Implementing Deep Learning for Atrial Fibrillation Diagnosis

Justine Mae S. Buenaventura [1], Carl Louise D. Rostata [2,*], Angel Ezekiel O. Anzures [3], Dr. Gracita O. Topacio [4], Dr. Arvin R. Yumul [5], Engr. Mark Montances [6]
[1,2,3] College of Engineering, Architecture and Technology, De La Salle University – Dasmariñas, Cavite, Philippines
[4,5] Internal Medicine, ManilaMed – Medical Center Manila, Manila, Philippines
[6] Department of Electronics Engineering, De La Salle University – Dasmariñas, Cavite, Philippines
Email: [1] bjs1388@dlsud.edu.ph, [2] rcd0891@dlsud.edu.ph, [3] aao0986@dlsud.edu.ph, [6] momontances@dlsud.edu.ph
*Corresponding Author

***Abstract*—Atrial fibrillation (AF) is a critical yet often underdiagnosed cardiac arrhythmia, particularly in low-resource environments where digital ECG systems and cardiology specialists may be inaccessible. This study presents a deep learning-based diagnostic tool designed to classify AF and normal sinus rhythm (NSR) from real-world, clinic-verified ECG images. Leveraging EfficientNet-B0, a convolutional neural network optimized for efficiency and accuracy, the model was trained on preprocessed images of 12-lead ECG printouts captured via mobile phone. Preprocessing included grayscale conversion, denoising, sharpening, and resizing, ensuring consistent input quality across variable capture conditions. The model achieved a 93% accuracy, balanced F1-scores of 0.93 for both classes, and an ROC-AUC of 0.98 on a validation dataset. Beyond validation, the model was integrated into a web-based interface and tested on newly acquired ECG images under varied real-world conditions, maintaining strong classification performance with cautious predictions. By combining accessible imaging workflows with deep learning inference, this system offers strong potential to serve as an accessible diagnostic aid in under-resourced healthcare environments.**

Keywords—atrial fibrillation, ECG, deep learning, image classification, EfficientNet, web development

I. INTRODUCTION

Atrial fibrillation (AF) is the most prevalent type of sustained cardiac arrhythmia, known to increase the risk of stroke, heart failure, and other life-threatening complications when left undetected. Despite its clinical significance, AF often remains undiagnosed, especially in community-level clinics and remote areas lacking access to advanced diagnostic tools or trained cardiologists. Conventional AF detection relies on digital ECG data interpreted by specialists or signal-based algorithms—resources that may be impractical or unavailable in many primary care settings.

This study addresses that gap by introducing an image-based deep learning solution capable of diagnosing AF from photographs of printed 12-lead ECG strips. These images, often captured via mobile phones, reflect the realities of front-line healthcare: inconsistent lighting, varied paper quality, and limited equipment. Instead of depending on raw digital signal inputs, this approach uses convolutional neural networks (CNNs) trained on preprocessed ECG images, offering a more accessible and cost-effective alternative.

The developed system builds upon the EfficientNet-B0 architecture, chosen for its strong balance of accuracy and computational efficiency. It was trained using real ECG data collected from a clinical setting and integrated into a web-based interface designed for real-time, point-of-care use. This paper outlines the full process—from dataset preparation and model development to deployment and real-world evaluation—demonstrating how deep learning can be practically applied to support AF screening in under-resourced environments.

II. METHODOLOGY

A. Dataset Acquisition and Preprocessing

The dataset used in this study was collected from ECG printouts sourced at Dr. Topacio's clinic in Manila, Philippines. Only ECGs classified as either atrial fibrillation (AF) or normal sinus rhythm (NSR) were included, and each was manually validated by a medical expert to ensure diagnostic accuracy. Three datasets were constructed in this study, each incrementally derived from the preceding one.

