

A Deep Learning Framework for Classifications of Pathological Heartbeats from ECG Signals

Muhammad Rashid ¹, Gul Sher Ali Khan ², Vijay Singh ³, Raja Vavekanand ^{3*}

¹Muhammad Nawaz Shareef University of Agriculture Multan.

²Buitems, Quetta.

³Benazir Bhutto Shaheed University Lyari.

***Corresponding Author:** Raja Vavekanand, Benazir Bhutto Shaheed University Lyari.

Received Date: December 13, 2024; **Accepted Date:** December 30, 2024; **Published Date:** January 06, 2024

Citation: Muhammad Rashid, Ali Khan GS, Vijay Singh, Raja Vavekanand, (2025), A Deep Learning Framework for Classifications of Pathological Heartbeats from ECG Signals, *J, Surgical Case Reports and Images*, 8(1); DOI:10.31579/2690-1897/231

Copyright: © 2025, Raja Vavekanand. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Cardiovascular diseases (CVDs) are a major global health concern, necessitating efficient and reliable diagnostic methods. Electrocardiograms (ECGs) play a vital role in detecting cardiac abnormalities, but their manual interpretation by experts can be labor-intensive and subject to variability. In this study, we propose a novel approach for ECG classification using convolutional neural networks (CNNs) and transfer learning. The model was first pre-trained on a dataset of aggregated ECG signals to capture global patterns and subsequently fine-tuned on individual patient data to tailor predictions to specific characteristics. We utilized the MIT-BIH Arrhythmia Database for training and evaluation. Individual patient-specific models achieved an average balanced accuracy of 94.6%, while transfer learning-based models reached 93.5%. While individual models demonstrated marginally superior performance, transfer learning offered significant advantages in data-scarce scenarios by leveraging pre-trained knowledge. Our findings highlight the potential of transfer learning in addressing challenges such as data scarcity and inter-patient variability.

Keywords: ecg signal; deep learning; pathology classifications; transfer learning

1. Introduction

Cardiovascular diseases (CVDs) are the leading cause of mortality globally, necessitating efficient diagnostic tools for early detection and management. Electrocardiograms (ECGs) are non-invasive, cost-effective instruments widely used to monitor and diagnose heart conditions such as arrhythmias and myocardial infarction. However, manual interpretation of ECG signals by medical professionals is labor-intensive and prone to human error, underscoring the importance of automated ECG classification systems [1].

Recent advancements in machine learning (ML) and deep learning (DL) have revolutionized the field of medical signal processing, enabling highly accurate and efficient ECG classification [2]. Convolutional Neural Networks (CNNs), in particular, have emerged as the state-of-the-art method for analyzing ECG signals due to their capability to learn hierarchical features from raw data [3,14]. Despite their potential, training CNNs for ECG classification presents challenges, including limited annotated datasets, inter-patient variability, and class imbalance caused by the rarity of pathological events [4]. Transfer learning offers a promising solution to these challenges by leveraging pre-trained models to adapt to specific tasks with

limited data [5,15]. This approach is particularly advantageous for personalized medicine, where ECG signals exhibit patient-specific characteristics [6]. The MIT-BIH Arrhythmia Database [7], a widely used benchmark dataset for ECG analysis, provides an excellent foundation for exploring the efficacy of transfer learning in classifying pathological heartbeats. In this study, we propose a two-stage approach for ECG classification: pre-training a generic CNN on a multi-patient dataset and fine-tuning individual models for specific patients. Our objectives are to evaluate the efficacy of transfer learning in capturing generic and patient-specific features and to compare its performance with traditional patient-specific classifiers.

2. Methods

The methodology involves preprocessing ECG signals followed by their fusion into a triple-channel input. These images are then fed into a CNN for classifying pathological heartbeats (Figure 1).

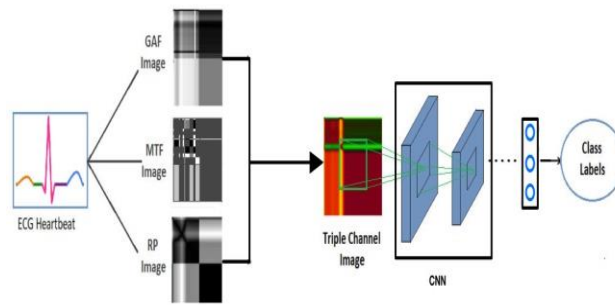


Figure 1: System architecture for ECG classification using GAF, MTF, and RP transformations with a CNN-based triple-channel approach.

