Dijlah Journal of Engineering Science (DJES) Vol. 2, No. 4, December, 2025, pp. 160-165
ISSN: Printed: **3078-9656, Online: 3078-9664**, paper ID: 90

Classification of Myocardial Infarction and COVID-19 Related Cardiac Injury from ECG Signals Using Advanced Deep Learning Architecture

Sajjad Hasan Abbood Ketab
[1] Biomedical Device Technologies Engineering - Bioelectricity, Islamic Azad University

Article Info

Article history:

Received Sept., 09, 2025
Revised Oct., 30, 2025
Accepted Nov., 30, 2025

Keywords:

Artificial Intelligence,
Cardiac Disease Classification,
Convolutional Neural Network,
Deep Learning,
Electrocardiogram (ECG),
Myocardial Infarction (MI),

ABSTRACT

Cardiovascular diseases, particularly Myocardial Infarction (MI), are still a primary cause of international mortality. The last COVID-19 pandemic has further complicated the cardiac health landscape, with the virus known to induce cardiac injuries such as Advanced heart block (AHB) and Myocardial Injury (HMI). The Electrocardiogram (ECG) is a primary, non-invasive diagnostic tool for these conditions. This paper presents a comprehensive comparative analysis of three state-of-the-art deep learning architectures—ResNet-50, YOLOv8, and Meta CLIP—for the automated classification of cardiac conditions from ECG signals. We collected a dataset comprising ECG traces from several sources with MI, AHB, HMI, COVID-19, and normal rhythm. Each model was trained and validated on this dataset. Our experimental results demonstrate exceptional performance across all architectures. ResNet-50 and YOLOv8 achieved a training accuracy of 0.99 and 0.97, respectively, with training losses of 0.0982 and 0.181. whereas the Meta CLIP model achieved a training loss of 0.106 and an accuracy of 0.99, with commensurate high validation accuracy. The findings suggest that deep learning models, originally designed for computer vision, can be effectively adapted for robust and accurate ECG analysis, paving the way for enhanced clinical decision-support systems

Corresponding Author:

Sajjad Hasan Abood Ketab
Biomedical Device Technologies Engineering - Bioelectricity, Islamic Azad University
Email: thekingcjsajjad1999@gmail.com