1. Dataset 1 – Baseline Synthetic Dataset: Comprised of 73 original ECG images per class, augmented to 365 per class for a total of 876 samples (438 AF, 438 SR). No preprocessing was applied aside from padding to square dimensions to preserve waveform proportions. All images were resized to 224×224 pixels for EfficientNet-B0 input.
2. Dataset 2 – Real-World ECG Images (Raw): Included 121 original images per class, augmented to 1936 samples. Images were captured from ECG printouts using phone cameras or scanners, manually cropped to isolate the waveform region. No preprocessing was applied, leaving noise and variability such as rotation, uneven lighting, grid overlays, paper folds, and camera artifacts.
3. Dataset 3 – Preprocessed Real-World ECG Images: Built on Dataset 2 with the same 121 original images per class, augmented to 1930 samples. A

Received: 22-8- 2025
Revised: 15-9-2025
Published: 30-6-2026

preprocessing pipeline was applied, consisting of grayscale conversion, fastNlMeansDenoising, morphological erosion, padding, sharpening, and resizing to 224×224 pixels.

Given the variability introduced by photographic capture—such as shadows, folds in the paper, inconsistent angles, and image noise—a preprocessing pipeline was designed to standardize the dataset while preserving relevant waveform information. Each image was first converted to grayscale to eliminate color inconsistencies and reduce computational complexity. Denoising using Fast Nl Means Denoising from OpenCV was used to remove visual noise without distorting key signal features. Sharpening were then used to enhance the contrast and visibility of ECG waveforms, making important components like P-waves and R-peaks more distinguishable.

To further emphasize the structure of the waveforms, morphological operations such as erosion were employed. These helped reinforce the line quality of the traces while suppressing background text and markings commonly found in ECG paper reports. Images were then padded to maintain a square aspect ratio and resized to 224×224 pixel dimensions aligned with the input requirements of the EfficientNet-B0 model architecture.

This preprocessing strategy ensured that despite inconsistencies in capture conditions, each ECG image fed into the model had consistent spatial formatting and enhanced waveform clarity. By bridging the gap between uncontrolled real-world image capture and deep learning requirements, this pipeline enabled the model to generalize well and maintain high classification performance across varied input conditions.

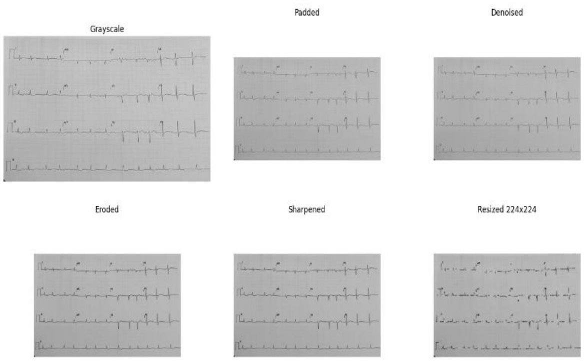

Fig. 1. ECG Image

B. *Model Development and Training*

1. Initial Training with K-Fold Cross-Validation

To establish a baseline, EfficientNet-B0 was trained using 5-fold cross-validation. Each fold consisted of approximately 80% training and 20% validation data. This method was intended for robust performance estimation. However, an error in methodology occurred when the best-performing fold model was repurposed for deployment rather than training a model on the complete dataset. Since fold-specific models are trained on only a subset of the data, this resulted in limited generalization capacity.

2. Full Dataset Model Training

To correct this limitation, a new model (efficientnet_b0_model_full) was trained on the entire dataset 1 (876 ECG images). The model utilized ImageNet-pretrained weights, with a two-class classifier head for Atrial Fibrillation (AF) and Normal Sinus Rhythm (NSR). Data augmentation included resizing, rotations, affine transformations, and color jittering. Training employed Adam (lr=1e-4), batch size 32, and Cross-Entropy Loss. Performance was moderate (Accuracy ≈ 82%), indicating that further refinements were required.

