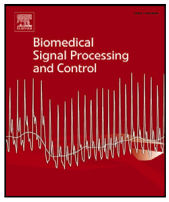
Biomedical Signal Processing and Control 89 (2024) 105714

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

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



ECGTransForm: Empowering adaptive ECG arrhythmia classification framework with bidirectional transformer

Hany El-Ghaish [a], Emadeldeen Eldele [b,*]

[a] Tanta University, Tanta, Egypt
[b] Nanyang Techological University, Singapore

ARTICLE INFO

Keywords:
ECG
Arrhythmia
Multi-scale Convolutions
Channel Recalibration Module
Bi-directional transformer
Context-Aware Loss

ABSTRACT

Cardiac arrhythmias, deviations from the normal rhythmic beating of the heart, are subtle yet critical indicators of potential cardiac challenges. Efficiently diagnosing them requires intricate understanding and representation of both spatial and temporal features present in Electrocardiogram (ECG) signals. This paper introduces ECGTransForm, a deep learning framework tailored for ECG arrhythmia classification. By embedding a novel Bidirectional Transformer (BiTrans) mechanism, our model comprehensively captures temporal dependencies from both antecedent and subsequent contexts. This is further augmented with Multi-scale Convolutions and a Channel Recalibration Module, ensuring a robust spatial feature extraction across various granularities. We also introduce a Context-Aware Loss (CAL) that addresses the class imbalance challenge inherent in ECG datasets by dynamically adjusting weights based on class representation. Extensive experiments reveal that ECGTransForm outperforms contemporary models, proving its efficacy in extracting meaningful features for arrhythmia diagnosis. Our work offers a significant step towards enhancing the accuracy and efficiency of automated ECG-based cardiac diagnoses, with potential implications for broader cardiac care applications. The source code is available at https://github.com/emadeldeen24/ECGTransForm.

1. Introduction

Arrhythmia is one cardiac condition that affects the normal functioning of the heart and can lead to serious health consequences [1]. Accurate detection and classification of arrhythmias are crucial for effective treatment and patient management. Electrocardiogram (ECG) is a widely used non-invasive technique for monitoring cardiac activities. The ECG signal provides important information about the electrical activities of the heart and is widely employed for diagnosing various cardiac conditions [2]. The task of arrhythmia classification involves analyzing the complex ECG signals and identifying the specific arrhythmia type based on its characteristic features [3]. Previously, classifying ECG signals was achieved by extracting engineered features, and then deploying traditional machine learning approaches like Support Vector Machines, Random Forests, k-nearest Neighbors, and Naive Bayes. Despite the interpretability and efficiency of these techniques, the process of manually extracting domain-specific features by medical professionals is often time-consuming, subjective, and prone to errors. Therefore, there has been a growing interest in developing automated approaches for arrhythmia classification using deep learning techniques.

In recent years, deep learning techniques have exhibited remarkable capabilities in automating the diagnosis of arrhythmias, leveraging the wealth of information contained in ECG signals [4]. The development of powerful and effective deep learning methods has the potential to significantly improve the accuracy and efficiency of arrhythmia diagnosis, and ultimately lead to better patient outcomes.

While existing literature has made substantial progress in arrhythmia classification using deep learning, several recurring challenges can be identified. Specifically, current models often struggle to effectively capture complex spatial patterns within ECG signals, particularly those that span multiple scales [5]. Another observation is that these models often may not fully capture the interdependencies among the extracted features, potentially missing crucial information that could enhance classification accuracy [6]. Furthermore, most approaches model temporal dependencies unidirectionally, overlooking the value of future context [7]. Last, class imbalance is a common problem that exists among various ECG arrhythmia classification datasets [8]. These challenges highlight the pressing need for an innovative solution that holistically addresses both spatial and temporal dimensions while effectively addressing the class imbalance problem.

* Corresponding author.
E-mail address: emad0002@ntu.edu.sg (E. Eldele).

https://doi.org/10.1016/j.bspc.2023.105714
Received 28 August 2023; Received in revised form 30 October 2023; Accepted 6 November 2023
Available online 9 November 2023

To address the aforementioned limitations, we propose ECGTransForm a deep learning model that incorporates four key components, each targeting a specific challenge in ECG-based arrhythmia classification. The first component, i.e., the Multi-scale Convolutions, aims to capture spatial patterns across varying scales within the ECG signal. This stage enhances the model's ability to detect intricate details present at different levels of granularity. The second component, i.e., the Channel Recalibration Module, is designed to rectify the lack of cross-channel interaction in the extracted features as inspired by [9]. By fusing spatial and channel-wise information at each layer, this module enables the model to effectively capture interdependencies among the extracted features. This holistic approach enhances the representation of spatial features and their relevance in arrhythmia classification. The third component, i.e., the Bidirectional Transformer, tackles the temporal dimension comprehensively. By leveraging bidirectional processing, our model effectively learns from both past and future contexts, enhancing its temporal feature extraction capabilities. This bidirectional approach allows the model to capture complex temporal dependencies that could be pivotal in distinguishing subtle variations between arrhythmia classes. Last, we introduce Context-Aware Loss (CAL), a dynamic weighting mechanism designed to address the class imbalance in the ECG datasets. By leveraging logarithmic weighting and considering the overall dataset distribution, it offers a nuanced approach to enhance the model's focus on underrepresented classes.

In summary, our contributions presented in this work are:

- **Enhanced Spatial Recognition:** We develop a deep learning framework that adeptly identifies and captures intricate spatial patterns within ECG signals across various scales, leading to improved detection of subtle arrhythmia signatures.
- **Holistic Feature Interactivity:** We implemented a recalibration mechanism that bridges the gap between spatial and temporal dimensions, ensuring the model fully capitalizes on the interdependencies among the extracted features.
- **Robust Temporal Processing and Imbalance Addressing:** We introduce a bidirectional Transformer to model both past and future temporal contexts in ECG data, coupled with a Context-Aware loss function designed to optimize focus on underrepresented classes, elevating overall classification efficacy.
- **Empirical Validation:** We conduct extensive experiments on two real-world ECG arrhythmia classification datasets, and the results underscore the superiority of our proposed model over existing methods.

2. Related work

ECG arrhythmia classification has gained significant attention in recent years, as evidenced by a number of recent works [10]. A key challenge in ECG arrhythmia classification is the extraction of meaningful features from ECG signals. This task can be approached through two primary methods: one entails the use of manually designed features in combination with traditional machine learning techniques like Random Forest (RF) [11] and Support Vector Machine (SVM) [12], while the other employs deep learning methods for automatic feature extraction. Next, we will delve into the exploration of hand-crafted feature techniques for ECG arrhythmia classification (Section 2.1) and investigate deep learning methods (Section 2.2).

2.1. Machine learning methods

Traditional approaches to ECG arrhythmia classification have primarily relied on handcrafted feature extraction techniques, with classifiers like Support Vector Machines (SVM) being employed to discern different arrhythmias. For instance, Majeed et al. [13] extracted time domain and frequency domain features obtained through an optimized Triple Band filter bank. Then, they selected the most discriminative features and fed to three classifiers: Least Square Support Vector Machine (LS-SVM), K-means, and K-nearest, where the LS-SVM achieved the best results. Also, Qin et al. [12] extracted low-dimensional ECG beat feature vectors using wavelet multi-resolution analysis. Following that, they used principal component analysis (PCA) to reduce features' dimensionality and input to an SVM.

In addition, Zabihi et al. [14] detected atrial fibrillation (AF) rhythm by extracting 491 hand-crafted features from time, frequency, and time–frequency domains. Those are then filtered to 150 features with a feature ranking procedure to be fed into an RF classifier. Similarly, Rouhi et al. [11] proposed an interpretable method that extracts hand-crafted features, and then selects the best of them for the RF classifier. Last, Wany et al. [15] extracted local and global characteristics based on clinical diagnosis, morphology features, and statistical features. Those were reduced with PCA and sent to an RF classifier.

