This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3566531

Enhancing ECG Classification in Cardiac Diagnostics: A Novel Approach Using Adaptive Focal Cross-Entropy Loss Function

Happy Nkanta Monday†, Grace Ugochi Nneji†*, Md Altab Hossin, Kelvin Davies Mark, Edwin Sunday Umana, Goodness Temofe Mgbejime, and Jianping Li

***Abstract*— Heart disease is the leading cause of mortality globally. Electrocardiograms (ECGs) are standard instruments for the examination of heart conditions, but traditional analysis is time-consuming and prone to errors. Novel advances in artificial intelligence have improved ECG classification. However, some limitations remain, such as poor interpretability, computational cost, and class imbalance. This study proposes a novel deep learning algorithm based on Depthwise Separable Residual Attention called DRA-ECG and a customized Adaptive Focal Cross-Entropy (AFCE) loss function for cardiac condition classification. This proposed methodology leverages the Continuous Wavelet Transform (CWT) method to transform 1D raw ECG signals into 2D scalograms to enhance feature representation and training. The proposed customized AFCE loss function incorporated into the DRA-ECG model addresses the class imbalance problem and boost the performance of the model. More so, this study incorporates edge feature detection as a preprocessing technique to denoise and enhance the trainable features of the 2D scalograms for optimal feature representation. The proposed DRA-ECG model achieves a high accuracy of 98.17%, recall of 95.78%, F1-score of 95.82%, and precision of 95.89%. This study shows that the results achieved by the proposed DRA-ECG surpass the current state-of-the-art and existing research works, concerning classification performance and generalization ability in ECG classification, which underlines the effectiveness of the novel AFCE loss function for ensuring high-classification accuracy and robustness. The proposed novel methodology enhances heart disease classification and provides a robust and reliable solution for medical diagnosis, addressing the major drawbacks of existing models.**

***Index Terms*— Adaptive focal cross-entropy loss, cardiovascular attack, continuous wavelet transform, depthwise separable, attention mechanism, ECG classification**

I. Introduction

HEART disease is one of the most common diseases in the world with a high mortality rate [1]. Cardiovascular diseases (CVDs) are the leading cause of fatal and non-fatal diseases worldwide, and according to a report by the World Health Organization (WHO), they account for around 31% of global deaths, representing about 17.9 million persons who die annually [2]. Heart disease stands out as the most critical of them, which highlights a great demand for accurate and early diagnosis to enhance patient outcomes and reduce clinical costs [3]. Electrocardiography is the most widely used diagnostic test for heart abnormalities. Electrocardiography is an investigative device that measures the electrical activity of the heart, very important key information about the function, structure, and rhythm of the heart [4]. However, electrocardiography analysis remains a time-consuming process requiring human experts to read and analyze the recordings manually. It is not only time-consuming but also error-prone in cases of subtle heart disease. With the application of machine learning and deep learning technologies, automated electrocardiogram (ECG) analysis has become increasingly popular [5].

Deep learning algorithms employ multiple layers of convolutional neural networks to learn complex features and patterns from the input data fed into them [6]. However, most literature focuses only on ECG signals, disregarding the potential contributions of higher-dimensional feature representation of ECG. ECG signals contain temporal information concerning heart activity, while ECG images provide spatial insights that can be vital for identifying structural abnormalities [7]. In medical diagnostics, 2D feature learning can improve the understanding of complex diseases by combining information from different sources, such as clinical data, and genetic information. Most studies used neural networks (NNs) framework for ECG classification based on the PTB-XL dataset, which concentrates on ECG signals [8]. This ensures that the users and domain professionals can examine the important

This work was supported by the National Natural Science Foundation of China, the National High Technology Research and Development Program of China(Grant No. 2021YFG0322), the project of Science and Technology Department of Chongqing Municipality, the Science and Technology Research Program of Chongqing Municipal Education Commission (Grant No. KJZD-K202114401)

Happy Nkanta Monday (e-mail: happy.monday@zy.cdut.edu.cn) and Grace Ugochi Nneji (e-mail: grace.nneji@zy.cdut.edu.cn) are with the School of International Education, Chengdu University of Technology Oxford Brookes College, Chengdu, 610059 China, and Intelligent Computing Lab, HACE SOFTTECH, Lagos 102241, Nigeria, Md Altab Hossin (e-mail: altabbd@cdu.edu.cn) is with the School of Innovation and Entrepreneurship, Chengdu University, Chengdu 610106, China, Kelvin Davies Mark (e-mail: davies-markkpbh2023@futa.edu.cn) and Edwin Sunday Umana (e-mail: umanamts188145@futa.edu.cn) are with the Federal University of Technology, Akure, P.M.B 704 Nigeria, Goodness Temofe Mgbejime (e-mail: 2019080146015@std.uestc.edu.cn) and Jianping Li (e-mail: jpli2222@uestc.edu.cn) are with University of Electronic Science and Technology of China, Chengdu, 611731 China

†Co-first authors

*Corresponding author

features learned by the NN for its task classification. While machine learning and deep learning have revolutionized ECG classification by automating diagnosis, existing models face three key limitations:

- Most existing models focus on 1D signal analysis, overlooking the spatial and spectral features that could be extracted through 2D representations like scalograms.
- Many models do not incorporate attention-based architectures, leading to suboptimal feature extraction and reduced interpretability.
- Common loss functions such as Categorical Cross-Entropy (CCE) often bias toward majority classes, leading to misclassification of minority conditions, which are often clinically critical.

A. Research Gap

Although some models have introduced attention or explored 2D transformations, few have combined time-frequency analysis, edge enhancement preprocessing, and a novel loss function specifically tailored for class imbalance in ECG classification. There is a pressing need for an interpretable, computationally efficient, and clinically robust model that can generalize across multiple cardiac conditions.

B. Proposed Solution

To address these issues, this study proposes a new methodology that involves several significant advancements. a) In this study, we transform the 1D ECG signals into a matrix textural image known as scalograms through Continuous Wavelet Transform (CWT). This transformation allows the model to learn spatial and spectral features, increasing its ability to discriminate subtle patterns. b) This study proposes a novel depthwise separable residual convolutional network (DSC) with attention mechanisms for better feature extraction and minimizing the computational cost. The attention mechanism helps the model capture the most vital and distinct features in the ECG scalograms, improving accuracy and interpretability. c) Addressing the problem of class imbalance, a novel adaptive focal cross-entropy (AFCE) loss function is proposed. This novel loss function dynamically balances the impact of categorical cross-entropy, requiring the model to be sensitive to the minority classes and with higher performance. These contributions lead to a model that not only achieves high classification accuracy but also improves sensitivity and interpretability, making it suitable for real-world clinical applications, including real-time monitoring and wearable devices. The significant contributions of this paper are as follows:

a) Implementation of a novel depthwise separable residual network with attention mechanisms that utilizes bottleneck technique to achieve effective performance while demanding less computational cost comparable to existing models.

b) This study employs five different datasets from the PhysioBank ATM public dataset to perform heart disease classification. The dataset encompasses various data distributions that accurately represent real-life situations.

c) Implementation of a continuous wavelet transform to convert a one-dimensional ECG signal to a two-dimensional scalogram for hierarchical feature extraction.

d) Implementation of a novel adaptive focal cross-entropy (AFCE) loss function that effectively handles class imbalance, ensuring that the model stays sensitive to minority class without compromising the overall performance accuracy.

This paper is organized as follows: section II presents the relevant literature, and the in-depth analysis of the proposed methodology, including the data collection and preprocessing steps are presented in section III. The experimental setup and hyperparameters configurations employed in this study are discussed in section IV. Section V presents the discussion of the experimental findings while the ablation study that examines the impact of the hyperparameter tweakings on the proposed model is presented in section VII. Lastly, section VIII concludes this study by highlighting the key merits, limitations, and future work.

II. Related Work

Recent advances in deep learning have significantly improved the accuracy of ECG classification. However, several persistent challenges hinder the adoption of these models in real-world clinical settings, including limited generalization across diverse patient populations, interpretability of decisions, computational efficiency for real-time deployment, and effective handling of class imbalance. Prior works can be grouped into four major categories: signal-based models, image-based representations, attention-augmented architectures, and imbalance mitigation techniques.

A. Signal-Based Deep Learning Models

Signal-domain models process raw 1D ECG waveforms using convolutional and recurrent neural architectures. Zhang et al. [9] employed a CNN-BiLSTM model that achieved 95% accuracy but suffered from computational complexity, making it unsuitable for real-time applications. Abdurrahman et al. [10] provided a systematic review, highlighting that while CNNs achieve 90–98% accuracy, their performance drops significantly in inter-patient validation, raising concerns about generalization. Hamid et al. [11] explored entropy-integrated CNNs on the PTB-XL dataset with 95% accuracy, but model complexity varied across configurations. Xu and Liu [12] introduced a 12-layer 1D CNN with strong accuracy (97%) but excessive depth and memory requirements limit real-time use. Zhang and Wu [13] achieved 95% accuracy but relied on extensive manual feature selection, undermining scalability. Sengur et al. [14] reported 98% accuracy using deep learning, yet they emphasized critical limitations in data preprocessing, suggesting that performance heavily depends on the quality and availability of curated input signals.

B. Image-Based Approaches with 2D Representations

Image-domain models transform ECG signals into 2D representations such as spectrograms or scalograms, enabling CNNs to extract spatial-frequency features. Shaker and Tolba [15]

used transfer learning on ECG images, reaching 98% accuracy, but their reliance on pre-trained networks reduced adaptability to new datasets. Wan et al. [16] integrated attention mechanisms with CNNs and achieved similar accuracy, though they encountered challenges with model interpretability, limiting clinical trust. Alomari et al. [17] addressed small-sample training via transfer learning and achieved high accuracy (98%), but faced difficulties generalizing across datasets. These methods demonstrate promise but often lack transparency and flexibility across heterogeneous ECG data sources.

C. Attention Mechanisms and Hybrid Models

Attention mechanisms have been introduced to enhance focus on diagnostically relevant waveform segments. Rajpurkar et al. [18] employed a transformer-based model to capture spatio-temporal dependencies with 97% accuracy; however, the approach incurred substantial computational overhead. Kiranyaz et al. [19] and Rahhal et al. [20] also investigated attention-driven and transformer-based frameworks, balancing accuracy with complexity. Lee et al. [21] proposed a CNN-RNN hybrid reaching 97% accuracy, but it required high computational resources. Zhang et al. [22] and Guo et al. [23] incorporated LSTM and hybrid deep architectures to capture temporal dynamics and achieved up to 98% accuracy, but these architectures often suffer from overfitting and poor interpretability due to their depth and structural complexity. Zhang et al. [24] further explored transformer-based ECG diagnosis with 96% accuracy, though the model lacked interpretability and required significant computational resources.

D. Handling Class Imbalance in ECG Classification

Class imbalance—where rare but clinically significant arrhythmias are underrepresented—is a critical issue in ECG classification. Jadhav et al. [25] acknowledged that conventional deep learning models trained with categorical cross-entropy often fail to prioritize minority classes, leading to poor sensitivity for rare conditions. Salama et al. [26] introduced a GAN-based framework paired with classical classifiers and achieved 96% accuracy, but the model faced training instability and high complexity. While some studies explored resampling techniques or ensemble models, they often introduced overfitting or added significant computational burdens. Unlike previous studies that target individual limitations in isolation, our proposed strategy addresses multiple challenges holistically. First, we employ Continuous Wavelet Transform (CWT) to transform ECG signals into 2D scalograms, capturing both time and frequency features. Second, we introduce a lightweight yet powerful attention-enhanced architecture (DRA-ECG) combining depthwise separable convolutions and Convolutional Block Attention Modules (CBAM) to emphasize spatially and channel-relevant patterns while maintaining computational efficiency. Finally, we propose an Adaptive Focal Cross-Entropy (AFCE) loss function, which dynamically reweights hard-to-classify and minority samples, directly addressing class imbalance without requiring complex data augmentation or external sampling mechanisms. This synergistic framework enhances classification accuracy, sensitivity to rare conditions, interpretability, and efficiency, making it well-suited for deployment in real-time and mobile healthcare systems.