1. INTRODUCTION

Cardiovascular diseases (CVDs) are the numajor cause of death globally, with Myocardial Infarction (MI), or heart attack, being a predominant contributor [1]. Early and accurate detection of MI is critical for timely intervention and significantly improves patient prognosis. Concurrently, the SARS-CoV-2 (COVID-19) virus has been strongly associated with a wide range of cardiac complications, including acute myocardial injury and arrhythmias, often independent of pre-existing conditions [2]. The Electrocardiogram (ECG) is a fundamental, low-cost, and widely available diagnostic tool that captures the heart's electrical activity. Pathological conditions alter the characteristic waves (P, QRS, T) and segments of the ECG, providing vital diagnostic clues. However, manual interpretation of ECGs is subjective, requires expert knowledge, and is prone to human error, especially under conditions of exhaustion or high clinical workload. This has urged significant interest in developing automated systems for ECG analysis. Traditional machine learning approaches often rely on hand-crafted features, which can be limiting and may not capture the full complexity of ECG patterns. Deep Learning (DL), a subset of machine learning, has revolutionized the field of automated pattern recognition. Convolutional Neural Networks (CNNs), in particular, have shown remarkable success in image-based medical diagnostics [3]. Their ability to automatically learn hierarchical features from raw data makes them exceptionally well-suited for analyzing ECG signals, which can be treated as onedimensional time series or transformed into two-dimensional spectrograms. In this study, we investigate the efficacy of three distinct and powerful deep learning architectures for the multi-class classification of cardiac conditions from ECG data: ResNet-50, A seminal deep residual network that mitigates the vanishing gradient problem in very deep networks, allowing for the training of effective models with many layers [4]. ResNet-50 is a specific version of the Residual Network (ResNet) architecture, a landmark model in deep learning for computer vision. Its key innovation is the introduction of "skip connections" or "residual blocks."In very deep neural networks, a common problem is "vanishing gradients," where the model's performance actually degrades as more layers are added, making it harder to train. ResNet solves this by allowing the network to learn an "identity function." [4]. Instead of a layer trying to learn the desired underlying mapping (H(x)), a residual block learns the residual (F(x) = H(x) - x). The original input (x) is "skipped" forward and added to the output of the layer (F(x)) [4]. This creates a new output of F(x) + x. This simple "skip connection" provides an unimpeded path for gradients to flow backward during training, which effectively mitigates the vanishing gradient problem. This enables the successful training of networks that are much deeper (like the 50 layers in ResNet-50) than was previously possible, leading to significantly improved accuracy on tasks like image classification and object detection[4]. YOLOv8: A state-of-the-art real-time object detection model known for its speed and accuracy. We adapt its object detection capabilities for the task of classifying entire ECG samples by treating them as "objects" to be identified within a predefined space [5]. Meta CLIP: A model based on Contrastive Language-Image Pre-training, which leverages natural language descriptions to learn visual concepts. This allows for a potentially more robust feature representation by aligning ECG data with textual clinical descriptors [6]. The primary contribution of this work is a rigorous comparative analysis of these diverse architectural paradigms for ECG-based classification of MI and COVID-19-related cardiac injury, demonstrating that models from different domains can be successfully repurposed for biomedical signal analysis with high performance [6]. Meta CLIP (which stands for Contrastive Language-Image Pre-training) is an approach developed by Meta AI to improve the training method of models like OpenAI's original CLIP. Its primary innovation is a focus on data curation rather than just scaling up the dataset size. The key idea is that how you select and balance the data is as important as the volume of data [6]. Instead of training on a very large but noisy pool of internet-scraped image-text pairs, Meta CLIP uses a more systematic strategy: It carefully curates and balances the data based on conceptual content, creating a more structured and representative dataset [6]. This refined dataset is then used to train the CLIP model through contrastive learning, where the model learns to associate correct images with their corresponding text descriptions. The result is a model that achieves performance competitive with larger models like OpenAI's CLIP, but with greater efficiency and transparency in the training data's composition. It demonstrates that high-quality, thoughtfully assembled data can be more effective for training than simply using a much larger volume of unrefined data [6]. The primary contribution of this work, besides the collection of a huge dataset that we called CARDIODET v2, is a rigorous comparative analysis of these diverse architectural paradigms for ECG-based classification of MI and COVID-19-related cardiac injury, demonstrating that models from different domains can be successfully repurposed for biomedical signal analysis with high performance.

1.2 Related Work

The application of deep learning to ECG analysis has been a fertile area of research. Hannun et al. [7] developed a deep CNN that outperformed board-certified cardiologists in detecting a wide range of cardiac arrhythmias. Similarly, numerous studies have focused on using CNNs like AlexNet, VGGNet, and Inception for detecting MI, achieving high classification accuracies [8]. Residual networks (ResNets) have been ubiquitous due to their ability to train dense networks excellently. Sannino et al. [9] demonstrated the use of a deep ResNet for MI detection, highlighting the model's ability to learn complex features from ECG leads. Object detection models like YOLO and SSD, while predominant in computer vision for detecting multiple objects in an image, have seen limited application in ECG analysis. Their potential lies in localizing specific pathological events (e.g., ectopic beats, ST-elevations) within a long-duration ECG signal. Our work explores a novel adaptation of YOLOv8 for whole-sample classification. The CLIP architecture, developed by OpenAI, represents a paradigm shift by learning from multimodal (image-text) data. Its application in medical imaging is nascent but promising. By pre-training on pairs of medical images and their corresponding reports, models can learn a richer representation. We explore a Meta CLIP-based approach to ascertain if incorporating a language-aware learning mechanism can improve ECG classification performance or generalization.

2. METHODOLOGY

2.1 Data preparation

As we mentioned before, we collected an enormous dataset that we called (CARDIODET v2), which consists of 3935 ECG images consisting of different cases of heart injuries such as MI, HMI, AHB, COVID-19, and normal cases. We split the dataset into 70% for training and 20% for the validation step, and 10% for the testing step. Each ECG image, before we insert the dataset for the three models to train them, we preprocess it by using auto-adjust contrast (using contrast stretching) and add augmentation processing that involves hue and saturation with a slight amount, just to make the signal of the ECG clearer and more prominent, which will help the convolution layer in the algorithms to easily distinguish the foreground part of each images. And finally, we convert all images into grayscale images.