While these methods have demonstrated efficacy, their reliance on hand-engineered features and manual feature selection could limit their capacity to adapt to diverse and intricate ECG patterns, necessitating more adaptive and automated feature extraction techniques. Consequently, more attention was shifted towards the deep learning methods that showed a capacity to automatically learn intricate patterns from raw ECG data

2.2. Deep learning methods

2.2.1. Automatic convolutional feature extraction

With the evolution of deep learning, there has been a surge in using neural networks for ECG arrhythmia classification. Convolutional neural networks (CNNs) have been the most popular architecture to use, because of their ability to automatically extract salient features from raw ECG signals and capture discriminative patterns. Zhi et al. [16] deployed a 1D residual CNN based on the ResNet architecture. By integrating two-way ECG signals with deep learning, they could differentiate between five heart rhythm classes. Building upon such frameworks, Srivastava et al. [17] unveiled a residual inception model enhanced by channel attention (RINCA). Their method emphasizes the importance of sequence segments and dominant channels in the classification of multi-lead ECG signals. Similarly, Kim et al. [18] amalgamated a residual network, squeeze-and-excitation block, and bidirectional Long-Short Term Memory (LSTM).

Some methods integrated CNN with wavelets to improve the feature extraction. For example, El Bouny et al. [19] amplified the extraction capabilities by introducing a Multi-Level Wavelet Convolutional Neural Network. This method synergizes 1D-CNN models with the Stationary Wavelet Transform to glean features from varied ECG signal scales. Houssein et al. [20] further propelled this idea by autonomously optimizing the CNN's hyper-parameters and using diverse feature extraction techniques directly on raw ECG signals.

Moreover, CNNs have been incorporated with other techniques to enhance performance. For instance, Hammad et al. [21] dovetailed deep neural networks with a genetic algorithm. Meanwhile, Al-Hadhrami et al. [22] ventured into a 2D-CNN domain using the DenseNet model, underscoring the adaptability of convolutional frameworks. On a parallel note, Nurmaini et al. [23] manifested the harmony between stacked denoising autoencoders and deep neural networks, elucidating the multifaceted nature of neural architectures in this domain.

While a myriad of CNN-based architectures has showcased promising results, a gap exists in the extraction of multi-scale patterns, as well as the attention to the discriminative features. Our approach, melding Multi-scale Convolutions with a Channel Recalibration Module, aims to address these challenges for a better arrhythmia classification performance.

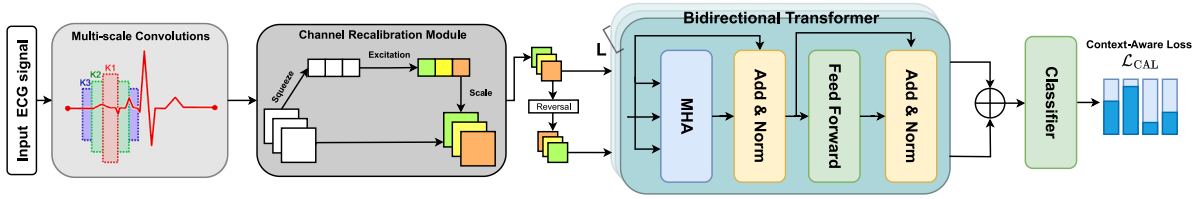


Fig. 1. A schematic diagram to the proposed framework ECGTransForm for ECG arrhythmia classification. The framework consists of a Multi-scale Convolutions module to capture spatial features at different scales. Following that, a Channel Recalibration Module highlights the interdependencies across channels. Next, we design a bidirectional Transformer module to learn efficient past and future temporal information. Last, we propose a Context-Aware Loss function to improve the model resilience to the class imbalance in ECG datasets.

2.2.2. Temporal dynamics and sequential models

The intrinsic temporal nature of ECG signals necessitates more focus on their temporal dynamics, leading to exploring sequential models, e.g., LSTMs and attention mechanisms.

Many studies integrated the power of CNNs with sequential models, as manifested in [24], where they fused CNNs with Recurrent Neural Networks (RNNs) for adept segmentation and classification of diverse cardiac rhythms from ECG recordings. Nevertheless, LSTM showed proficiency in learning temporal dynamics. Gao et al. [25] embarked on the LSTM frontier, proposing a model bolstered by focal loss, which disentangled time–space features and addressed category imbalance. In a similar vein, Mousavi et al. [26] amplified deep CNNs with sequence-to-sequence models to accentuate temporal insights. Adding another dimension, Jin et al. [27] conceived the dual-level attentional convolutional LSTM neural network to improve model interpretability.

Attention mechanisms, renowned for zeroing in on salient features, found their application in the works of Zhao et al. [28], who instituted a dual-channel CNN with such a mechanism. Similarly, Zhang et al. [29] presented a model that harnesses attention mechanisms across both spatial and temporal spectrums. The Transformer architecture was not to be left behind. As delineated by Xia et al. [30], a lightweight Transformer integrated with a CNN and a denoising autoencoder serves to amplify minority class performance, employing a unique seq2seq approach that bridges local and global ECG features.

While these models have made strides in temporal feature extraction, the bidirectional context provided by our proposed BiTrans showcases the potential for even more nuanced temporal understandings, especially when considering both past and future contexts.

2.2.3. Addressing class imbalance in ECG datasets

The skewed distribution of arrhythmia classes in many ECG datasets poses a challenge in achieving accurate and robust arrhythmia detection. Over the years, several strategies have emerged in the literature to counteract this imbalance and ensure fair representation of underrepresented classes.

Modifying loss functions has been one approach to place more emphasis on minority classes during training. Gao et al. [25] used an LSTM model complemented with a focal loss, specifically designed to handle imbalances by assigning more importance to hard-to-classify instances. Similarly, Lu et al. [8] incorporated focal loss into a depthwise separable CNN to counteract dataset skews. In addition, data augmentation has been another effective strategy. Ma et al. [31] tackled this by employing generative adversarial networks to artificially generate instances of sparse arrhythmia classes via a fusion model based on ResNet and BiLSTM. Furthermore, Peng et al. [32] employed the synthetic minority oversampling technique (SMOTE) to synthesize new minority sample ECG signals, emphasizing its capability to alleviate the detrimental effects of data imbalance on classification.

In light of these strategies, there still exists a quest for an optimal balance between model performance and sensitivity to minority classes. It is within this context that our Context-Aware Loss (CAL) component is introduced, aiming to offer a more nuanced and effective solution to the persistent challenge of class imbalance in ECG datasets.

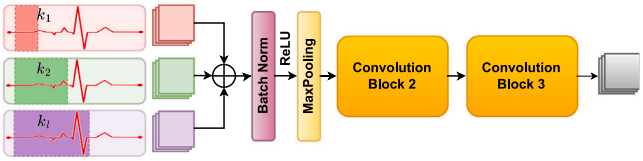


Fig. 2. The architecture of the multi-scale Convolutions module. Convolution Blocks 2 and 3 have the same structure, consisting of a convolutional layer followed by a Batch Normalization and MaxPooling layers.

3. Proposed methodology

3.1. Overview

In this section, we provide a comprehensive overview of the proposed framework ECGTransForm. The overall design of the proposed framework is depicted in Fig. 1. Our methodology comprises three core components, i.e., the Multi-scale Convolutions, the Channel Recalibration Module, and the Bidirectional Transformer. In addition, we include a Context-aware Loss function to address the class imbalance. Each component is designed to address specific limitations in the existing literature and contribute to the overall enhancement of arrhythmia classification performance. Next, we detail the design and functionality of each component, highlighting their synergistic interactions within the framework.

3.2. Multi-scale convolutions

The Multi-scale Convolutions (MSC) component is designed to capture spatial features within the ECG signal at various scales. The core concept involves employing multiple convolutional layers, each equipped with a distinct kernel size. Our objective is to enable the model to identify patterns across different spatial ranges, thereby enhancing the model's ability to capture both fine-grained and broader patterns inherent in arrhythmia-related ECG signals. The convolutional layers with small kernel sizes focus on capturing local details and fine-grained features in the input data. In contrast, the layers with larger kernel sizes gradually expand the receptive field, enabling the model to grasp more significant patterns and relationships between features. This operation is summarized in Fig. 2.

Formally, let $k = k_1, k_2, \ldots, k_l$ represent the set of l kernel sizes employed in the MSC component. Here, each k_i is associated with a particular scale such that k_i corresponds to capturing patterns within a temporal range of s_i milliseconds. For input ECG samples $x = (x_1, x_2, \ldots, x_m)$, where x_i is a univariate or multivariate sample and m is the number of samples, the output of the i-th convolutional layer C_i with kernel size k_i can be formulated as: $C_i = \text{Conv}(x, k_i, p_i)$, where p_i represents the padding required to balance the output of the multi-scale convolutional layers.

