Scientific African 28 (2025) e02687

Contents lists available at ScienceDirect

Scientific African

journal homepage: www.elsevier.com/locate/sciaf

GAF-GradCAM: Guided dynamic weighted fusion of temporal and frequency GAF 2D matrices for ECG-based arrhythmia detection using deep learning

Zakaria Khatar [a,*], Dounia Bentaleb [b], Noreddine Abghour [a], Khalid Moussaid [a]

[a] LIS Labs Faculty of Sciences Ain Chock, Hassan II University, Casablanca, Morocco
[b] Faculty of Science Techniques, University Hassan II, Mohammedia, Morocco

ARTICLE INFO

Editor name: DR B Gyampoh

Keywords:
ECG signal analysis
Gramian Angular Field
Grad-CAM
Dynamic fusion
Arrhythmia classification
Deep learning

ABSTRACT

This study introduces an innovative approach for arrhythmia classification that employs a Grad-CAM-guided dynamic weighted fusion of temporal and frequency features extracted from electrocardiogram (ECG) signals. By transforming ECG signals into two-dimensional Gramian Angular Field (GAF) matrices, the proposed method effectively captures temporal dynamics via Gramian Angular Summation Fields (GASF) and frequency dependencies from features extracted using Continuous Wavelet Transform (CWT) and refined through Principal Component Analysis (PCA). The Grad-CAM-guided dynamic fusion adaptively assigns importance to these complementary feature types based on their relevance for each input, enhancing both classification accuracy and interpretability. Optimizing this fusion process fine-tunes the balance between temporal and frequency information, thus focusing the model on the most critical ECG features. As a result, training accuracy reached 99.68% and validation accuracy 98.78%, alongside a substantial reduction in loss, underscoring the efficacy of Grad-CAM-guided fusion in integrating essential ECG features and advancing arrhythmia detection accuracy. Building on this fusion framework, this study further proposes a Hybrid Parallel-Residual Architecture specifically tailored for arrhythmia detection, integrating parallel and residual connections with Bidirectional Long Short-Term Memory (Bi-LSTM). This architecture ensures robust feature extraction and precise classification, achieving up to 98.75% accuracy, 99.14% sensitivity, and a 98.97% F1 score across multiple ECG leads, thereby surpassing traditional methods.

Introduction

Cardiac arrhythmias, irregularities in the heart's rhythm, remain a critical concern in cardiology due to their diverse manifestations and potentially severe health implications, including stroke or sudden cardiac death [1]. Accurate detection of arrhythmias from electrocardiogram (ECG) signals is paramount for timely diagnosis and treatment. However, traditional approaches that rely on manual inspection or basic signal processing techniques often exhibit limitations in efficiency and precision [2,3]. In recent years, the broader integration of artificial intelligence (AI) into healthcare has accelerated the adoption of machine learning (ML) and deep learning (DL) techniques for classifying cardiovascular diseases. These advanced data-driven methodologies hold promise for enhanced diagnostic accuracy, reduced manual overhead, and improved patient outcomes [4]. In particular, the rapid advancements

* Correspondence to: Laboratory of electrical engineering and smart systems, ENSET, Hassan II University of Casablanca, Morocco.
E-mail addresses: zakaria.khatar-etu@etu.univh2c.ma (Z. Khatar), dounia.bentaleb-etu@etu.univh2c.ma (D. Bentaleb), nabghour@gmail.com (N. Abghour), khmoussaid@gmail.com (K. Moussaid).

https://doi.org/10.1016/j.sciaf.2025.e02687
Received 10 December 2024; Received in revised form 22 March 2025; Accepted 3 April 2025
Available online 12 April 2025
2468-2276/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

in deep learning have revolutionized ECG analysis by enabling models to automatically learn salient features directly from raw signals, often surpassing traditional methods in diagnostic performance [5,6].

Despite these gains, a key challenge remains: **most existing methods do not fully leverage both temporal and frequency-domain features of ECG signals in a single adaptive framework**. Additionally, they often lack a targeted mechanism to highlight which aspects of the ECG (e.g., waveform intervals vs. spectral components) are most critical for classification in any given case. As a result, many current approaches either focus on a single domain (temporal or frequency) or rely on fixed-weight fusion, which may overlook subtle but diagnostic arrhythmia cues [7,8].

Aims of this study

Primary Objective. The primary objective of this work is to develop and evaluate a robust, accurate, and interpretable arrhythmia detection method by fusing both temporal and frequency-domain information derived from ECG signals. The study proposes transforming ECG time series into Gramian Angular Field (GAF) matrices to capture temporal dynamics, while leveraging wavelet transforms to extract salient frequency features. The integration of Gradient-weighted Class Activation Mapping (Grad-CAM) aims to enhance interpretability by highlighting key ECG regions that drive classification, thereby fostering clinical trust in deep learning methodologies.

Scope. The scope of this study resides within the broader context of applying AI-driven methods for cardiovascular disease classification. Although this experiment focuses on the MIT-BIH Arrhythmia Database as a benchmark for arrhythmia detection, the proposed methodology lends itself to a wide range of cardiovascular pathologies and varied ECG datasets. By combining temporal-frequency fusion with an interpretable mechanism like Grad-CAM, the complexity and heterogeneity often encountered in real-world clinical ECG signals are addressed in this study. The aim is to establish a versatile tool for healthcare practitioners, offering automated, accurate, and explainable support in diagnosing cardiovascular disorders.

Key contributions

- **Grad-CAM-Guided Dynamic Weighted Fusion:** We propose a novel mechanism that adaptively fuses temporal and frequency features based on their relevance for each input, as highlighted by Grad-CAM activation maps.
- **Hybrid Parallel-Residual Architecture:** A new deep learning design that combines parallel convolutional paths with residual connections and a Bi-LSTM layer, enabling robust extraction of multi-scale ECG features from the fused GAF matrices.
- **Improved Interpretability:** By incorporating Grad-CAM, our framework provides visual explanations pinpointing which ECG segments drive the classification, offering clearer insights for clinical validation.
- **State-of-the-Art Performance:** Our method achieves up to 98.75% accuracy on the MIT-BIH Arrhythmia Database, surpassing existing techniques in both accuracy and sensitivity.

Literature review

The analysis of ECG signals has traditionally relied on signal processing techniques such as wavelet transforms and Fourier analysis [9–11]. While these conventional methods have been instrumental in decomposing ECG signals into their frequency components, they often struggle to capture the complex morphology of arrhythmias and are sensitive to noise and inter-patient variability. For example, Khatar et al. (2024) [11,12] demonstrated that classical transformations may fail to highlight subtle signal features critical for accurate diagnosis.

With the advent of deep learning, particularly Convolutional Neural Networks (CNNs), ECG analysis has undergone a significant transformation [13,14]. CNNs are capable of autonomously learning hierarchical features directly from raw data, which has led to marked improvements in arrhythmia classification performance. However, despite these advances, CNN-based models sometimes struggle to capture both local details and global contextual information inherent in ECG signals.

Recent studies have addressed these shortcomings by integrating attention mechanisms into CNN architectures. Zhang et al. (2023) [15] and Liu et al. (2022) [16] have proposed spatial attention strategies that enable models to focus on the most informative parts of the signal, thereby enhancing classification accuracy. Furthermore, alternative representations of ECG signals—such as Gramian Angular Field (GAF) and Markov Transition Field (MTF)—have been introduced to capture complex temporal and probabilistic patterns. Li et al. (2021) [17] and Wang et al. (2022a) [18] demonstrated that transforming ECG signals into 2D images via these methods can improve both interpretability and performance, offering a novel perspective on ECG data.

Fusion techniques that combine multiple signal representations have emerged as a powerful strategy to further boost classification outcomes. Multi-view fusion approaches, as presented by Tan et al. (2023) [19], integrate complementary features from different representations to enhance arrhythmia detection. Additionally, Chen et al. (2022) [20] and Zhu et al. (2022) [21] have shown that combining deep learning features with handcrafted features can capture both detailed and abstract signal characteristics, leading to superior performance. Complementarily, Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, have been successfully applied to model the temporal dependencies in ECG signals [22], while recent CNN-Transformer architectures have set new benchmarks by merging local feature extraction with global context modeling [23].

Another crucial aspect of modern ECG analysis is model interpretability. Explainable AI techniques, such as Gradient-weighted Class Activation Mapping (Grad-CAM), have been developed to provide visual insights into the regions of the ECG that drive

Table 1
Comparison of techniques in existing methods versus our proposed approach (methodological overview).

Reference	Preprocessing/Feature Extraction	Fusion Mechanism	Interpretability	Architecture
Zhang et al. (2023) [15]	GAF (temporal only)	None (single-domain)	Not used	Simple CNN
Chen et al. (2022) [20]	Wavelet transform + handcrafted	Static combination	Partial (via attention)	CNN + Transformer
Tan et al. (2023) [19]	Multi-view signal transforms	Multi-view fusion	Basic activation maps	CNN + RNN
Ours (2025)	**CBD filtering + GAF (temporal & freq) + CWT + PCA**	**Dynamic weighting guided by Grad-CAM**	**Grad-CAM overlays**	**Parallel-Residual + Bi-LSTM**

classification decisions [24,25]. These methods help bridge the gap between complex deep learning models and clinical trust, ensuring that model predictions can be interpreted in a meaningful way.

Despite these significant advances, challenges such as class imbalance and dataset variability remain. Weimann et al. (2021) [26] highlighted the importance of balanced and diverse datasets to ensure that models generalize effectively across different patient populations. In response to these challenges, recent innovations have focused on transfer learning and multi-modal approaches to improve robustness and performance [27,28].

Comparison of existing techniques

Table 1 provides a concise summary comparing different ECG-based arrhythmia detection approaches. Rather than focusing on final performance metrics, this table highlights the primary preprocessing/extraction strategies, fusion mechanisms, interpretability methods, and architectural choices.

By integrating a hybrid CBD filter for preprocessing with a Grad-CAM-guided dynamically weighted fusion of temporal and frequency GAF matrices, followed by a hybrid parallel-residual architecture for classification, our method exploits the strengths of existing techniques to improve both the accuracy and interpretability of ECG-based arrhythmia detection.

Dataset and signal preprocessing

Dataset

The dataset used for this research is the well-established MIT-BIH Arrhythmia Database [29], a benchmark resource in cardiac research and one of the most widely used public datasets for ECG-based arrhythmia classification. The recordings were originally obtained from real-world ambulatory ECGs and are meticulously annotated by medical professionals, ensuring high accuracy in the diagnostic labels. These annotations cover a wide variety of arrhythmic conditions ranging from more common abnormalities such as premature ventricular contractions to more complex conditions like atrial fibrillation and ventricular tachycardia.

By including both normal and pathological cases, the MIT-BIH Arrhythmia Database ensures that machine learning models are exposed to a broad spectrum of ECG morphologies and rhythm patterns. In this study, this diversity serves as an ideal foundation for training and evaluating the proposed dynamic weighted fusion approach guided by Grad-CAM. Ultimately, the rich variability of the data provides a robust test bed for improving both the accuracy and interpretability of arrhythmia detection models.

