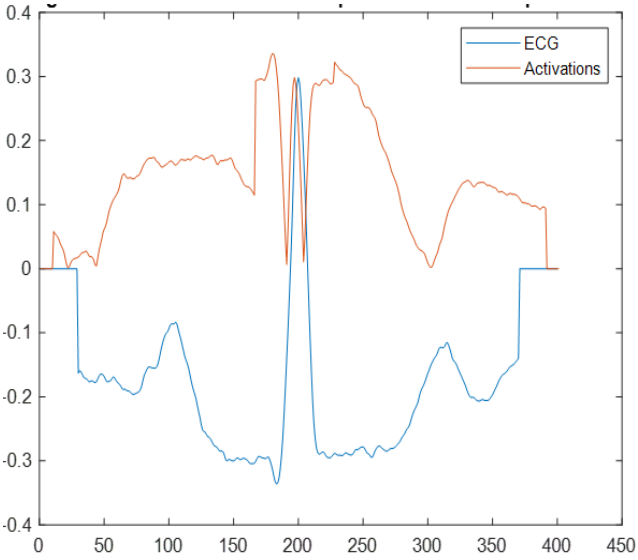


Figure 12: LBBB based ECG Sample vs CNN activated sample

A broad QRS complex form characterizes the traits of left bundle branch block (LBBB). This kind of signal shows a depressed ST segment and a T-wave appears after the QRS complex, following standard physiological patterns. In Figure 12, it is demonstrated that the network identifies the QRS complex accurately, as the activation region of the proposed model shows a high value in this region.

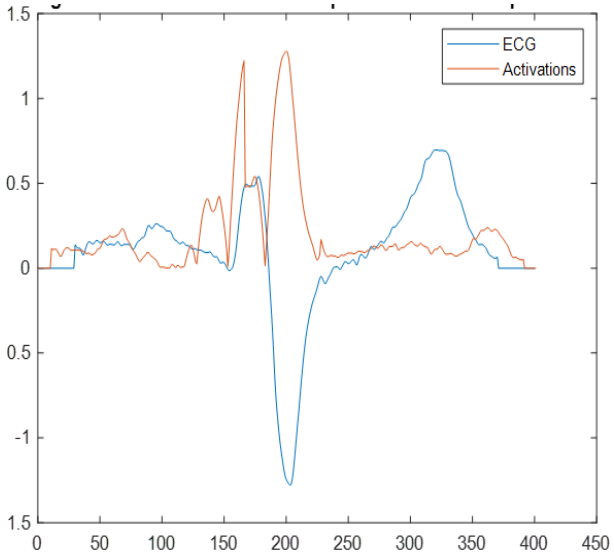


Figure 13: PAC-based ECG Sample vs CNN-activated sample

As shown in Figure 13, the ECG signal with a QRS complex and a general T wave, but showing an abnormal P wave, represents a Premature Atrial Contraction (PAC). Figure 13 illustrates the ECG signal and its corresponding activation map.

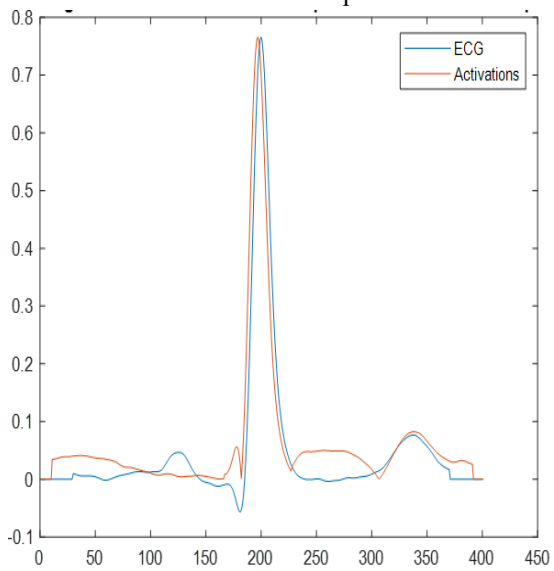


Figure 14: Normal-based ECG Sample vs CNN-activated sample

Let us consider a normal heart condition in which all the peaks, such as P, QRS, and T wave, look normal. The Network activation shown in Figure 14 provides appropriate regions for normal P, QRS, and T points.

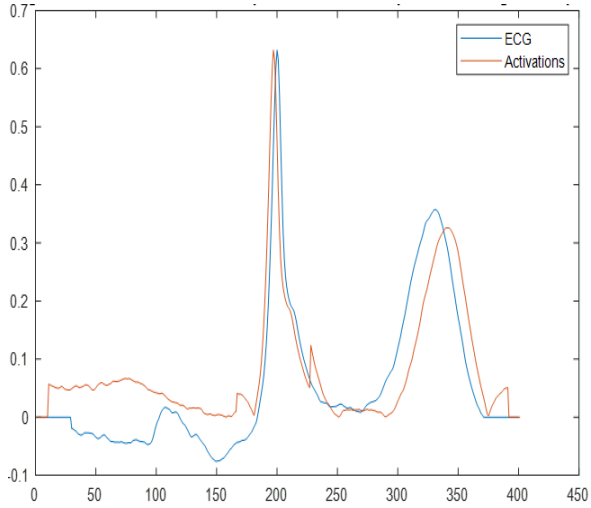


Figure 15: ST-Segment depression-based ECG Sample vs CNN activated sample

In the case of depression-based ST segment, the ST segments get depressed, and P & T waves will be normal. This is also illustrated in Figure 15 for the CNN activation of ST depression.