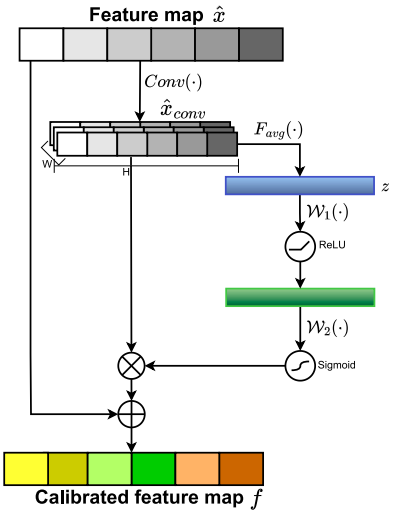


Fig. 3. The design of the Channel Recalibration Module..

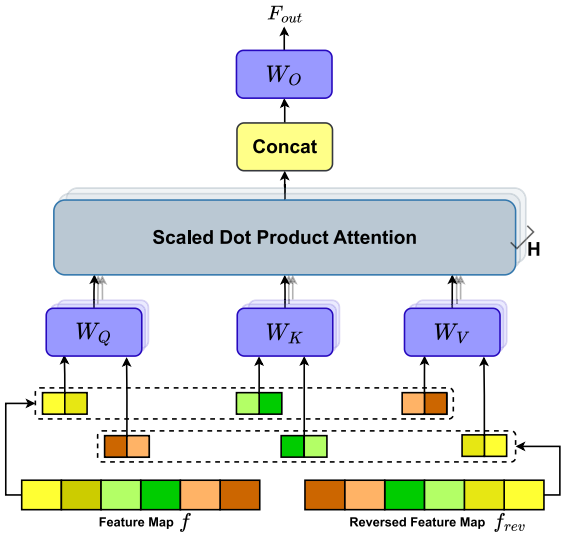


Fig. 4. The architecture of the Bi-directional Transformer module.

Subsequently, the features extracted from each convolutional layer are averaged element-wise to obtain a fused representation of multi-scale information:

$$\hat{x} = \frac{1}{l}\sum_{i=1}^{l} \text{Conv}(x, k_i, p_i), \tag{1}$$

where l is the number of convolutional layers in the MSC component, k_i represents the kernel size for the ith convolutional layer, and p_i represents the padding for the ith convolutional layer. Notably, we experimented with different techniques, but averaging showed the best performance (see supplementary materials).

While being followed by subsequent layers, i.e., convolutional, batch normalization, max-pooling, and dropout operations, the integration of multi-scale convolutional layers at the forefront is of paramount significance. The initial multi-scale convolutional layers function as feature preprocessors, extracting essential patterns that reflect physiological phenomena. This preprocessing enhances the effectiveness of downstream layers, which can then focus on refining and aggregating these features. Upon passing through the other convolutional blocks, the final output feature map $\hat{x} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m)$ is sent to the subsequent layer.

3.3. Channel recalibration module

The Channel Recalibration Module (CRM), inspired from [9], stands as a significant component towards an adaptive recalibration of channel-wise feature responses, with an explicit focus on modeling interdependencies among channels based on global context information. This component improves the model's capacity to discern and leverage complex relationships between different lead channels.

While the prior component in our framework, i.e., the MSC, primarily enhances the spatial features within the ECG signal, the CRM introduced another perspective, i.e., recalibrating channel-wise features, to enhance the quality of channel-based encodings across the hierarchy of features. This module undertakes the fusion of spatial and channel-wise information at each layer. In addition, the recalibration process adds the capability to discern and amplify channel-specific patterns that hold pivotal significance for accurate arrhythmia classification.

As we delve into the mechanics of the Channel Recalibration Module, its relationship with the other components of our framework becomes evident. Specifically, while the Multi-scale Convolutions module focuses on capturing and emphasizing the spatial nuances within the ECG signals, the Channel Recalibration Module refines these spatial features by optimizing channel-wise patterns. This integrated approach ensures that both spatial and channel-wise features are adequately addressed, thereby enabling the model to effectively differentiate between various arrhythmia classes.

Fig. 3 provides an illustration of the mechanism of the Channel Recalibration Module's operation. We deploy an initial convolutional layer to further refine and prepare the feature maps generated by the preceding MSC module for channel-wise recalibration, such that $\hat{x}_{conv} = Conv(\hat{x})$.

Next, we squeeze the spatial dimensions of the convolutional feature maps into a channel-wise descriptor by applying global average pooling. This reduces the feature maps into a vector of channel-wise statistics that capture the information across the entire spatial extent. Formally, this operation can be represented as follows:

$$z = F_{avg}(\hat{x}_{conv}) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} \hat{x}_{conv}^{i,j}, \tag{2}$$

where F_{avg} is an adaptive average pooling layer, H and W are the dimensions of the convolutional feature maps $\hat{x}$, and z is the channel-wise descriptor obtained through global average pooling.

In the next step, the channel-wise descriptor is fed into a small neural network consisting of fully connected layers. These layers act as excitation towards learning to model the inter-dependencies between different channels and capturing the importance of each channel. The network consists of a bottleneck layer followed by a gating ReLU activation function to limit model complexity and facilitate generalization. This layer is followed by another fully connected layer with a sigmoid activation that produces values between 0 and 1, indicating the degree of emphasis each channel should receive. The output of the network represents the channel-wise importance scores. These operations can be expressed as follows:

$$s = \sigma(\mathcal{W}_2\delta(\mathcal{W}_1(z))), \tag{3}$$

where s denotes the channel-wise importance scores, $\mathcal{W}_1$ and $\mathcal{W}_2$ represent the fully connected layers, σ is the sigmoid activation function, and δ is the ReLU activation function. It is worth noting that the sigmoid function was chosen to allow simultaneous emphasis on multiple ECG channels, while the ReLU function, used in the gating mechanism's bottleneck structure, aids in reducing model complexity and enhancing generalization.

The channel-wise importance scores s are then used to recalibrate the original feature maps $\hat{x}$ as follows:

$$f = s \odot \hat{x} \quad \in \mathbb{R}^{N \times d}, \tag{4}$$

where $\odot$ denotes element-wise multiplication and f is the calibrated feature map with new dimensions, i.e., N indicating the number of channels and d indicating the feature length. By incorporating these operations, our model can dynamically recalibrate the importance of features, leading to improved performance on various tasks.

3.4. Bidirectional transformer

Cardiac arrhythmias present intricate temporal patterns in ECG signals. These patterns, which can be subtle variations in the ECG waveform, manifest not just from previous cardiac cycles but can also be indicative of upcoming irregularities. Thus, capturing bidirectional temporal dependencies from both past and future contexts is pivotal to making accurate classifications.

The Transformer architecture, originally designed for sequence-to-sequence tasks in natural language processing [33], is inherently capable of modeling long-range dependencies in a sequence, limiting their capacity to grasp intricate contextual features. Therefore, we extend the Transformer network and introduce the Bidirectional Transformer (BiTrans) to enhance the learned temporal relations in the extracted features and harness information from both preceding and succeeding ECG waveforms. Considering that the input to the BiTrans is the extracted feature map from previous layers, which encode high-level spatial information, they will serve as a valuable context for the subsequent temporal relations learning task. Our Bidirectional Transformer, inspired by bidirectional LSTM (bi-LSTM), introduces an inversion mechanism to the feature map, generating an additional input that incorporates the reversed sequence of spatial information. This is illustrated in Fig. 4. This allows it to simultaneously consider past and future contexts when predicting a specific position. By combining outputs from both directions, our approach achieves a richer contextual understanding, effectively capturing intricate relationships within the input sequence. This is different from the standard Transformer encoder, which primarily focuses on the context leading up to a current position in the input sequence, neglecting the subsequent positions.