3. Retraining with Frozen Blocks

Building on the Fold 3 pipeline, retraining unfroze EfficientNet-B0 Blocks 4–7 to improve ECG-specific feature adaptation. Focal Loss (α=0.35, γ=2) was introduced to focus learning on harder samples. Dataset 3 (denoised and preprocessed ECGs) was used with advanced augmentations, including noise and shadow simulation. This improved model generalization with balanced performance across classes (Accuracy ≈ 82%, F1 ≈ 0.82).

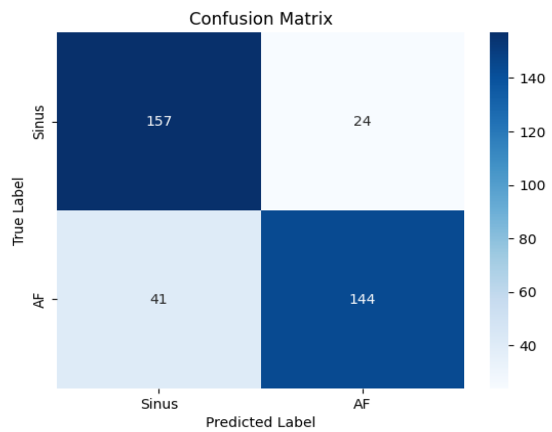

Fig. 2. ECG Data Set

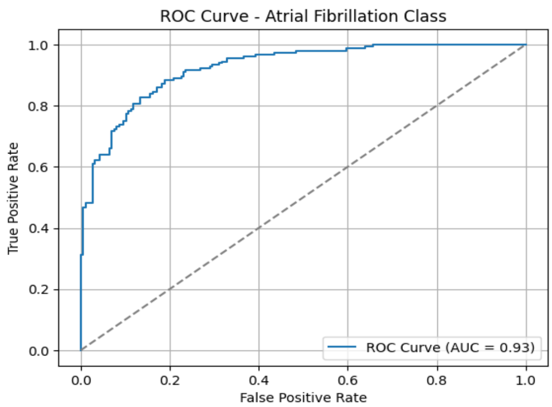

Fig. 3. Atrial Fibrillation Class (1)

4. Fine-Tuning Full Dataset Model

To further enhance sensitivity, especially for Normal Sinus Rhythm, Blocks 3–7 were unfrozen. A layer-wise learning rate strategy was applied: 5e-6 for the backbone, 1e-4 for the classifier. Macro-F1 was used as the early stopping metric to prevent class imbalance. This refinement improved accuracy to 89% and macro-F1 to 0.89, with ROC-AUC (AF) = 0.96, indicating stronger clinical applicability.

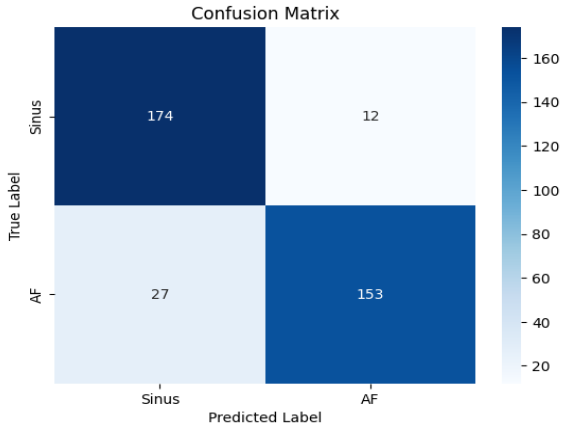

Fig. 4. A layer-wise learning rate strategy

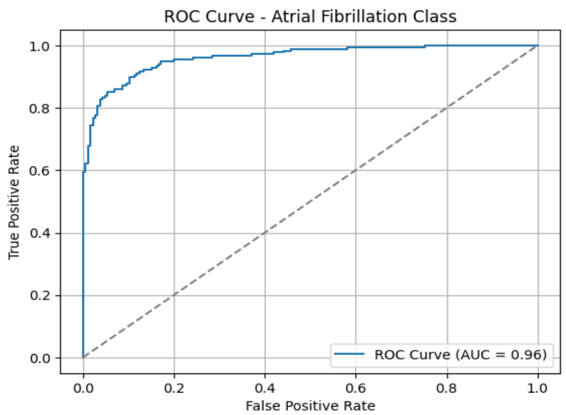

Fig. 5. Atrial Fibrillation Class (2)