ECG signal preprocessing

Preprocessing ECG signals is an essential step to ensure high-quality extraction of both temporal and frequency domain features, which are later transformed into *Gramian Angular Field (GAF)* matrices for arrhythmia classification. The main goal of this stage is to remove noise and artifacts—such as baseline wander, power line interference, and muscle noise—that can distort key cardiac features and degrade the quality of the GAF matrices. The sequence of filters applied during the preprocessing phase is illustrated in Fig. 1. This diagram shows the order in which each filtering technique is used to clean the ECG signal, ensuring optimal preparation for feature extraction as proposed in [30]:

In this study, a series of filters designed was applied to enhance the signal-to-noise ratio (SNR) and preserve the essential features of the ECG signal. These filters, each with specific purposes, ensure that the signals are ready for transformation into 2D matrices that represent both temporal and frequency domain features. The filtering methods used are summarized in Table 2:

Once this filtering pipeline is applied, the ECG signals are significantly cleaner, allowing for accurate extraction of temporal and frequency features. The effects of the filtering process are demonstrated in Fig. 2, where the signal before and after filtering is shown:

This improved signal clarity is crucial for generating high-quality GAF matrices that accurately capture both the temporal and frequency-based features of the ECG signal, ultimately enhancing the precision of arrhythmia classification.

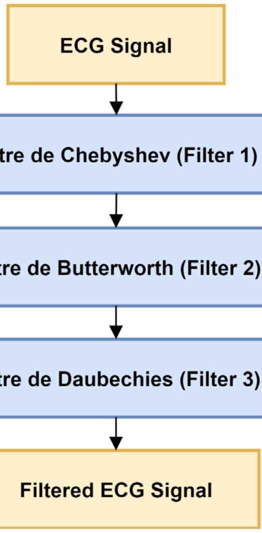

Fig. 1. Sequence of filters applied during ECG signal preprocessing.

Table 2
Filtering techniques used for ECG signal preprocessing.

Filter	Purpose and effect
Chebyshev filter	Removes high-frequency noise with sharp attenuation, ensuring that unwanted frequencies are eliminated while preserving key temporal features of the ECG signal.
Butterworth filter	Smoothens the signal with a flat frequency response, effectively eliminating low-frequency noise such as baseline wander, which enhances the quality of frequency domain features.
Daubechies wavelet filter	Decomposes the signal at multiple time scales, highlighting critical cardiac events (QRS complex, P wave, T wave) and enabling multi-resolution analysis for both temporal and frequency-based feature extraction.

Proposed model overview

This section provides a general overview of the proposed model for arrhythmia detection using ECG signals. The solution is motivated by the limitations of existing methods in capturing the complex interplay between temporal dynamics and frequency information inherent in ECG signals. Traditional approaches often struggle to integrate these complementary features, resulting in suboptimal classification performance and limited interpretability. To address these challenges, the proposed model leverages advanced signal transformations and deep learning techniques to combine both temporal and frequency domain features.

Specifically, the model begins with ECG signal preprocessing and transforms the data into 2D Gramian Angular Field (GAF) matrices. These matrices capture temporal dynamics through Gramian Angular Summation Fields (GASF) and encapsulate frequency dependencies via features extracted using the Continuous Wavelet Transform (CWT) and subsequently reduced through Principal Component Analysis (PCA). This rich, image-based representation enables the model to capture subtle and complex patterns in the ECG data.

The key novelty of the proposed approach is the Grad-CAM-guided dynamic weighted fusion mechanism, which adaptively assigns importance to the temporal and frequency features based on their relevance for each input. This dynamic fusion not only enhances classification accuracy by focusing on the most informative features but also improves interpretability by visually highlighting the critical regions in the ECG signal.

Following the fusion step, the fused features are fed into a hybrid deep learning architecture specifically designed for arrhythmia detection. This architecture incorporates parallel convolutional layers and residual connections alongside a Bidirectional Long Short-Term Memory (Bi-LSTM) network. The combination of these components ensures robust feature extraction and precise classification by capturing both local details and long-range dependencies in the signal.

A visual representation of the model is shown in Fig. 3:

The subsequent sections provide detailed descriptions of each component of the model, including the preprocessing pipeline, the Grad-CAM-guided fusion process, and the hybrid parallel-residual architecture.

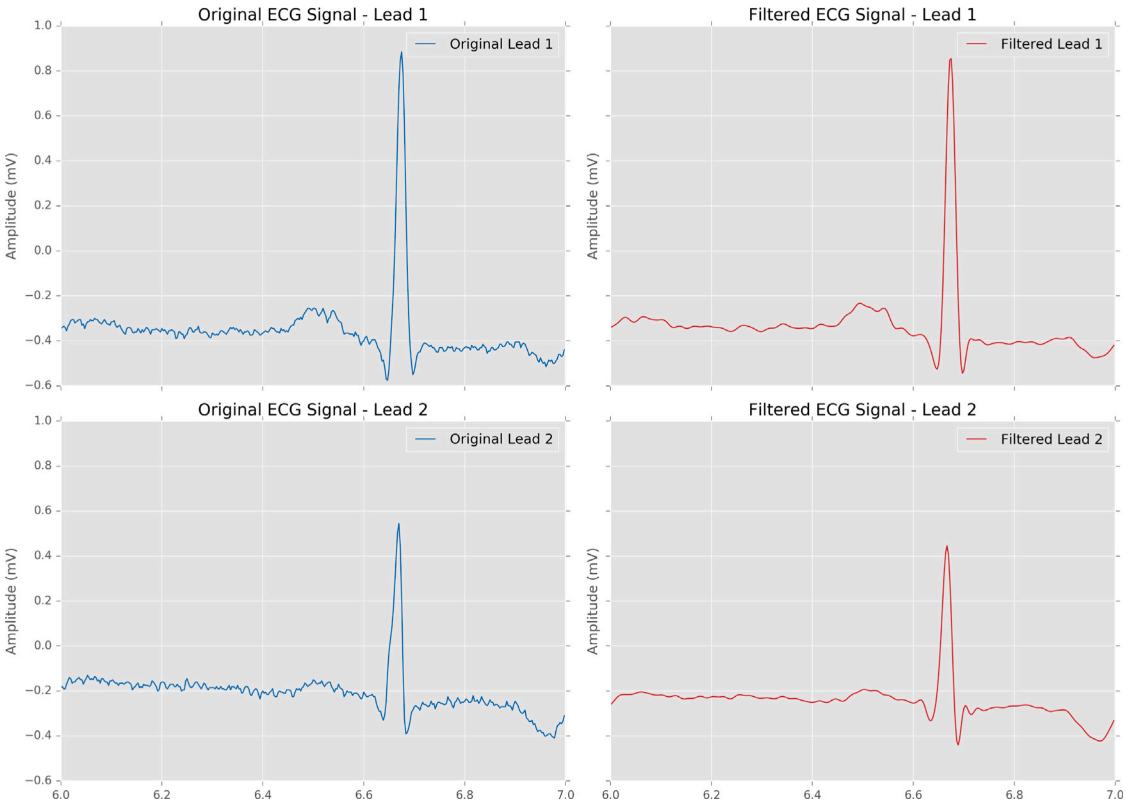

Fig. 2. ECG signal before and after filtering.

Feature extraction and GAF generation

In this section, the extraction of both temporal and frequency domain features was described from ECG signals, followed by their transformation into Gramian Angular Field (GAF) matrices. These matrices serve as a powerful visual tool for analyzing complex patterns within ECG signals, which is critical for accurate arrhythmia classification.

Temporal feature extraction

Temporal features were extracted by first filtering the ECG signal to remove noise, followed by transforming the signal into a Gramian Angular Summation Field (GASF) image.

Step 1: ECG signal filtering. The ECG signal was processed using a bandpass filter to retain the frequencies of interest (0.5 to 50 Hz) and remove noise. The filtered signal, denoted as $X(t)$, represents the cleaned version of the original ECG signal.

Step 2: Transformation into GASF. To capture the temporal dynamics of the ECG signal, the Gramian Angular Summation Field (GASF) transformation was applied. GASF is a technique that converts time-series data into a two-dimensional matrix by encoding the angular information of the signal into a polar coordinate system.

Given a normalized ECG signal $X(t)$ with values rescaled to the interval $[-1, 1]$, each value X_i can be mapped to an angle ϕ_i in polar coordinates:

$$\phi_i = \cos^{-1}(X_i), \quad \text{where} \quad X_i \in [-1, 1] \tag{1}$$

The GASF matrix is then computed as the summation of the cosines of these angles:

$$\text{GASF}(i, j) = \cos(\phi_i + \phi_j) \tag{2}$$

where i and j are indices representing different time points in the signal. This results in a symmetric matrix where each element reflects the temporal relationships between different parts of the ECG signal.

The final GASF matrix, representing the temporal features of the ECG signal, is resized to a 128×128 pixel image:

$$\text{GAF}_{\text{Temporal}} = \text{GASF}(X(t)) \tag{3}$$

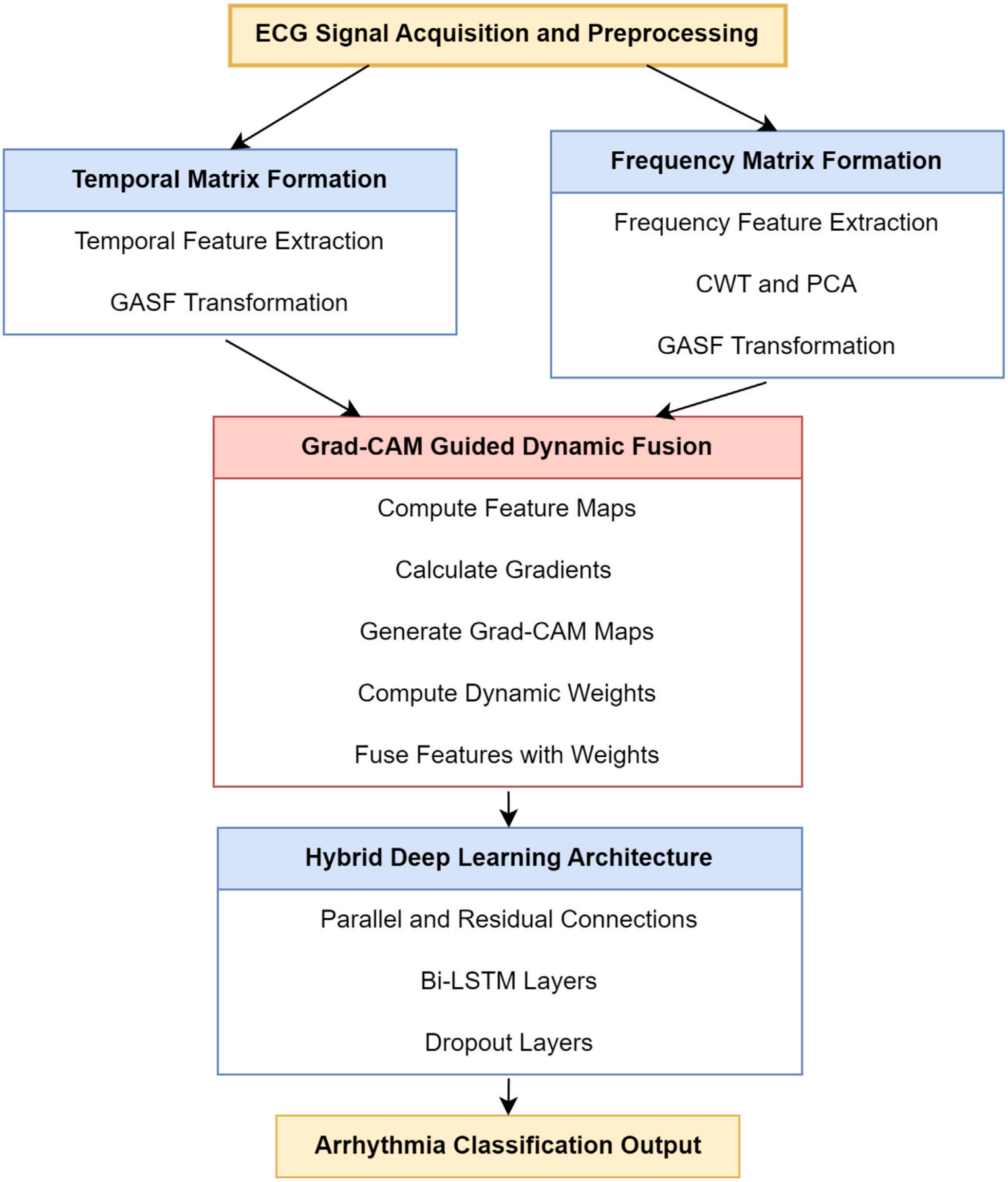

Fig. 3. Overview of the proposed model.