Formally, the self-attention mechanism is applied to the input, which consists of three duplicates of the same feature set f. This input is linearly projected into query, key, and value matrices for each head:

$$f_Q^h = W_Q^h \cdot f, \quad f_K^h = W_K^h \cdot f, \quad f_V^h = W_V^h \cdot f, \tag{5}$$

where f_Q^h, f_K^h, and f_V^h are the query, key, and value matrices for the hth head. Also, W_Q^h, W_K^h, and W_V^h are learnable projection matrices for the hth head.

The self-attention mechanism for each head is then applied as follows:

$$A^h = \text{Attention}(f_Q^h, f_K^h, f_V^h) = \text{softmax}\left(\frac{f_Q^h (f_K^h)^T}{\sqrt{d_k}}\right) f_V^h, \tag{6}$$

where d_k is the dimension of f_K^h for the hth head.

The multi-head attention mechanism is obtained by concatenating the results from all H heads and applying another linear projection to obtain the final multi-head attention output:

$$\text{MHA}(f, f, f) = \text{Concat}(A^1, \ldots, A^H) \cdot W_O, \tag{7}$$

where $\text{MHA}(f, f, f)$ represents the multi-head self-attention output, and W_O is the output projection matrix for the final multi-head attention output.

The key innovation in the Bidirectional Transformer network is incorporating bidirectional attention. To achieve this, we reverse the input feature map to take into account both past and future tokens, as follows.

$$f_{rev}[i, j] = x[i, d - j - 1] \tag{8}$$

for $i = 0, 1, \ldots, N-1$ and $j = 0, 1, \ldots, d-1$. Following that, we apply the reversed input to Eqs. (6) and (7) once again. Last, we add the outputs of the original feature map and its reversed counterpart as follows.

$$F_{out} = \text{MHA}(f, f, f) + \text{MHA}(f_{rev}, f_{rev}, f_{rev}) \tag{9}$$

In general, the encoder is composed of a stack of L identical layers. Each layer has two sub-layers, i.e., the multi-head self-attention mechanism, and the fully connected feed-forward network. Therefore, the above operations are being performed L times. Each layer in this stack sequentially processes the input data f, i.e., Layer_i includes operations represented by Eqs. (5) through (9), capturing increasingly abstract and contextualized representations. In this way, the relationships and dependencies within the input are learned hierarchically.

The output of each layer in the Bidirectional Transformer serves as the input to the subsequent layer, allowing for the gradual extraction of complex patterns. The final output of the Bidirectional Transformer Network, which we denote as $\text{BiTrans}(f)$, emerges by successively applying each layer to the initial input f. In essence, it can be expressed as:

$$\text{BiTrans}(f) = \text{Layer}_L(\text{Layer}_{L-1}(\ldots(\text{Layer}_1(f))\ldots))$$

3.5. Classification layer

3.5.1. Classifier

The classifier, as depicted in Fig. 1, serves as the final decision-making component. Its primary role is to take the feature representations extracted from the earlier components of our system and make predictions regarding the presence or absence of ECG arrhythmias. These feature vectors encapsulate the essential information extracted from the input ECG signals, capturing both fine-grained and multi-scale patterns. The role of the classifier is to analyze these feature vectors and determine which class or category the input ECG signal belongs to. In our implementation, the classifier is implemented as a single fully connected layer. The simple implementation decreases the model complexity and hence avoids overfitting, and it was found to be effective in capturing the discriminative information present in the feature vectors.

3.5.2. Context-aware loss

ECG data often encounters class imbalance issues due to the disproportionate presence of different cardiac arrhythmias. This imbalance poses a significant challenge for deep learning models, as they favor majority classes, i.e., the normal class, leading to suboptimal performance in accurately identifying rare events, i.e., the disease classes. The traditional method of assigning class weights in the cross-entropy loss typically involves directly inverting the class frequencies or normalizing them to ensure minority classes get more weight. However, this can sometimes lead to extremely high weights for very rare classes, which might cause the model to overfit these classes.

To overcome this limitation, we present the Context-Aware Loss (CAL). Inspired by [34], CAL provides a mechanism to handle class imbalance by assigning dynamic weights to each class based on their prevalence in the dataset. The weights are computed with three considerations. The first is incorporating the overall context. Instead of merely inverting the frequencies of classes, CAL considers the ratio of the total samples to the class with the maximum samples. This ensures that the weights are derived in the context of the overall dataset distribution. The second is logarithmic weighting, which aims to dampen the effect of extreme values, thus preventing overly high or low weights. This can be particularly useful to strike a balance between emphasizing minority classes and not causing overfitting. Last is flooring the weights by ensuring that the minimum weight is 1. In this way, we avoid assigning too low weights to any class, ensuring that all classes are considered by the model.

Given a dataset with C classes, let n_i represent the number of samples of class i, and $\mathcal{N}$ be the total number of samples in the dataset,

i.e., $\mathcal{N} = \sum_{i=1}^{C} n_i$. The class with the maximum number of samples can be represented as $n_{max} = \max_i n_i$. We define the parameter μ as the inverse ratio of the total samples to the class with the maximum number of samples, i.e., $\mu = \frac{n_{max}}{\mathcal{N}}$.

For each class i, the weight w_i is computed as:

$$w_i = \log(\mu \times \mathcal{N}/n_i). \tag{10}$$

However, we floor w_i if it is found to be less than 1, as follows:

$$w_i = \max(1, \log(\mu \times \mathcal{N}/n_i)). \tag{11}$$

This weight is then used to modify the cross-entropy loss, amplifying the loss contribution of under-represented classes, as follows.

$$\mathcal{L}_{CAL} = -\sum_{i=1}^{C} w_i y_i \log(\hat{y}_i). \tag{12}$$

4. Experiment setup

4.1. Datasets

Our proposed method is tested on two real-world datasets: the MIT-BIH arrhythmia database [35] and the PTB Diagnostic ECG Database [36]. These datasets are diverse in various aspects, which demonstrates the generality of our method.

4.1.1. MIT-BIH arrhythmia database

The MIT-BIH arrhythmia database [35] is a widely used dataset for research in arrhythmia detection and classification. The dataset contains 48 half-hour ECG recordings obtained from 47 subjects, with a total of over 110,000 individual beats annotated by a panel of cardiologists. The annotated beats are classified into 16 types of arrhythmias, including premature ventricular contractions (PVCs), atrial fibrillation (AF), and other less common types. The recordings have a sampling rate of 360 Hz and are digitized with 11-bit resolution. The dataset has been instrumental in the development and evaluation of various algorithms for arrhythmia detection and classification, and its availability has facilitated the advancement of research in the field of cardiac arrhythmia analysis. The MIT-BIH dataset consists of five distinctive classes. These classes are: **(1)** Normal Sinus Rhythm (N) representing the normal heart rhythms, **(2)** Supraventricular Premature or Ectopic Beat (S), which includes abnormal beats that originate above the ventricles, typically in the atria, **(3)** Ventricular Premature or Ectopic Beat (V), which represents abnormal beats that originate in the ventricles, **(4)** Fusion of Ventricular and Normal Beat (F), which indicates beats that are a combination of normal sinus rhythm and abnormal ventricular rhythms, **(5)** Unknown Beats (Q), which is used for beats that cannot be confidently categorized into any of the specific classes mentioned above.

4.1.2. The PTB diagnostic ECG database

The PTB Diagnostic ECG Database [36] is a publicly available dataset of ECG recordings that were collected at the Department of Cardiology, University Hospital Bonn, Germany. The dataset includes 5,388 ECG recordings from 549 patients, which were acquired using a 12-lead system with a sampling frequency of 1,000 Hz and a resolution of 16 bits. The recordings were collected from patients with various cardiac diseases, such as myocardial infarction, cardiac hypertrophy, and arrhythmia. This dataset consists of two classes: normal and abnormal ECG recordings. In addition to the ECG recordings, the dataset also includes diagnostic labels for each recording, which were assigned by expert cardiologists based on the clinical diagnosis of the patient.

The PTB dataset consists of two main classes, i.e., **(1)** Normal (N), which represents normal ECG recordings, and **(2)** the Myocardial Infarction (M), which represents ECG recordings that exhibit signs of myocardial infarction, indicating the presence of a heart attack.

In addition, Table 1 presents a detailed description of the number of samples in each class for both datasets.

Table 1
The detailed description of adopted datasets.

