Innovative ECG Classification Approach Utilizing a Transfer Learning-Driven Ensemble Architecture

HMLS Kumari[1#]
[1]Computer Center, Faculty of Engineering, University of Peradeniya, Sri Lanka
[#]lihinisangeetha99@gmail.com

ABSTRACT An electrocardiogram (ECG/EKG) is a vital methodology that is used for the diagnosis and monitoring of heart diseases by recording the electrical activity of the heart. However, manual analysis of ECGs shows limitations such as noise sensitivity, visual interpretation constraints and data imbalance. The proposed study a deep learning ensemble model combining DenseNet121, InceptionV3, and ResNet50 are implement to classify ECG images to improve diagnostic accuracy. The model is trained on two datasets: the National Heart Foundation 2023 ECG dataset and the ECG Dataset for Heart Condition Classification, focusing the main cardiac conditions such as abnormal heartbeat, myocardial infarction. The preprocessing techniques include background removal of ECG signal images, grayscale conversion, and data augmentation to enhance image quality and overfitting reduction. Stratified 5-Fold cross-validation was employed to demonstrate the generalization abilities of the proposed models. Early stopping and performance plots demonstrated that proposed model is not overfitting and two proposed models show consistent accuracy which suggests the model is not biased toward a specific dataset. While the ensemble models, as demonstrated in this study, produce better results than single models. The proposed study demonstrates validation accuracies of 98.62% and 96.75% for the National Heart Foundation 2023 dataset and the ECG dataset for heart condition classification, respectively, using 5-fold stratified cross-validation. There are still some limitations, such as the proposed ensemble models not being evaluated using Explainable AI, which reduces clinical trust. Additionally, small datasets can limit the model's generalizability. Therefore, this study demonstrates the potential of deep ensemble models with advanced preprocessing for ECG classification, but it also highlights the importance of greater transparency, better dataset diversity, and real-world validation in future research studies.

INDEX TERMS Data Augmentation, ECG (electrocardiogram), Ensemble Architectures, Preprocessing techniques, Transfer learning

I. INTRODUCTION

An electrocardiogram involves placing temporary electrodes on chest and limbs to record heart's electrical signals that regulate its beats. This information is analysed by a computer and is shown as a wave pattern the healthcare professional can further analyse for diagnostic insight. Figure 1 shows the components of the heart that are involved in an ECG. The internal pacemaker is a natural system in the heart that controls its rhythm and rate. The sinoatrial (SA) node, or what is commonly called the internal pacemaker of the heart. The SA node initiates the heartbeat by firing an electrical impulse. An ECG detects this electrical signal and tracks its path as heart contracts and then relaxes with each heartbeat. The electrical activity of the heart produces three distinct waves, each corresponding to specific actions within the heart. The first wave, the "P wave," is produced by the upper chambers of the heart, known as the atria, where the heartbeat originates. The "QRS complex" is the wave produced by the contraction of the lower chambers, the ventricles. Finally, the third wave, called the "T wave," represents the recovery or resting phase of the heart following a beat [1]. Figure 2 shows an ECG wave for a normal person, and Figure 3 depicts the general representation of the wave in a 1-heartbeat process.

Classification of abnormalities in the ECG signals is not an easy task in any way because several obstacles lay the process of classification. Firstly, analysis or identification directly on the paper recordings of the ECG is a huge problem.

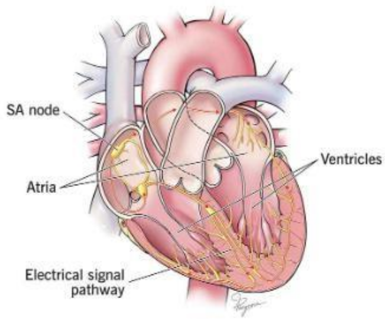

Figure 1. The diagram of heart with its components that helps to get ECG graphs [1]

Figure 2. The normal person ECG wave [1]

Figure 3. The normal person ECG wave [1]

In manual diagnosis, medical professionals often find it hard to interpret ECG signals using simple visual examination because waveforms can be complex and have subtle differences that can be missed. To help with that, deep learning models based on advanced image and signal processing have been developed. These models are considered to provide more accurate and reliable decisions than traditional human methods. However, there are some drawbacks in these models, such as data noise, unbalanced case types, and differences in patient conditions. Machine learning methods have shown improved results, but their accuracy is still limited, especially in handling changing situations based on patient conditions.