3.1 Dataset

The dataset utilized in this study is the MIT-BIH Arrhythmia Database [1], a benchmark dataset widely used for ECG analysis. It comprises 48 half-hour, two-channel ECG recordings sampled at 360 Hz. Each recording is accompanied by expert annotations categorizing beats into various classes. For this study, the data was pre-processed to normalize signals to the range (0,1), filter noise with a Butterworth bandpass filter (0.4–30 Hz), and segment into 1-second windows (360 samples). Each window was labeled as

normal (0) or abnormal (1). The final dataset consisted of 49,245 labeled samples from 29 patients, after excluding imbalanced samples.

3.2 CNN Architecture

A 1D CNN was designed for binary classification of ECG windows. The model comprises four convolutional blocks with 32 kernels (5×5), max pooling (5×5), and ReLU activation. Residual connections inspired by ResNet [2] enhance feature propagation. The flattened output of the convolutional layers feeds into a dense layer with 160 units, followed by a sigmoid-activated binary classifier (Figure 2).

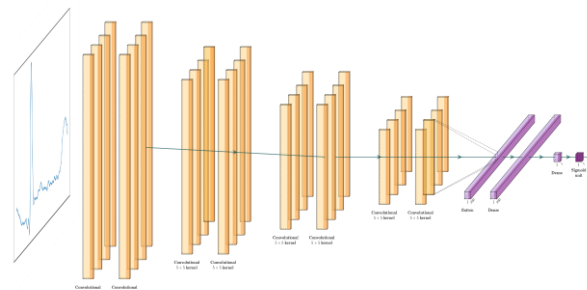


Figure 2: The architecture of our CNN

3.3 Training

This study employed a two-stage training. Pre-training, generic CNN was pre-trained on the aggregated dataset from all patients. The model optimized binary cross-entropy (BCE) loss using the Adam optimizer with a learning rate of 0.001 and weight decay of 0.0001. Fine-tuning, the pre-trained model was fine-tuned for each patient individually to capture personalized ECG

characteristics (Table 1). Hyperparameters were optimized using grid search for each patient.

$$H_p(q) = -\frac{1}{N} \sum_{i=1}^N [y_i \cdot \log(p(y_i)) + (1 - y_i) \cdot \log(1 - p(y_i))]$$

where y_i is the true label and $p(y_i)$ is the predicted probability.

Patient	Batch Size	Learning rate	Weighted Sampling	Balanced Accuracy
100	4	0.001	True	0.713
102	4	0.001	True	1.00
104	4	0.001	False	0.993
105	4	0.001	False	0.996
106	16	0.001	True	0.985
108	32	0.01	True	0.776
114	4	0.001	False	0.955
116	16	0.001	True	0.932
119	4	0.01	True	1.00
200	32	0.01	True	0.962
202	16	0.001	True	0.903
201	32	0.001	True	0.925
203	4	0.001	True	0.966

205	32	0.001	True	0.958
208	4	0.01	False	0.977
209	4	0.001	False	0.950
212	16	0.001	True	1.00
210	16	0.001	True	0.943
213	4	0.001	True	0.919
215	4	0.001	True	100
217	4	0.001	True	0.998
219	16	0.01	True	0.965
220	16	0.01	True	0.955
222	32	0.001	False	0.897
221	32	0.001	False	0.994
228	16	0.001	True	0.986
223	16	0.001	True	0.916
231	4	0.001	True	0.996
233	32	0.001	False	0.988

Table 1: This Table provides the optimal hyperparameters for each patient, found through grid search based on the balanced accuracy on the validation set.

4. Experimental Setup

4.1 Tasks and Dataset Splitting

The dataset was split into training, validation, and test sets using stratified sampling to preserve the class ratio. An 80-20 split was used for training-validation and test sets, followed by an 80-20 split of the training-validation set to create training and validation subsets. Weighted sampling was applied to handle class imbalance, ensuring the model adequately learned both normal and abnormal classes.

4.2 CNN Architecture and Parameters

The proposed CNN model (Figure 2) accepts 1×128 pulse windows as input. It consists of convolutional blocks with 32 kernels (5×5), max pooling, skip

connections, and ReLU activation. The architecture concludes with a dense layer (160 units) and a sigmoid-activated output layer for binary classification. The architecture has 67,169 trainable parameters, suitable for the dataset's size and complexity.

4.3 Training Protocols

The CNN was pre-trained on the aggregated dataset from all patients to learn generic features. The Adam optimizer was employed with a learning rate of 0.001, weight decay of 0.0001, and a batch size of 32. The training proceeded for 30 epochs, and the model with the lowest validation loss was retained. Fine-tuning: The pre-trained CNN was fine-tuned individually for each patient using the same architecture and hyperparameters optimized via grid search (Table 2).

