A Deep Convolutional Model for Heart Disease Prediction based on ECG Data with Explainable AI

SREEJA M. U., SUPRIYA M. H.
Department of Electronics,
Cochin University of Science and Technology,
Kochi-22,
INDIA

Abstract: - Heart disease (HD) prediction is crucial in realizing the notion of intelligent healthcare owing to the exploding number of heart diseases being reported on a daily basis. However, in a domain like healthcare, accountability is key for a medical practitioner to completely adopt the decisions of an intelligent model. Accordingly, the proposed model develops a convolutional model for heart disease prediction based on ECG data in a supervised manner. Moreover, the easily accessible and economical ECG data is utilized in the model in the form of image data. The incorporation of ECG data as images has provided amazing results in the recent researches compared to being considered as signals. The architecture follows a stacked Convolutional Neural Network for extracting features from ECG images followed by fully connected network for classification. The evaluation of the proposed model on customized public datasets demonstrates its ability to achieve impressive outcomes by leveraging the characteristics of convolutional neural networks (CNNs) and supervised learning. Similarly, Explainability in the form of interpretability has been incorporated into the framework thus ensuring accountability of the model which is crucial in medical domain. Detailed experiments for identification of ideal model architecture are conducted. Further, local and vision based Explainability has been explored in detail using LIME and Grad-CAM. The model could achieve a precision, recall and f1-score of 0.982, 0.982, and 0.981 respectively proving the superiority of the model. Moreover, Explainability visualization based on popular algorithms for true positive and false positive results have shown promising results on the PhysioNet ECG dataset.

Key-Words: - Explainability, ECG, Heart disease, Convolutional Neural network, Grad-CAM, LIME

Received: July 14, 2022. Revised: June 13, 2023. Accepted: July 25, 2023. Published: September 8, 2023.

1 Introduction

A heart disease prediction model can be a valuable tool in healthcare for several reasons, the prime one being early detection. Heart diseases, such as coronary artery disease or heart failure, can often be asymptomatic or exhibit subtle symptoms in the early stages. A prediction model can analyze various risk factors from any kind of data pinpoint individuals who have higher chances of ending with heart disease. Early detection allows for timely intervention and preventive measures to reduce the risk or slacken the disease progression. Another advantage is that healthcare resources can be allocated more effectively. High-risk patients can receive closer monitoring, targeted interventions, and appropriate follow-ups, while low-risk individuals can be reassured and receive general preventive advice. This optimization helps in managing limited healthcare resources efficiently. Heart disease prediction models can assist healthcare providers in tailoring treatment plans for a personalized treatment. By analyzing patient-specific risk factors, the model can provide insights into the potential drivers of the disease and suggest appropriate interventions. This personalized approach enhances patient outcomes and improves the overall quality of care.

Explainability refers to the capability to comprehend, interpret and infer the predictions of a predictive model. In healthcare, it is crucial to build trust between healthcare providers and patients. When using a heart disease prediction model, Explainability helps patients and providers understand the factors considered by the model and the reasons behind its predictions. This transparency fosters trust, increasing patient acceptance and engagement with the model's recommendations. Explainability enables healthcare providers to make informed clinical decisions leading to development of clinical decision support systems. By understanding the model's reasoning, clinicians can evaluate the validity and reliability of its

predictions. This information helps them consider additional clinical factors, interpret the results in the context of the patient's overall health, and make more accurate diagnoses and treatment decisions. It also helps identify potential biases or errors in the model's predictions, promoting fairness and equity in healthcare delivery. Explainable models provide an opportunity to educate patients about their risk factors, lifestyle choices, and the impact on their heart health. Patients who understand the reasons behind their risk predictions are more likely to actively participate in preventive measures, make necessary lifestyle changes, and adhere to treatment plans.

By combining predictive accuracy with Explainability, healthcare providers can enhance heart care and improve patient outcomes. The proposed model trains a deep convolutional model that aims at categorizing input ECG images into healthy and other unhealthy classes. The model further analyses the significance of Explainability with two popular algorithms Local Interpretable Model-agnostic Explanations (LIME) and Grad-CAM which helps to analyze a deep learning model and to make its predictions comprehensible. Grad-CAM have extreme application in image classification tasks which enables the model to interpret black box model by visualizing the class activation maps corresponding to various layers.
The objective of the proposed work is to design and train an explainable convolutional model for heart disease prediction by utilizing the most accessible and simplest form of heart data which is ECG. The work aims to identify the optimal architecture as well as the best explainable algorithm for the proposed problem. The main highlights of the paper are:

1. An explicit deep convolutional model for ECG based image prediction trained on customized datasets on most frequently occurring categories of heart disease in a supervised manner.
2. A comparative analysis of vision-based Explainability models on ECG data with Grad-CAM and LIME.
3. Detailed comparative analysis with baseline models based on popular evaluation metrics.

Section 2 outlines the recent literature followed by the critical gaps identified. The detailed methodology is elaborated in section 4 followed by analysis of results in section 5 and conclusion in section 6.

2 Related Works

This section reviews the recent related researches that have formulated models for ECG prediction based on deep learning and those models that have explored Explainability as well.