Recent research further enhances deep learning, which can automatically identify important features in ECG images and signals. Techniques like CNNs and RNNs have worked well, but they do not demonstrate high accuracy when used alone. Therefore, ensemble models where several models are combined to form a stronger system can be introduced, though this approach is not yet widely used. Data augmentation, which involves generating more varied training samples, is also useful for preventing overfitting and improving model performance, especially when the dataset is small. Stratified 5-fold cross-validation was used to evaluate the reliability and generalizability of the predicted decisions from the proposed models. Even as the proposed study claims novelty through a method of ensemble learning, the contribution is incremental, rather than revolutionary. The methodology employs an ensemble of well-known pre-trained CNN architectures such as DenseNet121, InceptionV3, and ResNet50, combined by a simple average ensemble approach with deep layer fine-tuning, in addition to the addition of dense layers, dropout, and batch normalization. Technically valid, the methodology in effect reproduces existing architectures without introducing a fundamentally new learning algorithm or architectural innovation. But while the ensemble approach may not constitute a conceptual innovation, it offers a solution to a key practical issue in medical image classification, enhancing robustness and generalizability via suitably tuned ensembles of pre-existing CNN models. The addition of stratified 5-fold cross-validation, dropout regularization, and equable performance across folds further reinforces the clinical credibility of the findings. Furthermore, parallel application of multiple architectures allows complementary feature extraction, offering an effective but practical means of improving ECG classification performance particularly in situations where resources are constrained or in real-time diagnostic contexts. Nevertheless, the study can be strengthened by exploring more innovative alternatives, i.e., Transformer-based methods, or comparison to different ensemble techniques and advanced fusion mechanisms for better basis to justify its novelty claims.

II. RELATED WORKS

Mohammed et al. [2] proposed a CNN that was especially designed for the classification of ECG signals. The study focuses on the classification of five types of ECG arrhythmic signals using the MIT-BIH Arrhythmia Dataset from PhysioNet. ECG preprocessing techniques such as denoising, ORS peak detection, and heart beat segmentation were performed. 1-D CNN ECG heart beat classifier was performed and achieved a result of an accuracy of 95.2% achieved. This study outlines how deep learning models depend on the prior quality of input data to function optimally through preprocessing steps in the medical imaging field. Using a single dataset makes it difficult to generalize the model's application. Without employing techniques like regularization, data modification, or outside validation, the chances of overfitting rise, which adversely affects the model's usefulness in actual clinical environments. The ensemble model, which was proposed by Essa et al. [3], combines a deep learning-based multi-model of a CNN with an LSTM network. As a result, can conclude that multi models have shown better performance than single models.

The other known system adheres to the hybrid classification technique of ECG images obtained from the "ECG Images dataset of Cardiac Patients" done by tariq et al., [4]. Those included ECG images with 12-lead ECG classified under abnormal heartbeat, Myocardial Infraction (MI), prior MI history, and normal ECG. With the integration of IoT-based data acquisition with a CNN-based classifier integrated with an attention module, excellent accuracy of 98.39% was realized, depicting the potential of this proposed system in identifying cardiac disorder diagnosis. Acharya et al. [5] proposed a 13-layer deep fully convolutional neural network for classification. Although the proposed methodology performed very efficiently, it involves complex computational time and is bound to have limitations in practice; the results would vary based on expertise. Similar to ANNs, the performance of a CNN also depends upon structure, weights, and preferences of earlier layers. To improve the efficiency and reduce overfitting, pooling was used that reduced the size of output generated from convolutional layers and consequently reduced computational demands. The final accuracy is 88.67%, the specificity 90.00%, and a sensitivity of 95.00% respectively. Fatema et al. [6] proposed a deep learning approach for cardiovascular diseases classification using the enhanced paper-based electrocardiogram images. The work focused on the optimization of precision with a minimal time complexity to classify heart disorders into five classes. Different image preprocessing techniques were used to improve the quality and eliminate artifacts of the images before training. Besides, model components and hyperparameter optimization can be done via an ablation study, thereby improving the performances even more. Indeed, the InRes-106 model recorded an impressive accuracy of 98.34%, higher than any individual models (InceptionV3: 90.56%, ResNet50: 89.63%, DenseNet201: 88.94%, VGG19: 87.87%, MobileNetV2: 80.56%).