Database	MIT-BIH Arrhythmia						PTB diagnostic ECG		
Classes	N	S	V	F	Q	**Total**	N	M	**Total**
Training	72 471	2223	5788	641	6431	**87 554**	3236	8400	**11 636**
Testing	18 118	556	1448	162	1608	**21 892**	809	2100	**2909**

4.1.3. Datasets preprocessing:

The preprocessing of ECG signals plays a pivotal role, as these signals serve as the primary input. To enhance the efficiency of our approach, we applied the following preprocessing procedure for ECG signals and the subsequent extraction of cardiac beats, following [37]. These steps are as follows:

1. *Segmentation of ECG Signal*: The continuous ECG signal is initially segmented into discrete 10-second windows, and a specific 10-second window is selected from the ECG signal for further processing.
2. *Amplitude Normalization*: The amplitude values within the chosen window are normalized to fall within the range of zero to one, ensuring uniformity in the representation of the ECG signal.
3. *Identification of Local Maxima*: The set of local maxima is determined by examining zero-crossings within the first derivative of the signal.
4. *Detection of R-Peak Candidates*: An essential aspect of the process involves identifying a set of ECG R-peak candidates. This identification is achieved by imposing a threshold of 0.9 on the normalized values associated with the local maxima.
5. *Nominal Heartbeat Period Estimation*: The median value of the R-R time intervals is computed, serving as an estimation of the nominal heartbeat period within the selected 10-second window.
6. *Signal Fragment Selection*: For each identified R-peak, a segment of the signal with a length equal to 1.2 times the estimated heartbeat period (1.2T) is selected.
7. *Zero Padding*: Each of the selected signal segments is zero-padded to achieve a predefined and consistent fixed length.

4.2. Implementation details

The datasets were split into 80% for training and 20% for testing to ensure a sufficient amount of data for both model training and evaluation. Further, extracting 20% of the training dataset for validation allows for model hyperparameter tuning and ensures the generalizability of the model. Experiments were repeated three times with three different seeds to assess the stability and robustness of our model. The average performance with standard deviation was reported to mitigate any anomalies due to random initializations. The training epochs were set to 60 based on empirical observations, where we noticed that the model's performance stabilized thereafter, indicating the convergence of training. We used Adam optimizer with a learning rate of 1e−3 and weight decay of 1e−4, and batch size of 128. The Adam optimizer was chosen due to its adaptability and efficiency in deep learning tasks, as seen in previous research. We set the learning rate to 1e−3 as this is a common starting point that balances training speed and stability. The weight decay of 1e−4 helps prevent overfitting, ensuring that weights do not grow disproportionately large. We selected a batch size of 128. We assigned three convolutional layers ($l = 3$) based on empirical observations which showed an ability to effectively capture multi-scale features in ECG data. The chosen kernel sizes ($k_1 = 5, k_2 = 9$, and $k_3 = 11$) span a range of scales, allowing the model to detect patterns of various lengths in the ECG signal. A kernel size of 8 was empirically determined to be effective for subsequent convolutional operations, providing a good balance between receptive field size and computational efficiency. The number of layers ($L = 3$) and heads ($h = 5$) were chosen based on their success in previous transformer-based applications in related healthcare data [34]. The dropout value of 0.5 is set to prevent overfitting, providing a balance between model complexity and generalization capability.

Table 2
The detailed results of ECGTransForm for each class for the MIT-BIH dataset.

Metric	Per-class performance					Macro-Avg
	N	S	V	F	Q	
Precision	99.36±0.09	91.67±2.20	95.74±1.46	91.30±5.32	99.30±0.27	95.47±1.59
Recall	99.46±0.27	86.91±2.35	97.65±0.17	82.68±3.88	99.06±0.04	93.15±1.16
F1-score	99.41±0.08	89.22±0.22	96.69±0.72	86.78±1.22	99.18±0.15	94.26±0.28

Table 3
The detailed results of ECGTransForm for each class for the PTB Diagnostic ECG dataset.

Metric	Per-class performance		Macro-Avg
	N	M	
Precision	98.96±0.08	99.74±0.06	99.35±0.04
Recall	99.32±0.17	99.59±0.03	99.46±0.08
F1-score	99.14±0.08	99.67±0.03	99.41±0.05

4.3. Evaluation metrics

To evaluate our models, we adopted four performance metrics that reflect the model's ability to classify samples correctly. The first metric is accuracy (ACC), which is a widely used evaluation metric in classification tasks. It measures the proportion of correctly classified samples among all samples, and can be calculated as follows:

$$\mathrm{ACC} = \frac{TP + TN}{TP + TN + FP + FN},$$

where TP, TN, FP, and FN represent the True Positive, True Negative, False Positive, and False Negative respectively.

However, since the datasets are imbalanced, reporting only the accuracy will not be representative of the true performance of the model. Therefore, we adopted the macro average F1-score (MF1) to highlight the performance of imbalanced datasets. MF1 calculates the weighted harmonic mean of precision and recall for each class, as follows:

$$\mathrm{MF1} = \frac{1}{C}\sum_{i=1}^{C}\frac{2 \cdot TP_i}{2 \cdot TP_i + FP_i + FN_i},$$

where C is the number of classes.

Additionally, we included two more metrics, i.e., Precision and Recall. Precision measures the accuracy of positive predictions. It is the ratio of true positive predictions to the total positive predictions. The equation for Precision is as follows: $\mathrm{Precision} = \frac{TP}{TP+FP}$. Recall measures the ability of the model to identify all relevant instances. It is the ratio of true positive predictions to the total actual positive instances. The equation for Recall is as follows: $\mathrm{Recall} = \frac{TP}{TP+FN}$.

5. Results

In the results section, we present the outcomes of various experiments that assess the performance of our model using different configurations.

5.1. Performance analysis: In-depth examination

The performance analysis of our proposed model reveals remarkable classification accuracy across various arrhythmia classes. Table 2 provides a comprehensive breakdown of the model's precision, recall, and F1-score for each class on the MIT-BIH dataset, offering an in-depth understanding of its ability to accurately classify ECG arrhythmias. Class *N* and Q notably stand out with exceptionally high Precision, Recall, and F1-scores, nearing a perfect score of 100%. The strong capability of ECGTransForm in correctly classifying these classes can be attributed to the high number of samples for class *N* and the distinctiveness of class Q. The Ventricular Premature or Ectopic Beat (V) also displays a high precision rate of 95.74%, reinforcing the model's adeptness in capturing and distinguishing even the complex ventricular arrhythmias. In contrast, classes S and F achieve lower performance than other classes. This can be regarded to the ambiguity of these classes. Specifically, class F represents fusion beats, which are a combination of normal and abnormal rhythms. Detecting these accurately can be challenging, as they may not fit neatly into either category, leading to mislabeling or ambiguity in the dataset [38]. In addition, fusion beats can vary widely among individuals and even within the same individual over time [39]. Similarly, Class S may include various subtypes of supraventricular arrhythmias, some of which might be more challenging to classify [40].

Table 3 also shows the analysis of the ECGTransForm performance on the PTB dataset. Our model demonstrates high performance across both classes in this dataset. For class N, the precision and recall stand at 98.96% and 99.32%, respectively, indicating an almost impeccable classification capability. Class M goes a step further with a near-perfect precision score of 99.74% and a recall rate of 99.59%. These values suggest that the model makes extremely accurate predictions for both classes while also capturing almost all the true positive instances. The near-perfect macro-average scores underscore the model's superior capability in generalizing and accurately classifying ECG data from the PTB Diagnostic ECG dataset.

5.2. Comparative study: Benchmarking against established baseline models

The comparative evaluation of our proposed framework against the state-of-the-art baseline methods reveals some insights about the performance of our framework. The results of this study are presented in Table 4, which shows a diversity in terms of methodologies being employed for ECG-based arrhythmia classification. Examining the performance metrics, our ECGTransForm model outperforms all the baselines with an accuracy of 99.35% and an MF1 score of 94.26%. This indicates that the integration of various advanced methods into a single framework contributes to the higher discriminative capability and effective arrhythmia classification. The closest competitors are the works of Nurmaini et al. (2020) [23] and Kim et al. (2022) [18]. Even though their accuracy is marginally below our framework, it is noteworthy that they both surpass 99%. However, their MF1 scores, while being commendable, still lag behind our method by approximately 2–3 percentage points, revealing a significant improvement by our model. This underlines the robustness of our framework in capturing complex spatial and temporal arrhythmia patterns, surpassing a method that incorporates competing components.