2.1 Heart disease Prediction using Deep Learning Models

Nonlinear parameters have been explored by various machine learning algorithms from ECG data, [1], [2]. The same has been exploited by Deep neural networks where the significance of metaheuristic optimization algorithm has shown faster results over ECG data, [3], [4], [5]. Chaotic error scatter map using master and slave systems created in combination with deep neural networks with disease recognition, [6]. Convolutional neural networks have also shown success for retrieving temporal information and definitely can be applied to ECG information transfer learning. Successful networks like Alex net have shown amazing results the one-dimensional uses signal needs to be converted to image data in order to train the CNN, [7]. Extreme learning machines (ELM) in combination with recurrent neural networks have been used for ECG classification, [8], [9], [10]. Clustering and fuzzy learning works have been experimented were gain ratio is employed for feature selection which is included as weights to the ELM, [11], [12]. Temporal information is crucial in ECG data which can be efficiently extracted from heartbeat signals using ELMs where wavelength transforms can be applied for transforming the signal for retrieving the time frequency features, [13], [14].

2.2 Explainable Models for ECG based Heart Disease Prediction

Explainability models like LIME and SHAP has resulted in better insights specifically on ECG data, [15]. The excellent performance of Grad-CAM for Explainability on visual information, it is widely used in heart disease classification on ECG data, [16], [17], [18], along with transfer learning, [19], and federated learning, [20]. Attention based models formulated on one dimensional CNNs are successful models for ECG data as proposed in [21]. Due to their simplicity of use, Holter monitors have recently attracted a lot of interest. Holter ECG data is made up of ECG data collected over an extended period of time, making it unlikely that an occasional arrhythmia would be missed. Holter monitor data have also been successfully used to apply gradient

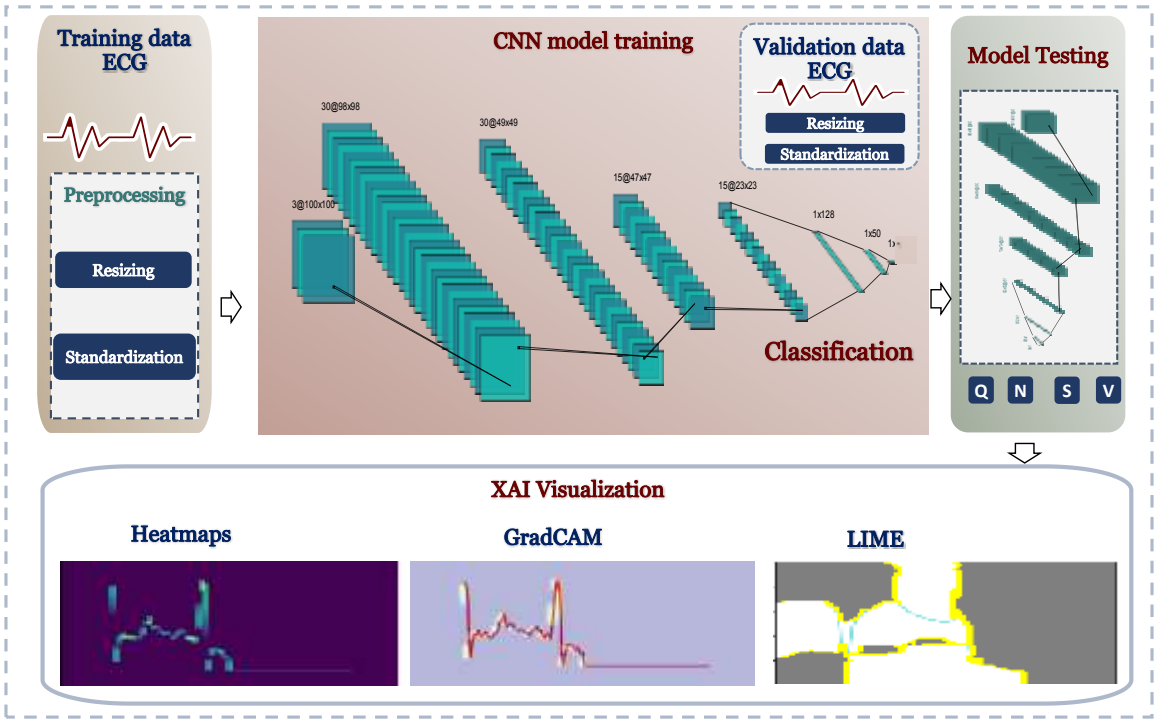


Fig. 1: Proposed Framework

activation mapping and Convolutional neural network, [22]. Recurrent neural networks have been investigated for the same, and based on ECG data, many variations have been tested for cardiovascular abnormalities, [23].

3 Motivation

Based on the review of recent works, the following points have been identified:

- Large number of heart abnormality detection models are proposed in the recent research works. However, supervised models have not been addressed much which are proven to produce better results.
- Models that address the most relevant categories of heart diseases in the form of ECG image signals are not addressed, ECG being the most economical form of input data for heart disease detection.
- Although, Explainability is crucial in a domain like heart care, local Explainability models like LIME and SHAP have been explored but, image specific models like Grad-CAM were not found especially for ECG signals.