This matrix captures the temporal structure of the ECG signal, including timing and amplitude relationships, which are crucial for identifying various arrhythmias.

Frequency feature extraction

In addition to temporal features, frequency features was extracted from the ECG signal using the Continuous Wavelet Transform (CWT), followed by dimensionality reduction through Principal Component Analysis (PCA), and then transformed the reduced data into a GASF image.

Step 1: Continuous Wavelet Transform (CWT). The CWT decomposes the ECG signal into a time–frequency representation, highlighting the frequency components of the signal over time. For a given signal $X(t)$, the CWT is defined as:

$$\mathrm{CWT}(a,b) = \int_{-\infty}^{\infty} X(t)\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)dt \tag{4}$$

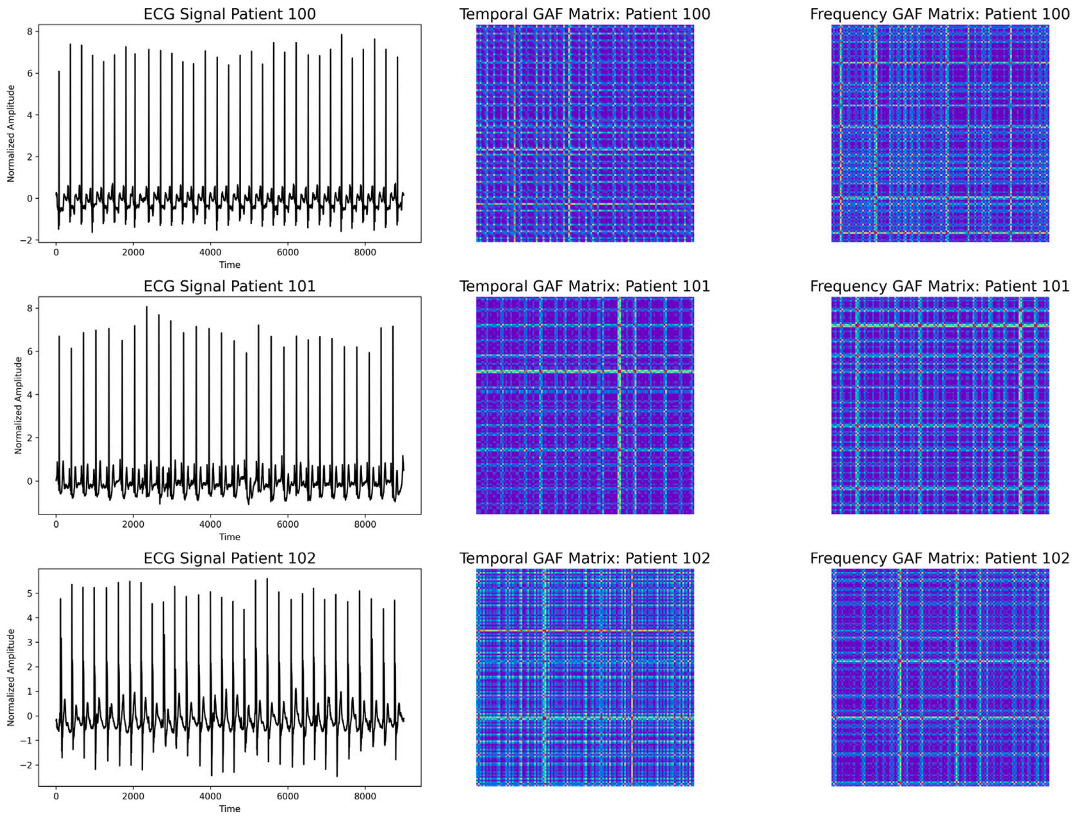

Fig. 4. Temporal and frequency-based GAF matrices of selected patients ECG signals.

where a and b are the scale and translation parameters, respectively, and ψ is the mother wavelet (in this case, the Ricker wavelet).

The result is a matrix where each element represents the signal's frequency content at a particular time and scale.

Step 2: Principal Component Analysis (PCA). Given the high dimensionality of the CWT matrix, PCA was applied to reduce the dimensionality while retaining the most significant frequency features. Let C be the CWT matrix, PCA transforms it into a reduced matrix C_{PCA}:

$$C_{\mathrm{PCA}} = \mathrm{PCA}(C) \tag{5}$$

Here, C_{PCA} is a one-dimensional vector that captures the essential frequency features of the signal.

Step 3: Transformation into GASF. The reduced frequency information C_{PCA} was then transformed into a Gramian Angular Summation Field (GASF) image, similar to the temporal GAF transformation:

$$\mathrm{GAF}_{\mathrm{Frequency}} = \mathrm{GASF}(C_{\mathrm{PCA}}) \tag{6}$$

This matrix provides a visual representation of the frequency content of the ECG signal, capturing how different frequency components interact over time.

Matrix formation and visualization

For each patient included in the study, both temporal and frequency-based GAF matrices were generated from the normalized ECG signals. These matrices encapsulate the complex patterns inherent in the ECG signals, making them suitable for advanced analysis and classification tasks.

- **Temporal GAF matrices:** Represent the temporal relationships within the ECG signal, capturing the timing and sequence of heartbeats.
- **Frequency GAF matrices:** Illustrate the frequency content over time, highlighting the distribution and interaction of different frequency components.

Fig. 4 presents a comprehensive view of the ECG signal analysis through three distinct visual representations for each patient. The first column on the left displays the original filtered ECG signals. These plots illustrate the raw temporal data, showcasing the amplitude variations and the overall waveform shape, which are crucial for identifying the rhythmic patterns and potential anomalies in the heart's activity.

The middle column showcases the temporal GAF matrices, which transform the time-series data into a visual format that encapsulates the timing and sequence of heartbeats. These matrices are essential for understanding the temporal relationships within the ECG signal, offering a unique perspective on the sequential dynamics of the cardiac cycle.

The rightmost column presents the frequency-based GAF matrices, which provide insights into the frequency content of the ECG signals over time. These matrices highlight how different frequency components interact and evolve throughout the cardiac cycle. By analyzing these frequency patterns, it becomes possible to detect subtle changes in the heart's rhythm that might not be evident in the time domain alone.

Together, these three columns offer a comprehensive visualization of both the temporal and frequency features of the ECG signals, facilitating more accurate and robust classification when applied in deep learning models. The combined analysis ensures that both aspects of the signal are thoroughly examined, improving the overall diagnostic performance.

Weighted fusion method guided by Grad-CAM

After extracting the temporal and frequency features from the ECG signals and converting them into the respective Gramian Angular Field (GAF) matrices, a weighted fusion method guided by Gradient-weighted Class Activation Mapping (Grad-CAM) was proposed to combine these features effectively. In this approach, Grad-CAM is applied to the feature maps generated from both the temporal and frequency domains, producing activation maps that highlight the most influential regions of the signal. These maps are subsequently used to adaptively adjust the weights assigned to each domain, allowing the model to focus on the most relevant aspects of the ECG data.

The central idea behind this dynamic weighting lies in determining how each domain—temporal or frequency—contributes to the classification outcome for a specific input. Whenever the temporal domain contains features strongly indicative of a particular arrhythmia (for example, abnormal intervals or wave morphologies), the method automatically allocates a higher weight to the temporal GAF representation. Conversely, if the ECG's frequency content reveals distinctive wavelet-based patterns, the system emphasizes the frequency GAF. In doing so, the fusion process becomes context-aware and selectively leverages the domain with the most discriminative cues for each individual instance.

By incorporating Grad-CAM in this manner, the proposed method not only improves classification accuracy but also offers a clear visualization of the decision-making process. The activation maps can be superimposed on the GAF matrices to illustrate which segments of the signal drive the classification most strongly. This transparency facilitates interpretability, allowing practitioners to verify that the regions spotlighted by the model align with clinically meaningful markers. Moreover, the adaptive nature of the fusion ensures that the model does not rely on one domain at the expense of another, thus mitigating the risk of overlooking subtle patterns that might be crucial for less prevalent or more complex arrhythmias.

Advantages of the proposed method

The integration of Grad-CAM into the fusion process offers several advantages:

- **Adaptive Weighting:** The fusion weights are dynamically adjusted based on the relevance of each feature type for the specific input, leading to a more effective combination of features.
- **Enhanced Interpretability:** Grad-CAM provides visual explanations for the model's predictions, which is crucial in medical applications.
- **Improved Performance:** By focusing on the most informative features, the model achieves higher classification accuracy, particularly in challenging cases.

Feature extraction

Let $\mathbf{X}_t \in \mathbb{R}^{n\times n}$ represent the temporal GAF image and $\mathbf{X}_f \in \mathbb{R}^{n\times n}$ represent the frequency GAF image for an ECG signal segment. Both matrices are processed separately by convolutional neural networks (CNNs) based on the Hybrid Parallel-Residual Architecture to extract specific features related to the temporal variations and frequency components of the ECG signal:

$$\mathbf{F}_t = \mathrm{CNN}_t(\mathbf{X}_t), \quad \mathbf{F}_f = \mathrm{CNN}_f(\mathbf{X}_f) \tag{7}$$

where $\mathbf{F}_t, \mathbf{F}_f \in \mathbb{R}^{h\times w\times d}$ are the feature maps extracted from the temporal and frequency GAF matrices, respectively. Here, h and w denote the height and width of the feature maps, and d is the number of feature channels.

Grad-CAM computation for feature importance

To guide the fusion process, the Grad-CAM maps for both temporal and frequency feature representations was computed, highlighting the regions that contribute most to the classification decision.

Step 1: Initial predictions. Perform a forward pass to obtain the initial class scores before fusion:

$$y_t = \mathrm{FC}_t(\mathrm{GAP}(\mathbf{F}_t)), \quad y_f = \mathrm{FC}_f(\mathrm{GAP}(\mathbf{F}_f)) \tag{8}$$

where $\mathrm{GAP}(\cdot)$ denotes Global Average Pooling, and FC_t, FC_f are fully connected layers for temporal and frequency branches, respectively.