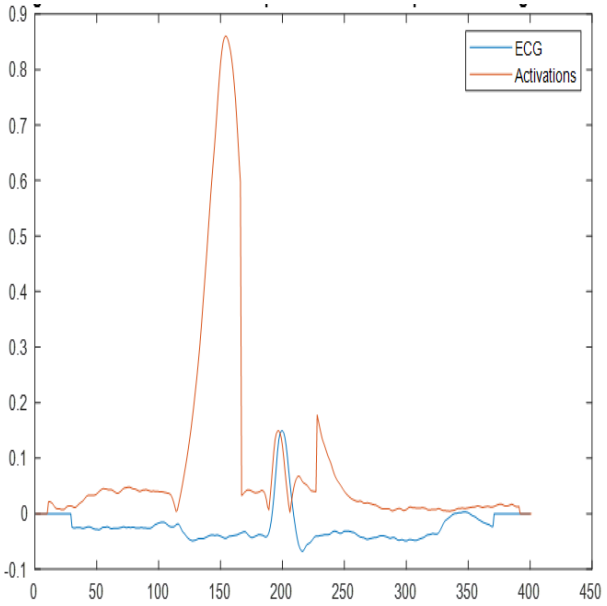


Figure 16: ST-Segment elevation-based ECG Sample vs CNN activated sample

A normal QRS complex and T-wave accompany the elevation of the ST segment signal. The activation of ECG signals using a CNN is illustrated in Figure 16, which

clearly shows that the signal is activated in the QRS complex, T wave, and elevated near the ST Segment. This indicates that the CNN classifies correct features in the case of ST Segment elevation.

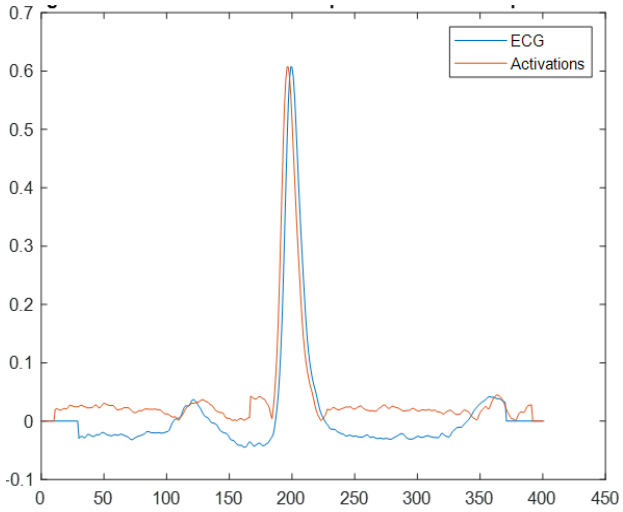


Figure 17: RBBB-based ECG Sample vs CNN-activated sample

The (RBBB) has a signal with a wide-spread S wave. This considers a specific lead like I, V5 and V6 features. Figure 17 shows the activation of the CNN in the case of RBBB, taken from lead V5.

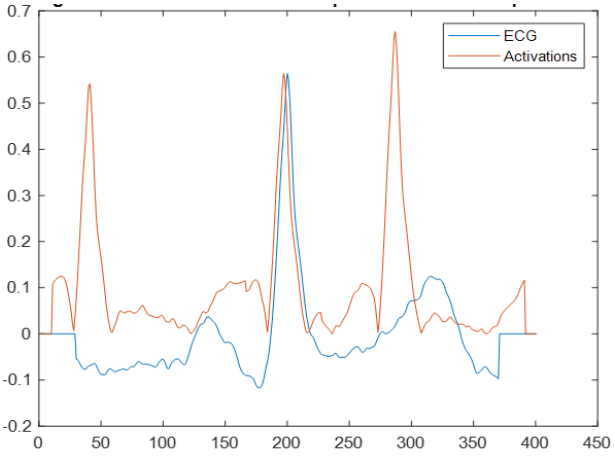
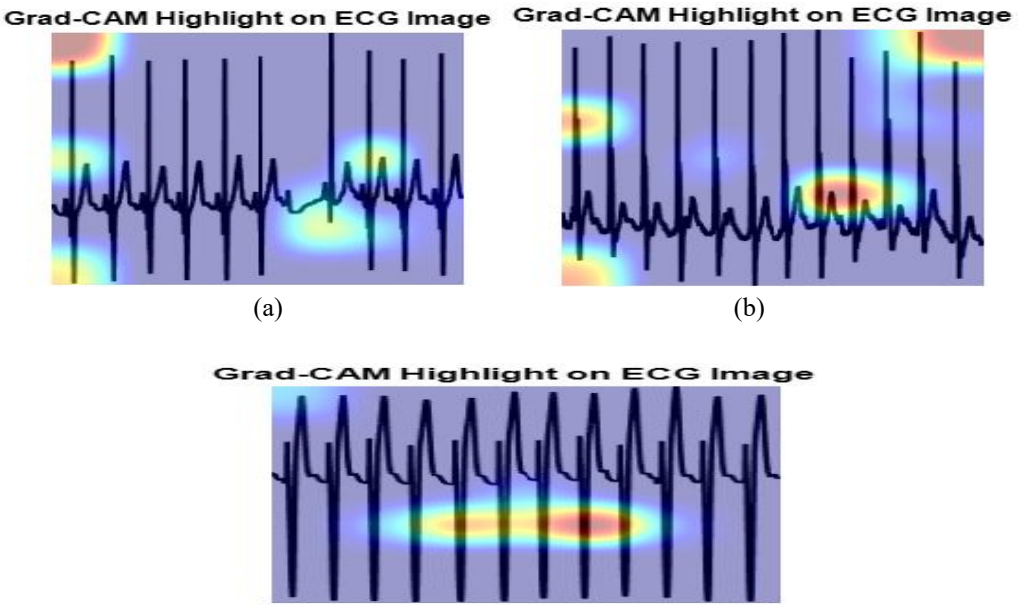