Furthermore, our model's performance exceeds the results of Pokaprakarn et al. [24], who utilized Seq2Seq and CRNN for classification. Similarly, our ACC and MF1 outperform Hammad et al. (2020) [21], and Jin et al. (2022) [27], showcasing the effectiveness of our proposed framework against diverse methods. The presented comparative study substantiates our model's position at the forefront of ECG arrhythmia classification methodologies.

Compared to these baselines, our model presents several advantages from the architectural perspective. Unlike Seq2Seq models, primarily designed for sequence-to-sequence tasks like machine translation, our model is purpose-built for classification, capturing long-term dependencies regardless of sequence order with its bidirectional nature. Compared to LSTM-based models, the BiTrans structure in our model

Table 4
Comparison of our proposed framework against baseline methods on MIT-BIH dataset.

Baseline	Method	ACC	MF1
Hammad et al. (2020) [21]	ResNet + LSTM + GA	98.00	89.70
Nurmaini et al. (2020) [23]	AE + DNN	99.34	91.44
Kim et al. (2022) [18]	ResNet+ SE block + biLSTM	99.20	91.69
Pokaprakarn et al. (2022) [24]	Seq2Seq + CRNN	97.60	89.00
Jin et al. (2022) [27]	DLA + CLSTM	88.76	80.54
Xia et al. (2023) [30]	AE + Transformer	97.66	Nan
ECGTransForm *(Ours)*	MSC + CRM + BiTrans + CAL	**99.35**±0.16	**94.26**±0.28

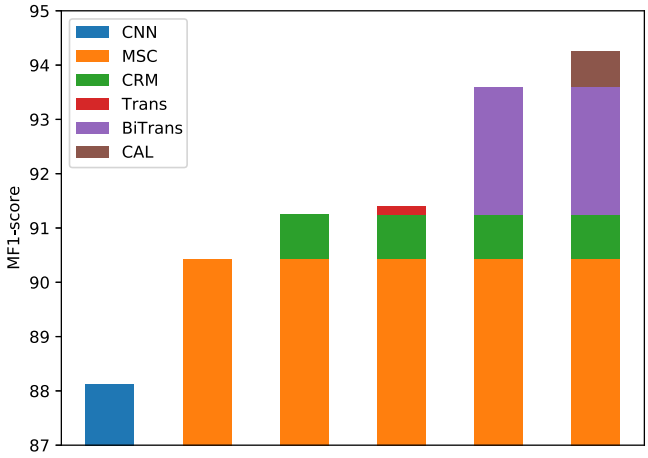


Fig. 5. Ablation study to the effect of each component on the performance.

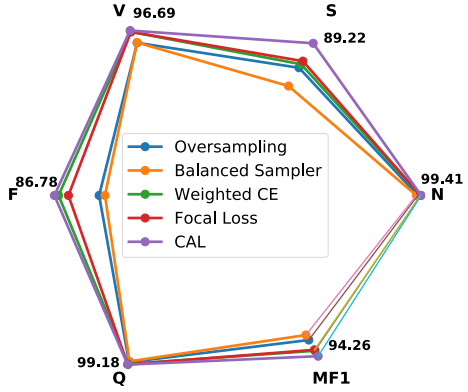


Fig. 6. Per-class F1-score performance for each class-imbalance handling method.

not only grasps complex temporal patterns from both past and future contexts but also offers faster parallel processing and addresses the vanishing gradient challenge prevalent in lengthy sequences. Differing from Auto-encoders (AE), which focus on data reconstruction, our design prioritizes the extraction of discriminative features, employing Multi-scale Convolutions and the Channel Recalibration Module for optimal representation. Additionally, it streamlines the learning process with end-to-end training. When set against the ResNet architecture, our model excels in extracting features at diverse scales, accentuating channel-wise interactions, and potentially achieving high performance with lesser complexity, underscoring its efficiency and effectiveness for the task at hand.

5.3. Component analysis: Ablation study and comprehensive insights

In our pursuit of dissecting the contributions of individual components within our framework, we conducted an ablation study, detailed in Fig. 5. We first show the performance of the traditional CNN model, which consists of 3 consecutive convolution blocks as [41] (without MSC module), to serve as a foundation for comparison. Notably, this model has the lowest F1-score of 88.1%. With the introduction of the Multi-scale Convolutions (MSC), the model's performance witnesses an uplift, as indicated by an improved F1-score to 90.4%. This enhancement underscores the significance of capturing multi-scale spatial patterns within the ECG signals. Further, the Channel Recalibration Module (CRM) enhances the model's classification prowess, leading to substantial improvements in F1-score by 0.9%. By dynamically fusing channel-wise information, this module refines the model's ability to harness class-specific features.

Expanding upon this foundation, the integration of our proposed Bidirectional Transformer (BiTrans) offers additional dimensions of insight. In fact, it leads to a significant improvement in performance by 2.35% if compared with the traditional Transformer, which only improves the performance by 0.15%. This signifies the substantial effect of leveraging future context information. The apex of our ablation study culminates with the incorporation of Context-aware loss (CAL). This component, tailored to address the class imbalance, brings forth a remarkable boost in the F1-score, underscoring its efficacy in enhancing classification across classes. The CAL mechanism's ability to dynamically adjust class weights based on context adds an invaluable layer of refinement to our framework.

5.4. Addressing class imbalance: Evaluating context-aware loss

We evaluate our Context-Aware Loss (CAL) mechanism against alternative data imbalance handling techniques, i.e., the oversampling with SMOTE [42], the balanced sampling, weighted cross-entropy loss [43], and focal loss [8].

The comparison in Fig. 6 provides a nuanced understanding of the efficacy of CAL in mitigating class imbalance in ECG datasets and enhancing classification performance. Notably, all the compared methods can perform well on the majority class N and the distinct class Q with the F1-score surpassing 99%. The differences among these methods are more notable in the relatively minor classes, i.e., F, V, and S. Illustratively, examining class S provides an even more compelling narrative. CAL yields an F1-score of 89.22%, substantially outperforming the 74.84% of oversampling and the 64.15% of Balanced Sampler. This difference can be attributed to CAL's adaptability in capturing the unique characteristics of minority classes.

Notably, we find that the balanced sampling technique, aimed at refining the class distribution during training, demonstrates limited effectiveness, with an MF1 score of 81.84% and the worst performance on the three minor classes. In addition, the oversampling approach, which artificially increases the instances of minority classes, increases the computational complexity and still exhibits comparably lower performance. On the other hand, the weighted Cross-Entropy (CE) and Focal Loss mechanisms, designed to assign different weights to classes, showcase notable improvements, yet the superiority of CAL remains evident, with MF1 differentials of 3.19% and 3.75% over them, respectively.

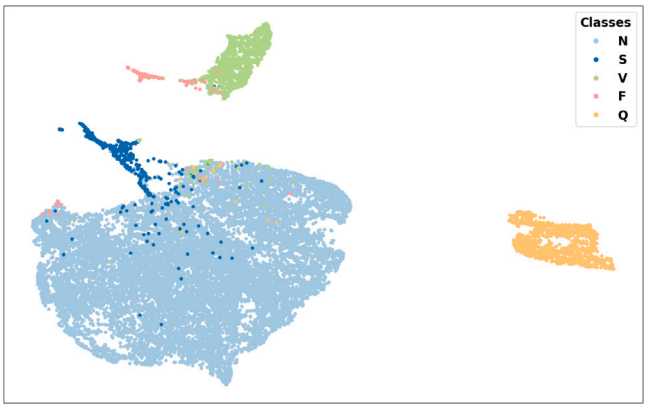


Fig. 7. UMAP visualization of the features generated by ECGTransForm on MIT-BIH dataset.

5.5. Visual insights: UMAP-based features visualization

Given the high dimensionality of the features extracted by our model, it is essential to project these features onto a lower-dimensional space to visualize them. UMAP [44] (Uniform Manifold Approximation and Projection) offers an effective approach for this purpose, reducing the feature dimensions while retaining the most meaningful structure.