Step 2: Gradient computation. Compute the gradients of the predicted score for the target class c with respect to the feature maps:

$$\frac{\partial y_t^c}{\partial \mathbf{F}_t}, \quad \frac{\partial y_f^c}{\partial \mathbf{F}_f} \tag{9}$$

Step 3: Weights calculation. Calculate the weights for each feature map channel by global average pooling the gradients:

$$\alpha_k^t = \frac{1}{h \times w} \sum_{i=1}^{h} \sum_{j=1}^{w} \left(\frac{\partial y_t^c}{\partial \mathbf{F}_t^{k,i,j}} \right), \quad \alpha_k^f = \frac{1}{h \times w} \sum_{i=1}^{h} \sum_{j=1}^{w} \left(\frac{\partial y_f^c}{\partial \mathbf{F}_f^{k,i,j}} \right) \tag{10}$$

where $\mathbf{F}_t^{k,i,j}$ and $\mathbf{F}_f^{k,i,j}$ represent the activation at position (i,j) in channel k of the feature maps.

Step 4: Grad-CAM map generation. Generate the Grad-CAM maps for both feature sets:

$$\mathrm{GradCAM}t = \mathrm{ReLU}\left(\sum k = 1^d \alpha_k^t \mathbf{F}_t^k\right), \quad \mathrm{GradCAM}f = \mathrm{ReLU}\left(\sum k = 1^d \alpha_k^f \mathbf{F}_f^k\right) \tag{11}$$

Grad-CAM heatmap visualization for temporal and frequency features

To further enhance the interpretability of the weighted fusion method, **Gradient-weighted Class Activation Mapping (Grad-CAM)** was employed to visualize the key areas of the ECG signals that contribute most to the classification decision. By overlaying the Grad-CAM heatmaps on both the temporal and frequency **Gramian Angular Field (GAF)** matrices, a better understanding of the critical signal regions that influence the model's predictions can be achieved.

Temporal GAF heatmap. The Grad-CAM heatmap for temporal GAF highlights which sections of the ECG's temporal features have the highest impact on the model's decision-making. The regions emphasized in the heatmap provide insights into the specific temporal patterns, such as heartbeat intervals, that play a significant role in the arrhythmia classification process.

Frequency GAF heatmap. Similarly, the Grad-CAM heatmap for the frequency GAF focuses on the frequency domain features. This allows us to see how different frequency components of the ECG signals are used by the model to make its classification. The Grad-CAM for frequency GAF is essential for identifying subtle frequency patterns, which might not be immediately visible in the time domain alone.

Visualization of grad-CAM overlays. Fig. 5 presents the Grad-CAM heatmaps overlaid on the temporal and frequency GAF matrices for a specific patient and lead. These heatmaps indicate which regions of the ECG signals contributed most significantly to the classification of arrhythmias. By combining both temporal and frequency heatmaps, the model ensures that all critical aspects of the signal are considered in the final prediction.

Grad-CAM heatmaps in 5 highlight the most informative regions in the temporal and frequency GAF matrices, emphasizing sections of the ECG signal that are crucial for arrhythmia classification. These visualizations allow us to observe how the model prioritizes key temporal dynamics and frequency components, ensuring that the most relevant patterns in the ECG signal are effectively leveraged for accurate arrhythmia detection. This provides valuable insights into the Grad-CAM-guided dynamic weighted fusion method, demonstrating how it integrates these critical features to enhance classification accuracy.

Dynamic weight calculation based on Grad-CAM

Using the Grad-CAM maps, the importance weights for the temporal and frequency features was computed.

Step 1: Normalization of Grad-CAM maps. Normalize the Grad-CAM maps to the range $[0,1]$:

$$\mathrm{Grad\tilde{C}AM}_t = \frac{\mathrm{GradCAM}_t - \min(\mathrm{GradCAM}_t)}{\max(\mathrm{GradCAM}_t) - \min(\mathrm{GradCAM}_t)} \tag{12}$$

$$\mathrm{Grad\tilde{C}AM}_f = \frac{\mathrm{GradCAM}_f - \min(\mathrm{GradCAM}_f)}{\max(\mathrm{GradCAM}_f) - \min(\mathrm{GradCAM}_f)} \tag{13}$$

Step 2: Importance score computation. Compute the overall importance scores for the temporal and frequency features:

$$S_t = \sum_{i=1}^{h} \sum_{j=1}^{w} \left(\mathrm{Grad\tilde{C}AM}_t^{i,j}\right), \quad S_f = \sum_{i=1}^{h} \sum_{j=1}^{w} \left(\mathrm{Grad\tilde{C}AM}_f^{i,j}\right) \tag{14}$$

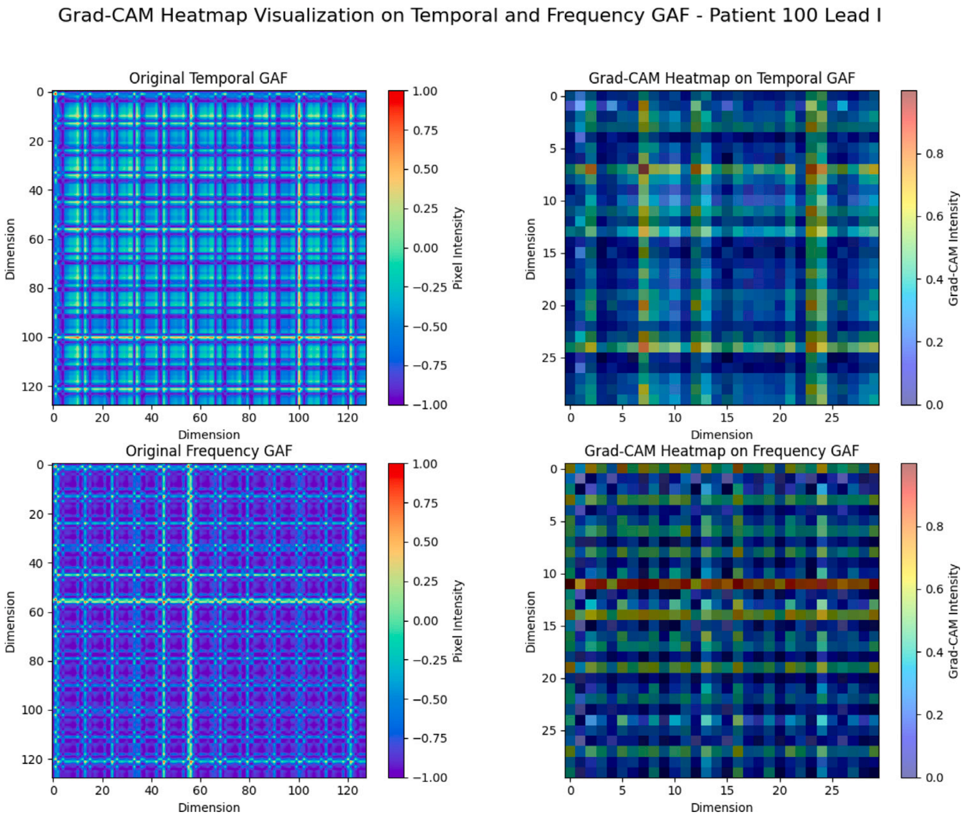

Fig. 5. Grad-CAM Heatmap Visualization for Temporal and Frequency GAF matrices of Patient 100 (Lead I), highlighting the most informative regions of these matrices for arrhythmia classification.

Step 3: Weight determination. Determine the dynamic weights based on the importance scores:

$$w_t = \frac{S_t}{S_t + S_f}, \quad w_f = \frac{S_f}{S_t + S_f} \tag{15}$$

These weights satisfy $w_t + w_f = 1$ and reflect the relative importance of each feature type for the current input.

Grad-CAM guided fusion

With the computed weights, the temporal and frequency features was fused.

Step 1: Feature vector extraction. Flatten the feature maps using Global Average Pooling:

$$\mathbf{c}_t = \mathrm{GAP}(\mathbf{F}_t) \in \mathbb{R}^d, \quad \mathbf{c}_f = \mathrm{GAP}(\mathbf{F}_f) \in \mathbb{R}^d \tag{16}$$

Step 2: Weighted fusion. Fuse the feature vectors using the dynamic weights:

$$\mathbf{c}_{\text{fusion}} = w_t \cdot \mathbf{c}_t + w_f \cdot \mathbf{c}_f \tag{17}$$

End-to-end learning

The Grad-CAM-guided fusion method is integrated into a complete deep learning architecture, where the entire model is optimized in an end-to-end manner.

Step 1: Classification layer. The fused feature vector $\mathbf{c}_{\text{fusion}}$ is passed through several fully connected layers for the final classification:

$$\hat{y} = \mathrm{Softmax}(\mathrm{FC}(\mathbf{c}_{\text{fusion}})) \tag{18}$$

where $\hat{y} \in \mathbb{R}^K$ is the probability vector for the K arrhythmia classes, and $\mathrm{FC}(\cdot)$ represents the fully connected layers with appropriate activation functions.

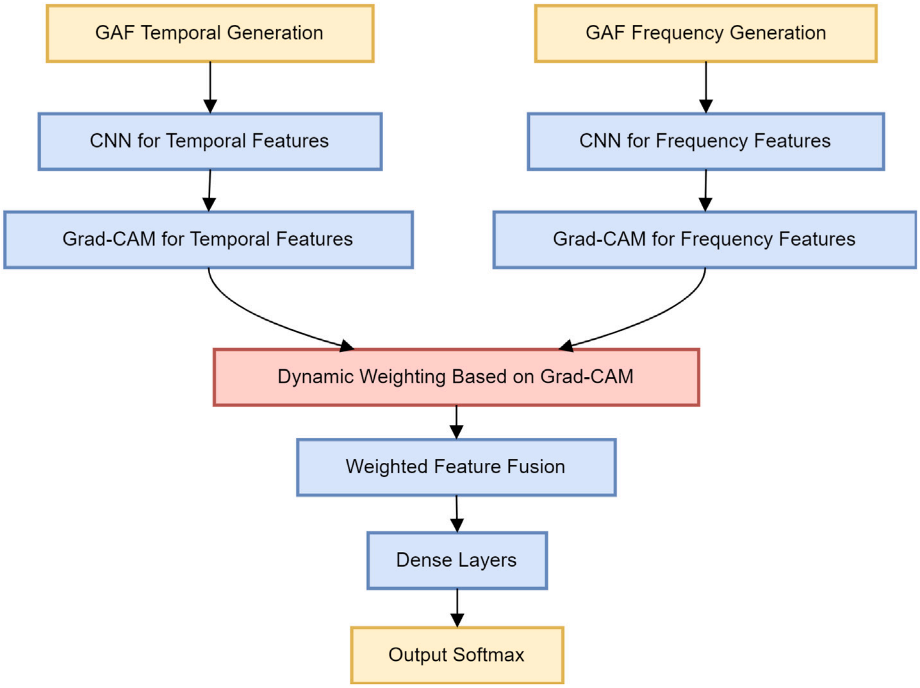

Fig. 6. Diagram of the proposed weighted fusion method using Grad-CAM for temporal and frequency GAF feature extraction and fusion.

Step 2: Loss function. The model is trained by minimizing the cross-entropy loss function:

$$\mathcal{L} = -\sum_{k=1}^{K} y_k \log(\hat{y}_k) \tag{19}$$

where y_k is the true label for class k, and $\hat{y}_k$ is the predicted probability for class k.