Figure 18: PVC-based ECG Sample vs CNN-activated sample

In the case of Premature Ventricular Contraction (PVC), the ECG signal exhibits a broad QRS complex. Figure 18 shows the performance of CNN activation for PVC, where the signal gets activated in the region of the broad QRS. Grad-CAM highlights the time regions in an ECG signal that are most influential for a model's decision. When applied to a CNN trained for ECG classification, these highlights often align with clinically significant features, such as the P-wave, QRS complex, and T-wave. The example of a Grad-CAM result is shown below.



(a) (b)

(c)

Figure 19: Grad-CAM Heatmap Output for (a) AF, (b) I-AV4B, (c) LBBB

4.3 Metrics evaluated

Four primary metrics evaluates the effectiveness of the suggested project. Accuracy metrics are the most important. False negative and False positive, True positive and True negative values provide the foundation for the accuracy finding. The numbers in this case indicate how well the CVD is anticipated. To obtain accurate data, the categorization must be flawless.

***Accuracy**:*

In this work, a test's accuracy is measured by how well it can distinguish between patient instances that are healthy and those that are sick. We should compute the percentage of true negative and true positive in each analysed instance in order to assess the accuracy of a test. In terms of math, this is expressed in eq. (3).

$$Acc = \frac{TP_j+TN_j}{TP_j+TN_j+FP_j+FN_j} \quad (3)$$

The precision and recall are evaluated using eq. (4) and eq. (5).

$$Precision = \frac{TP_j}{TP_j+FP_j} \quad (4)$$

$$Recall = \frac{TP_j}{TP_j+FN_j} \quad (5)$$

where TP_j, TN_j, FP_j, and FN_j of the amount of False negative and False positive, True positive and True negative samples, respectively. Of all nine classes, class j may be SNR, LBBB, RBBB, PAC, PVC, STD, AF, IVAB, and STE.

F1 score:

F1-score, is *a measure of a model's accuracy on a dataset*. It is used to evaluate the classification systems. The F1 score is determined by taking the consonant of recall and precision. The F1 score is evaluated using eq. (6).

$$F1 - Score = 2 * \frac{precision*recall}{precision+recall} \quad (6)$$

Hamming Loss (HL):

This parameter evaluates the portion of wrong classification w.r.t total available labels for individual test point. Finally, the HL is computed by considering the average of every test point and is given as [41]:

$$HL = \frac{1}{|X|}\sum_{x\epsilon X}\frac{1}{\tau}\sum_{j=1}^{l}\left[\left(L_j \epsilon HL(x)\right) \otimes \left(L_j \in y\right)\right] \quad (7)$$

In the case of a logical X-OR operation $\otimes$ the output got by the used model is HL(x). The j^{th}label is seen by L_j, and the test point's original classes appear in y.

The accuracy achieved using existing models is shown in table 2 and F1 score achieved using existing model is as shown in Table 3. The F1 score and accuracy are compared with the proposed model and achieves good results when compare to existing methods.

Table 2: Comparison of accuracy results

Author	Methodology	No of Leads	Database	Accuracy (%)
A.M Ladhi et al., [30]	CNN	12	N-31722	93.25
D. Sadhukhan et al., [31]	Logistic regression	03	N-5000	95.6
H. W. Lui [32]	CNN-LTSM	12	PTB and ATF challenge	94.6
Y.J Chen et al., [33]	ResNEt-LTSM	12	7704 samples	81
Hadiyoso et al., [34]	VGG-16	12	ECG signals	95
Lofti Mhamidi et al., [35]	MobileNet V2	12	ECG signals	94
P.G Giriprasad et al., [36]	AlexaNet	12	4 categories of signals	95.6
Proposed Model	Enhanced DenseNet-121 with Grad-CAM	12	CPSC 2018	97.45

Table 3: Comparison of F1 score results

Author	Methodology	Database	F1-Score (%)
Y. Li et al., [37]	117 features which are handcrafted	CPSC	60.8
Y. Cheng et al., [38]	Residual CNN- Single lead	MITDB	96.4
D. Jia et al., [39]	Multi-task CNN module	CPSC	87.2
J. Yoo et al., [40]	Residual CNN	PTBXL	61.2
M. Ganesh Kumar et al., [41]	DCNN; 12 lead	CPSC	96.7
Y. Jin et al., [42]	CNN+Bi-LSTM	MITDB, CPSC	80.56, 60.1
Z. Sun et al., [43]	Ensemble of Multiple Multi-label classifiers	CPSC	75.2
J. Cai et al., [44]	Multi ECG-Net	CPSC	86.3
H. Zhu et al., [45]	CNN- Sigmoid probability	CPSC	88.7
Proposed Model	Enhanced DenseNet-121 with Grad-CAM	CPSC 2018	97.75

The confusion matrix, as is shown in figure 20 and 21, is used to validate the results. For a look at of the outputs of the model and a summary of the classification's results problem, a confusion matrix is produced. It organizes, by class, the successful and unsuccessful predictions.