Parameter	Range
Learning Rate	{0.001, 0.01, 0.1}
Batch Size	{4, 16, 32}
Weighted Sampling	{True, False}

Table 1: Parameter Range for Grid Search in Individual Classifiers

4.4 Evaluation Metrics

Balanced accuracy was used to evaluate model performance. It is robust against class imbalance and calculated as:

$$\text{Balanced Accuracy} = \frac{\text{Sensitivity} + \text{Specificity}}{2}$$

5. Results

5.1 Pre-training Performance

The pre-trained CNN model demonstrated effective learning on the aggregated dataset. (Figure 2) shows the training and validation loss curves, with the lowest validation loss achieved at epoch 16. This model provided a robust starting point for patient-specific fine-tuning (Figure 3).

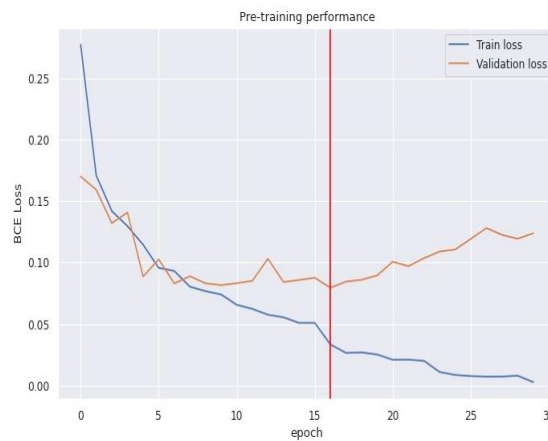


Figure 3: Training and validation loss curves over epochs for the pre-training phase.

5.2 Individual Classifiers

Baseline classifiers trained on individual patient data achieved an average balanced accuracy of 94.6% on the test set. (Figure 4) presents a histogram of the balanced accuracy for all individual classifiers, where 18 of 29

classifiers scored above 95%, and only one patient’s model performed below 75%. Figure 4 shows the training and validation loss and balanced accuracy curves for patient 203, indicating stable convergence and high performance (Figure 5).

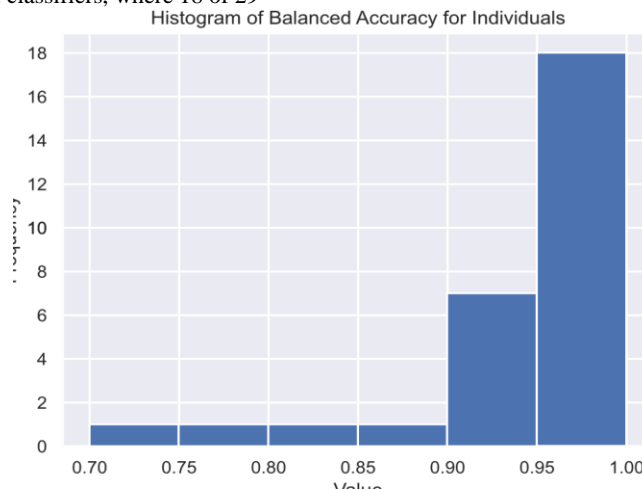


Figure 4: Histogram of the balanced accuracy of the individual classifiers on the test set.

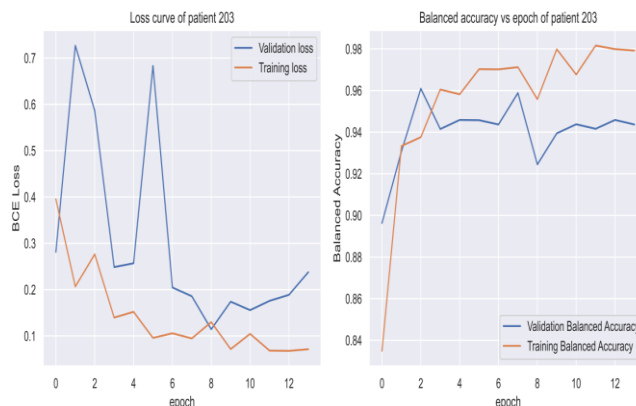


Figure 5. Loss curve (left) and balanced accuracy (right) plots over epochs for patient 203 for the individual classifier.

5.3 Transfer Learning

Fine-tuned models, leveraging pre-trained weights, achieved an average balanced accuracy of 93.5%, slightly below the individual classifiers. (Figure 6) shows the histogram of balanced accuracies for fine-tuned models, where

16 of 29 classifiers achieved above 95%. (Figure 7) compares the balanced accuracies of individual and fine-tuned classifiers, highlighting that fine-tuning boosted performance for certain patients but underperformed in others (Figure 8).