Step 3: Optimization. The **Adam** optimization algorithm Was employed, which is well-suited for handling sparse gradients and adaptive learning rates. Adam was used to update the model parameters, including those involved in the CNNs, fully connected layers, and the Grad-CAM computations, ensuring efficient and stable convergence of the training process. The optimizer adjusts the learning rate dynamically for each parameter based on estimates of lower-order moments, resulting in faster convergence and improved performance during the training phase.

After training, the model uses the vector $\hat{y}$ to predict the most likely arrhythmia class for a given segment of ECG signal. The Grad-CAM-guided fusion method effectively exploits the complementarity of the temporal and frequency features, focusing on the most informative aspects of each, which leads to robust and accurate arrhythmia classification.

Fig. 6 illustrates the architecture of the proposed weighted fusion method, guided by Grad-CAM, which dynamically integrates the temporal and frequency features extracted from ECG signals.

Optimization results: Loss and accuracy

The results of the training process reflect the effectiveness of the fusion mechanism used to combine temporal and frequency-based GAF matrices, guided by Grad-CAM. The objective was to optimize the fusion of these two domains to ensure that the most relevant features from both temporal and frequency information are dynamically weighted for improved feature extraction and representation.

The results in Fig. 7 demonstrate that the fusion method is highly effective. The model achieved a training accuracy of **99.68%** and a validation accuracy of **98.78%**, indicating that the fusion mechanism efficiently captures and combines the temporal and frequency features. The corresponding training loss decreased to **0.0308**, while the validation loss reached **0.0621**, confirming that the model successfully minimized the error in both domains.

These metrics highlight the robustness of the Grad-CAM-guided fusion process, ensuring that the fused GAF matrices are of high quality and ready to be used in more advanced stages of the model, specifically for arrhythmia classification. By reducing both the training and validation error rates, the fusion approach provides a strong foundation for accurate classification performance in subsequent phases.

The consistently high accuracy and low loss values observed during the training phase demonstrate the effectiveness of the Grad-CAM-guided fusion in extracting and weighting the most informative features from the ECG signals. These results ensure that the fused temporal and frequency features will provide a robust input for the classification of arrhythmias in future steps.

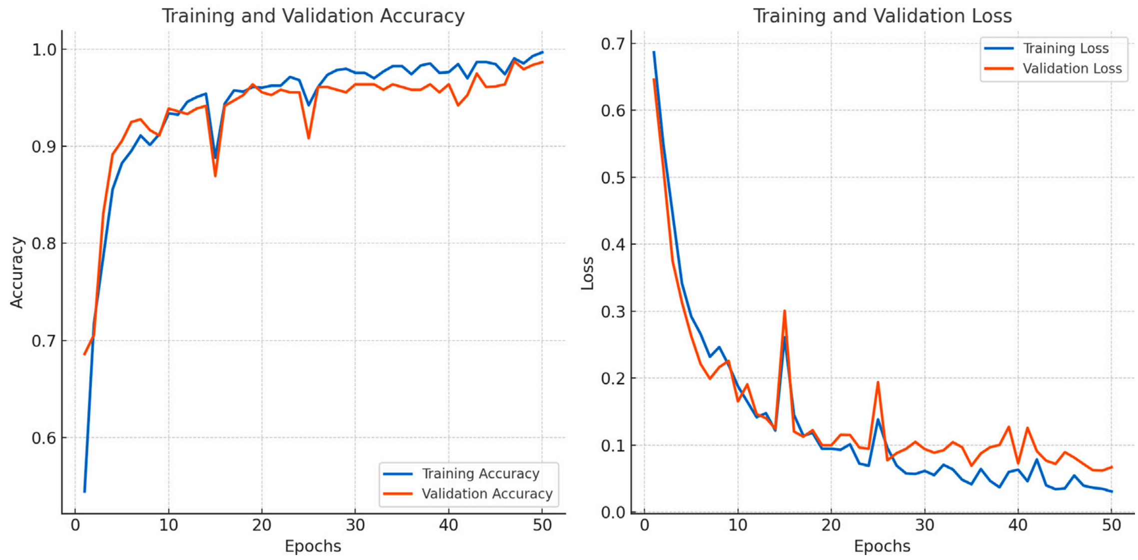

Fig. 7. Training and validation accuracy (left) and loss (right) curves over 50 epochs.

Interpretability and visualization

An added benefit of this method is the interpretability provided by the Grad-CAM maps. By visualizing these maps, clinicians can understand which parts of the ECG signal contributed most to the classification of arrhythmias, increasing confidence in the model's performance.

Fig. 8 illustrates the Grad-CAM maps for both temporal and frequency GAF matrices. The highlighted areas indicate the regions that have the most significant impact on arrhythmia classification, providing valuable insights into the critical features of the ECG signals that drive the model's decisions.

Proposed hybrid parallel-residual architecture for fused GAF matrices

In this section, a novel deep learning architecture developed was described specifically for arrhythmia classification using fused Gramian Angular Field (GAF) matrices. These GAF matrices combine temporal and frequency-based information from ECG signals, providing a rich 2D representation that captures both short-term dynamics and broader frequency content. To effectively leverage these fused matrices, the proposed architecture integrates **parallel convolutional layers**, **residual connections**, and **Bidirectional Long Short-Term Memory (Bi-LSTM)**. As illustrated in Fig. 9, this synergy addresses multiple challenges in ECG-based arrhythmia detection:

- **Multi-Scale Feature Extraction:** By using *parallel convolutional paths*, the model can simultaneously capture fine-grained details (e.g., subtle ECG waveform fluctuations) and larger-scale structures (e.g., global morphological patterns). This multi-scale approach is essential for identifying various types of arrhythmic events that manifest differently in time and frequency domains.
- **Efficient Training via Residual Connections:** Residual blocks help maintain stable gradient flow in deeper networks, mitigating vanishing or exploding gradients. This ensures that critical morphological features, such as QRS complexes and T-wave deviations, are not lost during deeper-layer processing. Consequently, the network can converge more reliably, learning robust representations of complex ECG patterns.
- **Long-Range Temporal Dependencies with Bi-LSTM:** Although the fused GAF matrices provide a two-dimensional snapshot of ECG data, the underlying signal still exhibits key temporal patterns. Incorporating *Bi-LSTM* layers allows the model to capture these long-range dependencies, enhancing its ability to detect arrhythmic events that unfold over extended periods of the heartbeat cycle.
- **Enhanced Interpretability Through Grad-CAM:** To ensure transparency in clinical settings, *Grad-CAM-guided fusion* was employed, overlaying attention maps on GAF matrices. This provides visual cues regarding which regions in the ECG (whether temporal segments or frequency components) drive the classification decision.

Overall, this hybridization of parallel, residual, and recurrent components is precisely motivated by the unique complexities of ECG-based arrhythmia detection. Parallel convolutions exploit the fused matrices at different scales, residual blocks stabilize learning in deeper networks, and Bi-LSTM layers capture temporal dependencies beyond what static 2D representations alone can convey. By leveraging these complementary strengths within a single framework, the proposed Hybrid Parallel-Residual Architecture achieves more robust and accurate arrhythmia detection while preserving interpretability through Grad-CAM visualizations.

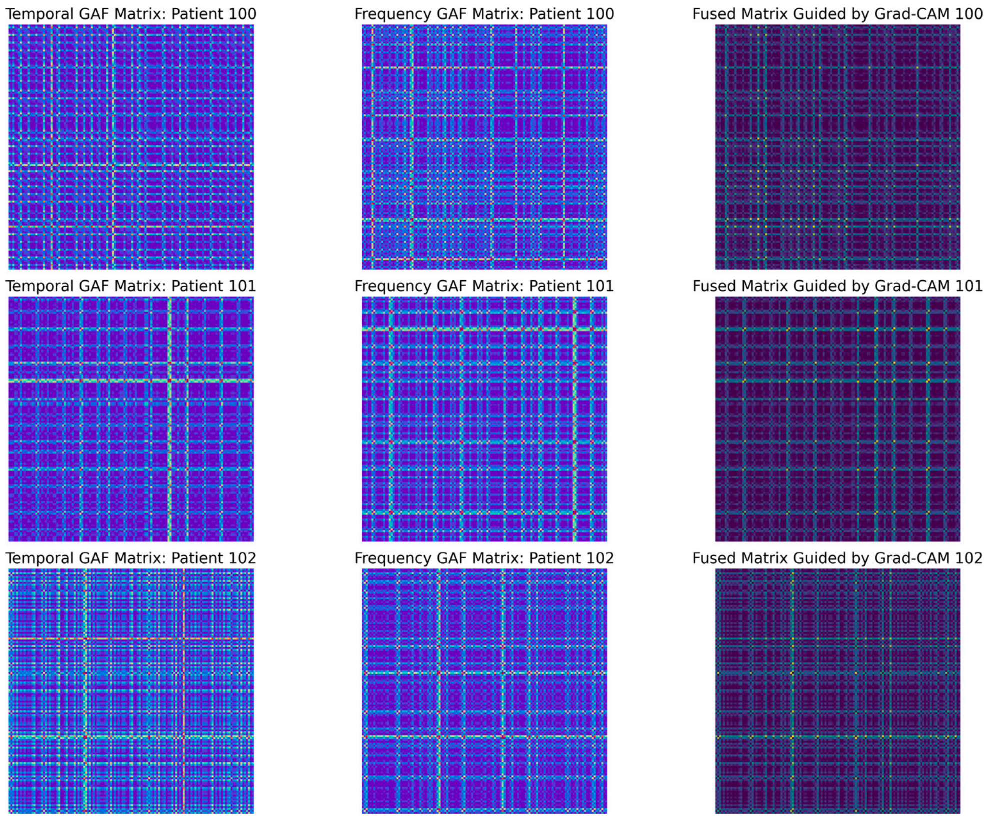

Fig. 8. Visualization of Grad-CAM maps overlaid on temporal and frequency GAF matrices, highlighting important regions contributing to the classification.

Fused GAF image input

The network receives fused GAF matrices as input. These matrices encapsulate both temporal and frequency-based features of ECG signals, providing a detailed representation of the signal's dynamics and facilitating the extraction of relevant features for arrhythmia classification.

Parallel-residual module

The core of the architecture is a Parallel-Residual module that combines parallel convolutional paths for multi-scale feature extraction and residual connections for efficient deep learning, ensuring robust feature representation from the fused GAF matrices.

Parallel convolution paths

The parallel pathways operate simultaneously, extracting features at different scales from the fused GAF matrices:

- **1 × 1 Convolution**: Captures fine-grained details from the input matrix.
- **3 × 3 Convolution**: Extracts medium-scale features important for identifying intermediate patterns.
- **5 × 5** Convolution: Focuses on capturing broader, large-scale structures within the GAF matrix.
- **Max Pooling**: Highlights the most prominent features by downsampling, emphasizing key areas of interest.

The outputs from these operations are concatenated to form a rich feature map that combines information across multiple scales.

Residual pathway

In parallel to the convolution paths, the residual pathway facilitates efficient learning by adding shortcut connections, allowing deeper networks to be trained effectively:

- **3 × 3 Convolution → BatchNorm → ReLU →** 3 × 3 **Convolution → BatchNorm**: A series of convolutions followed by batch normalization and ReLU activation ensures effective feature extraction while maintaining stability during training.