This led to the formulation of a simple deep CNN model for the above-mentioned categories of events for heart disease prediction.

4 Proposed Framework

The major phases of the projected work are indicated in Figure 1. The model aims at classifying ECG signals into four major classes the details of which are detailed in the next section. Since ECGs can be visualized as image data, Convolutional neural networks have been adopted for the same. The major phases for any Explainability based deep learning models have been adopted for the proposed framework which are:

4.1 Data Collection and Pre-Processing

The data for training is collected from customized collection of Physio net's MIT-BIH Arrhythmia Dataset in the form of images, [24]. The images are initially resized in the dimension of 100*100 and normalized by dividing with 255.0, the highest value possible to scale the values within the limits 0 and 1. Four classes of images have been considered and trained. The four categories of ECG signals along with their properties considered are (i) N with properties as Normal, Left/Right bundle branch block, Atrial escape and Nodal escape, (ii) S with properties as Atrial premature, Aberrant atrial premature, Nodal premature, Supra-ventricular premature, (iii) V with Premature ventricular contraction and Ventricular escape, and finally (iv) Q with properties Paced and Fusion of paced and normal.

4.2 Model Design

Figure 2 overviews the convolutional neural network (CNN) structure used in the work. The CNN architecture consists of two sets of convolutional layers, with 30 units and 15 units respectively, each having a size of (3x3). These layers are then followed by a 'batch normalization' layer and a 'max pooling' layer. The feature vector obtained from CNN is flattened and three dense layers follows with dimension 128, 50 and 4 units each. The purpose of the model is to learn to classify incoming ECG signal into 4 classes, hence the count 4 in the final layer. The architecture has been depicted in Figure 2. The intermediate Convolution layers follows ReLU activations as in (4).

$$Relu(x_{ij}) = max(0, ECG_{ij}) \quad (1)$$

where, *Relu* is the Rectified Linear Unit function and ECG_{ij} is the ij^{th} value in the input image.

Softmax is the activation function used for the network final layer which can be computed as in (2).

$$\hat{y}_i = \frac{e^{a_i}}{\sum_{j=0}^{N} e^{a_j}} \quad (2)$$

where the number of classes and the i^{th} output value are denoted as N and a_i respectively.

Loss function: The focus of the model is to classify electrocardiogram (ECG) images, which can be approached as a multiclass classification task. To measure the disparity between predicted probability values and the actual set of values, the categorical cross-entropy loss function is employed. In this context, The actual class is denoted by a one-hot encoded vector, and the model's performance is deemed superior when its prediction closely matches the one-hot vector, leading to a reduced loss. The computation of the loss function follows the formula mentioned in (3).

$$H(y, \hat{y}) = -\frac{1}{N}(\sum_{i=1}^{N} y_i . log(\hat{y}_i)) \quad (3)$$

where for *N* number of samples $\hat{y}_i$ and, y_i is the predicted value and the true value respectively.

Regularization: In order to avoid overfitting several techniques have shown best results. The proposed design uses dropout and early stopping to avoid overfitting.

*Dropout: T*he most frequently used regularization method is dropout. It dynamically chooses a few edges and discards them with all of their incoming and outgoing edges. A different set of output is obtained after this process. A dropout of 0.5 between the dense connections of the final fully connected layers, is applied in the proposed model.

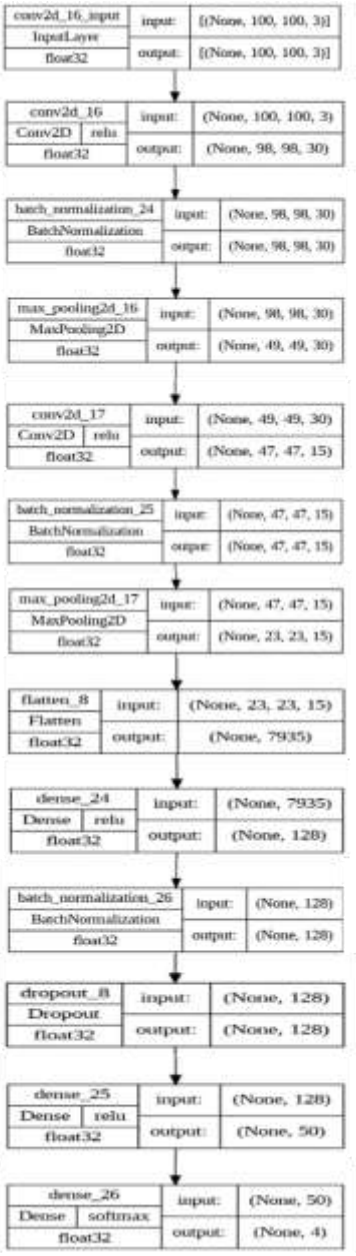


Fig. 2: CNN model architecture

Early stopping: Early stopping refers to a regularization technique where the model training is halted when validation results begins to go down. In this case, it is implemented by monitoring the validation loss and stopping training if it doesn't improve for a consecutive 10 epochs.

Optimization: For optimization, the Adam optimizer is utilized with an initial learning rate of 0.001. After conducting multiple experiments, it was observed that this combination led to improved convergence for the proposed model.