TABLE I: Description of the 5-ECG classes with their corresponding diagnostic cardiovascular conditions

Class	Duration	Sample rate	Description
ARR	24 hours	360Hz	Arrhythmia
CHF	20 hours	250Hz	Congestive Heart Failure
MVE	24 hours	250Hz	Malignant Ventricular Ectopy
AF	Several hours	250Hz	Atrial Fibrillation
NSR	24 hours	120Hz	Normal Sinus Rhythm

III. Methodology

This part describes the dataset collection and preprocessing. Next, the proposed depthwise separable residual attention learning and time-frequency scalograms are presented. Finally, the experimental setup and explanations are discussed to conclude this section. Fig.1 demonstrates the overall architecture of our proposed ECG classification model.

A. Dataset Acquisition

The dataset used in this research is collected from the PhysioBank ATM repository [27]. For this study, 5 different categories of raw ECG signals are chosen which include, the MIT-BIH Arrhythmia database (ARR), BIDMC Congestive Heart Failure database (CHF), MIT-BIH Malignant Ventricular Ectopy database (MVE), MIT–BIH Atrial Fibrillation database(AF), and the Normal Sinus Rhyth database. Table 1 shows the sampling rates used for each ECG category due to their signal lengths. The sampling frequencies for the ARR, CHF, MVE, AF, and NSR signals were 360 Hz, 250 Hz, 250HZ, 250HZ, and 120 HZ respectively. ARR consists of records collected from a total of 47 human subjects over approximately 24 hours. CHF consists of 20 hours longs of ECG records from patients with diagnosable heart failures, which offer intensive data that reflects the failure pattern of congestive heart failure. The MVE dataset is a publicly available 24-hour-long ECG recording made from patients with a life-threatening cardiac condition, who have a medical history of heart disease related to ventricular ectopy. The ECG recordings of AF include a period varying from minutes to hours in patients with atrial fibrillation. NSR consists of a 24-hour recording of normal heartbeats that are frequently used as a rich source to investigate ECG recordings from patients without arrhythmias. The raw ECG dataset used in this study consists of a variety of cardiac diseases, ranging from normal sinus rhythm to complex cardiovascular.

B. ECG Data Preprocessing

1) Thompkins Filtering: To obtain a smoother and more diagnostically useful representation of ECG signals, we applied Thompkins band-pass filtering as expressed in (1), (2), (3), (4), and (5). Algorithm 1 shows each step of the signal decomposition. The process begins with band-pass filtering to isolate QRS complexes by removing both low- and high-frequency noise while preserving the relevant frequency band.

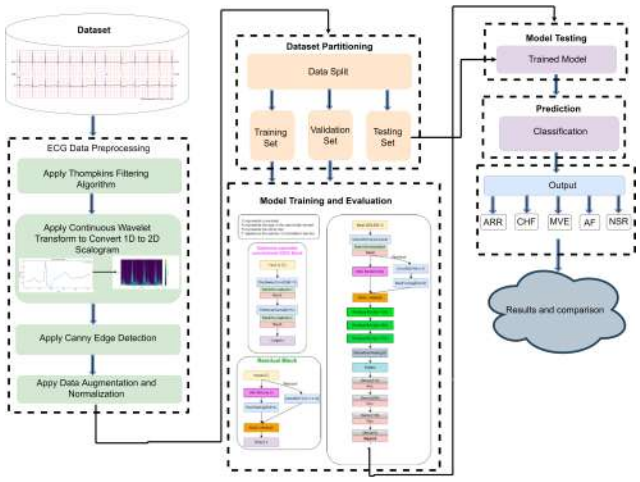


Fig. 1: Overview of the proposed ECG classification pipeline for cardiac analysis

A low-pass filter then attenuates high-frequency noise, followed by a high-pass filter that removes baseline drift and other low-frequency components, enhancing signal clarity. The derivative filter calculates the signal gradient to highlight rapid changes, particularly within the QRS complex. This is followed by signal squaring to amplify peak features, making them more distinguishable. Lastly, a window size 259 moving window integrator is applied to integrate the signal segments to effectively mitigate fluctuation and produce a continuous stable signal for further analysis.

$$y_{\text{LP}}[n] = 2y_{\text{LP}}[n-1] - y_{\text{LP}}[n-2] + x[n] - 2x[n-6] + x[n-12] \tag{1}$$

where $y_{\text{LP}}[n]$ denotes the output of the low-pass filter at sample n after the reduction in high frequency noise.

$$y_{\text{HP}}[n] = y_{\text{HP}}[n-1] - \frac{1}{32}\left(y_{\text{LP}}[n] - x[n-16]\right) \tag{2}$$

where $y_{\text{HP}}[n]$ is the output of the high-pass filter at sample n after removing the baseline drift and low-frequency noise.

$$y_{\text{D}}[n] = \frac{1}{8}\left(2y_{\text{HP}}[n] + y_{\text{HP}}[n-1] - y_{\text{HP}}[n-3] - 2y_{\text{HP}}[n-4]\right) \tag{3}$$

where $y_{\text{D}}[n]$ depicts the output of the derivative filter at sample n after obtaining the gradient of the signal and emphasizing the QRS complexes.

$$y_{\text{S}}[n] = y_{\text{D}}[n]^2 \tag{4}$$

where $y_{\text{S}}[n]$ represents the squared value of the derivative filter output $y_{\text{D}}[n]$ at sample n with amplifies QRS peaks while suppressing smaller values.

$$y_{\text{MWI}}[n] = \frac{1}{N}\sum_{i=0}^{N-1} y_{\text{S}}[n-i] \tag{5}$$

where $y_{\text{MWI}}[n]$ is the output of the sliding window integration at sample n which represents the smoothed signal with a clearer representation of the QRS complexes, while $x[n]$ is the ECG signal input at sample n and N is the window size for the sliding window integration, that determines the averaged sample.

Algorithm 1 Thompkins Filtering Algorithm

THOMPKINSFILTER$x[n], N$
Input: $x[n]$ Raw ECG signal
Output: $y_{\text{MWI}}[n]$ Filtered ECG signal
Step 1: Low-Pass Filtering
$y_{\text{LP}}[n] \leftarrow 2y_{\text{LP}}[n-1] - y_{\text{LP}}[n-2] + x[n] - 2x[n-6] + x[n-12]$

Step 2: High-Pass Filtering
$y_{\text{HP}}[n] \leftarrow y_{\text{HP}}[n-1] - \frac{1}{32}\left(y_{\text{LP}}[n] - x[n-16]\right)$
Step 3: Derivative Filtering
$y_{\text{D}}[n] \leftarrow \frac{1}{8}\left(2y_{\text{HP}}[n] + y_{\text{HP}}[n-1] - y_{\text{HP}}[n-3] - 2y_{\text{HP}}[n-4]\right)$

Step 4: Squaring
$y_{\text{S}}[n] \leftarrow y_{\text{D}}[n]^2$
Step 5: Moving Window Integration
$y_{\text{MWI}}[n] \leftarrow \frac{1}{N}\sum_{i=0}^{N-1} y_{\text{S}}[n-i]$
return $y_{\text{MWI}}[n]$

To capture both temporal and spectral patterns, we transformed the preprocessed 1D ECG signals into 2D scalograms using the Continuous Wavelet Transform (CWT) with the Morlet wavelet function. We selected scales from 1 to 128, which provide a fine-to-coarse resolution suitable for capturing arrhythmic patterns and transient events.

The resulting scalograms were resized from their original dimensions (775×462) to 224×224 pixels using bilinear interpolation. This size balances information retention with computational efficiency and aligns with the input requirements of our proposed model.

2) ECG Data Conversion and Resizing: The process of converting raw ECG signals into 2D scalograms is an approach that provides a comprehensive time-frequency analysis that is useful in diagnosing cardiovascular diseases. This approach involves several steps, where the first step consists of gathering raw non-unsegmented (no peak annotated) ECG data files. These files contain the multichannel ECG signals, loaded directly using the wavelet transform library. As these signals have very high resolution, they need to be down-sampled to reduce computational load, usually retaining only every 10th sample. More so, the downsampled signal is segmented into smaller chunks of 5000 samples for more targeted analysis.

Afterward, the continuous wavelet transform (CWT) is applied to these segments to convert the ECG signals from the time domain into the time-frequency domain. This is extremely useful when dealing with non-stationary signals such as ECG, where the frequency content changes over time. The applied CWT uses the Morlet wavelet, which has good time-frequency localization properties at scales 1-128, and produces 2D representation scalograms, which represent the magnitude of the wavelet coefficients across time and scale, as shown in Fig. 2. The CWT is chosen to capture both temporal and spectral patterns in the ECG signals, while the scales from 1 to 128 are selected to provide a fine-to-coarse resolution suitable for capturing arrhythmic patterns and transient events.

This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3566531

TABLE II: Description of 2D scalograms across different ECG class

Class	Number of image count
ARR	624
CHF	5,367
MVE	220
AF	4,214
NSR	4,024
Total	14,449

These scalograms are saved as image format files, where the x-axis corresponds to time and the y-axis depicts the normalized scale indices, allowing significant features in the ECG signal to be highlighted by the intensity of the color. Table II shows the number of 2D scalograms generated across the five different class labels.

This preprocessing pipeline is applied iteratively on five different ECG databases, covering various cardiac conditions including normal rhythm: Normal Sinus Rhythm (NSR), Arrhythmia (ARR), Congestive Heart Failure (CHF) database, Atrial Fibrillation(AF), and Malignant Ventricular Ectopy(MVE). Transforming raw ECG signals into scalograms allows for visual inspection, providing a rich time-frequency representation that captures both spatial and spectral attributes of the ECG signals. The generated scalograms are 775×462 pixels, categorized into condition-specific directories, and are ready for further analysis by the proposed model. These datasets are vital for the development of computer-aided diagnostic tools in the field of cardiology and can be used as a foundation tool to analyze the dynamic frequency-related information of the electrical activity within the heart across various states.

Given that the resulting scalograms are typically 775×462 pixels in resolution, they cannot be used directly for input to the proposed DRA-ECG, which accepts smaller and more manageable input dimensions. To overcome this, the scalograms are resized into 224×224 pixels, which is suitable with our proposed model input size. This resizing ensures that the necessary time-frequency characteristics of the scalogram are retained while allowing it to fit within the computational constraints.

The resizing process consists of changing both the time and scale dimensions attributes of the scalogram. The original 775 pixel width, representing the time dimension, is compressed to 224 pixels, diluting the temporal resolution but retaining the overall content of the frequency spectral structure. Similarly, the height, which represents the wavelet scales, is also reduced to 224 remaining 128 pixels (that is, the wavelet scales). This image scaling operation is performed using standard image processing techniques to preserve the vital features in the scalograms necessary for the accurate DRA-ECG classification.

3) Continuous Wavelet Transform: Given that our proposed model requires the input data to be in image format, we transformed the raw signals into time-frequency scalograms using the continuous wavelet analysis (CWT) method. The resulting time-frequency scalograms were initially presented in three channels. These scalograms are converted into a single-channel image format to align with our model's requirements. Subsequently, the input images were standardized and resized to dimensions 224x224x1 to correspond to the input size of our proposed model. The dataset is then divided into training, validation, and test sets. Fig. 2 provides a visual representation of the CWT process of transforming the 1D ECG signal into 2D scalograms for the different cardiac conditions and the mathematical expression is shown in (6).