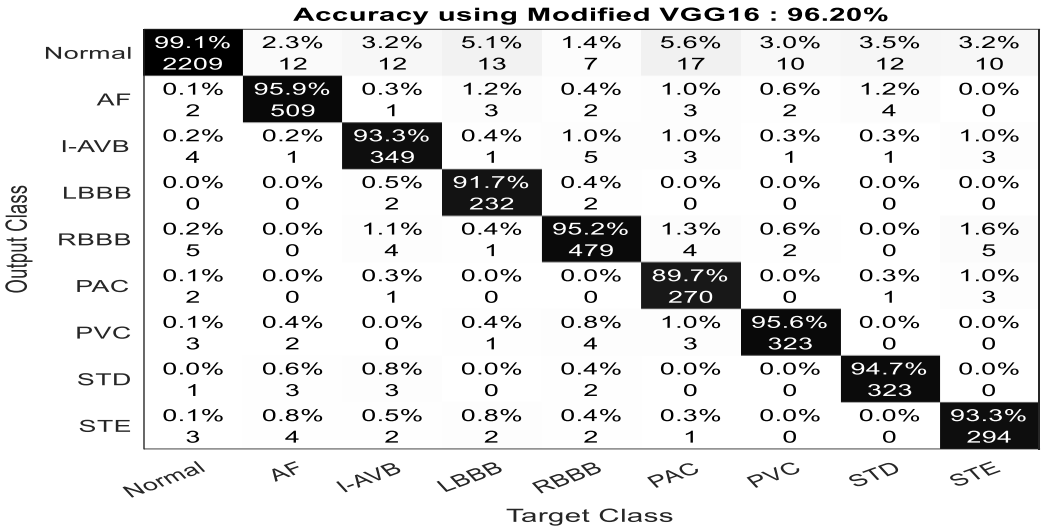


Figure 20: Confusion matrix achieved using VGG-16

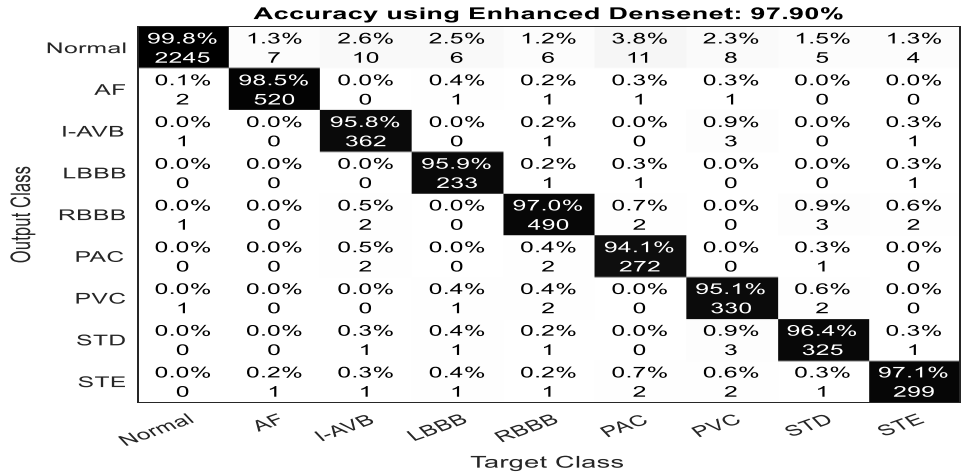


Figure 21: Confusion matrix achieved using Enhanced DenseNet-121

The ROC curve for a subset of classes AF, LBBB, PVC are evaluated for interpretability.