4.3 Model Training and Testing

The dataset has been split in the ratio 80:20 for training and testing respectively. The training has been performed for 100 epochs. Further, in order to save the best model, early stopping has been employed and a patience factor of 10 for saving the best model.

The different models experimented begins with a single CNN model, CNN model with dual stacked convolution layers and with batch normalization. Evaluation has been performed based on quantitative metrics and the best model was chosen as the optimal one. Results of ablation has been reported and performance evaluation based on classification metrics such as precision, recall, f1-score, accuracy, ROC curve, etc. are outlined in section 5. The model with the best results has been chosen as the optimal one from the experiments.

4.4 Model Explainability

For interpreting the decision of the trained model, algorithms for Explainable AI have been experimented. Explainability algorithms can be divided into local and global ones. The proposed model has experimented with both and have reported results for the same.

LIME: LIME (Local Interpretable Model-Agnostic Explanations) is an Explainability method which aids to interpret, explain or infer the outputs made by artificially intelligent models, including those applied to image classification tasks. LIME provides local explanations, meaning it focuses on explaining individual predictions rather than the model as a whole. When applying LIME to image classification, the basic idea is to understand which image parts have been influential in the model's prediction process. LIME accomplishes this by introducing perturbations to the input image and then monitoring how these perturbations affect the predictions. The major steps in LIME Explainability are:

a. Generate perturbed instances for the chosen image: Create slightly modified versions of the original image by perturbing its pixels. This can be done by applying random transformations or by systematically altering small patches of the image.

b. Select representative instances from model predictions: Choose a subset of the perturbed images and their corresponding predictions to create a local neighborhood. The selection can be based on proximity to the original image or using a sampling method.

c. Generate interpretable features: Convert the selected perturbed instances into a format that can be easily interpreted by human observers. For example, you might convert the images into superpixels (contiguous patches of pixels) and compute their average pixel values or intensities.

d. Train an interpretable model: Build an interpretable model, such as a linear model or decision tree, using the interpretable features obtained in the previous step and their corresponding predicted probabilities or scores. This model approximates the behavior of the underlying complex model within the local neighborhood.

e. Compute feature importance: Assess the importance of each interpretable feature (e.g., superpixel) by examining the learned weights or feature importance values of the interpretable model. The higher the weight/importance, the more influential the corresponding feature is for the model's prediction.

f. Generate explanation: Highlight the important features identified in the previous step on the original image. This can be done by overlaying the identified super pixels or using other visualization techniques.

.

Grad-CAM: Grad-CAM (Gradient-weighted Class Activation Mapping) is a technique used for visualizing and comprehending how the decisions of neural networks especially CNNs are made, when applied to image classification tasks. It provides insights into which parts in the input image were significant for the network's prediction. The main goal of Grad-CAM is to project the salient areas of an image that contribute to a specific class prediction made by the neural network. It generates a heatmap that indicates the importance of each pixel in the input image for the final prediction. The important steps in Grad-CAM calculation are:

- Gradient Calculation: The importance of each feature map in the final prediction is determined by calculating the gradient of the corresponding class with respect to the activations of the final convolutional layer.
- Global Average Pooling and weighted sum: The gradients are spatially pooled by taking their global average, resulting in a vector representing the importance of each feature map. These are weighted by their salience values obtained from the global average pooling step.
- Heatmap Generation: The weighted sum of the feature maps is passed through a ReLU activation function, and this activation map is up sampled to the dimension of the image, creating a heatmap.
- Visualization: The heatmap is overlaid on the input image to highlight the regions that contributed most to the prediction. The intense

regions in the heatmap indicate the areas that the network attended to when making the classification decision.
By using Grad-CAM, insights can be gained regarding how the decisions are made in the CNN and identify visual cues that influenced the predictions.

5 Results and Analysis

The model performance is analyzed in this section beginning with a brief into the evaluation metrics used followed by the results obtained and discussions based on the ablation study, quantitative evaluations, and quality of explanations.

5.1 Evaluation Metrics

The evaluation metrics considered for evaluation are the popular quantitative metrics used in a classification problem as shown below:

5.1.1 Accuracy

Accuracy is the ratio of total correct classifications made by the model and is calculated as in (4).

$$\text{Accuracy} = \frac{\text{TP}}{\text{N}} \tag{4}$$

where N is the sum of True Positives (TP), True Negatives (TN), False Positives (FP), and False negatives (FN).

5.1.2 Precision, Recall and F1-score

In multiclass classification, precision and recall takes up the same definition and are computed as in (5) and (6) respectively. F1-score balance the disadvantages of precision and recall and computed as in (7).

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

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

$$\text{F1\_score} = \frac{2*\text{Precision}*\text{Recall}}{\text{Precision+Recall}} \tag{7}$$

5.1.3 Area under Receiver Operating Characteristic Curve (AUC ROC)

Area under ROC curve is calculated from ROC curve which is plotted between Recall to False positive rate (FPR). It is also known as the ratio of wrong classification made by the model as in (8).

$$1 - \text{specificity} = \text{FPR} = \frac{\text{FP}}{\text{TN+FP}} \tag{8}$$

Apart from the above metrics, a confusion matrix also helps in evaluating the efficiency of the model with respect to each class.