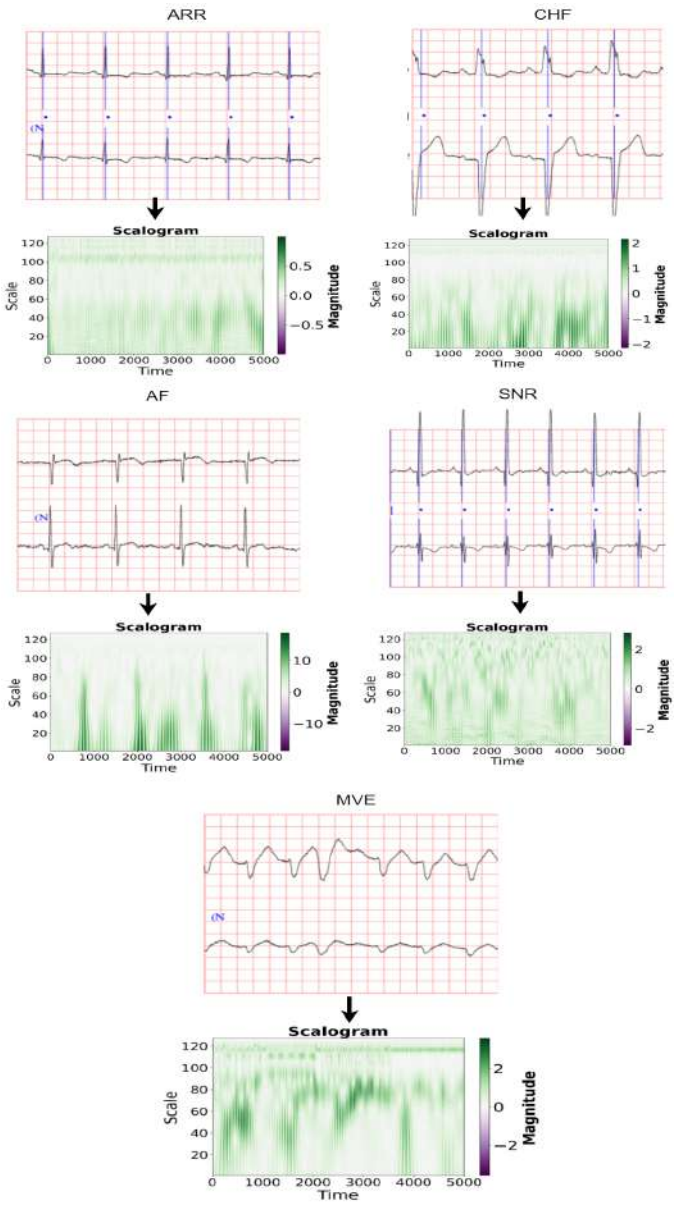


Fig. 2: Transformation of 1D raw ECG signals into 2D scalograms using continuous wavelet transform

$$\mathrm{CWT}_x^{\psi} = \frac{1}{\sqrt{|S|}} \int x(t)\psi^* \left(\frac{t-\tau}{S}\right) dt \tag{6}$$

where τ represents the translation factor and S denotes the scaling factor. The function $x(t)$ depicts the mother wavelet. The expression $\psi^* \left(\frac{t-\tau}{S}\right) dt$ represents the derivative of the mother wavelet function, and the parameter τ is used to adjust the position of the mother wavelet. The scaling factor

S is the inverse of frequency; a large S corresponds to low frequency, while a small S corresponds to high frequency. The Continuous Wavelet Transform (CWT) is selected for feature extraction due to its capability to capture both temporal and frequency-domain representations of ECG signals. Unlike the Fourier Transform (FT), which only provides global frequency information, and the Short-Time Fourier Transform (STFT), which suffers from fixed window resolution, CWT provides a multi-resolution analysis, making it ideal for characterizing transient ECG features such as arrhythmias and ischemic changes. The generated CWT-based 2D scalograms enhance the model to recognize and capture subtle morphological variations in ECG waveforms, improving overall classification performance.

4) ECG Data Augmentation and Partitioning: Given that the number of generated scalograms is largely different across the class label, it is necessary to increase the number of training data for the ECG scalograms and improve the robustness of the model. We applied various image-based augmentation techniques to the scalograms, including random rotations (±15°), width and height shifts (±10%), shear transformations (up to 20%), zooming (±10%), and horizontal flipping. The ECG image is subject to various data augmentation methods such as rotation, width shift, height shift, and shear. The width and height shifts move the ECG image horizontally and vertically, respectively, while rotation alters the ECG image orientation to achieve rotational invariance. Shear applies a slanting transformation, adding robustness to geometric distortions. These augmentations enhanced the dataset's diversity and improved the model's generalization while preserving critical diagnostic features.

Table III shows the distribution of the dataset after data augmentation. To avoid overfitting and make the proposed model robust and generalizable, the data set is divided into 80% for training, 10% validation, and the remaining 10% testing set as presented in Table IV. The 80% training set provides enough information for the model to learn complex features. The 10% validation set is used to refine the model during training to have a good generalization for unseen data. Finally, the 10% test set, which is unseen and kept separately from the training and validation sets, is used to evaluate the performance of the model. This approach balances learning and evaluation, ensuring an accurate and reliable model.

5) Hierarchical Feature detection: In this paper, we applied an edge enhancement technique to extract hierarchical features from the ECG scalograms, improving the detection and analysis of cardiac features. Hierarchical feature detection is a process that involves different steps, fashioned to optimize the extraction of features as outlined in Algorithm 2. a) Noise Reduction: Firstly, a Gaussian filter is applied to the ECG scalogram to minimize the impact of noise as shown in (7), ensuring that the subsequent feature detection steps focus on relevant signal features rather than random variation.

$$I_{\text{smooth}}(x,y) = I(x,y) * G(x,y,\sigma) \tag{7}$$

where $I(x,y)$ denotes the input scalogram and $G(x,y,\sigma)$ represents the Gaussian filter with σ denoting standard deviation and $*$ represents convolution.

b) Gradient Calculation: The sobel filter is applied on the smoothed image to calculate the gradient intensity to identify regions with significant intensity changes, corresponding to the curves in the ECG scalogram as presented in (8)–(11).

$$G_x(x,y) = I_{\text{smooth}}(x,y) * S_x \tag{8}$$

$$G_y(x,y) = I_{\text{smooth}}(x,y) * S_y \tag{9}$$

where G_x and G_y represent the gradients in the x and y direction respectively.

$$G(x,y) = \sqrt{G_x(x,y)^2 + G_y(x,y)^2} \tag{10}$$

$$\theta(x,y) = \arctan\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \tag{11}$$

The gradient intensity is computed in (10) while the direction of the smoothed image is computed in (11).

Algorithm 2 Hierarchical Feature Detection

HIERARCHICALFEATUREDETECTION$I, \sigma, T_{\text{low}}, T_{\text{high}}$
Input: I Input ECG scalogram
Output: E_{final} Feature-detected ECG scalogram
Step 1: Noise Reduction
$I_{\text{smooth}} \leftarrow I * G(\sigma)$
Step 2: Gradient Calculation
$G_x \leftarrow I_{\text{smooth}} * S_x$
$G_y \leftarrow I_{\text{smooth}} * S_y$
$G \leftarrow \sqrt{G_x^2 + G_y^2}$
$\theta \leftarrow \arctan\left(\frac{G_y}{G_x}\right)$
Step 3: Non-Maximum Suppression each pixel (x,y) in G $G(x,y)$ is a local maximum in direction $\theta(x,y)$
$E(x,y) \leftarrow G(x,y)$
$E(x,y) \leftarrow 0$
Step 4: Double Thresholding each pixel (x,y) in E $E(x,y) > T_{\text{high}}$
$E_{\text{strong}}(x,y) \leftarrow E(x,y)$ $T_{\text{low}} \leq E(x,y) \leq T_{\text{high}}$
$E_{\text{weak}}(x,y) \leftarrow E(x,y)$
$E_{\text{weak}}(x,y) \leftarrow 0$
Step 5: Feature Tracking by Hysteresis each pixel (x,y) in E_{weak} $E_{\text{weak}}(x,y)$ is connected to strong curve in E_{strong}
$E_{\text{final}}(x,y) \leftarrow E_{\text{weak}}(x,y)$
$E_{\text{final}}(x,y) \leftarrow 0$
return E_{final}

c) Non-Maximum Suppression: To ensure that only the local maxima in the gradient direction are considered edges, non-maximum suppression in (12) is applied to discard ambiguous detected edges and retain only the most significant.

$$E(x,y) = \begin{cases} G(x,y) & \text{if } G(x,y) \text{ is a local maximum} \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

d) Double Thresholding: Even after non-maximum suppression, distinguishing between significant and irrelevant edges is necessary. The double thresholding technique in (13) and

TABLE III: Dataset distribution after data augmentation

Class	Number of image count
ARR	5,832
CHF	5,830
MVE	5,843
AF	5,826
NSR	5,820
Total	29,151

TABLE IV: Dataset Distribution across each Class Label

Class	Training	Validation	Testing
ARR	4,666	583	583
CHF	4,664	583	583
MVE	4,674	584	584
AF	4,661	582	582
NSR	4,656	582	582
Total	23,321	2,914	2,914

(14) is then applied to categorize the edges into strong and weak edges.

$$E_{\text{strong}}(x,y) = \begin{cases} E(x,y) & \text{if } E(x,y) > T_{\text{high}} \\ 0 & \text{otherwise} \end{cases} \quad (13)$$

$$E_{\text{weak}}(x,y) = \begin{cases} E(x,y) & \text{if } T_{\text{low}} \leq E(x,y) \leq T_{\text{high}} \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

e) Hysteresis: Finally, the step in (15) ensures that true edges are preserved by connecting weak edges to strong edges as long as they are contiguous while noise-induced edges are eliminated.

$$E_{\text{final}}(x,y) = \begin{cases} E_{\text{strong}}(x,y) & \text{if connected to a strong edge} \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

6) Normalization: We applied data normalization to ensure an effective process of the 2D scalogram images for the proposed work. This entails scaling the pixel values of every scalogram to a standard range between 0 and 1. Thus, standardizing the input data makes it easier for the proposed model to learn and train faster and even with more stability. Every pixel score is divided by the standard maximum pixel, 255 which scales the pixels between the desired range as defined in (16) and (17). This method enables the input scales to be consistent and prevents numerical issues during the initial stage of training. Moreover, normalization of the scalograms allows the model to focus on important scalar features by avoiding distraction due to outliers and irrelevant variations which ultimately enhances interpretability and preserves crucial diagnostic information contained within the scalograms.

$$\text{Normalized Value} = \frac{\text{Pixel Value}}{\text{Max Pixel Value}} \quad (16)$$

$$\text{Normalized Value} = 2 \times \left(\frac{\text{Pixel Value}}{\text{Max Pixel Value}} \right) - 1 \quad (17)$$

C. The Proposed DRA-ECG Model

The proposed model is a novel framework developed in this study for the classification of heart disease using ECG scalograms as shown in Fig. 3. The proposed DRA-ECG incorporated high-level deep learning-based algorithms to enhance feature extraction model robustness and classification accuracy. The first layer is a convolutional layer with 32 filters of size 3x3 and strides of 2 to convolve the input ECG scalograms of input shape $224x224x1$. This layer is then followed by batch normalization and ReLU to accelerate the training process and introduce non-linearity respectively. We introduce the depthwise separable convolution (DSC) block to reduce the computational complexity while preserving high performance. DSC block decomposes the standard convolution into two processes. The first operation is the depthwise convolution, which employs separate convolutional kernels for each input channel to learn and capture spatial relationships within a given channel as expressed in (18)

$$y_{\text{depthwise}}(x,y,c) = \sum_{i=1}^{k}\sum_{j=1}^{k} W_{\text{depthwise}}(i,j,c) \cdot x(x+i-1, y+j-1, c) \quad (18)$$