5. Final Fine-Tuning with Label Smoothing

In the final iteration, Cross Entropy Loss with label smoothing (0.1) replaced Focal Loss to calibrate confidence scores. Training resumed from the best-performing checkpoint with lr=1e-6. This yielded the strongest performance: 93% accuracy, 0.93 macro-F1, and ROC-AUC (AF) = 0.98. Misclassifications were conservative (AF → NSR), which aligns better with screening scenarios where false positives are safer than false negatives.

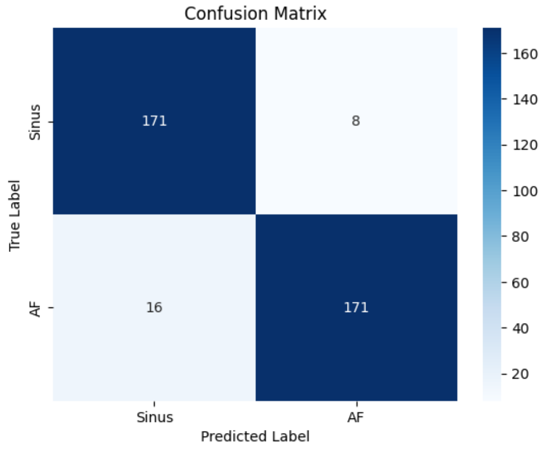

Fig. 6. Cross-Entropy Loss with label smoothing

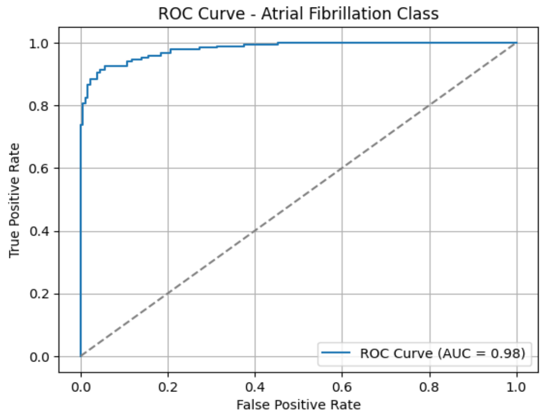

Fig. 7. Atrial Fibrillation Class (3)

6. Summary of Model Training Progression

Table 1. Summary of Model Training Progression

Training Stage	Loss Function	Unfrozen Blocks	Accuracy	ROC-AUC
K fold (Fold 3)	Cross Entropy	Head only	~85% (best fold)	0.93
Full Dataset (Baseline)	Cross Entropy	Head only	82%	0.91
Retraining	Focal Loss	4-7	82%	0.93
Fine-Tuning	Focal Loss	3-7	89%	0.96
Final Fine-Tuning	Cross Entropy + Label Smoothing	3-7	93%	0.98

C. System Integration and Development

The diagnostic platform was developed as a full-stack web system using Laravel and deployed on a LEMP (Linux, Nginx, MySQL, PHP) server, designed to function efficiently in real-world clinical workflows. The front end was built to be mobile-responsive and intuitive, allowing users to upload ECG images directly through file input or live capture via device camera. The layout prioritized simplicity while offering a detailed display of diagnostic results.

Upon image submission, the Laravel backend handles file validation—ensuring correct formats, resolution, and safe size limits—before invoking a Python-based inference pipeline using shell execution. The backend communicates with the trained EfficientNet-B0 model, which processes the image through the same preprocessing steps used during training, ensuring consistency.

Inference results include the predicted class (AF or NSR), confidence score, and a Grad-CAM-based heatmap for interpretability. These outputs are returned in real-time and rendered on a custom results page showing the input ECG, diagnosis summary, and corresponding visualization overlays.