5.2 Evaluation

The various experiments performed for evaluation are ablation study to identify the optimal model, comparison with baselines, quantitative evaluations and confusion matrix.

5.2.1 Ablation Study

Experiments have been conducted for identifying the optimal model. Ablation studies based on different modified architectures have been performed and results based on several performance metrics have been outlined in table 1. The various models experimented with are:

SingleCNNXAI: Batch normalization process and one layer of convolution has been removed in this model. Results indicate the relevance of the additional CNN layer and batch normalization in the proposed architecture. The significance using a convolution layer, for extracting features from ECG images are also demonstrated here.
DualCNNXAI: The DualCNNXAI model is almost similar to the proposed architecture. The batch Normalization layer has been ablated here. Results indicate the importance of including a batch normalization layer. Also, indicates how the results improved while adding an additional convolution layer. Convolutional layers can be applied to the output of other layers in addition to input data, such as raw pixel values. A hierarchical breakdown of the input is made possible by the stacking of convolutional layers.
DualCNNXAIwithBN: The overall proposed architecture including dual convolution and batch normalization has been experimented here. It could be seen that the incorporation of batch normalization has resulted in a steady increase in the results. Batch normalization enhances generalization performance, speeds up convergence, and permits greater learning rates, all of which result in better and more effective model training.

Results of ablation study establishes the importance of every component in the proposed architecture. Hence, the model with dual convolution followed by batch normalization has been chosen as the optimal model from the experiments.

5.2.2 Quantitative Evaluation

Model training and accuracy and loss plots

The accuracy and loss plots obtained during the training phase of the proposed deep learning model is shown in Figure 3 where 3(a) denotes the accuracy curves while 3(b) denotes the loss plots. Figure shows that the loss of training and testing is almost overlapping indicating that overfitting was addressed properly in the proposed architecture and that the results are consistent throughout the epochs. Similarly, the plots also demonstrate that the model has converged optimally.

Confusion matrix and ROC curves

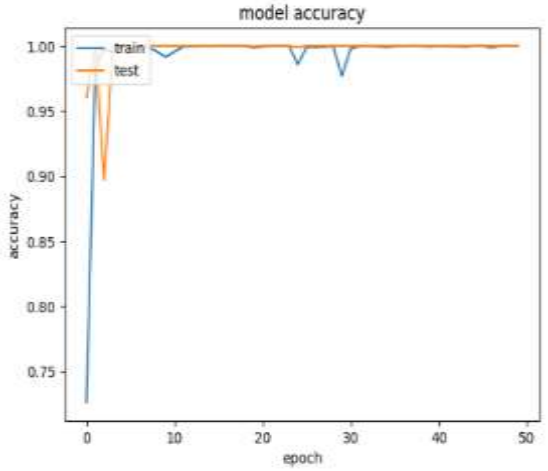
Figure 4 shows the confusion matrix report and the ROC curves obtained. The confusion matrix results in Figure 4a shows that the model classifies input with high accuracy for all the classes. The diagonal elements show the true positive results. It can be observed that there is a slight confusion in the classification results among classes 'V' and 'Q' which is negligible. The results obtained for all other entries are remarkable.

The Area under the Curve obtained is nearly 1 from Figure 4b, which indicates the high accuracy of classification achieved by the model for every classes. The high area under the ROC curve demonstrate that the model could perfectly distinguish between the different classes trained.

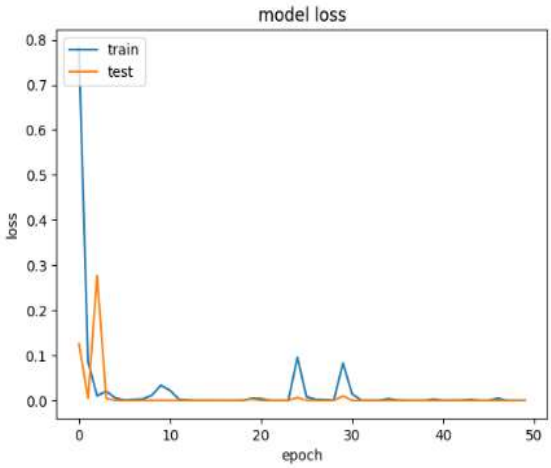
Accuracy, Precision, Recall, and f1-score

The accuracy, precision, recall, and f1-score values obtained for the proposed model are outlined in Table 1. The proposed model could achieve an elevated accuracy, precision, recall and f1-score of 0.985, 0.982, 0.982, and 0.981 owing to the dual convolution and batch normalization incorporated. The results also indicate that the model performs remarkably for the input data. Recall is often seen as a metric for quantifying performance, while precision is typically considered a measure of the quality of results.

High recall suggests that an algorithm retrieves a significant portion of relevant results, even if it occasionally includes some irrelevant ones. In contrast, higher precision indicates that an algorithm primarily provides relevant results and minimizes irrelevant ones. Here, the model could attain an equal elevated precision, and recall scores of 0.982 each which is remarkable and indicates the model performs in a balanced manner for all the classes.