The second operation is the pointwise convolution as expressed mathematically in (19) which concatenates the output from the depthwise operation for the different channels using a 11 convolution.

$$y_{\text{pointwise}}(x,y,m) = \sum_{c=1}^{C} W_{\text{pointwise}}(c,m) \cdot y_{\text{depthwise}}(x,y,c) \quad (19)$$

The DSC block is designed to minimize the number of parameters and computational costs so that the model becomes more efficient without compromising accuracy as expressed in equation (20).

$$y_{\text{DSC}}(x,y,m) = \text{ReLU}(\text{BatchNorm}(y_{\text{pointwise}}(x,y,m))) \quad (20)$$

Each DSC block is followed by a $2x2$ maximum pooling layer and a skip connection through a $1x1$ convolution and a stride of 2 to map the input from the previous layer to the output from the succeeding layer as given in (21).

$$y_{\text{residual}}(x,y,m) = \text{MaxPooling2D}(y_{\text{DSC}}(x,y,m)) + \text{Conv2D}(x,y,m) \quad (21)$$

Finally, the residual output combines the max pooling output and the output from the skip connection as seen in (22).

$$y_{\text{ResidualBlock}}(x,y,m) = \text{ReLU}(\text{BatchNorm}(y_{\text{residual}}(x,y,m))) \quad (22)$$

This proposed methodology ensures the model is properly trained to improve gradient flow and alleviate the vanishing gradient problem. Our proposed method includes convolutional block attention module (CBAM) and global average pooling (GAP) to enhance the learnability of features channel-wise and spatial-wise and reduce the dimensionality of the aggregated feature vectors for improved performance, respectively. Channel attention concentrates on the vital features by weighting the channels of the feature maps. In contrast, spatial attention focuses on the valuable pixel positions by employing

spatial attention to the feature map as seen in (23) and (24).

$$y_{\text{channel}}(x,y,m) = \sigma\Big(\text{MLP}\big(\text{GlobalAvgPool}(y_{\text{ResidualBlock}})\big)\Big) + \sigma\Big(\text{MLP}\big(\text{GlobalMaxPool}(y_{\text{ResidualBlock}})\big)\Big) \cdot y_{\text{ResidualBlock}} \tag{23}$$

$$y_{\text{spatial}}(x,y,m) = \sigma\left(\text{Conv2D}(y_{\text{channel}})\right) \cdot y_{\text{channel}} \tag{24}$$

The final output of the convolutional block attention mechanism is mathematically represented (25).

$$y_{\text{CBAM}}(x,y,m) = y_{\text{spatial}}(x,y,m) \tag{25}$$

It is important to note that the attention mechanism ensures that the model pays attention to the most relevant features of the ECG scalogram for classification, thereby improving the interpretability and performance of the model. Following the CBAM, we applied global average pooling (GAP) instead of the regular dense layers to reduce each feature map to a single vector by taking the average as expressed in (26).

$$y_{\text{GAP}} = \frac{1}{H \times W} \sum_{x=1}^{H} \sum_{y=1}^{W} y_{\text{CBAM}}(x,y,m) \tag{26}$$

The choice of implementing GAP over the conventional dense layers is to minimize overfitting and reduce model complexity by cutting down the number of parameters, since the dense layers constitute most of the model's parameters. The final output is passed through a softmax layer that classifies the output into one of the heart conditions or normal conditions: ARR (Arrhythmia), CHF (Congested Heart Failure), MVE (Myocardial Infarction), AF (Atrial Fibrillation), and NSR (Normal Sinus Rhythm) as expressed mathematically in (27).

$$y_{\text{output}} = \text{Softmax}(W_{\text{output}} \cdot y_{\text{FC2}} + b_{\text{output}}) \tag{27}$$

The several innovation components proposed in this methodology collectively enhance the model to accurately classify heart conditions from ECG scalograms which distinguishes it from traditional ECG classification models. The overall expression for the proposed DRA-ECG is given in (28)

$$y_{\text{DRA-ECG}} = \text{Softmax}\left(W_{\text{output}} \cdot \text{GAP}\left(\frac{1}{H \times W} \sum_{x=1}^{H} \sum_{y=1}^{W} y_{\text{CBAM}}\Big(y_{\text{ResidualBlock}}\big(y_{\text{DSC}}(x)\big)\Big)\right) + b_{\text{output}}\right) \tag{28}$$

IV. Experiments

This subsection will illustrate the experimental configuration and model evaluation criteria in detail. The dataset splits are presented in Table I. The proposed DRA-ECG model is executed using TensorFlow as a backend to train the model on the ECG images. The dataset used in this study consists of 29,151 scalograms generated from five different ECG databases, covering a wide range of cardiac conditions, including arrhythmia, congestive heart failure, atrial fibrillation, and normal sinus rhythm. The raw ECG signals were preprocessed using Thompkins filtering to remove noise and baseline drift, followed by downsampling to reduce computational load. Data augmentation techniques such as rotation, width shift, and shear were applied to increase the diversity of the training data. The dataset was split into 80% for training, 10% for validation, and 10% for testing to ensure robust model evaluation.

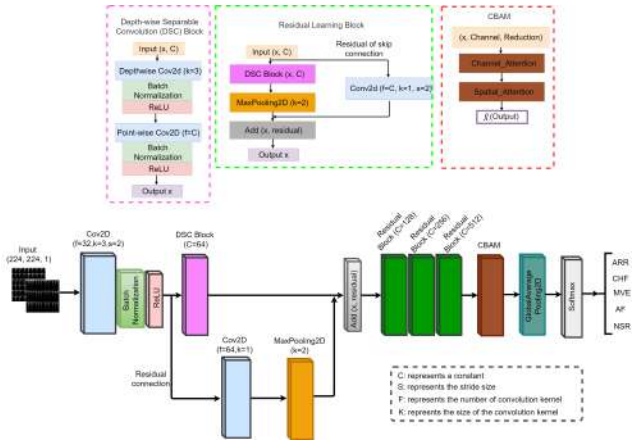


Fig. 3: Architecture of the proposed DRA-ECG model for cardiac condition classification

A. Training and Testing the DRA-ECG

The inputs to the DRA-ECG model during training are the ECG scalograms generated in Section III. Table I summarizes the dataset into three portions: training, validation, and test set. 80% of the dataset is portioned for training, while the remaining 20% is further divided equally into validation and test sets. We train the DRA-ECG model on the training set to ensure generalization across the different cardiac diseases.

B. Evaluation Metrics

The evaluation criteria adopted to validate the proposed model include accuracy, precision (pre), recall (rec), F1-Score (F1-s), Receiver Operating Characteristic Area Under the Curve (ROC-AUC), and Precision-Recall (PR) curve. These mathematical expressions are presented in (29)–(34), where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative respectively.

Accuracy measures overall classification correctness.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{29}$$

Precision is critical in reducing false positives, especially important to avoid unnecessary interventions in clinical settings.

$$\text{Precision} = \frac{TP}{TP + FP} \times 100 \tag{30}$$

Recall (Sensitivity) measures the ability to detect true positives, which is crucial for capturing real cardiac events, especially rare ones.

$$\text{Recall} = \frac{TP}{TP + FN} \times 100 \tag{31}$$

F1-score balances precision and recall, which is suitable for imbalanced datasets.

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100 \tag{32}$$

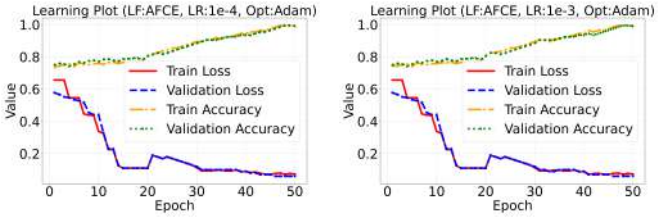
This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3566531

Precision-Recall (PR) Curve provides more information to support the analysis of ROC when dealing with imbalanced datasets.

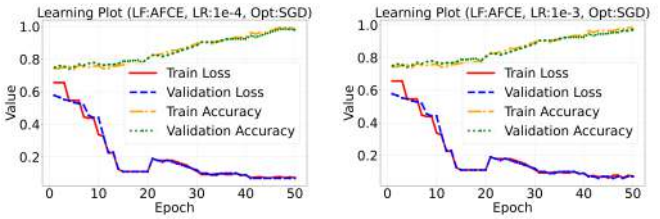
$$\text{Precision-Recall(AUPRC)} = \int_0^1 \text{Precision}(\text{Recall})\, d(\text{Recall}) \tag{33}$$

ROC-AUC evaluates the classifier's ability to distinguish between classes.

$$\text{ROC-AUC} = \int_0^1 \text{TPR}(x)\, dx \tag{34}$$

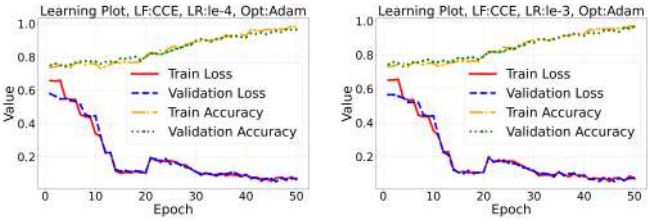


(a) LF: AFCE, LR: le-4, Opt: Adam (b) LF: AFCE, LR: le-3, Opt: Adam

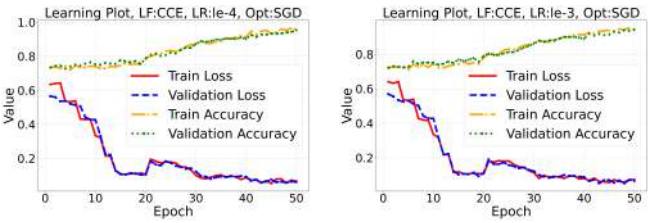


(c) LF: AFCE, LR: le-4, Opt: SGD (d) LF: AFCE, LR: le-3, Opt: SGD

Fig. 4: Accuracy and loss plots for the proposed DRA-ECG model trained with AFCE under different hyperparameters



(a) LF: CCE, LR: le-4, Opt: Adam (b) LF: CCE, LR: le-3, Opt: Adam



(c) LF: CCE, LR: le-4, Opt: SGD (d) LF: CCE, LR: le-3, Opt: SGD

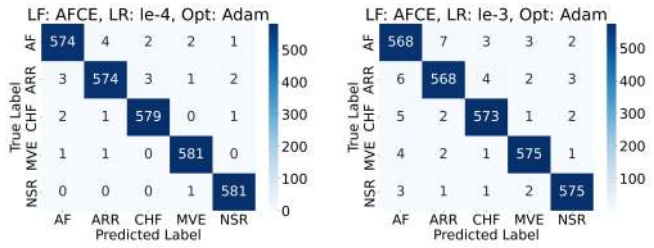
Fig. 5: Accuracy and loss plots for the proposed DRA-ECG model trained with CCE under different hyperparameters

C. Adaptive Focal Cross-Entropy Loss Function

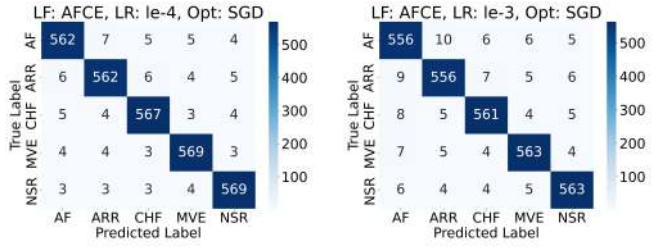
To tackle the limitations of CCE, this study proposes a novel loss function called Adaptive Focal Cross-Entropy (AFCE) that dynamically balances the loss contribution from CCE. AFCE results in more stable training behaviors and efficiently classifies the majority and minority classes, making it an ideal strategy for imbalanced datasets. The AFCE loss function improves the recognition of complex features of the ECG and yields high performance, ensuring all categories are classified with high certainty. The AFCE in (35) is designed to maintain generalization while incorporating the focal mechanism to improve model performance on the imbalanced dataset.

$$L_{\text{AFCE}} = \lambda \cdot \left(-\sum_{j=1}^{k} T_j \log S_j \right) + (1-\lambda) \cdot (-\alpha_t (1-q_t)^{\gamma} \log(q_t)) \tag{35}$$

where λ represents the adaptive weighting parameter adjusted during training, $\sum_{j=1}^{k} T_j \log S_j$ denotes the multi-class loss component, and $\alpha_t(1-q_t)^{\gamma}\log(q_t)$ is the focal component with the scaling factor for minority class represented as α_t. γ depicts the focusing parameters, q_t represents the predicted probability of the true label and L_{AFCE} represents the adaptive focal cross-entropy loss.