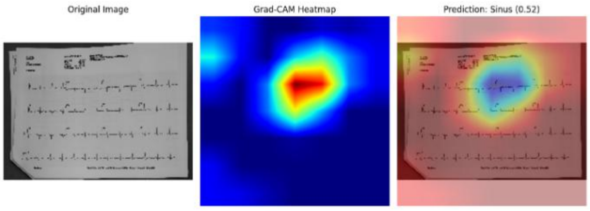

Fig. 8. Example Output with Grad-CAM overlay

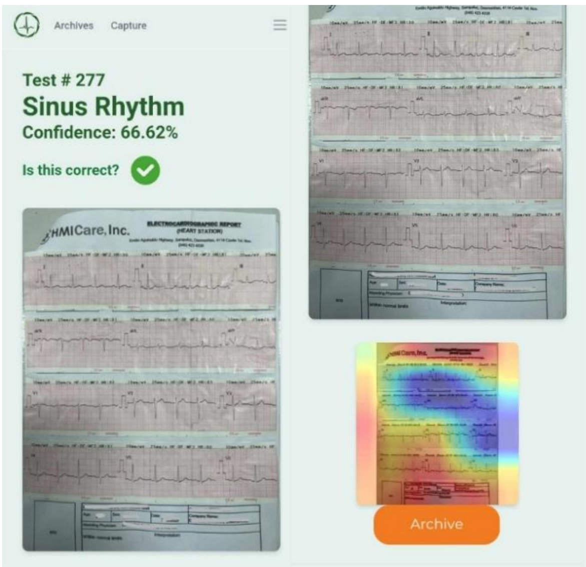

Fig. 9. Web Application Result from NSR unseen data

The system includes an admin panel with user authentication and activity logs for tracking uploaded files and test history. During testing and deployment, attention was also given to error handling and network latency, ensuring a smooth experience from capture to result. This system was optimized not only for technical performance but also for clinical practicality.

III. Results

A. Performance Evaluation

The final EfficientNet-B0 model underwent comprehensive evaluation using a validation dataset (20% of Dataset 3) composed of preprocessed ECG images unseen during training. The model achieved an overall accuracy of 93%, with an F1-score of 0.93 for both AF and NSR classes, demonstrating its balanced sensitivity and specificity. The area under the receiver operating characteristic curve (AUC) reached 0.98, indicating the model's strong discriminative capability between the two classes.

Additional performance metrics, including precision and recall, were closely aligned across both categories, supporting the conclusion that the model was neither overfitting nor favoring one class over another. A confusion matrix generated from validation results further revealed a low number of false positives and false negatives, solidifying the model's reliability in capturing the subtle variations in ECG waveform patterns associated with AF and NSR.

B. Preliminary Real-World Testing Through AFsense

To explore practical deployment feasibility, the final fine-tuned model was tested using 30 clinically validated ECG images (15 AF, 15 Normal Sinus), captured under three real-world conditions: (1) recommended setup (ideal distance and lighting), (2) too close with shadows, and (3) too far. These images were independent of the training and validation datasets and were validated by a licensed cardiologist.