(a)



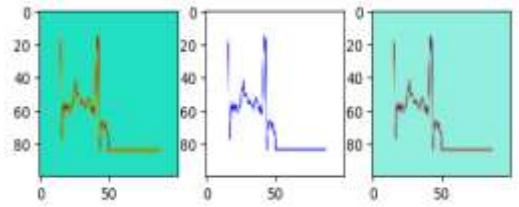
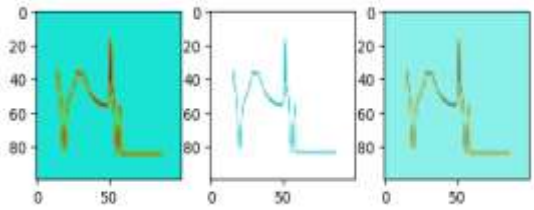
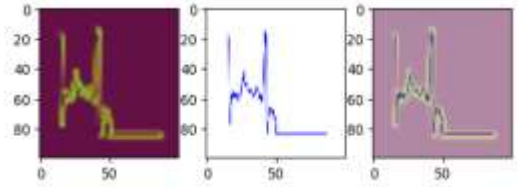
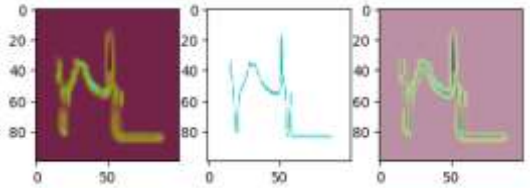
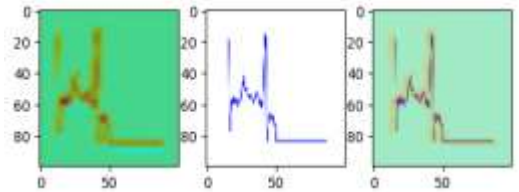
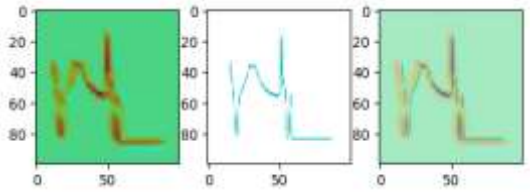
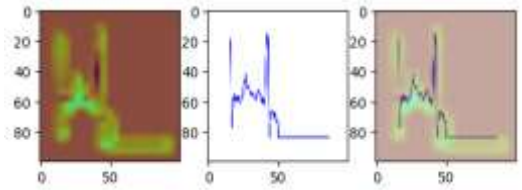
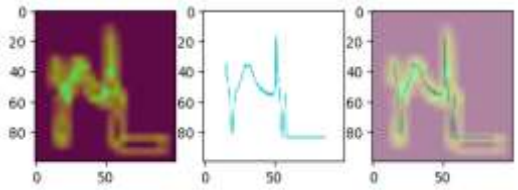
(b)

Fig. 3: Training performance of optimal models (a) Accuracy plots of train and test (b) Loss plots of train and test

Table 1. Ablation study based on quantitative metrics (Bold indicates best results)

Performance metrics	DualCNNXAI withBN	DualCNN XAI	SingleCNN XAI
Accuracy	**0.985**	0.778	0.653
Precision	**0.982**	0.735	0.541
Recall	**0.982**	0.813	0.762
F1 score	**0.981**	0.772	0.632
ROC AUC	**1.000**	0.884	0.735

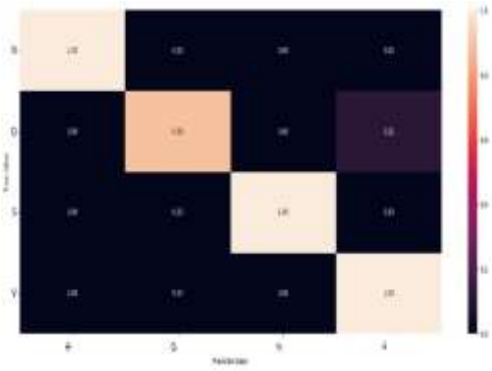
Table 2. Grad-CAM Visualization for True positive along with heatmap and activation map

Layer	(a) True Positive			(b) False positive		
	Heatmap	Image	Grad-CAM	Heatmap	Image	Grad-CAM
Conv1						
Max pool1						
Conv2						
Max pool2						

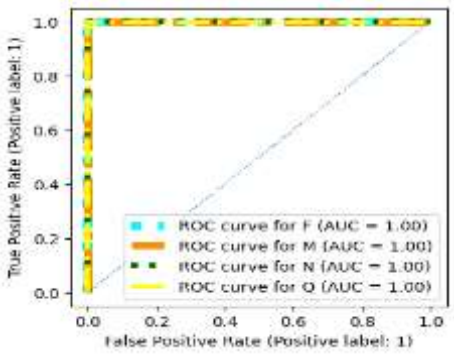
Quality of Explanations

Figure 5 shows the explanation generated for a true positive and false positive prediction using LIME algorithm. First, the model generated neighborhood data by randomly perturbing features from the instance. Later, the locally weighted models on this neighborhood data are learned to explain each of the classes in an interpretable way. The size of the neighbourhood to learn the model is set to 1000 samples. The samples considered are from class V and Q which demonstrated slight confusions in the classification accuracies according to the confusion matrix report.