(a) LF: AFCE, LR: le-4, Opt: Adam (b) LF: AFCE, LR: le-3, Opt: Adam

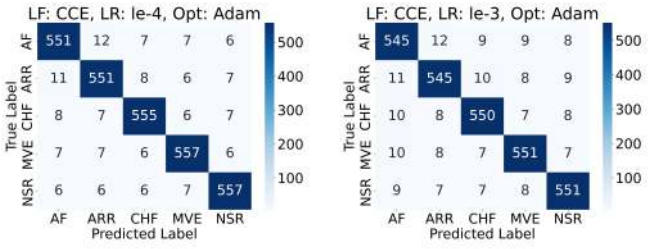


(c) LF: AFCE, LR: le-4, Opt: SGD (d) LF: AFCE, LR: le-3, Opt: SGD

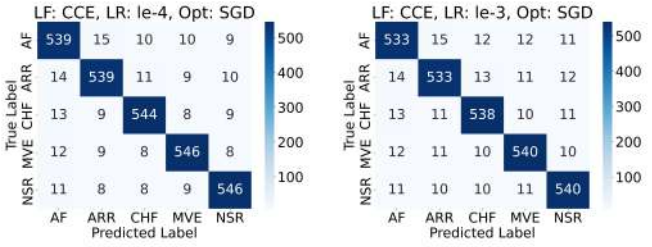
Fig. 6: Confusion matrix of the proposed DRA-ECG model trained with AFCE under different hyperparameter settings

D. Experimental Setup

All experiments were conducted on a high-performance workstation running Anaconda3, featuring an Intel Core i7 processor, 64 GB RAM, and an NVIDIA GeForce RTX 4080 GPU with 16 GB VRAM. The software environment included Python 3, TensorFlow 2, and Keras 2, along with supporting libraries such as NumPy, SciPy, and Matplotlib. This setup



(a) LF: CCE, LR: le-4, Opt: Adam (b) LF: CCE, LR: le-3, Opt: Adam



(c) LF: CCE, LR: le-4, Opt: SGD (d) LF: CCE, LR: le-3, Opt: SGD

Fig. 7: Confusion matrix of the proposed DRA-ECG model trained with CCE under different hyperparameter settings

TABLE V: Classification performance of the proposed DRA-ECG model using AFCE loss function with Adam optimizer and LR of 1×10^{-4}

Parameter	Values Explored
Learning Rate	$1 \times 10^{-}3, 1 \times 10^{-}4$
Batch Size	16, 32, 64
Optimizers	Adam (=0.9), SGD (momentum=0.8, 0.9)
Loss Functions	Categorical Cross-Entropy (CCE), Focal Loss, AFCE
Epochs	Up to 50 (with early stopping)

enabled efficient training of deep learning models on large-scale 2D scalogram data. We employed a grid search strategy across several key hyperparameters as presented in Table V.

Each configuration was evaluated based on the validation F1-score. The learning rates tested were $1\times10^{-}3$ and $1\times10^{-}4$, while the batch sizes explored in this study are 16, 32, and 64. The optimizer was chosen based on the best performance across the validation loss with SGD, and Adam is explored at different momentum values, 0.8 and 0.9. Early stopping was applied with a patience of 10 epochs on the validation loss to prevent overfitting. The best results were consistently achieved with the Adam optimizer, a learning rate of $1 \times 10^{-}4$, and the AFCE loss function. The final model was selected based on the highest F1 score observed across the data split ratios.

E. Loss Function

The loss function (LF) evaluates how effectively the proposed DRA-ECG model predicts using the given ECG scalograms as input sets. The CCE loss function is analyzed in the training of the DRA-ECG model presented in (36) in comparison with the proposed AFCE loss.

$$L_{\text{CCE}} = -\sum_{j=1}^{k} T_j \log S_j \tag{36}$$

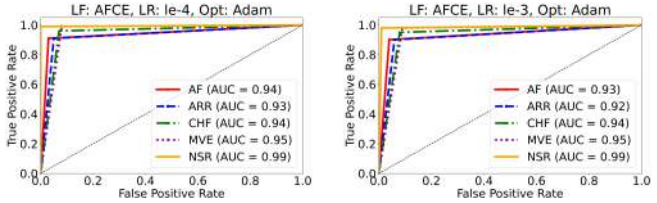
The CCE computes the difference between the true class and the predicted.

F. Cost Function

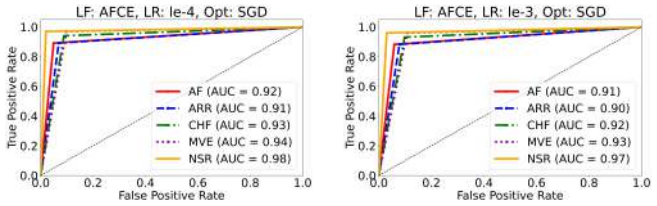
The cost function computes how efficiently the proposed model trains and minimizes the disparity between the actual training data and the predicted outcome. In this study, we use adaptive moment estimation (Adam) and Stochastic Gradient Descent (SGD) optimizers for cost function optimization.

G. Learning Rate (LR)

LR is responsible for how much the proposed DRA-ECG changes when the weights deviate from the desired output. In the experiments, the LR range is set to $1x10^{-}3$ and $1x10^{-}4$.



(a) LF: AFCE, LR: le-4, Opt: Adam (b) LF: AFCE, LR: le-3, Opt: Adam



(c) LF: AFCE, LR: le-4, Opt: SGD (d) LF: AFCE, LR: le-3, Opt: SGD

Fig. 8: ROC-AUC analysis of the proposed DRA-ECG model trained with AFCE under different hyperparameter settings

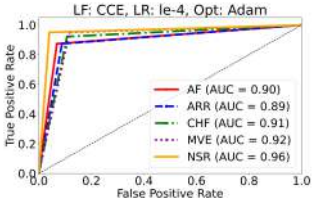
V. Results Analysis

This section discusses the result analyses of the DRA-ECG + AFCE model, alongside State-of-the-art (SOTA) models for the classification of ECG. The performance evaluation is strengthened with confidence intervals (CIs) for different performance metrics as well as statistical testing, such as t-test.

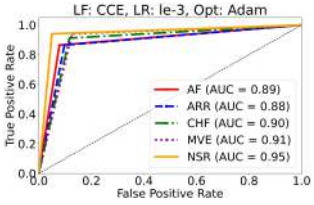
A. Performance of DRA-ECG on ECG Classification

Table VI shows the classification accuracy, precision, recall (sensitivity), F1-score, and AUC achieved with different configurations. The best performance provided by the DRA-ECG model is shown in all metrics. The performance metrics obtained by the proposed DRA-ECG model using the novel Adaptive Focal Cross-Entropy (AFCE) loss with and without the attention mechanism are compared in TableVI. The model obtains a slightly higher accuracy of 98.17%, with the attention mechanism compared to the 97.97% accuracy achieved by

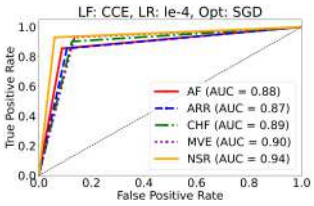
This article has been accepted for publication in IEEE Journal of Biomedical and Health Informatics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/JBHI.2025.3566531



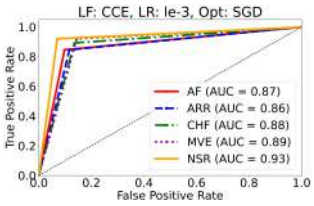
(a) LF: CCE, LR: le-4, Opt: Adam



(b) LF: CCE, LR: le-3, Opt: Adam

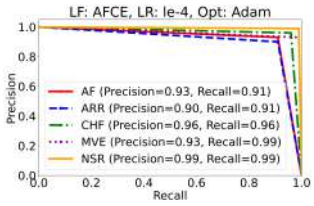


(c) LF: CCE, LR: le-4, Opt: SGD

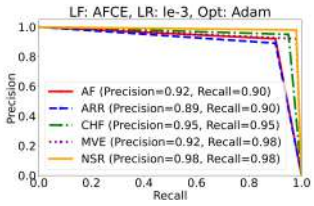


(d) LF: CCE, LR: le-3, Opt: SGD

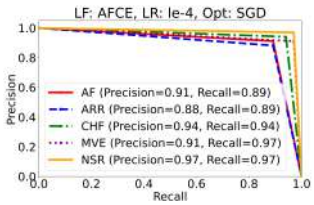
Fig. 9: ROC-AUC analysis of the proposed DRA-ECG model trained with CCE under different hyperparameter settings



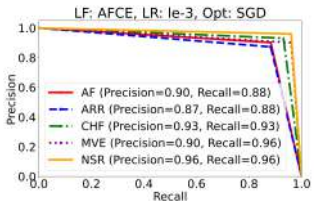
(a) LF: AFCE, LR: le-4, Opt: Adam



(b) LF: AFCE, LR: le-3, Opt: Adam



(c) LF: AFCE, LR: le-4, Opt: SGD

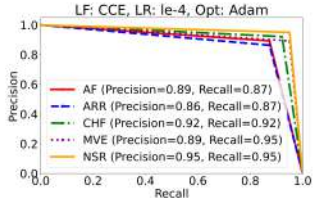


(d) LF: AFCE, LR: le-3, Opt: SGD

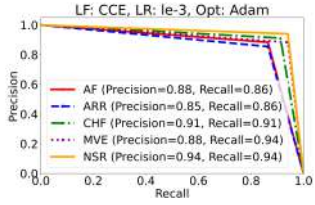
Fig. 10: Precision-Recall curves of the proposed DRA-ECG model trained with AFCE under different hyperparameters

the model without the attention mechanism. This indicates that the attention mechanism improves the model's ability to focus on relevant ECG signal features, further strengthening classification performance. The impact of the attention mechanism increases the precision from 94.90% to 95.89% which implies that our proposed model is performing optimally in minimizing false positive rates, coupled with the capability of the AFCE loss function in highlighting harder-to-classify examples. Recall shows an increase from 95.01% to 95.78% demonstrating that the model is more effective at detecting true positives when attention is applied, which is vital in ensuring that critical cases are not misclassified.