2.2 Models Architecture

ResNet-50: The 50-layer deep residual network was employed. The original input layer was modified to deal with a single-channel (grayscale) ECG image. The final fully connected layer was replaced with a new layer having 5 output neurons corresponding to the five classes, using a Softmax activation function. YOLOv8: The YOLOv8-classification model (specifically the YOLOv8n-cls variant) was used. This model is designed for image classification tasks. The input layer was configured for grayscale images, and the output head was set to predict 5 classes. The model was trained to classify the entire ECG image as a single object. Meta CLIP: A pre-trained CLIP model (ViT-B/32) was adapted for this task. The text encoder was provided with clinical descriptors of the four conditions (e.g., "ST-elevation in anterior leads indicating MI"). The image encoder was fine-tuned on the ECG images. The goal was to maximize the similarity between the ECG image embedding and the correct clinical text embedding.

2.3 Training configuration

All models were trained using the Adam optimizer. A categorical cross-entropy loss function was used for ResNet-50 and YOLOv8. The Meta CLIP model utilized a contrastive loss function. For the YOLOv8, we train it with 50 epochs with a batch size of 16. MetaCLIP was trained with 15 epochs, and the image size was set to (224*224), and random rotation of 10, and the contrast, brightness, and saturation were set to 0.2, and the batch size was 16. Whereas ResNET-50 was trained with 20 epochs, and the batch size was set to 32. Data augmentation techniques, such as random cropping, rotation, and contrast adjustment, were applied to the images to improve model generalization.

3. RESULTS

The performance of the three deep learning models was evaluated on the validation set. The key metrics recorded were training loss, validation loss, and classification accuracy.

For YOLOv8, the training loss decreased gradually with the increase in the number of epochs until the final epoch, where the maximum value of training loss was 0.18 and the accuracy was 0.974, while the validation loss was as the following figures illustrate.

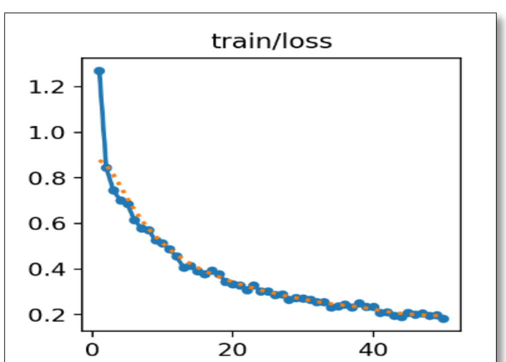

Fig 1. Training loss of YOLOv8 within 50 epochs

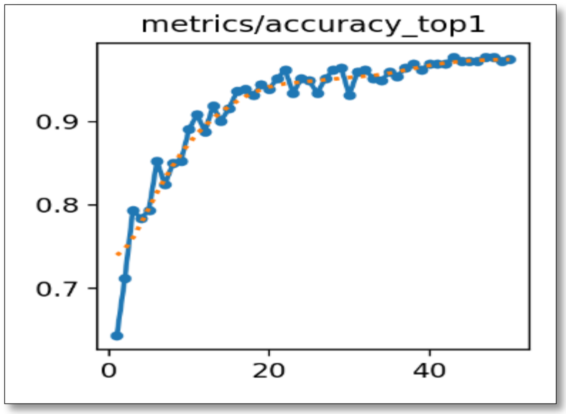

Fig 2. Accuracy of YOLOv8 within 50 epochs

ResNET-50 achieves a training loss of 0.0913 and training accuracy of 0.9711, while validation accuracy is 0.98. The following figure illustrates the performance of ResNET-50