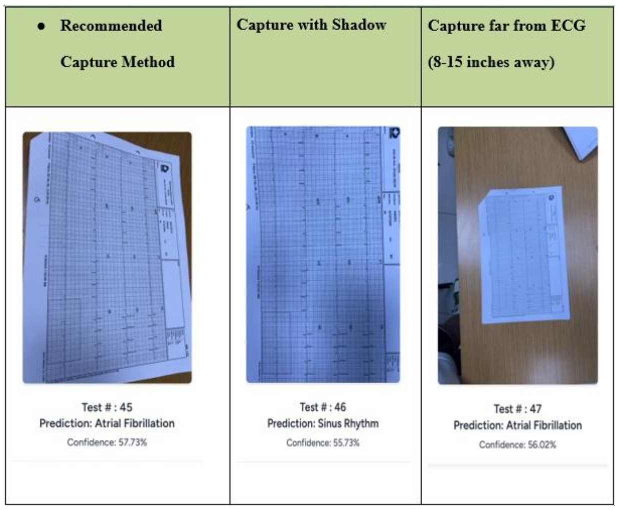

Fig. 10. Various capture methods for ECG

Under the recommended setup, the model correctly classified 11/15 AF cases and 15/15 NSR cases, though with modest confidence scores (50–70%) due to label smoothing. Performance degraded under non-ideal conditions, particularly when shadows obscured waveforms or when images were captured at excessive distances.

It is important to note that this evaluation is only a preliminary pilot test, not a robust validation. The sample size of 30 images, drawn from a single clinic and environment, is insufficient to support broad generalizations. Instead, these findings serve as a proof-of-concept case study, demonstrating both the feasibility and limitations of real-world image capture for atrial fibrillation detection.

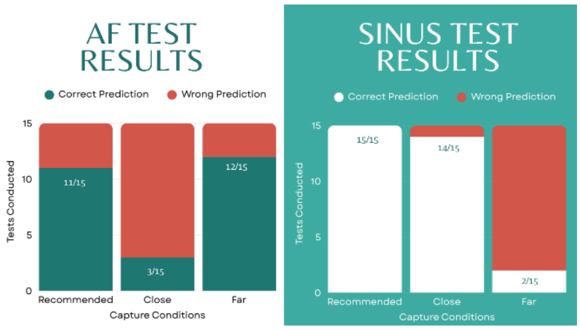

Fig. 11. AF Test and SINUS Test Results

A larger, multi-center dataset with more diverse acquisition conditions will be required to establish statistically reliable clinical performance.

IV. Discussion

This study demonstrates that atrial fibrillation can be reliably classified from photographs of printed 12-lead ECG strips using a lightweight CNN model. By training EfficientNet-B0 on progressively refined datasets, we showed that preprocessing techniques—including denoising, morphological operations, and sharpening—play a critical role in enhancing waveform clarity and improving classification accuracy. The consistent performance across varied image conditions highlights the practicality of this image-based approach for deployment in settings where digital ECG signals or cardiologist expertise may be unavailable.

Integration into a web-based interface further strengthened accessibility and usability, offering features such as Grad-CAM overlays that provide interpretability and support clinical decision-making. Label smoothing and calibration strategies contributed to more cautious, reliable outputs, aligning well with real-world diagnostic needs. While the model occasionally defaulted conservatively toward Normal Sinus Rhythm in ambiguous cases, this behavior may be clinically acceptable in screening contexts and could be further refined through uncertainty estimation or repeat-capture recommendations.

Nonetheless, several limitations remain. The datasets were relatively small and sourced from a single clinical site, which constrains generalizability. The initial reliance on K-Fold cross-validation for deployment underscored methodological challenges that were later corrected through full-dataset training and fine-tuning. Real-world testing on new ECGs, though encouraging, was limited in scale and should be regarded as a preliminary pilot rather than robust validation.

Taken together, these findings underscore both the promise and the challenges of image-based AF screening. With broader datasets, cross-clinic collaborations, and further architectural exploration, this approach has strong potential to serve as an accessible diagnostic aid in under-resourced healthcare environments.

V. Conclusion

This study demonstrated that deep learning, specifically the EfficientNet-B0 architecture, can effectively classify Atrial Fibrillation (AF) and Normal Sinus Rhythm (SR) from 12-lead ECG photographs. Using a cardiologist-verified dataset, the model achieved strong diagnostic performance with 93% accuracy, balanced F1-scores of 0.93, and an ROC-AUC of 0.98. Progressive refinements—including Focal Loss, block-wise fine-tuning, and preprocessing—addressed class imbalance and improved recognition of NSR cases, though image quality and capture conditions remained influential factors.