The 2D UMAP plot shown in Fig. 7, represents the extracted features from our model, just before the final classification layer. This ensures that the features hold rich discriminative information learned through the various layers. The figure reveals some notable remarks. First, classes Q and V are more distinct in their ECG patterns and tend to form well-separated clusters. In addition, the majority class *N* also forms a big cluster, but it overlaps with samples from other classes. This confirms the capability of our model to effectively learn and differentiate various arrhythmia types.

Counterpart, we find that the fusion beats (class F) may share characteristics with other arrhythmias, such as ventricular ectopic beats (Class V) or normal sinus rhythm (Class N). This overlap can lead to misclassification, as the algorithm may struggle to distinguish between them. Also, we notice that class S is more overlapping with class *N* since supraventricular ectopic beats arise from the atria or the atrioventricular junction, which is closer to the natural origin of the heartbeat (the sinoatrial node) compared to ventricular beats. Therefore, the waveforms might exhibit features that are more similar to normal rhythms compared to beats originating from the ventricles.

6. Limitations and future work

Our model, while showcasing promising results, can still suffer from some challenges. These can be considered as potential future works and summarized as follows:

Detecting Specific Arrhythmias: A notable limitation is its struggle with specific arrhythmia classes, particularly Fusion beats (Class F) and Supraventricular Premature beats (Class S). Fusion beats present a unique challenge due to their variable nature, resulting from the merging of natural and premature beats. Addressing this requires classifiers capable of discerning such nuanced differences. Likewise, for Class S, there might be underlying characteristics that our model has yet to grasp fully.

Integrating Advanced ECG Metrics: We believe that several paths can further improve our model's performance and applicability. One such avenue is the integration of additional features beyond the conventional ECG waveform. Incorporating metrics like heart rate variability or P-wave characteristics might offer a more comprehensive view of the cardiac cycle.

Optimizing Transformer Complexity: Another worthwhile consideration is addressing the computational complexity of our Bidirectional Transformer architecture. Leveraging pretrained modules could help in reducing the number of trainable parameters, which can streamline the training process, and facilitate faster deployment.

Strengthening Model Generalization: Finally, as the medical community continuously seeks models with broad applicability, future efforts should be channeled towards ensuring our model's robust generalization across diverse datasets (see supplementary materials). Enhancing this generalization capability via techniques, e.g., unsupervised domain adaptation [45,46], may provide pathways to bridge the distributional gaps between datasets and enable more effective cross-domain deployment.

7. Conclusions

In this work, we have presented a comprehensive solution to ECG arrhythmia classification by introducing the ECGTransForm framework, which emphasizes the power of bidirectional Transformers in capturing intricate temporal relationships. Our approach departs from conventional methods, targeting both the spatial and temporal intricacies inherent in ECG signals. Key insights from this work include the impact of our proposed Bidirectional Transformer (BiTrans) mechanism, which allows for the effective extraction of temporal features by harnessing the context from both past and future time instances. We also showed the impact of the integration of Multi-scale Convolutions and the Channel Recalibration Module in capturing varying scales of features and their cross-channel interdependencies. Last, our Context-Aware Loss (CAL) introduces a dynamic way to handle class imbalance, ensuring that the model retains sensitivity towards underrepresented classes. Visualization of the learned features, using techniques like UMAP, further underscores the capability of our model in delineating distinct arrhythmic classes, which is pivotal for practical applications.

CRediT authorship contribution statement

Hany El-Ghaish: Conception and design of the study, Data analysis, Supervision, Writing – original draft. **Emadeldeen Eldele:** Conception and design of the study, Methodology, Results analysis, Writing – review & editing, Visualization.

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.

Data availability

No data was used for the research described in the article.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.bspc.2023.105714.

References

[1] C. Antzelevitch, A. Burashnikov, Overview of basic mechanisms of cardiac Arrhythmia, Cardiac Electrophysiol. Clin. 3 (1) (2011) 23–45.
[2] S. Kaplan Berkaya, A.K. Uysal, E. Sora Gunal, S. Ergin, S. Gunal, M.B. Gulmezoglu, A survey on ECG analysis, Biomed. Signal Process. Control 43 (2018) 216–235, http://dx.doi.org/10.1016/j.bspc.2018.03.003.
[3] O. Faust, R.J. Martis, L. Min, G.L.Z. Zhong, W. Yu, Cardiac Arrhythmia classification using electrocardiogram, J. Med. Imag. Health Inform. 3 (2013) 448, http://dx.doi.org/10.3389/fncom.2020.564015.

[4] Q. Xiao, K. Lee, S.A. Mokhtar, I. Ismail, A.L.b.M. Pauzi, Q. Zhang, P.Y. Lim, Deep learning-based ECG Arrhythmia classification: A systematic review, Appl. Sci. 13 (8) (2023) http://dx.doi.org/10.3390/app13084964.

[5] R. Wang, J. Fan, Y. Li, Deep multi-scale fusion neural network for multi-class Arrhythmia detection, IEEE J. Biomed. Health Inf. 24 (9) (2020) 2461–2472, http://dx.doi.org/10.1109/JBHI.2020.2981526.

[6] N. Sinha, R. Kumar Tripathy, A. Das, ECG beat classification based on discriminative multilevel feature analysis and deep learning approach, Biomed. Signal Process. Control 78 (2022) 103943, http://dx.doi.org/10.1016/j.bspc.2022.103943.

[7] B. Ganguly, A. Ghosal, A. Das, D. Das, D. Chatterjee, D. Rakshit, Automated detection and classification of Arrhythmia from ECG signals using feature-induced long short-term memory network, IEEE Sensors Lett. 4 (8) (2020) 1–4, http://dx.doi.org/10.1109/LSENS.2020.3006756.

[8] Y. Lu, M. Jiang, L. Wei, J. Zhang, Z. Wang, B. Wei, L. Xia, Automated Arrhythmia classification using depthwise separable convolutional neural network with focal loss, Biomed. Signal Process. Control 69 (2021) 102843, http://dx.doi.org/10.1016/j.bspc.2021.102843.

[9] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2018, URL https://openaccess.thecvf.com/content_cvpr_2018/papers/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.pdf.

[10] E.J. da S. Luz, W.R. Schwartz, G. Cámara-Chávez, D. Menotti, ECG-based heartbeat classification for Arrhythmia detection: A survey, Comput. Methods Programs Biomed. 127 (2016) 144–164, http://dx.doi.org/10.1016/j.cmpb.2015.12.008.

[11] R. Rouhi, M. Clausel, J. Oster, F. Lauer, An interpretable hand-crafted feature-based model for atrial fibrillation detection, Front. Physiol. 12 (2021) 657304.

[12] Q. Qin, J. Li, L. Zhang, Y. Yue, C. Liu, Combining low-dimensional wavelet features and support vector machine for Arrhythmia beat classification, Sci. Rep. 7 (1) (2017) 6067, http://dx.doi.org/10.1038/s41598-017-06596-z.

[13] R.R. Majeed, S.K. Alkhafaji, ECG classification system based on multi-domain features approach coupled with least square support vector machine (LS-SVM), Comput. Methods Biomech. Biomed. Eng. 26 (5) (2023) 540–547, http://dx.doi.org/10.1080/10255842.2022.2072684.

[14] M. Zabihi, A.B. Rad, A.K. Katsaggelos, S. Kiranyaz, S. Narkilahti, M. Gabbouj, Detection of atrial fibrillation in ECG hand-held devices using a random forest classifier, in: Computing in Cardiology, CinC, IEEE, 2017, pp. 1–4, http://dx.doi.org/10.22489/CinC.2017.069-336.

[15] Z. Wang, H. Li, C. Han, S. Wang, L. Shi, Arrhythmia classification based on multiple features fusion and random forest using ECG, J. Med. Imag. Health Inform. 9 (8) (2019) 1645–1654, http://dx.doi.org/10.1166/jmihi.2019.2798.

[16] Z. Li, D. Zhou, L. Wan, J. Li, W. Mou, Heartbeat classification using deep residual convolutional neural network from 2-lead electrocardiogram, J. Electrocardiol. 58 (2020) 105–112, http://dx.doi.org/10.1016/j.jelectrocard.2019.11.046.