Deep Learning framework for cardio vascular disease prediction using ECG Images done by Muthu Meena et al. [7] proposed a study on the ECG image dataset to predict four primary cardiac conditions. These include an abnormal heartbeat, myocardial infarction, history of myocardial infarction, and normal cases using deep learning techniques. A hybrid model combining Inception-V3 and VGG-19 is proposed for cardiovascular disease prediction. Inception-V3 has excellent feature extraction capability with much efficiency, while VGG-19 extracts spatial information in minute detail from ECG images. This hybrid model performs the task with better results by considering the complementary strengths of these architectures. The proposed model significantly outperforms previous studies in terms of an accuracy of 96.0%, precision of 90.90%, recall of 100%, and F1-score of 95.23%.

III. METHODOLOGY

To classify ECG to detect heart diseases, Averaging Ensemble of DenseNet121, InceptionV3, and ResNet50 models is utilized. Before that, several preprocessing methods has been applied to those ECG images and then classification done.

A. Data Acquisition

1. National Heart Foundation 2023 ECG dataset

National heart foundation 2023 ECG dataset [8] is a free open dataset available on the Kaggle website. This is an ECG image dataset from the National heart foundation of Bangladesh, and it presents a collection of ECG images with various categories related to cardiac health. These images represent the heart's electrical activity versus time and convey critical information about the health status of the heart. Basically, there are four classes of subjects in the dataset: abnormal heartbeats, myocardial infarction, normal, and previous myocardial infarction.

Abnormal Heartbeat Patients: This category contains ECG images taken from patients who have irregular heartbeats or arrhythmias. Examples include Atrial fibrillation, bradycardia, tachycardia, and ectopic beats and other abnormal rhythm conditions [9]. These image types help analyse and diagnose various heart-related ailments. Myocardial Infarction class represents ECG images that are grouped into this category and are from patients diagnosed with myocardial infarction, commonly referred to as a heart attack. The recordings often reveal distinct changes in the waveform, such as ST-segment elevation or depression, T-wave inversion, or pathological Q-waves, signifying myocardial damage [10]. Normal Individuals represent group comprises ECG readings from individuals without detectable cardiac abnormalities. These recordings serve as standard reference patterns, enabling comparison with abnormal ECGs to identify deviations from normal cardiac activity [8]. The category Myocardial Infarction class history includes ECG images of patients with a previous history of myocardial infarction. Images depict changes that may be permanent relative to prior myocardial damage or transient variations seen at follow-up studies or during continuous cardiac monitoring [8]. The total amount of ECG images is 2898, and further details of classes in the above dataset are shown in Table 1. As shown in table1 above dataset is balanced.

Table 1. The brief details of number of ECG mages in each class of dataset National heart foundation ECG 2023 dataset

Class	No. ECG images
Abnormal Heartbeat	814
Myocardial Infraction	716
Normal Person	852
Patients that have history of Myocardial Infraction	516

2. ECG dataset for Heart condition classification

The proposed study is not biased to a single dataset. Hence, another dataset considered here is the ECG Dataset for Heart Condition Classification [11]. This is freely available in the Kaggle website. The categories included in this dataset are three classes of ECG signals: normal signals, abnormal rhythms, and cardiac signals arising due to some diseases. It includes records that are labelled and collected from healthy individuals and patients with different heart conditions. The data will be used for machine learning-based applications that can allow one to continuously monitor a person's health and predict cardiac diseases in real time. The dataset also maintains an appropriate balance of good-quality ECG images across classes. Therefore, the dataset offers access to powerful tools with

AI-enabled diagnostic support in healthcare. There are 707 images in the combined dataset. Table 2 demonstrates the brief details of number of ECG mages in each class of ECG for Heart condition classification dataset

Table 2. Number of ECG images in each class of the dataset ECG for Heart condition classification

Class	**Number of ECG images**
Abnormal heart beat	241
History of Myocardial Infraction	171
Normal Person	295

3. ECG Signal Preprocessing

The original image of national heart foundation 2023 ECG dataset is shown in Figure 4 and represents additional background in the ECG image. Therefore, the proposed approach followed the background removal preprocessing step of the ECG image so that only the ECG signal would be considered. Figure 5 presents the pre-processed ECG image.