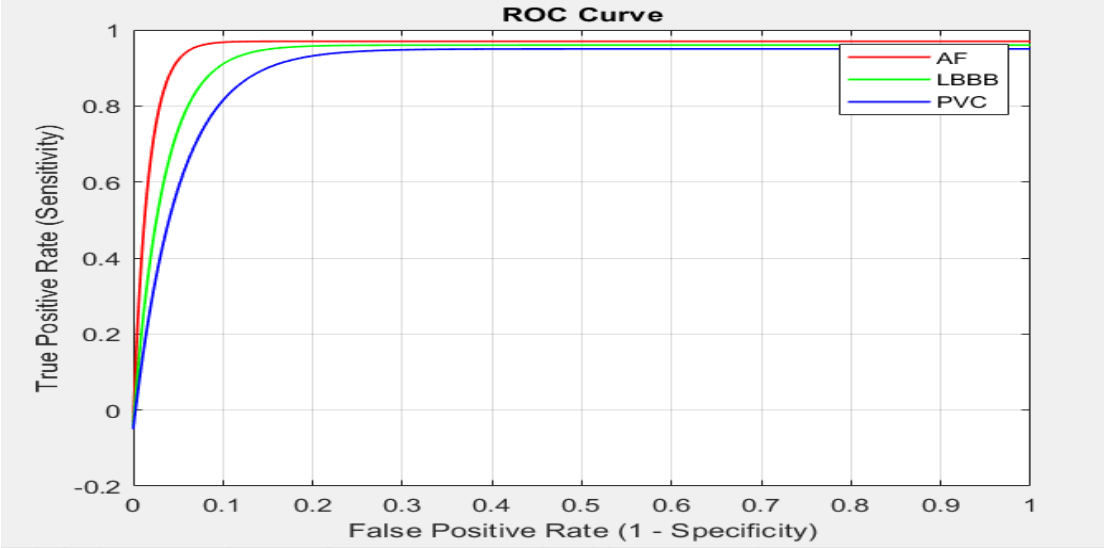


Figure 22: ROC curve

All the parameters evaluated using the proposed Enhance DenseNet-121 and VGG-16 is shown in table 4.

Table 4: Results obtained using proposed model

Parameter/Method	VGG-16	Enhanced DenseNet-121
Accuracy (%)	96.20	97.90
Recall (%)	96.52	97.34
Precision (%)	96.94	98.33
F1 Score (%)	96.49	97.75
Hamming Loss	0.27	0.18

The limitation of study refers as; our suggested technique can only detect eight serious cardiac conditions that fall into the categories of rhythm abnormalities and morphological abnormalities. Our model has the potential to be enhanced in the future to detect a broader range of cardiac disorders by training it using ECG recordings from other datasets. This makes it an all-inclusive automated approach for diagnosing heart disease.

Table 5: Per-Class Precision, Recall, and F1-Score for the Proposed Enhanced DenseNet-121 Model

Class	**Precision (%)**	**Recall (%)**	**F1-score (%)**
Normal (N)	98.0	97.4	97.7
Atrial Fibrillation (AF)	97.6	97.1	97.3
First-Degree AV Block (I-AVB)	96.0	95.4	95.7
Left Bundle Branch Block (LBBB)	97.5	96.9	97.2
Right Bundle Branch Block (RBBB)	97.8	97.1	97.4
Premature Atrial Contraction (PAC)	95.5	94.9	95.2
Premature Ventricular Contraction (PVC)	97.2	96.5	96.8
ST-T Change (STTC)	97.7	97.0	97.4
Macro Average	**97.41**	**97.08**	**97.75**

Table 5: Per-class precision, recall, and F1-score for each of the eight categories of cardiovascular disease It can be seen from the results that majority classes such as Normal, AF, and STTC have high values regardless of time, while minority classes such as PAC and I-AVB have relatively lower values. The macro averages are in close agreement with the total accuracy of 97.45% and the total F1-score of 97.75% as reported above and hence confirming that the proposed Enhanced DenseNet-121 model is balanced and robust across all conditions.

Table 6: Ablation Study Showing the Impact of CBAM and SMOTE on the Performance of the Proposed Model

Model Variant	**Accuracy (%)**	**F1-score (%)**	**Recall (%)**
DenseNet-121 (Base)	96.15	96.32	96.10
DenseNet-121 + CBAM	96.90	97.05	96.78
DenseNet-121 + CBAM + SMOTE (Proposed EDN-121)	**97.45**	**97.75**	**97.34**

Table 6 ablation study in contributions of CBAM and SMOTE to Enhanced DenseNet-121 architecture The base DenseNet-121 model has a comparatively strong performance but incorporating CBAM provides a noticeable improvement due to its effective feature refinement capabilities. But by further adding the SMOTE which finally increases the performance and reinforces the enhancement of classification process with respect to class imbalance by increasing the accuracy, recall and F1-score concisely.

5 Discussion

The framework of the proposed Enhanced DenseNet-121 (EDN-121) achieves better performance than the baseline VGG-16 model and the most state-of-the-art approaches listed in Table 1. The attributes sensitivity of 97.90% and F1-score of 97.75% are notable upgrades from previous deep learning models for the same task, with reported F1-scores in the 81–87% range (1D-ResNet34 [15]; SE-ResNet [16]; Multi-ECGNet [44]).

This boosts performance due to a combination of key factors. (i) The concatenation of the Convolutional Block Attention Module (CBAM) allows the CNN to focus on the most salient spatial and channel features of the ecg signal. This attention mechanism improves the discrimination of fine anatomical changes that are important for differentiating similar hereditary cardiovascular disorders. On the contrary, VGG-16 and other standard models based on CNN did not have a feature specificity refinement process at this time between CNNs, thus lowering accuracy.

Secondly, in data preparation, we implement Synthetic Minority Oversampling Technique (SMOTE) to solve the class imbalance issue painful for clinical ECG datasets which ensures diversity and reliability to the model. Published previous work suffers from publication bias wherein highly skewed data distributions lead to less sensitivity for under populated classes such as PAC and I-AVB. SMOTE creates synthetic samples based on the similarities of the target variable, where the one hot encoding that you initially generate could lead to unbalanced representations thus, leading the model to a lack of generalizability with particular classes, this generates a concern for per-class recall.

Finally, we supplement our model with Grad-CAM, a visual explanation method to give additional model interpretability and enable clinicians to more selectively verify model decisions by looking at the most influential ECG segments that played a major role in prediction. In contrast to previous works that used “black-box” classifiers, which are constrained in a clinical setting. Clinically relevant sections such as the P-wave, QRS complex, and ST-segment are consistent with interpretations given by a cardiologist, which Grad-CAM visual appeal highlights in Figure 19.