Table 2 shows the figures after Grad-CAM from each prominent layer in the proposed architecture for the same true positive and false positive samples as mentioned above. The illustration depicts a contrast between the initial image from class V and Q, the Grad-CAM-generated heatmap, and the significance of the regions on the original image by superimposing it with the heatmap. Grad-CAM identifies the regions in the ECG image that contribute most to the classification of a particular class. It accomplishes this by collecting the gradients of the convolutional layers and generating heatmaps which highlight these discriminative regions. The method involves weakly-supervised localization. The algorithm starts by performing classification and obtaining the prediction for the first class. Grad-CAM maps are then generated for each predicted class. Subsequently, these maps undergo conversion into a binary format through the application of a 15% threshold relative to the maximum intensity. This operation generates interconnected clusters of pixels, followed by the establishment of a bounding box encompassing the most substantial cluster, which corresponds to the region of interest. It is important to highlight that this method is classified as weakly supervised localization because the models were not trained using bounding box annotations. In the absence of visualizations, humans may find it difficult to explain certain predictions made by the model. Nevertheless, Grad-CAM helps justify these errors.

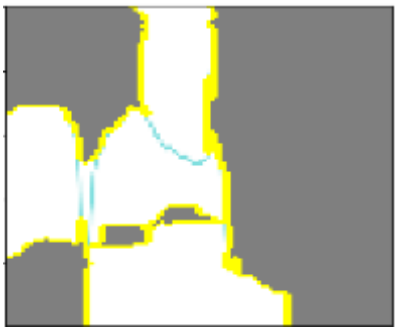


(a)

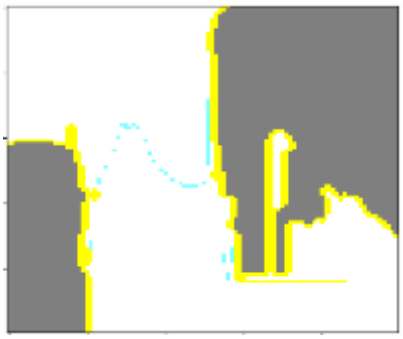


(b)

Fig. 4: Evaluation results (a) Confusion matrix for different classes (b) ROC curve for the different classes



(a)



(b)

Fig. 5: Explainability visualization for sample images for the second convolution layer with LIME explanation (a) True positive (sample from class V) and (b) False positive (sample from class Q)

6 Conclusion

The proposed model implements an intelligent explainable heart disease prediction model based on a deep convolutional neural network along with LIME and Grad-CAM activation map visualizations for better interpretability. The model could attain a higher accuracy, precision, recall and f1-score of 0.985, 0.982, 0.982, and 0.981 respectively. The model could perform in a superior manner compared to the baseline models such as single and dual convolutions without batch normalizations. In future, incorporating RNNs for time series feature extraction, and vision transformers are potential research directions. Similarly, employing echo state networks on patient medical records have shown amazing results recently which can be extended to ECG data as well. Additionally, combining chaos theory with ECG data as context information will lead to better results.

Acknowledgement:
The authors acknowledge the Chief Minister NavaKerala post-doctoral fellowship for funding the research work under the project titled "A smart system for Ischemic heart disease prediction through Explainable AI".No. KSHEC-A3/344/ Govt. Kerala-NKPDF/2022 dated 16.05.2022 and File no. CUSAT/AC(C) C2/3871/2022.

References:

[1] Deng, M., Huang, X., Liang, Z., Lin, W., Mo, B., Liang, D., ... & Chen, J, Classification of cardiac electrical signals between patients with myocardial infarction and normal subjects by using nonlinear dynamics features and different classification models, *Biomedical Signal Processing and Control*, 79, 2023,104105.

[2] Patro, S. P., Nayak, G. S., & Padhy, N, Heart disease prediction by using novel optimization algorithm: a supervised learning prospective, *Informatics in Medicine Unlocked*, 26, 2021, 100696.

[3] Sabrine, B. A., & Taoufik, A, Arrhythmia Classification Using Fractal Dimensions and Neural Networks, *In 2nd International Conference on Industry 4.0 and Artificial Intelligence (ICIAI 2021)*, 2022, February, pp. 182-187, Atlantis Press.

[4] Mazaheri, V., & Khodadadi, H., Heart arrhythmia diagnosis based on the combination of morphological, frequency and nonlinear features of ECG signals and metaheuristic

feature selection algorithm, *Expert Systems with Applications*, 161, 2020, 113697.

[5] Zarei, E., Barimani, N., & Nazari Golpayegani, G., Cardiac Arrhythmia Diagnosis with an Intelligent Algorithm using Chaos Features of Electrocardiogram Signal and Compound Classifier, *Journal of AI and Data Mining*, 10(4), 2022, pp.515-527.

[6] Wang, M. H., Huang, M. L., Lu, S. D., & Ye, G. C. (2020). Application of Artificial Neural Network and Empirical Mode Decomposition with Chaos Theory to Electrocardiography Diagnosis, *Sensors and Materials*, 32(9), 2020, pp. 3051-3064.