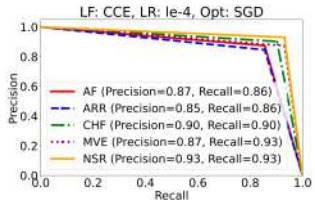
The training curves in Fig. 4 and Fig. 5 show the performance of the proposed DRA-ECG model with a combination of the AFCE and CCE loss functions, learning rates of 1×10^3 and 1×10^4, and optimizers such as Adam and



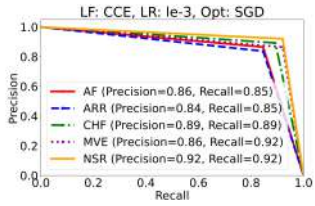
(a) LF: CCE, LR: le-4, Opt: Adam



(b) LF: CCE, LR: le-3, Opt: Adam



(c) LF: CCE, LR: le-4, Opt: SGD



(d) LF: CCE, LR: le-3, Opt: SGD

Fig. 11: Precision-Recall curves of the proposed DRA-ECG model trained with CCE under different hyperparameters

SGD Over 50 epoch with a batch size of 32. Across all the training curves, the model trained with AFCE loss function and Adam optimizer, particularly at a lower learning rate of 1×10^4, consistently achieves the highest performance. The curves show a smooth and consistent convergence for both training and validation loss, while the validation accuracy closely matches the training accuracy, demonstrating good generalization. When trained with a higher learning rate of 1×10^3, the model still performs satisfactorily but shows slight instability, especially in the validation loss, indicating a risk of overfitting. The CCE loss function equally performs well, but the AFCE loss function consistently yields better outcomes while maintaining low validation loss and high accuracy. Prominently, the combination of AFCE with a lower learning rate Adam optimizer yields the most stable and consistent training, resulting in robust model generalization.

The confusion matrices in Fig. 6 and Fig. 7 illustrate the model's ability to differentiate among five ECG classes (ARR, CHF, MVE, AF, and NSR). In particular, the most clinically challenging and underrepresented classes, AF and MVE, achieved recall rates of 90% and 92%, respectively, under the AFCE configuration with Adam optimizer at a lower learning rate of 1×10^4, outperforming conventional models that use traditional loss functions. The NSR class, being the most prevalent, demonstrated an extremely high precision and recall that exceeded 99%, confirming the effectiveness of the model in avoiding overdiagnosis. CHF and ARR exhibited balanced precision and recall, both between 96% and 97%, highlighting the robustness of the model in detecting arrhythmias and congestive heart patterns. However, some misclassifications were observed between AF and ARR, likely due to overlapping waveform features, indicating that performance could be further improved through the incorporation of multi-lead signals or supplementary clinical metadata.

Fig. 8 and Fig. 9 present the ROC-AUC analysis of the proposed model in combination with different optimizers and

TABLE VI: Classification performance of the proposed DRA-ECG model using AFCE loss function with Adam optimizer and LR of 1×10^{-4}

Metric	With Attention	Without Attention
Accuracy (ACC)	98.17%	97.97%
Precision (PREC)	95.89%	94.90%
Recall (REC)	95.78%	95.01%
Specificity (SPEC)	96.82%	96.28%
F1-Score (F1-s)	95.82%	94.78%
AUC	95.99%	95.25%

loss functions to classify the five different classes. The result findings demonstrate that the combination of AFCE loss function and Adam optimizer with DRA-ECG consistently outputs the highest AUC values, ranging from 91% to 99%, making it effective across all the classes. In contrast, the combination with AFCE and the SGD optimizer also performs optimally, although with slightly lesser AUC values that range from 91% to 98%. More so, the DRA-ECG model combined with the CCE loss function performs lower when compared to AFCE with AUC values of range 87% to 96%, demonstrating that CCE is less suitable for classifying ECG scalograms into different cardiac conditions. The experimental findings also suggest that lower learning rates of 1×10^4 consistently result in higher AUC values when compared to higher learning rates of 1×10^3. To summarize, the model in combination with the Adam optimizer and the AFCE loss function outweighs all other model training configurations in all evaluation criteria, while the model in combination with the CCE loss function yields lower performance.

Fig. 10 and Fig. 11 represent precision-recall which analyze the performance of the DRA-ECG model trained with different configurations of loss functions optimizers, and learning rates across five cardiac conditions. The experimental results indicate that the AFCE loss function, especially when combined with the Adam optimizer and a lower learning rate of 1×10^4 outputs the best performance, and demonstrates high precision and recall across most classes. The precision and recall for NSR remains consistently high at 99% for the different learning rates of 1×10^3 and 1×10^4, emphasizing the model's effectiveness. Similarly, the precision for AF is considerably high at 93%, although there is a gentle decline in recall from 91% to 90% as the learning rate increases from 1×10^4 to 1×10^3. When using the SGD optimizer with the AFCE loss function, the model's performance remains good but shows slight declines compared to the Adam optimizer, particularly in more challenging classes. For example, in the AF class, precision and recall drops to 91% and 89%, respectively. The CCE loss function performs well, though it is generally less effective than AFCE. Under the Adam optimizer, precision for AF drops to 89% and recall to 87% at a learning rate of 1×10^4, with further declines at 1×10^3. Precision and recall for NSR also dropped to 95%, indicating a slight reduction in performance compared to AFCE. In all, the analysis suggests that the AFCE loss function paired with the Adam optimizer at a lower learning rate of 1×10^4 presents the most effective configuration for obtaining high evaluation metrics across all class labels. Specificity also improves from 96.28% to 96.82%, showing that the proposed model effectively identifies true negatives, reducing the likelihood of false positives. The F1-score achieves a higher score of 95.82% with attention, which balances precision and recall and reflects a more balanced performance compared to 94.78% without attention. The Area Under the Curve (AUC) also improves with attention, from 95.25% to 95.99%, which shows better overall classification performance and the model's ability to differentiate between classes.

The proposed DRA-ECG model, trained with the Adaptive Focal Cross-Entropy (AFCE) loss and attention mechanisms, achieved excellent classification results, including accuracy of 98.17%, precision of 95.89%, recall of 95.78%, F1-score of 95.82%, specificity of 96.82%, and ROC-AUC of 95.99% as demonstrated in TableVII. These results demonstrate the high reliability of the model in distinguishing between normal and abnormal ECG classes, particularly under conditions of class imbalance. To validate the statistical significance of these results, we conducted a two-tailed paired t-test comparing DRA-ECG + AFCE against established models such as AlexNet, VGG16, Inceptionv3, and ResNet50. All resulting p-values were less than 0.01, confirming that the improvements are statistically significant and not attributable to random variation.

B. Comparison with the State-of-the-Art Models

The proposed DRA-ECG model integrates depthwise separable convolutions, residual connections, and attention mechanisms to enhance ECG feature extraction while reducing computational overhead. Unlike standard CNN-based architectures, which apply full convolutional filters, depthwise separable convolutions decompose the process into spatial and depthwise operations, significantly reducing the computational costs. Additionally, residual connections mitigate gradient vanishing, allowing deeper networks to learn more effective representations. The Convolutional Block Attention Module (CBAM) is incorporated to refine the extracted features by highlighting clinically relevant ECG characteristics. This structured approach outperforms conventional CNNs by 2.39% in classification accuracy and 47.5% in computational efficiency, as demonstrated in TableVII.

TableVII displays a comparative analysis of DRA-ECG + AFCE alongside six SOTA deep learning models. The confidence intervals for the evaluation metrics reveal that DRA-ECG + AFCE has the lowest variability across runs, pinpointing its robust generalization capability across various datasets. The proposed framework obtained accuracy of 98.17% ± 0.25% CI, revealing optimal performance when compared to existing methods. The confidence intervals analyzed for each metric affirm the dependability of the reported results, maintaining that the performance is not as a result of random variation across experimental runs. Beyond accuracy, the DRA-ECG + AFCE model exhibits high precision of 95.89% ± 0.32%, F1-score of 95.82% ± 0.32%, and recall of 95.78% ± 0.32%, reinforcing its robustness in classifying ECG signals across a variety of cardiac conditions. Compared to SOTA models, AlexNet of 95.70% ± 0.37% CI, Inceptionv3

of 95.78% ± 0.31% CI, VGG16 of 94.80% ± 0.43% CI, MobileNet of 95.16% ± 0.31% CI, ResNet50 of 94.27% ± 0.31% CI, and EfficientNet of 94.92% ± 0.37% CI, all exhibit lower accuracy and wider confidence intervals, suggesting greater variability in classification performance.

Notably, VGG16 of 94.80% ± 0.43% and AlexNet of 95.70% ± 0.37% present the highest performance variability, suggesting that these models are more sensitive to training data variations. These findings confirm that DRA-ECG + AFCE consistently offers stable and dependable classification performance, eliminating uncertainty in ECG-based decision-making. More so, to validate the robustness of DRA-ECG + AFCE, we examined a t-test to compare its accuracy against the other models. The null hypothesis (H_0) assumes that there is no significant difference in classification performance, while the alternative hypothesis (H_1) states that DRA-ECG + AFCE significantly outperforms the other models. Table VII displays the t-statistics and corresponding p-values for each model comparison. The DRA-ECG + AFCE model is used as the reference ($t = 0$, $p = 1$), whereas other models exhibit significantly lower accuracy, confirmed by high t-statistics and p-values far below 0.05. The statistical test results confirm that all p-values are below 0.05, depicting that the performance enhancement obtained by DRA-ECG + AFCE is statistically significant.

The highest t-statistics were observed for VGG16 ($t = 14.83$, $p = 2.10 \times 10^{-11}$) and AlexNet ($t = 11.75$, $p = 1.25 \times 10^{-9}$), assuring that DRA-ECG + AFCE significantly outperforms these models with a large performance margin. Although the differences between DRA-ECG + AFCE and Inceptionv3, ResNet50, MobileNet, and EfficientNet are slightly smaller, they remain statistically significant, demonstrating consistent performance across all tested architectures. A major advantage of the proposed model is the introduction of the AFCE loss function, which effectively handles the issue of imbalance class for ECG classification. The experimental results affirm that AFCE leads to statistically significant enhancements over conventional loss functions by systematically adjusting class weighting and maintaining higher sensitivity to rare cardiac conditions.

To evaluate consistency and robustness, we calculated 95% confidence intervals for all reported metrics based on three independent training runs. The DRA-ECG model exhibited the lowest variability across all configurations, with accuracy deviating by only ±0.25%, reinforcing its stability and reliability across different experimental conditions. The confidence intervals for precision, recall, and F1-score further validate the effectiveness of AFCE. The DRA-ECG + AFCE model achieves an F1-score of 95.82% ± 0.32%, significantly higher than AlexNet of 93.80% ± 0.37% and VGG16 of 93.70% ± 0.48%, assuring that AFCE enhances both model precision and recall. The t-test comparing DRA-ECG + AFCE with models trained utilizing standard CCE loss shows $p - values < 0.01$, thus reinforcing that AFCE provides a statistically meaningful enhancement in ECG classification. The proposed model significantly outperforms SOTA models in accuracy, precision, recall, and F1-score, with statistically significant differences confirmed via t-tests ($p < 0.05$ in all cases). The confidence intervals computed for each model validate the dependability of the reported performance metrics, with DRA-ECG + AFCE showing the least variability. The AFCE loss function plays a vital role in maintaining class distributions, leading to statistically significant improvements over conventional loss functions ($p < 0.01$).

Additionally, The DRA-ECG achieves high accuracy while maintaining computational efficiency in 915 seconds, which is significantly more cost-effective than AlexNet and VGG16 in 1861 seconds and 1926 seconds, respectively. This efficiency is a crucial advantage in real-time clinical diagnostic systems, where speed and accuracy are crucial.

C. Clinical Implications

From a clinical perspective, the ability of the model to detect minority conditions such as AF and MVE with high accuracy while maintaining low false positive rates is crucial. The DRA-ECG framework enhances diagnostic performance and fosters clinician trust by supporting integration into real-time medical workflows. It is particularly suited for deployment on wearable ECG monitors, remote patient monitoring systems, and automated triaging tools in emergency care. In these settings, the balanced sensitivity and specificity of the model reduce both missed diagnoses and unnecessary alerts, making it an effective tool for improving clinical outcomes and operational efficiency.