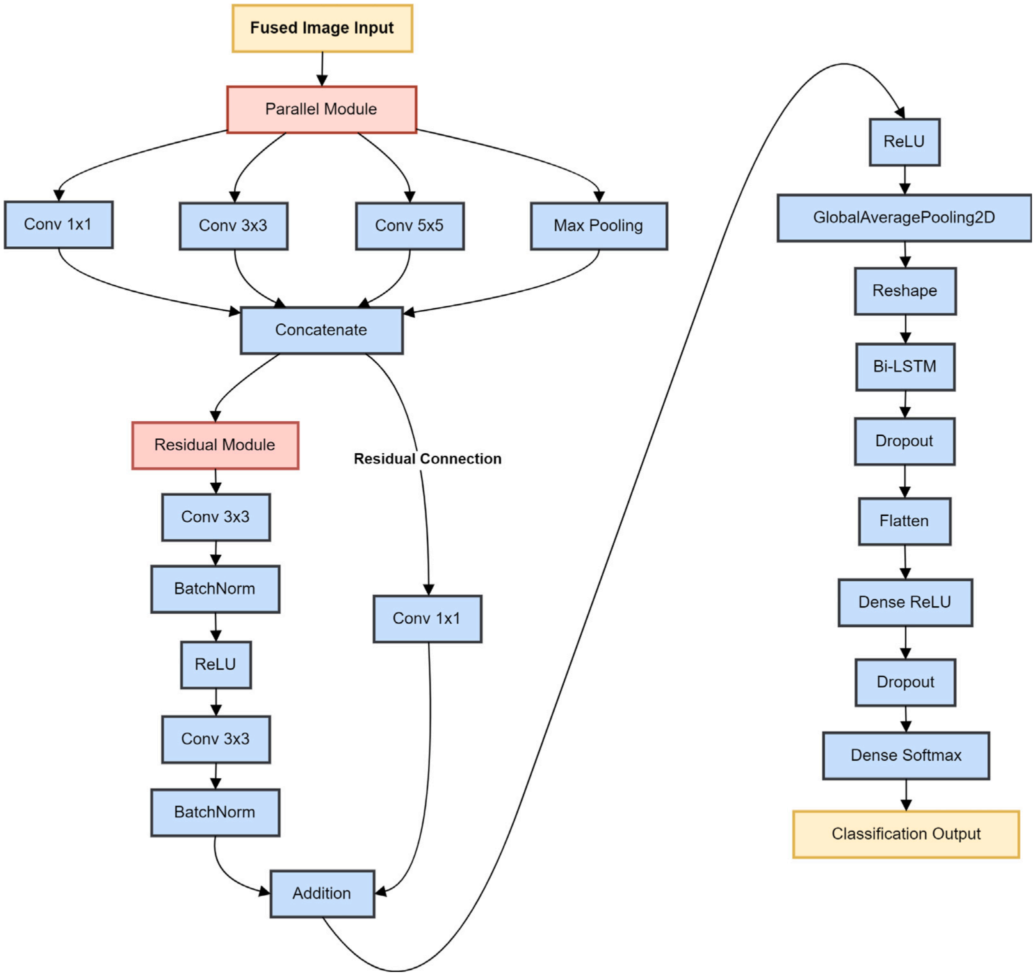

Fig. 9. Diagram of the Hybrid Parallel-Residual Architecture for Fused GAF Matrices.

- **Residual Connection**: A shortcut connection (1 × 1 convolution) helps preserve the identity of the input, improving the gradient flow in the network and making it easier to train deep layers.
- **Addition → ReLU**: Combines the output of the main path with the residual connection and applies a ReLU activation to introduce non-linearity.

Global feature pooling

After the Parallel-Residual module, the extracted features are globally pooled to reduce dimensionality, followed by a Grad-CAM-guided mechanism that dynamically adjusts the importance of the most salient features:

- **Global Average Pooling (2D)**: Reduces the spatial dimensions of the feature maps by averaging, while retaining the most critical information.
- **Grad-CAM Fusion**: A Grad-CAM-guided fusion mechanism dynamically adjusts the contribution of different features by highlighting the regions most relevant for arrhythmia classification.

Bi-LSTM and dropout layers

The network includes Bi-LSTM layers to capture temporal dependencies across the fused GAF matrix, ensuring that both short- and long-term patterns in the signal are considered. Dropout layers are applied to prevent overfitting:

- **Bi-LSTM**: A bidirectional LSTM captures dependencies in both forward and backward directions, offering a complete understanding of the temporal patterns in the ECG signal.

Table 3
Performance comparison between the traditional method and the proposed GAF-GradCAM method for arrhythmia classification across 12 ECG leads. Metrics include Accuracy (ACC), Sensitivity (SNS), and F1 score (F1).

Lead	Traditional method			Proposed GAF-GradCAM method		
	ACC (%)	SNS (%)	F1 (%)	ACC (%)	SNS (%)	F1 (%)
I	97.71	97.22	97.13	97.82	97.13	97.04
II	97.56	98.03	97.53	98.75	98.52	98.73
III	97.04	97.170	96.97	97.33	98.55	98.02
V1	97.73	98.34	97.66	98.61	99.14	98.73
V2	97.76	97.70	97.49	98.55	98.36	98.27
V3	97.05	97.06	96.96	97.38	97.43	97.41
V4	96.80	97.31	97.11	97.02	97.46	97.54
V5	97.47	97.75	97.88	98.28	98.85	98.97
V6	96.82	96.66	96.71	97.04	96.64	97.59
aVL	97.22	97.03	97.07	97.27	97.38	97.38
aVR	97.02	97.03	96.78	97.74	97.74	97.09
aVF	97.83	97.85	96.64	98.71	98.06	98.16

- **Dropout**: Applied to reduce overfitting by randomly deactivating neurons during training, ensuring that the model generalizes well to new data.

Dense layers for final classification

The classification is performed through a series of fully connected layers, which refine the feature representations before producing the final output:

- **Flatten**: Flattens the feature maps into a one-dimensional vector to prepare them for dense layers.
- **Dense (ReLU)**: A fully connected layer with ReLU activation refines the features extracted by the previous layers.
- **Dropout**: Another dropout layer is included to further prevent overfitting and improve generalization.
- **Dense (Softmax)**: The final dense layer uses softmax activation to output the probability distribution over the arrhythmia classes.

Classification output

The final output of the network is the prediction of the arrhythmia class, based on the fused GAF matrices and the features extracted throughout the network. The Grad-CAM mechanism ensures that the most relevant areas of the input are given appropriate weight in the decision-making process.

This architecture combines the strengths of parallel convolution and residual learning, while the Grad-CAM-guided fusion enhances interpretability, providing robust predictions for arrhythmia classification from fused temporal and frequency GAF matrices.

Results and discussion

In this section, the proposed GAF-GradCAM method was compared against a *traditional method* that employs only Gramian Angular Field (GAF) representations of ECG signals for arrhythmia classification, without incorporating the Grad-CAM-guided dynamic weighted fusion. Specifically, this traditional approach is inspired by Zhang et al. [31], where ECG time series were transformed into GAF images for classification (in their study, targeting myocardial infarction). In this case, this GAF-based technique is used as a baseline for arrhythmia detection, providing a point of comparison to highlight the benefits introduced by the proposed Grad-CAM-guided fusion scheme.

The Table 3 presents the detailed performance comparison between the two methods. For each ECG lead, the metrics of interest were calculated, and the results demonstrate the advantages of the GAF-GradCAM approach, particularly in leads that are crucial for detecting specific types of arrhythmias. As shown in Table 3, the proposed GAF-GradCAM method achieves higher scores in several leads, indicating its effectiveness in improving arrhythmia detection.

Comparison of performance

Fig. 10 shows the graphs generated from the results comparing the traditional method and the GAF-GradCAM method for arrhythmia classification across 12 ECG leads. The graphs provide a visual representation of the performance in terms of Accuracy, Sensitivity, and F1 score for both methods.

From the results shown in Fig. 10, the following key insights can be drawn:

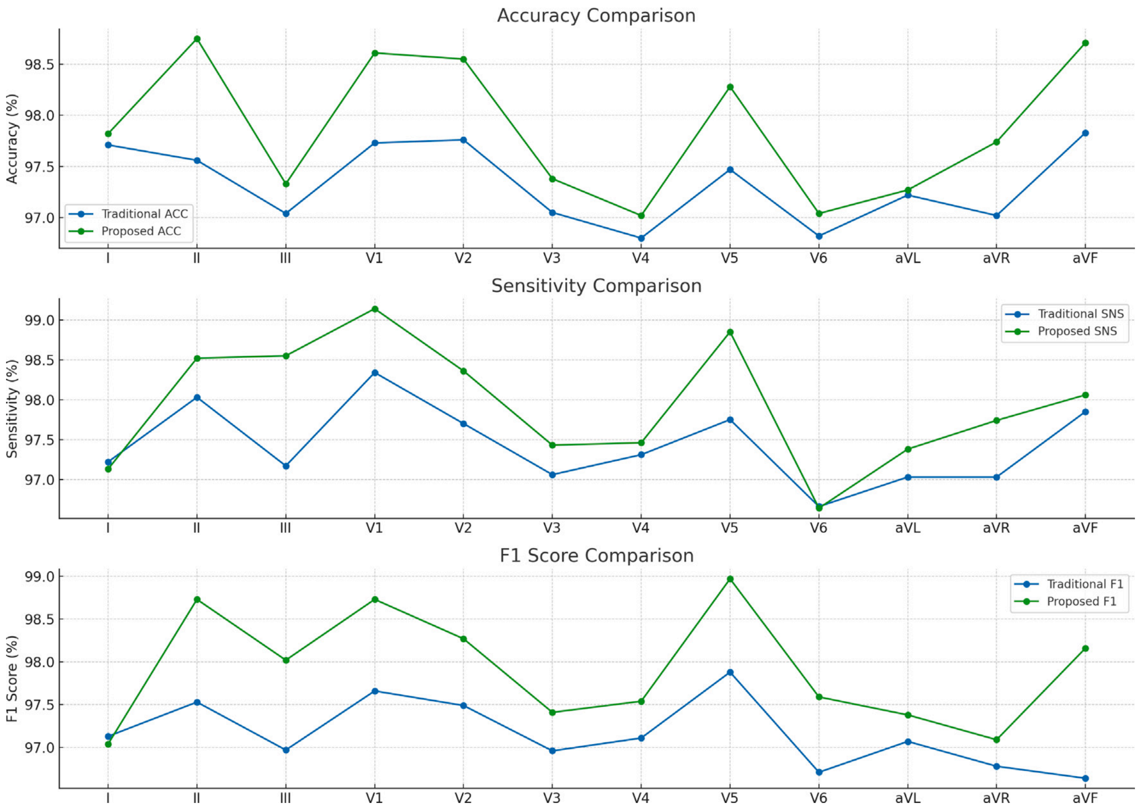

Fig. 10. Performance comparison between the traditional method and the GAF-GradCAM method for arrhythmia classification.