[17] A. Srivastava, S. Pratiher, S. Alam, A. Hari, N. Banerjee, N. Ghosh, A. Patra, A deep residual inception network with channel attention modules for multi-label cardiac abnormality detection from reduced-lead ECG, Physiol. Meas. 43 (6) (2022) 064005, http://dx.doi.org/10.1088/1361-6579/ac6f40.

[18] Y.K. Kim, M. Lee, H.S. Song, S.-W. Lee, Automatic cardiac Arrhythmia classification using residual network combined with long short-term memory, IEEE Trans. Instrum. Meas. 71 (2022) 1–17, http://dx.doi.org/10.1109/TIM.2022.3181276.

[19] L.E. bouny, M. Khalil, A. Adib, An end-to-end multi-level wavelet convolutional neural networks for heart diseases diagnosis, Neurocomputing 417 (2020) 187–201, http://dx.doi.org/10.1016/j.neucom.2020.07.056.

[20] E.H. Houssein, M. Hassaballah, I.E. Ibrahim, D.S. AbdElminaam, Y.M. Wazery, An automatic Arrhythmia classification model based on improved marine predators algorithm and convolutions neural networks, Expert Syst. Appl. 187 (2022) 115936, http://dx.doi.org/10.1016/j.eswa.2021.115936.

[21] M. Hammad, A.M. Iliyasu, A. Subasi, E.S.L. Ho, A.A.A. El-Latif, A multitier deep learning model for Arrhythmia detection, IEEE Trans. Instrum. Meas. 70 (2021) 1–9, http://dx.doi.org/10.1109/TIM.2020.3033072.

[22] T. Al-Hadhrami, H. Ullah, M.B. Bin Heyat, et al., An end-to-end cardiac Arrhythmia recognition method with an effective DenseNet model on imbalanced datasets using ECG signal, Comput. Intell. Neurosci. (2022) http://dx.doi.org/10.1155/2022/9475162.

[23] S. Nurmaini, A. Darmawahyuni, A.N. Sakti Mukti, M.N. Rachmatullah, F. Firdaus, B. Tutuko, Deep learning-based stacked denoising and autoencoder for ECG heartbeat classification, Electronics 9 (1) (2020) 135, http://dx.doi.org/10.3390/electronics9010135.

[24] T. Pokaprakarn, R.R. Kitzmiller, R. Moorman, D.E. Lake, A.K. Krishnamurthy, M.R. Kosorok, Sequence to sequence ECG cardiac rhythm classification using convolutional recurrent neural networks, IEEE J. Biomed. Health Inf. 26 (2) (2022) 572–580, http://dx.doi.org/10.1109/JBHI.2021.3098662.

[25] J. Gao, H. Zhang, P. Lu, Z. Wang, An effective LSTM recurrent network to detect Arrhythmia on imbalanced ECG dataset, J. Healthc. Eng. (2019) http://dx.doi.org/10.1155/2019/6320651.

[26] S. Mousavi, F. Afghah, Inter- and intra- patient ECG heartbeat classification for Arrhythmia detection: A sequence to sequence deep learning approach, in: IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, 2019, pp. 1308–1312, http://dx.doi.org/10.1109/ICASSP.2019.8683140.

[27] Y. Jin, J. Liu, Y. Liu, C. Qin, Z. Li, D. Xiao, L. Zhao, C. Liu, A novel interpretable method based on dual-level attentional deep neural network for actual multilabel Arrhythmia detection, IEEE Trans. Instrum. Meas. 71 (2022) 1–11, http://dx.doi.org/10.1109/TIM.2021.3135330.

[28] R. Zhao, R. He, ECG-based Arrhythmia detection using attention-based convolutional neural network, in: Data Science, Springer Nature Singapore, Singapore, 2021, pp. 481–504, http://dx.doi.org/10.1007/978-981-16-5940-9_41.

[29] J. Zhang, A. Liu, M. Gao, X. Chen, X. Zhang, X. Chen, ECG-based multi-class Arrhythmia detection using spatio-temporal attention-based convolutional recurrent neural network, Artif. Intell. Med. 106 (2020) 101856, http://dx.doi.org/10.1016/j.artmed.2020.101856.

[30] 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, http://dx.doi.org/10.1016/j.bspc.2023.105271.

[31] S. Ma, J. Cui, W. Xiao, L. Liu, Deep learning-based data augmentation and model fusion for automatic Arrhythmia identification and classification algorithms, Comput. Intell. Neurosci. 2022 (2022) http://dx.doi.org/10.1155/2022/1577778.

[32] X. Peng, W. Shu, C. Pan, Z. Ke, H. Zhu, X. Zhou, W.W. Song, DSCSSA: A classification framework for spatiotemporal features extraction of Arrhythmia based on the Seq2Seq model with attention mechanism, IEEE Trans. Instrum. Meas. 71 (2022) 1–12, http://dx.doi.org/10.1109/TIM.2022.3194906.

[33] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L.u. Kaiser, I. Polosukhin, Attention is all you need, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 30, Curran Associates, Inc., 2017, URL https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf.

[34] E. Eldele, Z. Chen, C. Liu, M. Wu, C.-K. Kwoh, X. Li, C. Guan, An attention-based deep learning approach for sleep stage classification with single-channel EEG, IEEE Trans. Neural Syst. Rehabil. Eng. 29 (2021) 809–818, http://dx.doi.org/10.1109/TNSRE.2021.3076234.

[35] G. Moody, R. Mark, The impact of the MIT-BIH Arrhythmia database, IEEE Eng. Med. Biol. Mag. 20 (3) (2001) 45–50, http://dx.doi.org/10.1109/51.932724.

[36] R. Bousseljot, D. Kreiseler, A. Schnabel, Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet, Walter de Gruyter, Berlin/New York Berlin, New York, 1995, http://dx.doi.org/10.1515/bmte.1995.40.s1.317.

[37] M. Kachuee, S. Fazeli, M. Sarrafzadeh, ECG heartbeat classification: A deep transferable representation, in: IEEE International Conference on Healthcare Informatics, ICHI, 2018, pp. 443–444, http://dx.doi.org/10.1109/ICHI.2018.00092.

[38] J.L. Baez-Escudero, Chapter 37 - Ventricular Arrhythmias, in: G.N. Levine (Ed.), Cardiology Secrets, Fifth Edition, Elsevier, 2018, pp. 337–343, http://dx.doi.org/10.1016/B978-0-323-47870-0.00037-4.

[39] I. Neves, D. Folgado, S. Santos, M. Barandas, A. Campagner, L. Ronzio, F. Cabitza, H. Gamboa, Interpretable heartbeat classification using local model-agnostic explanations on ECGs, Comput. Biol. Med. 133 (2021) 104393, http://dx.doi.org/10.1016/j.compbiomed.2021.104393.

[40] F. Morady, Catheter Ablation of Supraventricular Arrhythmias, Blackwell Science Inc, 2004, http://dx.doi.org/10.1046/j.1540-8167.2004.03516.x.

[41] E. Eldele, M. Ragab, Z. Chen, M. Wu, C.K. Kwoh, X. Li, C. Guan, Time-series representation learning via temporal and contextual contrasting, in: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, 2021, pp. 2352–2359, http://dx.doi.org/10.24963/ijcai.2021/324.

[42] N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: Synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357, http://dx.doi.org/10.1613/jair.953.

[43] E. Essa, X. Xie, An ensemble of deep learning-based multi-model for ECG heartbeats Arrhythmia classification, IEEE Access 9 (2021) 103452–103464, http://dx.doi.org/10.1109/ACCESS.2021.3098986.

[44] L. McInnes, J. Healy, N. Saul, L. Großberger, UMAP: Uniform manifold approximation and projection, J. Open Source Softw. 3 (29) (2018) 861, http://dx.doi.org/10.21105/joss.00861.

[45] M. Chen, G. Wang, Z. Ding, J. Li, H. Yang, Unsupervised domain adaptation for ECG Arrhythmia classification, in: 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, 2020, pp. 304–307, http://dx.doi.org/10.1109/EMBC44109.2020.9175928.

[46] G. Wang, M. Chen, Z. Ding, J. Li, H. Yang, P. Zhang, Inter-patient ECG Arrhythmia heartbeat classification based on unsupervised domain adaptation, Neurocomputing 454 (2021) 339–349, http://dx.doi.org/10.1016/j.neucom.2021.04.104.