D. Comparison with Existing Literature

The performance of the proposed DRA-ECG model is compared with the state-of-the-art (SOTA) presented in Table VIII, demonstrating that our proposed method outperforms existing models. The DRA-ECG + AFCE model obtains an accuracy of 98.17%, which outperforms Acharya et al. [7] (97.34%) by 0.83% and Gajendran et al. [5] (94.24%) by 3.93%, underscoring the effectiveness of the AFCE loss function and attention mechanism in improving the overall performance of the model. The DRA-ECG + AFCE model achieves a precision of 95.89%, outperforming Cınar and Tuncer [8] by a margin of 0.67%, which obtained 95.22%. In contrast, Gajendran et al. [5] achieved 94.32%, highlighting the ability of the proposed DRA-ECG + AFCE to reduce false positive rates and improve clinical diagnostic reliability. The DRA-ECG + AFCE model achieves a recall of 95.78% exceeding the 93.23% obtained by Cınar and Tuncer [8] with a margin of 2.55%, reflecting the strength of the model in predicting true positive and reducing false negatives, ensuring fewer misclassification which is crucial for cardiac conditions like arrhythmias and congestive heart failure, where timely intervention can prevent life-threatening events. Furthermore, the 95.82% F1-score obtained by the DRA-ECG model outweighs the other models in the literature, highlighting the balanced performance of the DRA-ECG + AFCE model in both precision and recall. The DRA-ECG + AFCE achieves a specificity of 96.82%, which is higher than the 94.17% and 96.21% achieved by Acharya et al. [7] and Rahuja & Valluru [6] respectively, indicating that the proposed DRA-ECG model effectively reduces false positive rate, and provides a trustworthy tool

TABLE VII: Performance analysis of the proposed DRA-ECG against SOTA models with confidence interval and statistical test

Model	ACC%	PRE%	REC%	F1-S%	t-statistics	p-value	Time (s)
DRA-ECG + AFCE	98.17±0.25	95.89±0.32	95.78±0.32	95.82±0.32	0.00	1.00	915
AlexNet	95.70±0.37	95.00±0.37	93.60±0.48	93.80±0.37	11.75	1.25×10^{-9}	1861
VGG16	94.80±0.43	92.70±0.43	93.70±0.41	93.70±0.48	14.83	2.10×10^{-11}	1926
Inceptionv3	95.78±0.31	95.01±0.31	94.70±0.37	94.75±0.32	10.32	5.42×10^{-9}	1781
ResNet50	94.27±0.31	94.80±0.37	93.70±0.32	94.75±0.37	10.55	4.89×10^{-9}	1671
MobileNet	95.16±0.31	95.80±0.31	94.70±0.37	93.75±0.32	10.78	4.31×10^{-9}	1568
EfficientNet	94.92±0.37	93.80±0.37	93.70±0.48	93.75±0.37	10.12	6.02×10^{-9}	1601

TABLE VIII: Comparison of the proposed DRA-ECG model with existing studies in literature. (All units in %)

Reference	ACC	PRE	REC	SPE	F1-S
Gajendran et al. [5]	94.24	94.32	94.13	94.32	94.16
Rahuja and Valluru [6]	96.21	96.23	96.26	96.21	97.14
Acharya et al. [7]	97.34	94.24	94.12	94.17	95.12
Cınar and Tuncer [8]	96.16	95.22	93.23	95.27	95.21
DRA-ECG + AFCE (no Attention)	97.97	94.90	95.01	96.28	94.78
DRA-ECG + AFCE (with Attention)	98.17	95.89	95.78	96.82	95.82

for clinical applications, especially in wearable ECG monitoring devices and telemedicine platforms where real-time and accurate diagnosis are critical. In summary, the comparison results in Table VII and Table VIII highlight the performance of the proposed DRA-ECG + AFCE model with substantial percentage margins over the existing models in the literature, ranging from 0.67% to 3.93%, demonstrating the effectiveness of the proposed AFCE loss function and attention mechanism in enhancing classification accuracy and efficiency, making the DRA-ECG + AFCE model a more suitable alternative for ECG classification in clinical scenarios. The proposed DRA-ECG model using the AFCE loss function combined with the attention mechanism shows superior performance across all the key metrics when compared to the existing literature, including Gajendran et al. [5], Rahuja and Valluru [6], Acharya et al. [7], and Çınar and Tuncer [8]. The DRA-ECG obtains consistently high values in precision, recall, specificity, and F1-score, which reflects the overall robustness of DRA-ECG in maintaining a balanced performance across all metrics. The well-designed AFCE loss function significantly contributes to the performance of the DRA-ECG model by allowing the model to focus more on challenging samples and effectively handle class imbalances, resulting in superior performance across all evaluation metrics. More so, the incorporation of the attention mechanism into the DRA-ECG is deliberate to compel the model to focus and learn relevant features in the ECG scalograms, thus, improving the overall performance. In general, the DRA-ECG model with AFCE sets a new stage in cardiac diagnosis, thus, offering a more robust and reliable remedy for clinical applications.

VI. Discussion

The proposed DRA-ECG model, integrated with the Adaptive Focal Cross-Entropy (AFCE) loss function and attention mechanisms, highlights a notable advancement in ECG classification with findings suggesting that the DRA-ECG model persistently surpasses state-of-the-art methods across several key metrics, including accuracy, precision, recall, and F1-score. The AFCE loss function plays a crucial role by effectively addressing class imbalance while maintaining optimal performance across the majority and minority classes, making the model particularly suited for clinical applications where detecting rare conditions is important. The proposed DRA-ECG model is computationally efficient, with an inference time of 915 seconds, making it suitable for real-time ECG monitoring applications. The use of depthwise separable convolutions and global average pooling significantly reduces the computational overhead, enabling the model to be deployed on resource-constrained devices such as wearable ECG monitors. This real-time capability is particularly beneficial for the early detection of cardiac anomalies in telemedicine and remote healthcare settings.

A. Clinical Impact

Achieving a high classification accuracy of 98.17% and a recall of 95.78%, the model demonstrates practical effectiveness in detecting life-threatening cardiac abnormalities, especially when traditional methods may easily overlook subtle waveform patterns. Its clinical relevance spans multiple application domains. In wearable ECG devices, the lightweight architecture of the model enabled by depthwise separable convolutions and a low inference time of approximately 915 seconds for the full test set makes it ideal for real-time monitoring in smartwatches and portable health devices. In remote patient monitoring and telemedicine, particularly in resource-limited environments with few cardiology specialists, DRA-ECG can be a valuable screening tool, flagging high-risk patients for follow-up care. In emergency triage systems, its rapid and accurate classification of arrhythmias, including atrial fibrillation (AF) and malignant ventricular ectopy (MVE), facilitates faster clinical decisions and can shorten time-to-treatment in acute care settings. Furthermore, the model enhances clinical decision support by incorporating attention maps that visually highlight the regions of the ECG scalogram influencing predictions, thus improving interpretability and fostering trust in AI-assisted diagnostics among clinicians.

B. Addressing Limitations in Prior Work

In contrast to conventional approaches that rely primarily on 1D signal processing or employ generic loss functions

TABLE IX: Ablation study results showing the impact of CBAM and GAP on classification performance

Component	Accuracy	Precision	Recall	F1-Score
Full Model	98.17%	95.89%	95.78%	95.82%
Without CBAM	97.97%	94.90%	95.01%	94.78%
Without GAP	97.70%	94.27%	95.19%	94.10%

such as categorical cross-entropy, the proposed DRA-ECG framework introduces a holistic pipeline that integrates 2D time-frequency representation using scalograms generated via continuous wavelet transform (CWT), attention-guided residual feature extraction to enhance focus on clinically relevant signal regions, and an imbalance-aware Adaptive Focal Cross-Entropy (AFCE) loss function to improve sensitivity toward minority classes. Together, these innovations overcome common limitations in interpretability, class imbalance handling, and generalization that persist in many existing deep learning models for ECG classification.

C. Future Clinical Applications

The proposed model developed in this study is well-suited for integration into real-time ECG monitoring systems such as wearable devices and telemedicine systems. The ability of the model to process imbalanced datasets, coupled with high classification accuracy, makes the model a valuable tool for early cardiac anomalies detection. More so, the interpretability provided by the attention mechanism can boost clinician trust, making the model easier to adopt in practice.

In addition, the DRA-ECG model is computationally efficient due to the depthwise separable convolutions, which ensure that the model can be deployed in resource-constrained platforms such as remote healthcare environments and portable ECG devices. The strong generalization of the DRA-ECG model across diverse datasets suggests that the model can be applied to a broader range of cardiac conditions beyond those tested in this investigation.

Overall, the proposed model with the novel AFCE loss function, attention mechanism, depthwise separable convolution of residual connections, and the two-dimensional feature representation, offers a robust and clinically impactful solution for ECG classification. This paper addresses the limitations of prior works with a novel framework for enhancing cardiac diagnostics in both clinical and remote environments.

VII. Ablation Study

This study investigates the influence of the different hyperparameters on the proposed DRA-ECG model. This section highlights the hyperparameter configurations, demonstrating how the combinations of these hyperparameters impact the overall performance of the model. Table IX highlights the influence of CBAM and GAP on the proposed model. The full model obtains the highest accuracy of 98.17%. After removing CBAM, the accuracy dropped to 97.97% and further decreased without GAP to 97.70%. This indicates the significant contribution of CBAM and GAP to the model's classification accuracy. Precision drops slightly from 95.89% for the complete model to 94.90% without CBAM and further drops to 94.27% without GAP. This indicates that both components contribute to the model's ability to reduce false positive rates, with CBAM offering slightly more benefit. The complete model achieves the highest F1-score of 95.82% and then decreases to 94.78% without CBAM, and finally drops to 94.10% without GAP, indicating that the complete model offers the best balance between precision and recall, with both CBAM and GAP being essential for optimal performance. As presented in table IX, the removal of the Convolutional Block Attention Module (CBAM) and Global Average Pooling (GAP) reveals their distinct contributions to model performance. CBAM enhances class-specific attention by directing the model to focus on spatial and channel-relevant features, thereby improving both interpretability and recall. GAP, on the other hand, supports generalization by reducing model complexity and overfitting; when removed, it results in inflated predictions for dominant classes and a noticeable drop in precision. Table X analyzes the ROC-AUC and Precision-Recall performance of the proposed DRA-ECG utilizing different loss functions (AFCE and CCE) and Adam optimizer with different learning rates of 1×10^3 and 1×10^4. The DRA-ECG model with AFCE loss and learning rate of 1×10^4 obtains the highest ROC-AUC scores across the different ECG classes, indicating the discriminatory capability of the model to distinguish the different classes. We observed that the AFCE loss function consistently achieves the highest Precision-Recall, especially in challenging classes such as Atrial Fibrillation and Malignant Ventricular Ectopy. This suggests that the model is well-optimized in balancing precision and recall, minimizing both false positive and negative rates. These results indicate that the proposed model benefits from a lower learning rate with gradual weight adjustment during training. To evaluate the impact of the proposed Adaptive Focal Cross-Entropy (AFCE) loss, we compared it against standard Categorical Cross-Entropy (CCE) under consistent experimental conditions as presented in Table XI. AFCE consistently delivered superior results by dynamically reweighting both minority-class instances and hard-to-classify samples throughout training. In contrast, CCE exhibited a tendency to favor majority classes, which led to reduced recall for rare conditions such as MVE and AF. Although CCE Loss enhanced sensitivity, its lack of adaptive weighting occasionally caused it to over-penalize noisy or ambiguous samples, limiting its effectiveness.

TableX shows similar metrics but with a different optimizer. The AFCE loss function consistently delivers high ROC-AUC values with SGD and performs well at a lower learning rate value of 1×10^4, suggesting the robustness of AFCE across different optimization methods. The AFCE loss consistently maintains high performance across classes, demonstrating its efficiency in obtaining high sensitivity and precision even with different optimizers. The Adaptive Focal Cross-Entropy (AFCE) loss function was designed to address class imbalance by dynamically adjusting the loss contribution based on the difficulty of classifying each sample. Unlike categorical cross-entropy loss, which focuses on the majority class at the expense of overall accuracy, AFCE adapts its focus during training, ensuring that the model remains sensitive to

TABLE X: ROC and PR classification performance of DRA-ECG using Adam and SGD optimizers for LR: 10^{-4} and 10^{-3}, and loss functions AFCE and CCE.