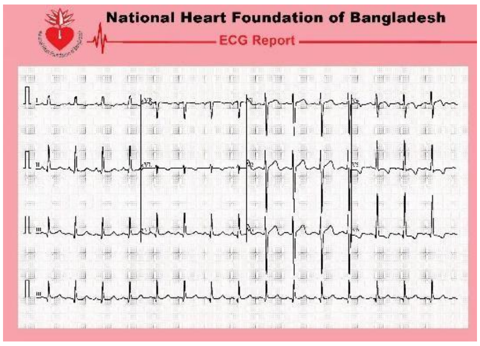

Figure 4. The ECG image of Abnormal heart beat class of National Heart Foundation 2023 ECG dataset.

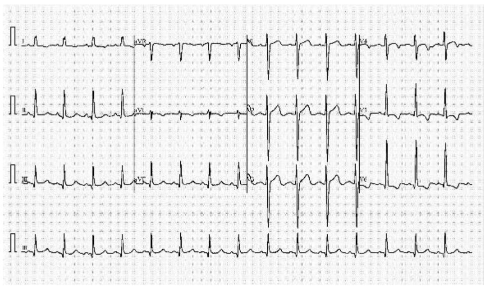

Figure 5. The cropped ECG image

ECG image extraction by removing the background was done only for the National Heart Foundation 2023 ECG dataset.
The preprocessing technique is applied on the ECG image by changing into grayscale[12], normalizing for increasing the contrast of the ECG image, after which Gaussian Blur[13] was applied in reducing noise using the Gaussian kernel size of 5, Otsu's thresholding [14] used to change the blurred image into a binary image, resize the image to 256X 256.Use the above preprocessing techniques to make 1 image from the original dataset and add those images to the original datasets. As a result, each image makes new 5 images according to the preprocessing steps. Figure 6 shows the samples of pre-processed steps used for all classes of dataset National heart foundation ECG 2023 dataset.
After splitting new dataset into train70%, test 20% and validation 10% by total dataset and train the proposed Ensemble CNN model. Few data augmentation steps are used to reduce overfitting before training.

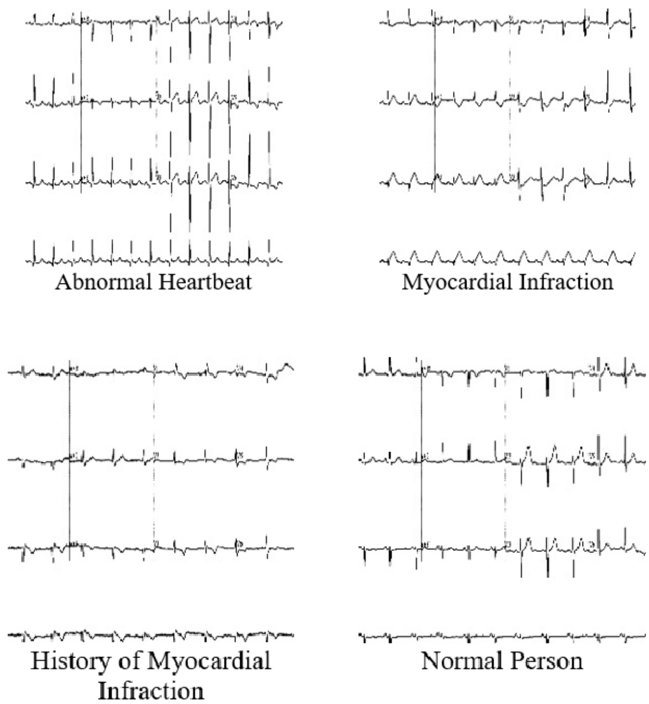

Figure 6. The resulted ECG images after preprocessing techniques.

4. Ensemble Model for CNN Transfer Learning

As shown in Figure 7, The proposed study proposed an ensemble model of DenseNet121, InceptionV3 and Resnet50.

The top layer of each transfer learning model is removed (include_top=False), and ImageNet weights are used. To retain the pre-trained features of the models, 80% of the layers are frozen. A global average pooling layer, a dense layer with 1024 neurons using the ReLU activation function are added to each model's output. These steps are followed individually for all three transfer learning models. The outputs from each model are averaged using the Average() layer. The resulting fused model is passed through a Dense layer (1024 neurons with ReLU activation), followed by a Dropout layer (0.4) for regularization and Batch

Normalization for training stability. The final layer is an output layer consisting of a Dense layer, with the number of units depending on the number of classes in the classification problem.