Our framework offers a unique combination of high accuracy, stability with respect to imbalanced multi-label data, and interpretability, setting it apart from other methods. These features make the EDN-121 framework especially well placed for use in real-world ECG-based screening and diagnosis systems, especially for settings where transparency and trust are critical. These are strengths of the study, however it is worth acknowledging that the study is solely limited to the CPSC 2018 dataset and only eight diagnostic classes. Further research will validate cross-dataset and incorporate additional cardiovascular conditions to increase the system clinical applicability.

While Grad-CAM successfully identifies parts of the ECG signal that contribute to model decision making, it is limited in a few ways. These heat maps are of lower resolution, making it difficult to accurately localize important waveform features for diagnosis. In addition, Grad-CAM essentially reflects the spatial feature relevance rather than considering the temporal dependencies of sequential ECG signals. In future work, we will investigate temporal attention mechanisms that could mitigate this effect, as well as complementary explainability methods (SHAP and LIME) for more expression multi-faceted understanding of model predictions, as well as increased clinical interpretability.

5.1 Limitations of the study

Although the proposed Enhanced DenseNet-121 framework shows promising performance, there are some limitations that need to be acknowledged in order to contextualize its clinical and research applications. Unsupervised training − The model is trained only on the

eight cardiovascular conditions from the CPSC 2018 data set. While these classes reflect the broadly defined rhythm and morphologic abnormalities, numerous other clinically relevant conditions (eg, rare arrhythmias, ischemic event, congenital heart disease) are not incorporated. Consequently, this system cannot be used directly for clinical diagnostics without further data and retraining in a large variety of disease classes.

Second, the majority of patients in the real-world setting display several coexisting conditions (i.e., comorbidities). This can cause overlapping and more difficulty classification in ECG patterns [137]. The current investigation focuses on singleton manifestations of the eight targeted abnormalities and is not designed to provide explicit modeling of these more intricate comorbid presentations. This may limit its use in the most diverse patient populations, where mixed pathologies frequently occur.

Third, the Synthetic Minority Oversampling Technique (SMOTE) was used to deal with the class imbalance problem, which has some limitations. While synthetic examples may not accommodate complicated physiological differences of minority classes, they may have high potential for overfitting and capturing rare pattern metrics wrong. Some performance improvements were seen for underrepresented conditions (for example, premature atrial contractions [PAC] and first-degree atrioventricular block [I-AVB]), and other strategies (eg, cost-sensitive learning, or curated data collection beyond our subset) could further increase model robustness.

Fourth, noting that this study directly compares EDN-121 with VGG-16 over the same datasets under exactly the same experimental conditions, we also recognize the lack of a direct re-implementation over more recent state of the art methods inc1`luding 1D-ResNet34, SE-ResNet, and Multi-ECGNet. Future work will explore this limitation to perform a more complete comparison of the performance

Finally, our study is based off of one dataset, limiting its generalizability. Without external validation on datasets, such as PTB-XL or MIT-BIH, it is challenging to comprehend how the model is likely to behave among heterogeneous populations, routine clinical settings, and different equipment. Next steps will emphasize cross-dataset validation, incorporation of less common pathologies, and integration with electronic health record data in order to increase usability and clinical accessibility.

6 Conclusion

Our proposal in this study was to investigate the Deep Learning area of Transfer Learning utilising pre-trained Convolutional Neural Networks to increase the accuracy of the classification of CVDs based on the 12-lead ECG signal dataset. After performing EDN-121, the model established explainable deep learning (EDL) for ECG classification task. The model designed outperformed compared with other existing models which are available of multilabel ECG signals. This process includes detection of depression and elevation of ST segment. The EDL checks the performance of EDN-121 in obtaining right features for the purpose of classification. The rate of accuracy obtained using the designed model is 97.90% and F1-Score is 97.75%. The results show the efficiency and effective of the model designed on CPSC 2018 dataset. To increase the accuracy rate meta heuristic algorithms can be initiated in future. Further the proposed technique can be applied to other signals like EMG, EEG to find related diseases with respect to the signals.

References

[1] World Health Organization, Prevention of Cardiovascular Disease. Pocket Guidelines for Assessment and Management of Cardiovascular Risk. Africa Who/Ish Cardiovascular Risk Prediction Charts for the African Region. Geneva, Switzerland: World Health Organization, 2007.

[2] H. Wang *et al.*, “Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: A systematic analysis for the Global Burden of Disease Study 2015,” *Lancet*, vol. 388, no. 10053, pp. 1459–1544, 2016.

[3] M. Naghavi, H. Wang, R. Lozano, A. Davis, X. Liang, and M. Zhou, “GBD 2013 mortality and causes of death collaborators. Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013,” *Lancet*, vol. 385, no. 9963, pp. 117–171, 2015.

[4] G. B. Moody and R. G. Mark, “The impact of the mit-bih arrhythmia database,” IEEE Engineering in Medicine and Biology Magazine, vol. 20, no. 3, pp. 45–50, 2001.