1. **Accuracy:** The GAF-GradCAM method demonstrates a notable improvement in accuracy across most leads, particularly in leads II, V1, and aVF. For instance, the accuracy for lead II increases from 97.56% to 98.75%, and for lead V1, it improves from 97.73% to 98.61%. Similarly, lead aVF sees an increase from 97.83% to 98.71%. This suggests that the method captures the relevant features of the ECG signal more effectively for these leads, which are often critical for detecting arrhythmias.
2. **Sensitivity:** Sensitivity is also improved in leads V1 and II, demonstrating the GAF-GradCAM method's superior ability to detect arrhythmias. For lead II, sensitivity rises from 98.03% to 98.52%, and lead V1 achieves a particularly high sensitivity of 99.14%, up from 98.34%. This high sensitivity in lead V1 is crucial for accurately detecting ventricular arrhythmias, which are often challenging to identify.
3. **F1 Score:** The improvement in F1 score is significant in leads V1, V5, and II. In lead V1, the F1 score increases from 97.66% to 98.73%, while in lead V5, it improves from 97.88% to 98.97%, and in lead II, it rises from 97.53% to 98.73%. This shows a better balance between precision and recall in these leads, which is especially important for avoiding false negatives in arrhythmia detection, ensuring that critical anomalies are not missed.

These results indicate that the GAF-GradCAM method generally enhances the performance of arrhythmia classification, with significant gains in key leads such as II, V1, and aVF, which are crucial for identifying complex anomalies.

Statistical significance analysis

To further validate the performance improvements observed in Table 3, a paired t-test was conducted on each of the three evaluation metrics (Accuracy, Sensitivity, and F1) across the 12 ECG leads. The aim was to determine whether the gains achieved by the proposed GAF-GradCAM method over the traditional GAF-only approach are statistically significant.

Data preparation. For each lead ℓ, the difference was computed

$$d_\ell = M^{(\ell)}_{\text{GAF-GradCAM}} - M^{(\ell)}_{\text{Traditional}},$$

where M represents the metric value (ACC, SNS, or F1). This yielded 12 paired differences per metric.

Test procedure. A paired two-tailed t-test was applied under the null hypothesis H_0: "*There is no significant difference between the traditional method and the GAF-GradCAM method*". The alternative hypothesis H_1 states: "*The GAF-GradCAM method outperforms the traditional method*".

Table 4
Paired t-test results comparing the proposed GAF-GradCAM method with the traditional method over 12 leads for each metric. Mean difference ($\bar{d}$) is reported in percentage points.

Metric	Mean difference (%)	*t-Statistic*	*p*-value
Accuracy (ACC)	0.54	4.97	<0.001
Sensitivity (SNS)	0.51	3.89	0.002
F1 Score (F1)	0.75	5.12	<0.001

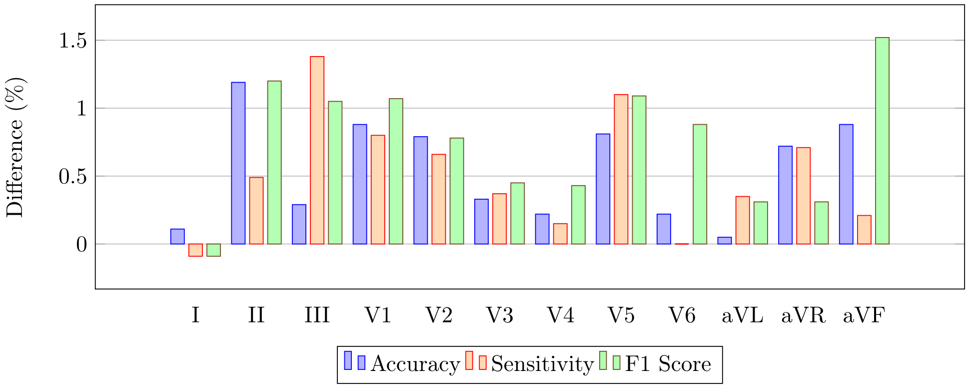

Fig. 11. Difference in performance (%): Proposed Method minus Traditional Method for each lead and each metric. Bars above 0 indicate an improvement by GAF-GradCAM.

Results. Table 4 summarizes the mean difference ($\bar{d}$) for each metric, along with the resulting t statistic and p-value. In all three cases, the improvements brought by GAF-GradCAM are statistically significant ($p < 0.05$):

All mean differences are positive, indicating higher performance for the GAF-GradCAM method. Moreover, the p-values are well below the typical threshold of 0.05, confirming that these performance gains are unlikely to be due to chance. Thus, statistical analysis supports the claim that incorporating Grad-CAM-guided dynamic weighted fusion yields a significant improvement in arrhythmia classification.

Visualizing the Differences. To illustrate these improvements more clearly, Fig. 11 shows the difference in performance (Proposed minus Traditional) for each lead and each metric. Bars above zero indicate an improvement conferred by GAF-GradCAM:

Comparison with state-of-the-art studies

Evaluating the performance of the proposed method in the context of existing approaches is essential to highlight its effectiveness in ECG-based arrhythmia classification. In this section, the proposed approach was compared with state-of-the-art methods that utilize various preprocessing techniques and deep learning architectures for ECG signal classification. The comparison focuses on the methods used, the dataset employed, and the achieved classification accuracy.

The results in Table 5 demonstrate that the proposed method outperforms existing approaches in terms of classification accuracy. By integrating Grad-CAM-guided dynamic weighted fusion of temporal and frequency GAF matrices with a hybrid parallel-residual architecture, the proposed model effectively captures both time-domain and frequency-domain features, leading to more robust arrhythmia detection and classification.

Discussion

The results presented in this study demonstrate the significant improvements achieved through the proposed GAF-GradCAM method for arrhythmia classification. By leveraging both temporal and frequency domain features from ECG signals, the method successfully captures intricate details that traditional approaches might overlook. The use of Gramian Angular Field (GAF) matrices, particularly the Gramian Angular Summation Field (GASF) for temporal features, enables a two-dimensional representation of ECG signals that highlights both time-dependent and frequency-based patterns.

In the temporal domain, GASF captures angular relationships between time points, providing detailed insights into the heart's electrical activity dynamics. For the frequency domain, the application of Continuous Wavelet Transform (CWT) followed by Principal Component Analysis (PCA) preserves the most relevant frequency components, transforming them into GAF matrices. The Grad-CAM-guided dynamic weighted fusion method then adaptively balances the contribution of temporal and frequency features, ensuring that the model focuses on the most critical cues for arrhythmia detection. This targeted emphasis on salient features explains the high accuracy and sensitivity observed in leads such as II, V1, and aVF, which often exhibit the most diagnostically informative waveforms.

Table 5
Comparison of the proposed method with state-of-the-art studies for ECG arrhythmia classification.

Reference	Methods used	Dataset	Accuracy (%)
Zeng et al. (2023) [32]	TQWT + CEEMD (preprocessing); CNN + LSTM (classification)	MIT-BIH Arrhythmia Database	97.20
Xia et al. (2023) [33]	Denoising Autoencoder (preprocessing); Transformer + CNN (classification)	MIT-BIH Arrhythmia Database	97.93
Anand et al. (2024) [34]	Improved ResNet-18 (classification)	MIT-BIH Arrhythmia Database	98.14
Kuila et al. (2024) [35]	DEA with LSTM (focal loss optimization)	MIT-BIH Arrhythmia Database	98.23
Alamatsaz et al. (2024) [36]	Resampling + Baseline Wander Removal (preprocessing); Hybrid CNN-LSTM (classification)	MIT-BIH/Long-term AF	98.24
The proposed method (2024)	**Hybrid CBD Filter + Grad-CAM Guided Dynamic Weighted Fusion of Temporal and Frequency GAF (preprocessing); Hybrid Parallel-Residual Architecture for Fused GAF Matrices (classification)**	**MIT-BIH Arrhythmia Database**	**98.75**

However, it is noted that certain leads, such as V3 and V4, showed more modest gains. This could be attributed to their unique waveforms, where further optimization or the integration of additional feature extraction techniques might yield improvements. Nonetheless, the interpretability provided by Grad-CAM highlights key ECG segments that drive the classification outcome, offering clinicians a transparent view of the model's decision-making. While Grad-CAM improves interpretability, the current study does not fully elaborate on how these visual cues are validated by clinicians. Demonstrating that the highlighted segments align with real-world cardiological assessments would further solidify the model's clinical utility.

Critical and elaborated discussion of methodological contributions

To understand precisely how each component contributes to the final performance, the provides a more detailed analysis:

- **Gramian Angular Field (GAF) Transformations:** Converting ECG signals into GAF matrices facilitates spatial feature extraction with CNN-based architectures, effectively revealing both local (beat-level) and global (rhythmic) patterns essential for arrhythmia detection.
- **Continuous Wavelet Transform (CWT) and PCA:** CWT highlights subtle frequency-dependent anomalies, while PCA reduces dimensionality, retaining only the most discriminative frequency components. This combination ensures that both dominant and transient frequency features are captured with less computational overhead.
- **Grad-CAM-Guided Dynamic Weighted Fusion:** Instead of assigning a fixed weight to temporal and frequency features, the model relies on Grad-CAM to pinpoint the most relevant feature maps for a given ECG. This adaptive mechanism increases classification precision by focusing on high-importance zones of the ECG signal.
- **Hybrid Parallel-Residual Architecture:** Parallel convolutional paths enable multi-scale feature extraction, capturing both fine-grained and broader morphological cues. Residual connections promote stable gradients in deeper layers, thereby learning robust representations of complex arrhythmias.
- **Bi-LSTM Integration:** Incorporating Bi-LSTM further strengthens the temporal modeling of ECG sequences. By considering forward and backward dependencies, the network can detect arrhythmic events that span multiple beats, complementing the CNN's spatial feature extraction.

Collectively, these methodological elements create a synergistic effect that boosts performance across diverse leads. The most substantial performance jumps—particularly in leads with strong frequency or temporal signatures—highlight the benefit of fusing domain-specific representations (temporal vs. frequency) in an adaptive manner. Equally important, the Grad-CAM overlays offer a transparent explanation for the model's classification decisions, thereby facilitating clinical adoption.

Advantages and disadvantages of the proposed method

Advantages:

- *High Accuracy and Sensitivity:* The fusion of temporal and frequency features, guided by Grad-CAM, notably improves classification metrics, especially in leads II, V1, and aVF.
- *Multi-Scale Learning:* Parallel convolutions and residual connections allow the network to capture both fine-grained ECG waveforms and broader morphological structures.
- *Robust Temporal Modeling:* Bi-LSTM layers incorporate long-range dependencies, essential for detecting arrhythmias evolving over multiple heartbeats.

Table 6
List of primary notations and symbols used throughout the paper.

Notation	Description
$X(t)$	Filtered ECG signal in the time domain
ϕ_i	Angular representation of the normalized signal value X_i in GAF
$\mathrm{GASF}(\cdot)$	Gramian Angular Summation Field transformation
$\mathbf{X}_t$	GAF image capturing temporal features, dimension $n \times n$
$\mathbf{X}_f$	GAF image capturing frequency features, dimension $n \times n$
$\mathbf{F}_t, \mathbf{F}_f$	Feature maps from CNNs for temporal and frequency GAF, size $h \times w \times d$
$\frac{\partial y_t^c}{\partial \mathbf{F}_t}$	Gradient of class score y_t^c w.r.t. feature map $\mathbf{F}_t$
α_k^t, α_k^f	Channel-wise weights derived from Grad-CAM for temporal and frequency domains
$\mathrm{GradCAM}_t, \mathrm{GradCAM}_f$	Grad-CAM activation maps for temporal and frequency feature maps
S_t, S_f	Importance scores computed from the Grad-CAM maps
w_t, w_f	Dynamic weights for temporal and frequency features, $w_t + w_f = 1$
$\mathbf{c}_{\text{fusion}}$	Final fused feature vector
$\hat{y}$	Predicted class probability distribution, dimension K

- *Explainability:* Grad-CAM heatmaps highlight critical ECG segments, aiding clinical interpretation and trust.