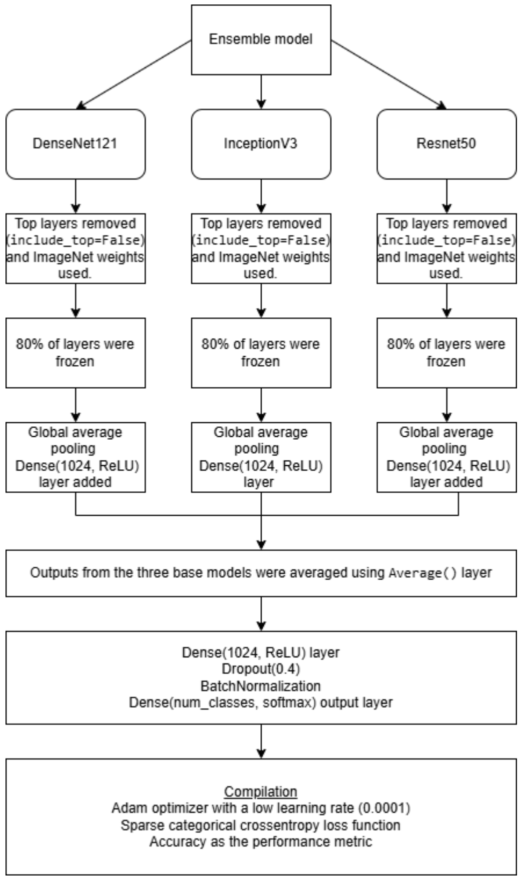

Figure 7. The proposed ensemble model

Using an Adam optimizer with a learning rate of 0.0001, the model was trained as per the schedule outlined in the provided performance evaluation method employing sparse categorical cross entropy loss. For better class balance within folds, given the small size of the datasets and to ensure robust evaluation with various data splits, a 5-fold Stratified K-Fold Cross-Validation was implemented. During both the training and evaluation phases, each iteration of the folds was trained as a fresh model to reduce overfitting, incorporating early stopping and learning rate reduction callbacks to optimize performance.

Model performance was evaluated on the training and validating sets, and the training accuracy, validation accuracy, precision, recall, and F1 score for validation were calculated for each fold.

IV. RESULTS AND DISCUSSION

Using the ensemble model of the transfer learning approach, suboptimal accuracy was achieved. This performance could be mainly due to the limited size of the dataset. Small datasets often result in overfitting of the model on the training data while failing on the unseen test data [15]. To overcome this, early stopping and 5-Fold Stratified cross validation are used while training the model

The proposed model is based on an ensemble model of Densenet121, Inception V3, and Resnet50 for the above two datasets. Other than that, we used to train individual models like Densenet121, InceptionV3, and Resnet50 for the above dataset separately. But finally, we got high accuracy for the ensemble model other than an individual model of Densenet121, Resnet 50, and InceptionV3. This observation can be seen in both datasets used in proposed work. The proposed model was trained by taking the 50 epochs, batch size 32 with learning rate 0.0001, dense layer 1024 along with the 'relu' activation layer, and the dropout 0.4 National Heart Foundation 2023 ECG dataset with 4 classes and the ECG dataset to classify Heart conditions with 3 classes, respectively.

Table 3 demonstrates the metrics of the best performance achieved at validation in 5-fold stratified cross validation which are validation accuracy, validation precision, validation F1-score, and validation recall for both the National Heart Foundation 2023 ECG dataset and another ECG dataset used for heart condition classification.

Table 3. Highest validation metrics of the ensemble model (DenseNet121, InceptionV3, ResNet50) on two datasets

Ensemble Model	Validation Accuracy	Validation Precision	Validation Recall	Validation F1 Score
National Heart Foundation	0.9862	0.9869	0.9862	0.9863
Heart condition classification	0.9675	0.9687	0.9675	0.9674

National Heart Foundation 2023 ECG Dataset shows maximum accuracy of 98.62% which is highest testing accuracy among proposed ensemble models.

The Loss graph over epochs, accuracy graph over epochs and confusion matrix for proposed model for National Heart Foundation 2023 ECG Dataset and ECG dataset for Heart condition classification is shown in Figures 8,9,10 respectively.