To enable real-world deployment, the system was operationalized through AF Sense, a Laravel-based web platform integrated with Python inference. Real-time testing confirmed high accuracy under proper capture conditions, while also revealing performance degradation in poorly lit or misaligned images. These results highlight both the feasibility and limitations of using deep learning for automated AF detection in practical settings.

While the datasets were limited in size and sourced from a single clinical site, the results highlight the strong potential of image-based approaches for arrhythmia screening. Expanding datasets, incorporating additional clinical partners, and exploring advanced architecture will be important next steps to ensure broader generalizability. With these improvements, the proposed system could serve as a practical, low-cost diagnostic aid in under-resourced environments, supporting earlier AF detection and improved patient outcomes.

Acknowledgment

The authors would like to express their sincere appreciation to De La Salle University – Dasmariñas for cultivating an academic environment that fosters innovation, research, and collaboration. We are also deeply grateful to Engr. Mark Montances, Dr. Gracita O. Topacio, and Dr. Arvin R. Yumul whose dedicated mentorship, expert guidance, and unwavering support were vital throughout the course of this project. Their encouragement and insight helped shape this work into a meaningful contribution to the field of AI-assisted healthcare.

References

[1] Y. Baek et al., "A deep learning algorithm for detecting AF during sinus rhythm," Scientific Reports, vol. 11, 12818, 2021.
[2] A. Król-Józaga, "Visual transformation of ECGs for CNN classification," Computing in Cardiology, vol. 49, 2022.
[3] R. Ribeiro et al., "Automatic diagnosis of ECGs at scale," Nature Communications, vol. 11, no. 1, p. 1760, 2020.
[4] J.W. Hughes et al., "Performance of CNNs versus physicians in ECG interpretation," JAMA Cardiology, vol. 6, no. 11, pp. 1285–1293, 2021.
[5] R. Ao and G. He, "Image-based deep learning for ECG interpretation," Frontiers in Artificial Intelligence, vol. 5, 2023.
[6] D. Patwari et al., "A Ramanujan-based CNN for AF detection," IEEE Journal of Biomedical and Health Informatics, 2024.

[7] J. Lin et al., "AI-assisted ECG analysis for AF diagnosis," IEEE Access, vol. 9, pp. 98765–98773, 2021.
[8] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Mach. Learn., 2019, pp. 6105–6114.
[9] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. Int. Conf. Learn. Represent., 2015.
[10] C. Szegedy et al., "Rethinking the inception architecture for computer vision," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2016, pp. 2818–2826.
[11] G. Li et al., "Classification of ECG signals based on 1D convolutional neural network and attention mechanism," IEEE Access, vol. 9, pp. 34349–34357, 2021.
[12] F. S. Yen et al., "Application of convolutional neural network in the diagnosis of arrhythmia based on ECG signals," Sensors, vol. 20, no. 18, p. 5219, 2020.
[13] M. Abedi et al., "Interpretation of 12-lead ECGs using deep convolutional neural networks," Journal of Electrocardiology, vol. 58, pp. 66–70, 2020.
[14] Z. Zhang et al., "ECG image classification using deep CNNs with enhanced spatial resolution," Biomedical Signal Processing and Control, vol. 72, 2022.
[15] B. Yildirim, "A novel wavelet sequence-based on deep bidirectional LSTM network model for ECG signal classification," Computers in Biology and Medicine, vol. 96, pp. 189–202, 2018.
[16] J. Hannun et al., "Cardiologist-level arrhythmia detection and classification in ambulatory ECG using a deep neural network," Nature Medicine, vol. 25, pp. 65–69, 2019.
[17] A. M. Ayyad and A. B. Nassif, "A deep learning framework for detecting arrhythmia using 2D ECG images," IEEE Access, vol. 10, pp. 87438–87450, 2022.
[18] Y. LeCun et al., "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.
[19] J. Esteva et al., "A guide to deep learning in healthcare," Nature Medicine, vol. 25, no. 1, pp. 24–29, 2019.