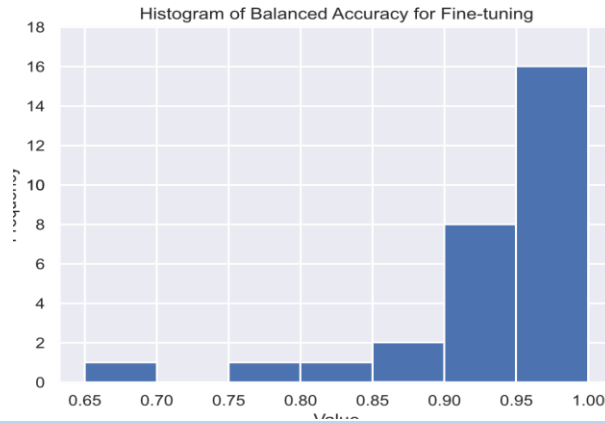


Figure 6: Histogram of the balanced accuracy of the fine-tuned classifiers on the test set.

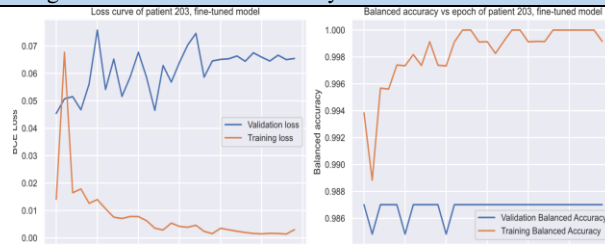


Figure 7: Loss curve (left) and balanced accuracy (right) plots over epochs for patient 203 for the fine-tuned classifier.

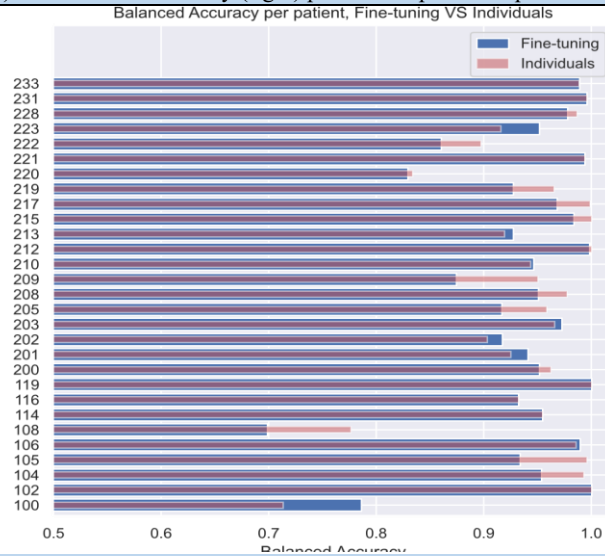


Figure 8: Bar plot of the balanced accuracy for each patient on the test set (individuals, fine-tuning).

The results indicate that while individual classifiers marginally outperform fine-tuned models on average, transfer learning is beneficial in scenarios with limited data. Hyperparameter optimization specific to fine-tuning could further enhance performance.

6. Discussion

The results of this study highlight the effectiveness of CNNs in classifying ECG signals into normal and abnormal categories. Individual classifiers achieved higher average balanced accuracy compared to fine-tuned transfer learning models. This suggests that patient-specific models are adept at capturing the unique characteristics of individual ECG signals. However, the transfer learning approach demonstrated competitive performance, with particular benefits for patients with limited training data. The slightly lower accuracy of fine-tuned models may be attributed to the reuse of hyperparameters optimized for individual classifiers during fine-tuning. Fine-tuning-specific hyperparameter tuning could address this limitation and enhance model performance. Additionally, the performance variation among

patients reflects the inter-patient variability in ECG morphology, a known challenge in ECG classification. While the study primarily focused on binary classification, the methodology could be extended to multi-class classification of specific arrhythmia types. Future work could also explore integrating data augmentation techniques, such as adding noise or varying sampling rates, to increase the robustness of the models. Overall, this research demonstrates the potential of transfer learning in ECG analysis and its utility in addressing the challenges of data scarcity and variability.