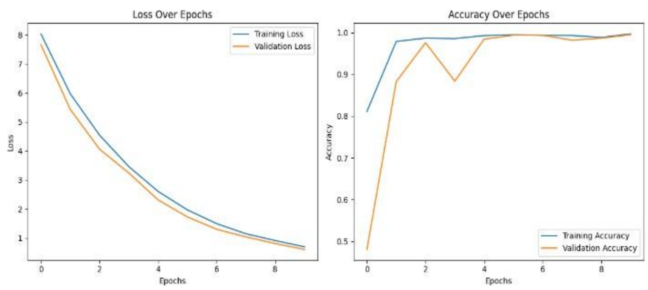

Figure 8. Loss curve and accuracy curve for model with accuracy 98.62% in National Heart Foundation 2023 ECG Dataset

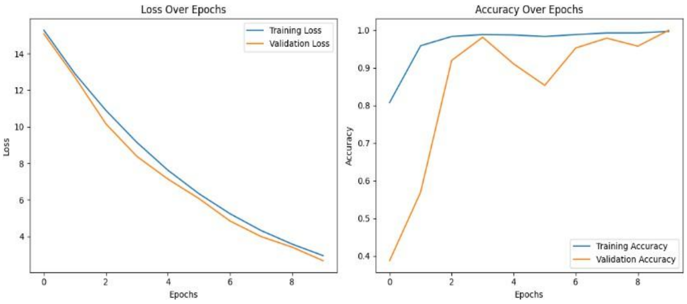

Figure 9. Loss curve and accuracy curve for model with accuracy 96.75% in ECG dataset for Heart condition classification

Model is not overfitting and generalize and the proposed model is not biased for one dataset. The proposed model performs well than recent work. Table 4 briefly describe how proposed model performs well than recent work that described in recent work.

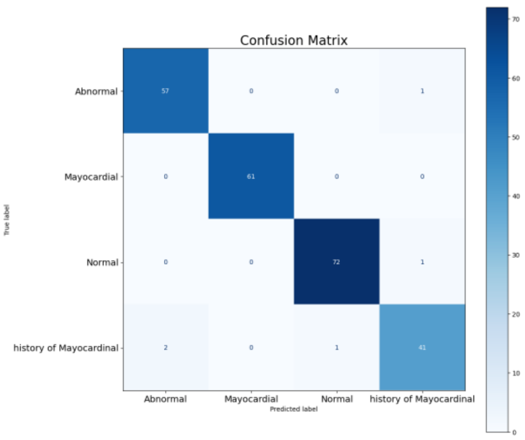

Figure 10. Confusion Matrix for model with

Table 4. Best training and validation metrics of the ensemble model (DenseNet121, InceptionV3, ResNet50) on two datasets.

Reference model	Accuracy
Mohammed et al. [2]: Tailored 1-D CNN ECG heartbeat classifier	95.2%
Essa et al. [3]: Deep learning-based multi-model ensemble (CNN + LSTM)	Not specified
Tariq et al. [4]: CNN-based classifier with IoT data acquisition and attention module	98.39%
Acharya et al. [5]: 13-layer fully convolutional neural network	88.67%
Fatema et al. [6]: InRes-106 model (Enhanced paper-based ECG images)	98.34%
Fatema et al. [6]: Individual models (InceptionV3, ResNet50, DenseNet201, VGG19, MobileNetV2)	90.56%, 89.63%, 88.94%, 87.87%, 80.56% respectively
Muthu Meena et al. [7]: Hybrid model combining Inception-V3 and VGG-19	96.0%
Proposed model	98.62%

V. CONCLUSION

Abnormality classification in ECG signals is still challenging due to the various limitations imposed by the nature of the signal, deficiency of conventional visual analysis methodologies, external noise, and imbalance of datasets. Advanced computer systems and machine learning methods have enormously improved the accuracy of diagnosis, but conventional approaches suffer from problems like false positives and bounded accuracy. This encourages deep learning algorithms that could ensure improved classification performance, promising minimal diagnostic errors.

Recent works have focused on approaches concerning single models, while the accuracy was mostly moderate. Simultaneously, ensemble modelling methods that will take advantage of several models still have great underexplored potential in ECG image classification tasks. The most possible future research directions also involve systematic integration between data augmentation and ensemble methods. Addressing these gaps may provide a path toward more robust, more accurate, and generalizable systems for ECG signal classification with improved patient outcomes and more reliable diagnostic tools.