[7] Eltrass, A. S., Tayel, M. B., & Ammar, A. I, Automated ECG multi-class classification system based on combining deep learning features with HRV and ECG measures, *Neural Computing and Applications*, 34(11), 2022, pp. 8755-8775.

[8] Kuila, S., Dhanda, N., & Joardar, S, ECG signal classification and arrhythmia detection using ELM-RNN, *Multimedia Tools and Applications*, 81(18), 2022, pp. 25233-25249.

[9] SAYGILI, A., A novel approach to heart attack prediction improvement via extreme learning machines classifier integrated with data resampling strategy, *Konya Mühendislik Bilimleri Dergisi*, 8(4), 2020, 853-865.

[10] Fathurachman, M., Kalsum, U., Safitri, N., & Utomo, C. P., Heart disease diagnosis using extreme learning based neural networks, *In 2014 International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA)*, 2014, August, pp. 23-27, IEEE.

[11] Irene, D. S., Sethukarasi, T., & Vadivelan, N., Heart disease prediction using hybrid fuzzy K-medoids attribute weighting method with DBN-KELM based regression model, *Medical Hypotheses*, 143, 2020, 110072.

[12] Safii, I., Kamisutara, M., & Faahrudin, T. M., Imam Safii Heart Disease Classification using Gain Ratio Feature Selection with Hidden Layer Modification in Extreme Learning Machine. *IJCONSIST JOURNALS*, 2(02), 2021, pp. 71-76.

[13] Xu, Y., Zhang, S., Cao, Z., Chen, Q., & Xiao, W., Extreme learning machine for heartbeat classification with hybrid time-domain and wavelet time-frequency features, *Journal of Healthcare Engineering*, 2021.

[14] Singh, R. S., Saini, B. S., & Sunkaria, R. K. (2018). Detection of coronary artery disease by reduced features and extreme learning machine. *Clujul Medical*, 91(2), 166.

[15] Anand, A., Kadian, T., Shetty, M. K., & Gupta, A.,Explainable AI decision model for ECG data of cardiac disorders, *Biomedical Signal Processing and Control*, 75, 2022, 103584.

[16] Tzou, H. A., Lin, S. F., & Chen, P. S., Paroxysmal atrial fibrillation prediction based on morphological variant P-wave analysis with wideband ECG and deep learning, *Computer Methods and Programs in Biomedicine*, 211, 2021, 106396.

[17] Duffy, G., Jain, I., He, B., & Ouyang, D, Interpretable deep learning prediction of 3d assessment of cardiac function, *In PACIFIC SYMPOSIUM ON BIOCOMPUTING 2022*, 2022, pp. 231-241.

[18] Ganeshkumar, M., Ravi, V., Sowmya, V., Gopalakrishnan, E. A., & Soman, K. P, Explainable deep learning-based approach for multilabel classification of electrocardiogram, *IEEE Transactions on Engineering Management*, 2021.

[19] Apama, C., Rohini, P., & Pandiyarasan, V., Interpretation of ResNet50 model for MI related cardiac events using Explainable Grad-CAM approach, *In Current Directions in Biomedical Engineering*, 2022, September, Vol. 8, No. 2, pp. 723-726. De Gruyter.

[20] Raza, A., Tran, K. P., Koehl, L., & Li, S., Designing ecg monitoring healthcare system with federated transfer learning and explainable AI, *Knowledge-Based Systems*, 236, 2022, 107763.

[21] Le, K. H., Pham, H. H., Nguyen, T. B., Nguyen, T. A., Thanh, T. N., & Do, C. D, LightX3ECG: A Lightweight and eXplainable Deep Learning System for 3-lead Electrocardiogram Classification, *arXiv preprint arXiv*:2207.12381, 2022.

[22] Taniguchi, H., Takata, T., Takechi, M., Furukawa, A., Iwasawa, J., Kawamura, A., ... & Tamura, Y., Explainable artificial intelligence model for diagnosis of atrial fibrillation using holter electrocardiogram waveforms, *International Heart Journal*, 62(3), 2021, pp. 534-539.

[23] Sanjana, K., Sowmya, V., Gopalakrishnan, E. A., & Soman, K. P., Explainable artificial intelligence for heart rate variability in ECG signal, *Healthcare Technology Letters*, 7(6), 2020, 146.

[24] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... & Stanley, H. E., PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals,

Circulation [Online]. 101 (23), 2020, pp. e215–e220

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

-Sreeja M. U. carried out the Conceptualization, investigation, methodology and validation, writing, draft and editing.

-Supriya M. H. has supervised, reviewed and validated the work and manuscript.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

The research work is funded under the Chief Ministers NavaKerala post-doctoral fellowship for the project titled "A smart system for Ischemic heart disease prediction through Explainable AI".No. KSHEC-A3/344/ Govt. Kerala-NKPDF/2022 dated 16.05.2022 and File no. CUSAT/AC(C) C2/3871/2022.

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0
https://creativecommons.org/licenses/by/4.0/deed.en_US