7. Conclusion

This study demonstrates the efficacy of CNNs for ECG classification and evaluates the impact of transfer learning in a binary classification task. Individual classifiers, tailored to specific patients, achieved the highest balanced accuracy, averaging 94.6%. Transfer learning, while slightly less accurate at 93.5%, provided a scalable solution for data-scarce scenarios. The pre-training phase effectively captured global ECG patterns, enabling transfer learning models to adapt to individual patient data. While individual

classifiers had a slight edge in performance, transfer learning proved advantageous for patients with limited data, highlighting its potential for practical applications. Future research should focus on optimizing hyperparameters specific to fine-tuning and incorporating multi-class classification. Expanding the dataset by integrating additional sources or applying advanced data augmentation could further improve model robustness. These advancements will enhance the utility of AI-driven solutions in personalized healthcare, offering scalable and accurate diagnostic support in real-world settings.

References

- Schläpfer, J., & Wellens, H. J. (2017). Computer-Interpreted Electrocardiograms. *Journal of the American College of Cardiology*, 70(9), 1183–1192.
- Steinhubl, S. R., Waalen, J., Edwards, A. M., Ariniello, L. M., Mehta, R. R. et al. (2018). Effect of a Home-Based Wearable Continuous ECG Monitoring Patch on Detection of Undiagnosed Atrial Fibrillation. *JAMA*, 320(2), 146.
- Ma, J., & Dong, M. (2014). R&D of versatile distributed e-home healthcare system for cardiovascular disease monitoring and diagnosis. *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI . . .)*, 12, 444–447.
- Minami, K., Nakajima, H., & Toyoshima, T. (1999). Real-time discrimination of ventricular tachyarrhythmia with Fourier-transform neural network. *IEEE Transactions on Biomedical Engineering*, 46(2), 179–185.
- Inan, O., Giovangrandi, L., & Kovacs, G. (2006). Robust Neural-Network-Based Classification of Premature Ventricular Contractions Using Wavelet Transform and Timing Interval Features. *IEEE Transactions on Biomedical Engineering*, 53(12), 2507–2515.
- Coast, D., Stern, R., Cano, G., & Briller, S. (1990). An approach to cardiac arrhythmia analysis using hidden Markov models. *IEEE Transactions on Biomedical Engineering*, 37(9), 826–836. <https://doi.org/10.1109/10.58593>
- Ebrahimi, Z., Loni, M., Daneshtalab, M., & Gharehbaghi, A. (2020). A review on deep learning methods for ECG arrhythmia classification. *Expert Systems With Applications X*, 7, 100033. <https://doi.org/10.1016/j.eswax.2020.100033>
- Hadaeghi, F. (2019). Reservoir Computing Models for Patient-Adaptable ECG Monitoring in Wearable Devices. *arXiv*. <https://doi.org/10.48550/arxiv.1907.09504>
- Moody, G., & Mark, R. (2001). The impact of the MIT-BIH Arrhythmia Database. *IEEE Engineering in Medicine and Biology Magazine*, 20(3), 45–50. <https://doi.org/10.1109/51.932724>
- Mit. (2019). MIT-BIH Arrhythmia Database [Dataset]. In Zenodo (CERN European Organization for Nuclear Research). <https://doi.org/10.5281/zenodo.3477542>
- Azizi, S., Mustafa, B., Ryan, F., Beaver, Z., Freyberg, J. et al. (2021). Big Self-Supervised Models Advance Medical Image Classification. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 3458–3468. <https://doi.org/10.1109/iccv48922.2021.00346>
- Kotar, K., Ilharco, G., Schmidt, L., Ehsani, K., & Mottaghi, R. (2021). Contrasting Contrastive Self-Supervised Representation Learning Pipelines. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). <https://doi.org/10.1109/iccv48922.2021.00980>
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). <https://doi.org/10.1109/cvpr.2016.90>
- Vavekanand, R., Sam, K., Kumar, S., & Kumar, T. (2024). CardiacNet: A Neural Networks Based Heartbeat Classifications using ECG Signals . *Studies in Medical and Health Sciences*, 1(2), 1–17. <https://doi.org/10.48185/smhs.v1i2.1188>
- Vavekanand, R., & Kumar, T. (2024). Data augmentation of ultrasound imaging for non-invasive white blood cell in vitro peritoneal dialysis. *Biomedical Engineering Communications*, 3(4), 17. <https://doi.org/10.53388/bmec2024017>



This work is licensed under Creative Commons Attribution 4.0 License

To Submit Your Article Click Here:

[Submit Manuscript](#)

DOI:10.31579/2690-1897/231

Ready to submit your research? Choose Auctores and benefit from:

- fast, convenient online submission
- rigorous peer review by experienced research in your field
- rapid publication on acceptance
- authors retain copyrights
- unique DOI for all articles
- immediate, unrestricted online access

At Auctores, research is always in progress.

Learn more <https://auctoresonline.org/journals/journal-of-surgical-case-reports-and-images>