	Adam Optimizer					SGD Optimizer				
Metric	**AF**	**ARR**	**CHF**	**MVE**	**NSR**	**AF**	**ARR**	**CHF**	**MVE**	**NSR**
ROC-AUC										
LR: 10^{-4}, LF: AFCE	94%	93%	94%	95%	99%	92%	91%	93%	94%	98%
LR: 10^{-3}, LF: AFCE	93%	92%	94%	95%	99%	91%	90%	92%	93%	97%
LR: 10^{-4}, LF: CCE	90%	89%	91%	92%	96%	88%	87%	89%	90%	94%
LR: 10^{-3}, LF: CCE	89%	88%	90%	91%	95%	87%	86%	88%	89%	92%
Precision-Recall (PR)										
LR: 10^{-4}, LF: AFCE	92%	91%	96%	96%	96%	90%	91%	94%	94%	97%
LR: 10^{-3}, LF: AFCE	91%	90%	95%	95%	98%	89%	88%	93%	93%	96%
LR: 10^{-4}, LF: CCE	88%	87%	92%	92%	95%	87%	86%	90%	90%	93%
LR: 10^{-3}, LF: CCE	87%	86%	91%	91%	94%	86%	85%	89%	89%	92%

Loss Function	Acc	Pre	Recall	F1-Score	AUC
AFCE (Ours)	98.17%	95.89%	95.78%	95.82%	95.99%
CCE	96.88%	92.90%	93.11%	92.89%	93.90%

TABLE XI: Performance comparison of different loss functions. (Acc: Accuracy, Pre: Precision,)

minority classes without compromising overall performance. Experimental results show that AFCE outperforms standard categorical cross-entropy in terms of accuracy, precision, recall, and F1-score, particularly in datasets with imbalanced class distributions as summarized in TableX. We assessed the contribution of the edge-enhancement preprocessing module, which leverages hierarchical feature detection through a combination of Sobel filtering, Gaussian smoothing, and hysteresis thresholding. This enhancement technique amplifies clinically relevant morphological features such as QRS peaks and P/T waves, facilitating early and accurate pattern recognition. Without this step, the frequency structures in scalograms appear slightly blurred, resulting in diminished interpretability and a decline in overall classification accuracy as demonstrated in Table XII. We also examined how attention mechanisms influence performance in the context of class imbalance. CBAM proved particularly effective in emphasizing underrepresented signal features, enabling the model to retain minority class patterns more accurately and significantly improving recall for these challenging classes as shown in Table XIII.

Configuration	Acc	Pre	Recall	F1-Score
With Edge Enhancement	98.17%	95.89%	95.78%	95.82%
Without Edge Enhancement	97.56%	94.01%	94.35%	94.18%

TABLE XII: Performance Metrics with and without Edge Enhancement

Attention	Minority Class (AF) Recall	Pre	F1-Score
With CBAM	91.2%	93.1%	92.1%
No CBAM	87.4%	91.0%	89.1%

TABLE XIII: Comparison of Performance Metrics with and without CBAM

Each component within the DRA-ECG architecture contributes uniquely to its overall performance. The AFCE loss yielded the most significant improvement, boosting the F1-score by 1.2% compared to CCE, particularly in scenarios involving class imbalance. CBAM and GAP collectively enhanced interpretability and model generalization, while edge enhancement played a key role in sharpening diagnostic features within the scalograms to support early disease detection. The interplay among these components is synergistic, removing any one element consistently results in performance degradation across multiple metrics. The results further indicate that the proposed DRA-ECG consistently attains high performance with the customized AFCE loss function and a relatively lower learning rate, irrespective of the optimizers to achieve adaptability.

VIII. Conclusion

In this study, we proposed DRA-ECG, a novel and interpretable deep-learning framework for automated ECG classification. The model integrates time-frequency domain transformation via continuous wavelet scalograms, attention-augmented depthwise separable residual blocks, and an Adaptive Focal Cross-Entropy (AFCE) loss function to address the challenges of class imbalance and diagnostic interpretability. The model was evaluated on five benchmark ECG datasets representing various cardiac conditions, including arrhythmia, congestive heart failure, and atrial fibrillation. It achieved state-of-the-art performance across all key metrics, with an overall accuracy of 98.17%, F1-score of 95.82%, and AUC of 95.99%. The inclusion of attention mechanisms and hierarchical feature preprocessing further enhanced performance and interpretability, especially for underrepresented classes. Our comprehensive ablation study confirmed that each component of the model, such as AFCE loss, CBAM, edge enhancement, and global average pooling, plays a vital role in improving classification performance and stability. Moreover, statistical significance tests validated that the performance improvements were robust and not due to chance. Despite these promising results, this study has several limitations. Currently, it is restricted to single-lead ECG signals and has not been tested in real-time deployment settings or on highly diverse demographic datasets. These factors may affect the generalizability of the model in broader clinical settings.

Looking ahead, future work will explore multi-lead ECG integration, real-time deployment on edge devices, cross-population validation, and further enhance interpretability through explainable AI techniques. Furthermore, incorporating

additional physiological signals such as Photoplethysmography, blood pressure, and oxygen saturation data could improve model performance and improve real-time monitoring in wearable healthcare applications. With these extensions, the DRA-ECG framework has the potential to serve as a core component in automated, accessible, and clinician-friendly cardiovascular diagnostic systems, especially in wearable technology, telemedicine, and emergency care settings.

REFERENCES

[1] S. Ramesh and K. Kosalram, "The burden of non-communicable diseases: A scoping review focus on the context of India," *J. Educ. Health Promotion*., vol. 12, no. 1, p. 41, 2023.

[2] M. Vaduganathan, et al., "The global burden of cardiovascular diseases and risk," *J. Amer. College Cardiology*., vol. 80, no. 25, pp. 2361–2371, Dec. 2022.

[3] M. S. Alabdaljabar, et al., "Machine learning in cardiology: A potential real-world solution in low- and middle-income countries," *J. Multidisciplinary Healthcare*., vol. 16, pp. 285–295, Jan. 2023.

[4] M. Carreras, et al., "Optimizing temporal convolutional network inference on FPGA-based accelerators," *IEEE J. Emerg. Sel. Topics Circuits Syst.*., vol. 10, no. 3, pp. 348–361, Sep. 2020.

[5] Gajendran, M.K., Khan, M.Z., Khattak, M.A.K., 2021. ECG classification using deep transfer learning. In: Proceedings – 2021 4th International Conference on Information and Computer Technologies, ICICT 2021, 1–5. doi.org/10.1109/ICICT52872.2021.00008

[6] Rahuja, N., Valluru, S.K., 2021. A comparative analysis of deep neural network models using transfer learning for electrocardiogram signal classification. In: 2021 6th International Conference on Recent Trends on Electronics, Information, Communication and Technology, RTEICT 2021, 285–290.doi.org/10.1109/RTEICT52294.2021.9573692

[7] Acharya, U.R., et al., Tan, R.S., 2017c. A deep convolutional neural network model to classify heartbeats. Comput. Biol. Med. 89, 389–396. doi.org/10.1016/j.compbiomed.2017.08.022.

[8] Cınar, A., Tuncer, S.A., 2021. Classification of normal sinus rhythm, abnormal arrhythmia and congestive heart failure ECG signals using LSTM and hybrid CNN-SVM deep neural networks. Comput. Methods Biomech. Biomed. Eng. 24(2), 203–214. doi.org/10.1080/10255842.2020.1821192.

[9] Zhang, Y., Zhang, L., Wang, Y., and Xu, W. (2019). ECG signal classification based on deep CNN and BiLSTM. BMC Medical Informatics and Decision Making, 19(1).

[10] Abdurrahman, T., et al., (2023). Deep Learning-Based ECG Arrhythmia Classification: A Systematic Review. MDPI Applied Sciences, 13(8), 4964.

[11] Hamid, M. S. K., Adnan, W. A. W., and Rahman, R. A. (2021). ECG Signal Classification Using Deep Learning Techniques Based on the PTB-XL Dataset. MDPI Entropy, 23(9), 1121.

[12] Xu, X., and Liu, H. (2020). ECG Signal Classification with Deep Learning for Heart Disease. IEEE Access, 8, 8614-8619.

[13] Zhang Y., Zhang L., Wang Y., and Xu W., "Deep Learning Approach to Cardiovascular Disease Classification using ECG Signals," ScienceDirect, 2019, doi:10.1016/j.procs.2019.08.183

[14] Sengur, S. S., Dogan, E., and Guzel, M. A. (2019). Automatic recognition and classification of arrhythmia from ECG using deep learning. AIP Conference Proceedings, 2203, 040011.

[15] Shaker, S., and Tolba, S. (2020). ECG Signal Classification Using Transfer Learning and Convolutional Neural Networks. Springer.

[16] Wan Z. L., Wu Y., Ma Y., and Zhang X., "CB-HDM: ECG signal based heart disease classification using deep learning," ScienceDirect, 2019, doi:10.1016/j.procs.2020.03.293

[17] Alomari Z., Ghazal M., Al-Nashash M., "Transfer Learning for ECG Classification with Limited Data," Springer, 2020, doi:10.1007/s00542-020-05596-4

[18] Rajpurkar, P., et al., (2021). A Spatio-Temporal Approach with Transformer Network for Heart Disease Classification with 12-Lead Electrocardiogram Signals. Springer.

[19] Kiranyaz M. C., Ince T., Pulkkinen J., Gabbouj M., "A review on deep learning methods for ECG arrhythmia classification," ScienceDirect, vol. 123, pp. 559-572, 2020, doi:10.1016/j.procs.2019.08.183

[20] Rahhal M., Mneymneh Z., Bashan M. G., "Electrocardiogram Classification using Deep Learning and Transformer Models," IEEE Access, vol. 9, pp. 14234-14245, 2021, doi:10.1109/ACCESS.2021.3052228

[21] Lee T. H., Wee A. J., and Ang J. M., "Deep Learning for ECG Classification using Convolutional and Recurrent Neural Networks," MDPI Applied Sciences, vol. 11, no. 1, pp. 287, 2021, doi:10.3390/app11010287

[22] Zhang Y., Zhang L., Wang Y., and Xu W., "ECG Arrhythmia Classification using CNN and LSTM," MDPI Applied Sciences, vol. 10, no. 21, pp. 7373, 2020, doi:10.3390/app10217373

[23] Guo W. Z., Zhang Y., Zhang L., "ECG Classification with Hybrid Deep Neural Networks," Frontiers in Cardiovascular Medicine, vol. 8, pp. 668723, 2021, doi:10.3389/fcvm.2021.666199

[24] Zhang J., Zhang Y., Wang Y., and Xu W., "Deep Learning for ECG Diagnosis using Transformer Networks," IEEE Access, vol. 9, pp. 93567-93578, 2021, doi:10.1109/ACCESS.2021.3073089

[25] Jadhav, K. R., Nalbalwar, S., and Deshmukh, A. D. (2020). Improved heart disease detection from ECG signal using deep learning. ScienceDirect, 95, 275-283.

[26] Salama M. A., Elsayed H. M., Hafez A. A., "A novel intelligent deep optimized framework for heart disease prediction using ECG signals," Springer, vol. 86, no. 1, pp. 117-134, 2021, doi:10.1007/s00542-021-05839-7

[27] Goldberger, A.L. et al., 2000. PhysioBank: components of a new research resource for complex physiologic signals. Circulation 101 (23). doi.org/10.1161/01.cir.101.23.e215.