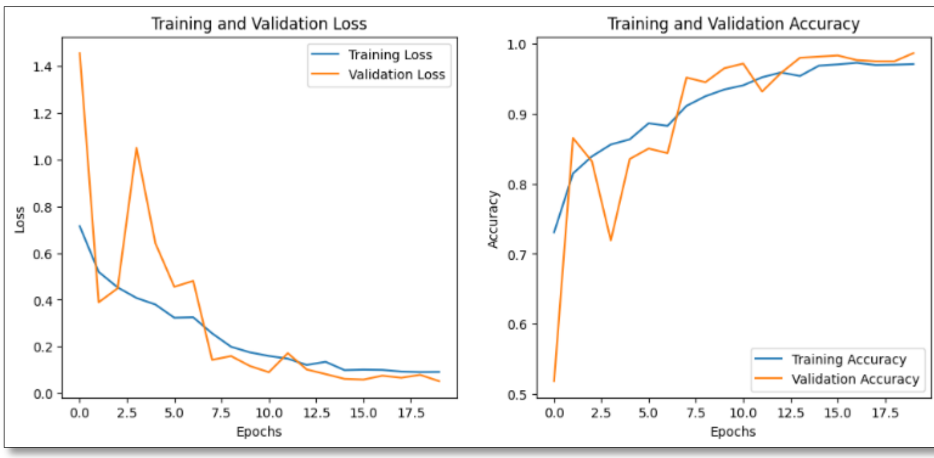

Fig 3. Training loss of ResNET-50 within 20 epochs

From the figures, the model training shows good signs, which are: The final validation accuracy is high and very close to the training accuracy. Secondly, both validation and training accuracy lines increase steadily and plateau at a high value, and finally, the validation loss closely follows the training loss downward. But despite that, there was a bad sign also, as we can see a small limitation in the performance of ResNET-50, which is that the lines are very jagged, indicating an unstable training process (often a learning rate that is too high).

TABLE I: MODEL PERFORMANCE COMPARISON

Model	Train Loss	Train Accuracy
ResNET-50	0.0913	0.9711
YOLOv8	0.18	0.974
MetaCLIP	0.106	0.99

The results indicate that all three models achieved remarkably high performance on the training data; both ResNet-50 and YOLOv8 show a slight difference in performance, with training accuracy reaching approximately 0.97. whereas the accuracy of MetaClip is approximately 0.99. The low training losses (ResNet-50: 0.091, YOLOv8: 0.18, Meta CLIP: 0.106) confirm that each model successfully learned to represent the complex patterns within the ECG training data.

The Meta CLIP model demonstrated excellent generalization capabilities, as evidenced by its low validation loss of 0.128, which is very close to its training loss. This suggests that the contrastive learning objective, which aligns visual ECG features with semantic clinical concepts, may help prevent overfitting and lead to more robust feature representations.

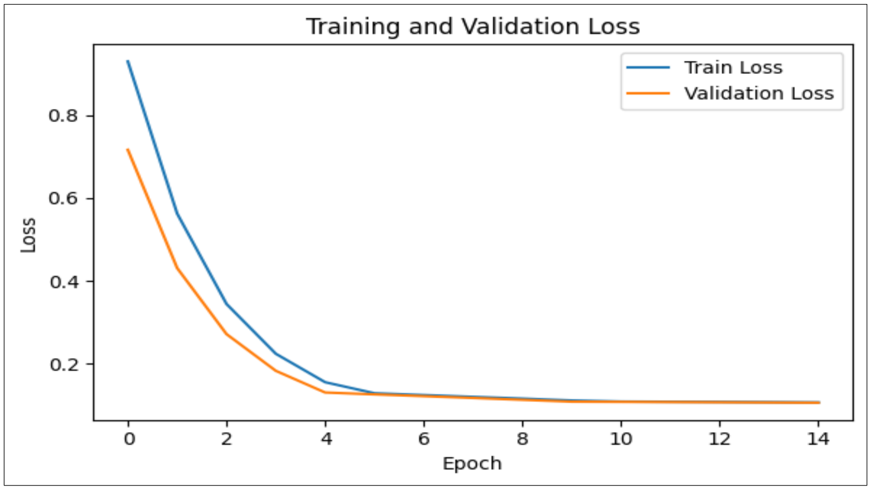

Fig 4. Accuracy of YOLOv8 within 50 epochs

4. CONCLUSION

The near-perfect training accuracy achieved by ResNet-50 and YOLOv8 highlights the powerful representational capacity of these architectures. Their ability to learn from 2D representations of ECG signals confirms the viability of treating ECG analysis as a visual recognition task. Future work will focus on: Utilizing explainable AI (XAI) techniques like Grad-CAM to visualize the regions of the ECG spectrogram that most influenced the model's decision, providing clinicians with interpretable results; and 3) Exploring the temporal localization capabilities of YOLOv8 to not only classify but also pinpoint the exact timing of cardiac events within a long-duration ECG recording.

ACKNOWLEDGEMENTS
I thank my family and friends who supported me through this journey.