[5] G. D. Clifford, C. Liu, B. Moody, H. L. Li-wei, I. Silva, Q. Li, A. Johnson, and R. G. Mark, “Af classification from a short single lead ecg recording: the physionet/computing in cardiology challenge 2017,” in 2017 Computing in Cardiology (CinC). IEEE, 2017, pp. 1–4.

[6] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu, Y. Liu, C. Ma, S. Wei, Z. He et al., “An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection,” Journal of Medical Imaging and Health Informatics, vol. 8, no. 7, pp. 1368–1373, 2018.

[7] P. Wagner, N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F. I. Lunze, W. Samek, and T. Schaeffter, “Ptb-xl, a large publicly available

electrocardiography dataset," Scientific Data, vol. 7, no. 1, pp. 1–15, 2020.

[8] Karrar AL-Jammali, "Prediction of Heart Diseases Using Data Mining Algorithms," Informatica 47 (2023), 57–62.

[9] Jeena Joseph, K Kartheeban, "Visualizing the Full Spectrum Optimization of K-Nearest Neighbors from Data Preprocessing to Hyperparameter Tuning and K-Fold Validation for Cardiovascular Disease Prediction," Informatica **49** (2025) 355–374.

[10] Liu, X.; Wang, H.; Li, Z.; Qin, L. Deep learning in ECG diagnosis: A review. Knowl. -Based Syst. 2021, 227, 107187.

[11] Egger, J.; Gsaxner, C.; Pepe, A.; Pomykala, K.L.; Jonske, F.; Kurz, M.; Li, J.; Kleesiek, J. Medical deep learning—A systematic meta-review. Comput. Methods Programs Biomed. 2022, 221, 106874.

[12] Siontis, K.C.; Noseworthy, P.A.; Attia, Z.I.; Friedman, P.A. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nat. Rev. Cardiol. 2021, 18, 465–478.

[13] Sulaiman Somani, Adam J Russak, Felix Richter, Shan Zhao, Akhil Vaid, Fayzan Chaudhry, Jessica K De Freitas, Nidhi Naik, Riccardo Miotto, Girish N Nadkarni, Jagat Narula, Edgar Argulian, and Benjamin S Glicksberg. Deep learning and the electrocardiogram: review of the current state-of-the-art. *EP Europace*, 23:1179–1191, 2021.

[14] U Rajendra Acharya, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, Muhammad Adam, Arkadiusz Gertych, and Ru San Tan. A deep convolutional neural network model to classify heartbeats. *Computers in Biology and Medicine*, 89, 08 2017.

[15] Dongdong Zhang, Samuel Yang, Xiaohui Yuan, and Ping Zhang. Interpretable deep learning for automatic diagnosis of 12-lead electrocardiogram. *iScience*, 24(4):102373, 2021.

[16] Zhaowei Zhu, Xiang Lan, Tingting Zhao, Yangming Guo, Pipin Kojodjojo, Zhuoyang Xu, Zhuo Liu, Siqi Liu, Han Wang, Xingzhi Sun, and Mengling Feng. Identification of 27 abnormalities from multi-lead ecg signals: an ensembled se-resnet framework with sign loss function. *Physiological Measurement*, 42, 06 2021.

[17] Qihang Yao, Xiaomao Fan, Yunpeng Cai, Ruxin Wang, Liyan Yin, and Ye Li. Time-incremental convolutional neural network for arrhythmia detection in varied-length electrocardiogram. In *2018 IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, pages 754–761, 2018.

[18] Mehmet Ozdemir, Gizem Ozdemir, and Onan Guren. Classification of covid-19 electrocardiograms by using hexaxial feature mapping and deep learning. *BMC Medical Informatics and Decision Making*, 21, 05 2021.

[19] Younghoon Cho, Joon-Myoung Kwon, Kyung-Hee Kim, Jose Medina- Inojosa, KI-Hyun Jeon, Soohyun Cho, Soo Lee, Jinsik Park, and Byung-Hee Oh. Artificial intelligence algorithm for detecting myocardial infarction using six-lead electrocardiography. *Scientific Reports*, 10, 11 2020.

[20] J. Hughes, Jeffrey Olgin, Robert Avram, Sean Abreau, Taylor Sittler, Kaahan Radia, Henry Hsia, Tomos Walters, Byron Lee, Joseph Gonzalez, and Geoffrey Tison. Performance of a convolutional neural network and explainability technique for 12-lead electrocardiogram interpretation. *JAMA Cardiology*, 6, 08 2021.

[21] Atul Anand, Tushar Kadian, Manu Shetty, and Anubha Gupta. Explainable ai decision model for ecg data of cardiac disorders. *Biomedical Signal Processing and Control*, 75:103584, 05 2022.

[22] Maksims Ivanovs, Roberts Kadik, is, and Kaspars Ozols. Perturbation based methods for explaining deep neural networks: A survey. *Pattern Recognition Letters*, 150:228–234, 10 2021.

[23] Ali Raza, Kim Phuc TRAN, Ludovic Koehl, and Shujun Li. Designing ECG monitoring healthcare system with federated transfer learning and explainable ai. *Knowledge-Based Systems Journal*, 236:107763, 11 2021.

[24] Ganeshkumar M., Vinayakumar Ravi, Sowmya V, Gopalakrishnan E.A, and Soman K.P. Explainable deep learning-based approach for multilabel classification of electrocardiogram. *IEEE Transactions on Engineering Management*, pages 1–13, 2021.