Disadvantages:

- *Computational Complexity:* Multiple stages (GAF transformation, CWT, PCA, deep architecture) require higher computational resources than simpler models.
- *Performance Variability Across Leads:* Certain leads (e.g., V3, V4) exhibit only moderate improvement, indicating that additional lead-specific optimizations may be needed.
- *Potential Overfitting:* While dropout layers and PCA help mitigate overfitting, very deep or complex models may still require careful regularization and larger training datasets.

Overall, the proposed GAF-GradCAM method presents a robust and interpretable approach to arrhythmia detection. While some leads exhibit marginally lower performance gains, the general trend underscores the effectiveness of the fused, multi-scale, and attention-driven strategy. Future refinements could focus on customizing feature extraction for specific leads or incorporating additional signal transformations to further enhance performance in more challenging scenarios.

Conclusion

This study introduces a novel approach for arrhythmia classification by leveraging a dynamic weighted fusion method, guided by Gradient-weighted Class Activation Mapping (Grad-CAM), to optimally combine temporal and frequency domain features extracted from ECG signals. The proposed methodology transforms ECG signals into Gramian Angular Field (GAF) 2D matrices, capturing both temporal and frequency features. Temporal features are encoded using Gramian Angular Summation Field (GASF) transformations, while the frequency domain is analyzed using Continuous Wavelet Transform (CWT) followed by Principal Component Analysis (PCA). The dynamic weighted fusion of these GAF matrices, guided by Grad-CAM, enhances classification accuracy by adaptively assigning importance to the most relevant features, while also improving the interpretability of the model's predictions. This optimization resulted in notable improvements in training and validation accuracies, underscoring the effectiveness of the fusion process.

The proposed method is implemented in a Hybrid Parallel-Residual Architecture, integrating parallel convolutional layers and residual connections with Bidirectional Long Short-Term Memory (Bi-LSTM) and dropout layers to ensure robust feature extraction and precise arrhythmia detection. The model achieved up to 98.75% accuracy, 99.14% sensitivity, and 98.97% F1 score across multiple ECG leads, significantly surpassing traditional methods.

This study demonstrates the effectiveness of combining temporal and frequency features through Grad-CAM-guided dynamic fusion, providing both high accuracy and enhanced interpretability for ECG-based arrhythmia classification. Future research could focus on refining this fusion technique and extending its application to other cardiac abnormalities or medical conditions, broadening its potential impact in clinical diagnostics.

Table of mathematical notations

See Table 6.

CRediT authorship contribution statement

Zakaria Khatar: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Visualization. **Dounia Bentaleb:** Conceptualization, Validation, Resources, Writing – review & editing, Supervision, Project administration. **Noreddine Abghour:** Data curation, Formal analysis, Supervision, Writing – review & editing. **Khalid Moussaid:** Data curation, Formal analysis, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] S.S. Chugh, R. Havmoeller, K. Narayanan, D. Singh, M. Rienstra, E.J. Benjamin, R.F. Gillum, Y.-H. Kim, J.H. McAnulty Jr., Z.-J. Zheng, et al., Worldwide epidemiology of atrial fibrillation: a global burden of disease 2010 study, Circ. 129 (8) (2014) 837–847.

[2] P. Rajpurkar, A.Y. Hannun, M. Haghpanahi, C. Bourn, A.Y. Ng, Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, Nature Med. 25 (1) (2019) 65–69.

[3] S. Hong, Y. Zhou, J. Shang, C. Xiao, J. Sun, Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review, Comput. Biol. Med. 122 (2020) 103801.

[4] S.J. Alonso, D. Langan, S. Windle, Big data in emergency medicine: a big deal, J. Emerg. Med. 51 (5) (2016) 689–696.

[5] A.Y. Hannun, P. Rajpurkar, M. Haghpanahi, G.H. Tison, C. Bourn, M.P. Turakhia, A.Y. Ng, Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network, Nature Med. 25 (1) (2019) 65–69.

[6] D. Bentaleb, Z. Khatar, Multi-criteria Bayesian optimization of empirical mode decomposition and hybrid filters fusion for enhanced ECG signal denoising and classification: Cardiac arrhythmia and myocardial infarction cases, Comput. Biol. Med. 184 (2025) 109462.

[7] A.H. Ribeiro, M.H. Ribeiro, G.M. Paixão, D.M. Oliveira, P.R. Gomes, J.A. Canazart, M.P. Ferreira, C.R. Andersson, P.W. Macfarlane, M. Wagner Jr., et al., Automatic diagnosis of the 12-lead ECG using a deep neural network, Nat. Commun. 11 (1) (2020) 1–9.

[8] Z. Khatar, D. Bentaleb, M. El Mansouri, Integrating advanced combined numerical filters for ECG denoising and cardiovascular disease classification using deep learning, in: International Conference on Digital Technologies and Applications, Springer, 2024, pp. 539–547.

[9] B. Pourbabaee, M.J. Roshtkhari, K. Khorasani, Deep convolutional neural networks and learning ECG features for screening paroxysmal atrial fibrillation patients, IEEE Trans. Syst. Man Cybern.: Syst. 51 (6) (2021) 3543–3554.

[10] W. Jiang, S. Dou, L. Zhu, G. Wang, A novel 1D-CNN architecture for cardiac arrhythmia classification, IEEE J. Biomed. Heal. Inform. 25 (6) (2021) 1984–1993.

[11] Z. Khatar, D. Bentaleb, Enhanced ECG signal features transformation to RGB matrix imaging for advanced deep learning classification of myocardial infarction and cardiac arrhythmia, Multimedia Tools Appl. (2024) 1–21.

[12] Z. Khatar, D. Bentaleb, S. Drissi, Bayesian methods for multi-objective optimization of hybrid numerical filters in ECG signal processing for accurate arrhythmia classification, Stat. Optim. Inf. Comput. (2024).

[13] S. Hong, Y. Zhou, J. Shang, C. Xiao, J. Sun, Opportunities and challenges of deep learning models for ECG interpretation: A comprehensive review, Artif. Intell. Med. 123 (2022) 102228.

[14] Z. Zhu, H. Wang, T. Zhao, Y. Guo, Z. Wang, X. Xu, et al., Classification of cardiac arrhythmia using angle distance and cosine similarity-based features from ECG signals, IEEE Access 9 (2021) 97537–97547.

[15] H. Zhang, D. Yao, X. Zhao, Z. Gao, A novel spatial attention mechanism for ECG arrhythmia classification, IEEE Trans. Biomed. Eng. 70 (2) (2023) 682–692.

[16] F. Liu, X. Zhou, J. Cao, Z. Wang, H. Wang, Y. Zhang, A CRNN-based channel-wise attention method for ECG arrhythmia classification, IEEE J. Biomed. Heal. Inform. 26 (6) (2022) 2693–2702.

[17] X. Li, Y. Zhang, Q. Zhang, J. Chen, Q. Wu, ECG arrhythmia classification using GAF and MTF images with deep learning, Comput. Methods Programs Biomed. 205 (2021) 106062.

[18] J. Wang, X. Yang, H. Cai, W. Tan, C. Jin, L. Li, Automatic cardiac arrhythmia classification using combination of CNN and LSTM techniques with variable length heart beats, Comput. Biol. Med. 141 (2022) 105017.

[19] J.H. Tan, Y. Hagiwara, W. Pang, I. Lim, S.L. Oh, M. Adam, et al., Automated diagnosis of cardiac arrhythmia using a novel deep learning technique with multi-lead ECG signals, Comput. Biol. Med. 154 (2023) 106534.

[20] C. Chen, Z. Hua, R. Zhang, G. Liu, W. Wen, Attention-based multi-scale feature fusion for ECG classification, IEEE J. Biomed. Heal. Inform. 26 (7) (2022) 3230–3241.

[21] Y. Zhu, J. Xu, K. Ren, J. Chen, H. Zhang, Y. Zhang, Self-supervised learning based domain adaptation for robust heartbeat classification, IEEE Trans. Instrum. Meas. 71 (2022) 1–11.

[22] Y. Chen, Y. Tao, Q. Zhang, H. Wen, Bi-LSTM for ECG arrhythmia classification with attention mechanism, IEEE Access 9 (2021) 42319–42329.

[23] S.S. Xu, S.C. Shen, M.W. Li, M.W. Mak, ECG arrhythmia classification using a convolutional transformer network, Eng. Appl. Artif. Intell. 117 (2023) 105523.

[24] N. Strodthoff, P. Wagner, T. Schaeffter, W. Samek, Deep learning for ECG analysis: Benchmarks and insights from PTB-XL, IEEE J. Biomed. Heal. Inform. 25 (5) (2021) 1519–1528.

[25] A. Mustaqeem, S.M. Anwar, A.R. Khan, M. Majid, A statistical analysis based interpretable deep learning approach for arrhythmia classification using ECG signals, Comput. Biol. Med. 141 (2022) 105051.

[26] K. Weimann, T.O.F. Conrad, Transfer learning for ECG classification, Sci. Rep. 11 (1) (2021) 5251.

[27] S. Raghavan, A. Batra, R. Ranjan, Transfer learning for improved ECG arrhythmia classification, IEEE J. Biomed. Heal. Inform. 27 (5) (2023) 2138–2147.

[28] X. Zhai, C. Tin, Automated ECG classification using dual path network with bidirectional LSTM, IEEE J. Biomed. Heal. Inform. 26 (6) (2022) 2523–2534.

[29] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circ. 101 (23) (2000) e215–e220.

[30] Z. Khatar, D. Bentaleb, O. Bouattane, Advanced detection of cardiac arrhythmias using a three-stage CBD filter and a multi-scale approach in a combined deep learning model, Biomed. Signal Process. Control. 88 (2024) 105551.

[31] G. Zhang, Y. Si, D. Wang, W. Yang, Y. Sun, Automated detection of myocardial infarction using a gramian angular field and principal component analysis network, IEEE Access 7 (2019) 171570–171583.

[32] W. Zeng, B. Su, Y. Chen, C. Yuan, Arrhythmia detection using TQWT, CEEMD and deep CNN-LSTM neural networks with ECG signals, Multimedia Tools Appl. 82 (19) (2023) 29913–29941.

[33] Y. Xia, Y. Xiong, K. Wang, A transformer model blended with CNN and denoising autoencoder for inter-patient ECG arrhythmia classification, Biomed. Signal Process. Control. 86 (2023) 105271.

[34] R. Anand, S.V. Lakshmi, D. Pandey, B.K. Pandey, An enhanced ResNet-50 deep learning model for arrhythmia detection using electrocardiogram biomedical indicators, Evol. Syst. 15 (1) (2024) 83–97.

[35] S. Kuila, N. Dhanda, S. Joardar, ECG signal classification using DEA with LSTM for arrhythmia detection, Multimedia Tools Appl. 83 (15) (2024) 45989–46016.

[36] N. Alamatsaz, L. Tabatabaei, M. Yazdchi, H. Payan, N. Alamatsaz, F. Nasimi, A lightweight hybrid CNN-LSTM explainable model for ECG-based arrhythmia detection, Biomed. Signal Process. Control. 90 (2024) 105884.