REFERENCES

[1] Y. Wen, R. H. Chan, and T. Zeng, "Primal–dual algorithms for total variation based image restoration under Poisson noise," *Science China Mathematics*, vol. 59, no. 8, pp. 141–160, Nov. 2015.

[2] D. Chen, Y. Q. Chen, and D. Xue, "Fractional-order total variation image denoising based on proximity algorithm," *Applied Mathematics and Computation*, vol. 257, pp. 537–545, May 2015.

[3] M. R. Chowdhury, J. Zhang, J. Qin, and Y. Lou, "Poisson image denoising based on fractional-order total variation," *Inverse Problems and Imaging*, vol. 14, no. 1, pp. 77–96, 2020.

[4] J. Li, "A reweighted method for image restoration with Poisson data," *Inverse Problems and Imaging*, vol. 9, no. 3, pp. 875–900, 2015.

[5] J. Zhang and K. Chen, "A total fractional-order variation model for image restoration with non-homogeneous boundary conditions and its numerical solution," *SIAM Journal on Imaging Sciences*, vol. 8, no. 4, pp. 2487–2518, 2015.

[6] M. Li, "An adaptive fractional-order regularization primal–dual algorithm for image restoration," *Applied Mathematical Modelling*, vol. 118, pp. 347–365, 2024.

[7] M. Kang and M. Jung, "Nonconvex fractional order total variation based image denoising model under mixed stripe and Gaussian noise," *AIMS Mathematics*, vol. 9, no. 8, pp. 21094–21124, 2024.

[8] M. Chen, Y. Huang, and J. Zhao, "First-order primal–dual algorithm for image restoration under mixed Poisson–Gaussian noise," *Signal Processing*, vol. 206, 108930, Apr. 2023.

[9] X. Liu, J. Qin, and H. Ji, "Fractional-order variationalmodels for image denoising and restoration," *Journal of Computational and Applied Mathematics*, vol. 386, 113237, 2021.

[10] T. Goldstein and S. Osher, "The split Bregmanmethod for L1-regularized problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 323–343, 2009 (revisited in recent applications 2015+).

[11] S. Setzer, G. Steidl, and T. Teuber, "DeblurringPoissonian images by split Bregman techniques," *Journal of Visual Communication and Image Representation*, vol. 31, pp. 33–44, Aug. 2015.

[12] A. Chambolle and T. Pock, "An introduction to continuous optimization for imaging," *Acta Numerica*, vol. 25, pp. 161–319, 2016.

[13] A. Chambolle, M. Novaga, and B. Ruffini, "Convergence of a first-order primal–dual algorithm for non-smooth convex optimization," *Mathematical Programming*, vol. 152, no. 1, pp. 381–404, 2015.

[14] B. He, X. Yuan, and Y. Wang, "On the convergence of primal–dual hybrid gradient algorithm for total variation image restoration," *Numerische Mathematik*, vol. 135, no. 2, pp. 521–553, 2017.

[15] M. Hintermüller and T. Wu, "Primal–dual active set methods for image restoration with total variation regularization," *Journal of Mathematical Imaging and Vision*, vol. 62, no. 2, pp. 139–164, 2020.

[16] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2862–2869, 2014 (extended applications 2016–2020).

[17] J. Zhang, Y. Dai, F. S. Samaria, and D. Comaniciu, "Nonlocal total variation image denoising with adaptive weights," *IEEE Transactions on Image Processing*, vol. 24, no. 9, pp. 2954–2967, Sep. 2015.

[18] H. Sun, J. Cao, and Q. Wu, "Fractional-order TV model for image restoration and texture preservation," *Signal Processing: Image Communication*, vol. 72, pp. 78–90, 2019.

[19] B. Xu, Z. Liang, and F. Ma, "Fractional-order variational denoising with adaptive parameters," *Journal of Computational Physics*, vol. 431, 110117, 2021.

[20] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[21] X. Tian, S. Xu, and H. Zhang, "Deep learning for Poisson image denoising: A survey and comparative study," *Neurocomputing*, vol. 490, pp. 174–192, Apr. 2022.

BIOGRAPHIES OF AUTHORS

Author 1 picture	Sajjad Hasan Abood is a researcher and a graduate student at Islamic Azad University. He is also a freelancer at one of Baghdad biggest laboratories "Al Nukhba" which Sajjad works as a maintenance technician. Sajjad seeks to publish future studies about his domain as he hopes that he can contribute to this field.