It describes the development and evaluation of an ensemble deep learning model for ECG signal classification, where DenseNet121, InceptionV3, and ResNet50 architectures are considered. With advanced preprocessing and data augmentation techniques, the proposed model shows very promising results in view of single models and the recent literature. After being trained using 5-fold stratified cross validation, this model achieved an almost 99.95% accuracy on training and a validation accuracy of 98.62% for the National Heart Foundation 2023 ECG Dataset and 96.75% of validation accuracy on ECG dataset for Heart Condition Classification. This has indicated that ensemble learning reinforces the performance by handling key challenges such as overfitting and class imbalance of the model.

The model was further generalized and stabilized by key enhancements: dropout layers, and batch normalization. Moreover, the comparison with the recent works allows the proposed model to show state-of-the-art accuracy, revealing the potential for practical applications of cardiovascular disease diagnosis. Furthermore, as much as this study has shown positive results and future work should assess the proposed model with Explainable AI methods. Achieving this would show how deep learning models can be applied in the healthcare field by fostering trust in AI systems among medical professionals.

REFERENCES

[1] Professional, C. C. medical, “What’s an EKG?,” Cleveland Clinic, Feb. 19, 2025. [Online]. Available: https://my.clevelandclinic.org/health/diagnostics/16953-electrocardiogram-ekg

[2] M. M. R. Khan et al., “Electrocardiogram heartbeat classification using convolutional neural networks for the detection of cardiac Arrhythmia,” in Proc. 2020 Fourth Int. Conf. I-SMAC (IoT in Social, Mobile, Analytics and Cloud), 2020, pp. –, doi: 10.1109/I-SMAC49090.2020.xxxxx.

[3] E. Essa and X. Xie, “An ensemble of deep learning-based multi-model for ECG heartbeats arrhythmia classification,” IEEE Access, vol. 9, pp. 103452–103464, 2021.

[4] T. Sadad et al., “Efficient classification of ECG images using a lightweight CNN with attention module and IoT,” Sensors, vol. 23, no. 18, p. 7697, 2023.

[5] U. R. Acharya et al., “Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals,” Computers in Biology and Medicine, vol. 100, pp. 270–278, 2018.

[6] K. Fatema et al., “A robust framework combining image processing and deep learning hybrid model to classify cardiovascular diseases using a limited number of paper-based complex ECG images,” Biomedicines, vol. 10, no. 11, p. 2835, 2022.

[7] S. Muthumeena and L. Priya, “Deep Learning Framework for Cardio Vascular Disease Prediction Using ECG Images,” in 2024 Int. Conf. Smart Syst. Electr., Electron., Commun. Comput. Eng. (ICSSEECC), 2024, pp. –, doi: 10.1109/ICSSEECCxxxxx.

[8] D. K. Mohsin, “National Heart Foundation 2023 ECG dataset,” Kaggle, Mar. 27, 2024. [Online]. Available: https://www.kaggle.com/datasets/drkhaledmohsin/national-heart-foundation-2023-ecg-dataset

[9] “Types of arrhythmias - AFIB institute.” [Online]. Available: https://afibinstitute.com.au/about-us/other-arrhythmias/

[10] N. Ojha, “Myocardial infarction,” StatPearls [Internet], Aug. 8, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK537076/

[11] S. Fazeli, “ECG Heartbeat Categorization Dataset,” Kaggle, May 31, 2018. [Online]. Available: https://www.kaggle.com/datasets/shayanfazeli/heartbeat

[12] R. C. Gonzalez, Digital Image Processing, Pearson Education India, 2009.

[13] J. Canny, “A computational approach to edge detection,” IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-6, pp. 679–698, 1986.

[14] N. Otsu, “A threshold selection method from gray-level histograms,” Automatica, vol. 11, pp. 285–296, 1975.

AUTHOR BIOGRAPHY

H.M.L.S. Kumari

H.M.L.S. Kumari, an Instructor at the Computer Center, Faculty of Engineering, University of Peradeniya, completed a Bachelor of Science Honours degree in Computer Science at the Faculty of Applied Sciences, Vavuniya Campus, University of Jaffna. Her research interests include Deep Learning, Computer Vision, Artificial Intelligence and Explainable AI, particularly in the health sector.