[25] F. Liu, C. Liu, L. Zhao, X. Zhang, X. Wu, X. Xu, Y. Liu, C. Ma, S. Wei, Z. He et al., "An open access database for evaluating the algorithms of electrocardiogram rhythm and morphology abnormality detection," Journal of Medical Imaging and Health Informatics, vol. 8, no. 7, pp.1368–1373, 2018.

[26] G. Begg et al., "Electrocardiogram interpretation and arrhythmia management: a primary and secondary care survey," British Journal of General Practice, vol. 66, no. 646, pp. 291-6, May 2016.

[27] V. G. Sujadevi and K. P. Soman, "Towards identifying most important leads for ECG classification. A data driven approach employing

deep learning," *Procedia Comput. Sci.*, vol. 171, pp. 602–608, 2020.

[28] Lodhi AM, Qureshi AN, Sharif U, Ashiq Z. A novel approach using voting from ECG leads to detect myocardial infarction. Adv Intell Syst Comput. 2018; 869:337–52.

[29] Sadhukhan D, Pal S, Mitra M. Automated identification of myocardial infarction using harmonic phase distribution pattern of ECG Data. IEEE Trans Instrum Meas. 2018;67(10):2303–13.

[30] Lui HW, Chow KL. Multi-class classification of myocardial infarction with convolutional and recurrent neural networks for portable ECG devices. Inform Med Unlocked. 2018;13:26–33.

[31] Chen YJ, Liu CL, Tseng VS, Hu YF, Chen SA. Large-scale classification of 12-lead ECG with deep learning. USA: IEEE EMBS International Conference on Biomedical & Health Informatics; 2019. p. 1–4.

[32] Hadiyoso, Sugondo; Fahrozi, Farell; Yuli Sun Hariyani; Sulistiyo, Mahmud Dwi, "Image Based ECG Signal Classification Using Convolutional Neural Network.," *International Journal of Online & Biomedical Engineering,* vol. 16, no. 4, pp. 64-78, 2022.

[33] Lotfi Mhamdi, Oussama Dammak, François Cottin Imed Ben Dhaou, "Artificial Intelligence for Cardiac Diseases Diagnosis and Prediction Using ECG Images on Embedded Systems," *MDPI,* vol. 10, no. 8, p. 12, 2022.

[34] P. Giriprasad Gaddam, A Sanjeeva reddy and R.V. Sreehari, "Automatic Classification of Cardiac Arrhythmias based on ECG Signals Using Transferred Deep Learning Convolution Neural Network," *Journal of Physics: Conference Series,* vol. 2089, p. 8, 2021.

[35] Y. Li et al., "multi-label classification of arrhythmia for long-term electrocardiogram signals with feature learning," IEEE Trans. Instrum. Meas., vol. 70, pp. 1-11, Jun 2021.

[36] Y. Cheng et al., "multi-label arrhythmia classification from fixed-length compressed ECG segments in real-time wearable ECG monitoring," 42nd Int. Conf. IEEE Eng. Med. Biol. Soc., pp. 580-583, Jul 2020.

[37] D. Jia et al., "An ensemble neural network for multi-label classification of electrocardiogram," In Proc. Int. Workshop Mach. Learn. Med. Eng. Cardiovasc. Healthcare, Springer, pp. 20-27, Oct 2019.

[38] J. Yoo et al., "k-Labelsets Method for Multi-Label ECG Signal Classification Based on SE-ResNet," App. Sci., vol. 11, no. 16, p7758, 2021.

[39] M. Ganeshkumar et al., "Explainable Deep Learning-Based Approach for Multilabel Classification of Electrocardiogram," IEEE Trans. Instrum. Meas., Sep 2021.

[40] Y. Jin et al., "A novel interpretable method based on dual level attentional deep neural network for actual Multi label Arrhythmia detection," IEEE Trans. Instrum. Meas., Dec 2021.

[41] G. M., V. Ravi, S. V, G. E.A and S. K.P, "Explainable Deep Learning-Based Approach for Multilabel Classification of Electrocardiogram," in *IEEE Transactions on Engineering Management*, vol. 70, no. 8, pp. 2787-2799, Aug. 2023.

[42] J. Cai, W. Sun, J. Guan and I. You, "Multi-ECGNet for ECG Arrythmia Multi-Label Classification," in *IEEE Access*, vol. 8, pp. 110848-110858, 2020.

[43] Zhu, J.; Lv, J.; Kong, D. CNN-FWS: A Model for the Diagnosis of Normal and Abnormal ECG with Feature Adaptive. Entropy, 2022, 24, 471.

[44] Agostinelli A, Marcantoni I, Moretti E, Sbrollini A, Fioretti S, Di Nardo F, Burattini L. Noninvasive Fetal Electrocardiography Part I: Pan-Tompkins' Algorithm Adaptation to Fetal R-peak Identification. Open Biomed Eng J. 2017 Mar 31; 11:17-24. doi: 10.2174/1874120701711010017. PMID: 28567128; PMCID: PMC5418929.

[45] Alex Kummer, Tamás Ruppert, Tibor Medvegy, János Abonyi, Machine learning-based software sensors for machine state monitoring - The role of SMOTE-based data augmentation, Results in Engineering, Volume 16, 2022, 100778, https://doi.org/10.1016/j.rineng.2022.100778.