DynaECG-Net: Dynamic Margin Metric Learning for Arrhythmia Classification Using Single-Lead Electrocardiogram

Amnah Nasim, Taesik Go

Jeonbuk National University, Jeonju, Republic of Korea

Abstract

Automated cardiovascular disease classification is crucial to enabling real-time and continuous monitoring using wearable electrocardiogram (ECG) devices. However, due to the limited number of pathological class samples and difficulty in separating some hard-to-discriminate ECG classes (such as normal sinus rhythm (N) and premature supraventricular contraction (S)) due to their morphological similarity, existing deep learning models frequently fail to ensure sufficient inter-class separation in the latent space, limiting their discriminative power. The proposed deep metric learning framework with dynamic margin triplet loss (DynaECG-Net) extracts feature embeddings that maximize latent space separability between N, S, and premature ventricular contraction (V) heartbeats. Disease-specific experiments conducted on MIT-BIH ECG Arrhythmia dataset for 3-class classification with 50% data used for training, yield overall: accuracy 98.97%, sensitivity (Sen) 97.76%, and, F1-score (F1) 96.82% and classwise: N (Sen=99.47%, F1=99.35%), S (Sen=95.23%, F1=93.36%) and V (Sen=99.44%, F1=98.19%) achieving better performance than the state-of-the-art in low inter-class separability and data-scarce conditions. Additionally, a t-SNE visualization demonstrated well-separated embedding clusters for each class.

1. Introduction

Arrhythmia, a cardiovascular disorder characterized by irregular heart rhythms, contributes significantly to the global burden of cardiovascular diseases (CVDs), which account for 32% of all deaths worldwide [1]. Early and accurate arrhythmia detection is essential for timely treatment, particularly with the rise of wearable single-lead ECG devices like KardiaMobile and EMAY that enable real-time, continuous monitoring. However, automated classification remains challenging due to the limited number of pathological samples and the morphological similarity between hard-to-separate classes such as normal sinus rhythm (N) and supraventricular ectopic beats (S).

Deep learning (DL) has emerged as a powerful tool to automate arrhythmia classification by learning relevant features directly from ECG data. DL models can be broadly categorized as pure models (e.g., CNNs, transformers) and hybrid models (e.g., CNN-LSTM, CNN-transformer). Convolutional neural networks (CNNs), due to their ability to learn spatial hierarchies of features, have become a foundational architecture for ECG analysis. Kiranyaz et al. [2] utilized 1D CNNs for patient-specific arrhythmia classification with up to 99% accuracy. In other studies, CNNs have been extended with recurrent layers such as LSTMs and GRUs to capture temporal dependencies [3]. Recent works have also integrated attention mechanisms to improve feature selection and focus on relevant heartbeat segments. Models like spatial-temporal attention networks and dual-level attention LSTMs demonstrated improvements in macro F1 score and interpretability, though many still rely on multi-lead ECG inputs or ignore class imbalance [4, 5]. Transformer-based models, initially developed for natural language processing, are increasingly applied to ECG due to their ability to model long-term dependencies in parallel [6]. Hybrid CNN-transformer architectures have shown promising performance, yet often involve complex, multi-lead input data unsuitable for practical deployment [7]. Most existing deep learning models struggle with these constraints, often failing to ensure adequate inter-class separation in the latent space and requiring post-processing to reach acceptable performance [37], [38]. This highlights the need for accurate, lightweight, and interpretable models designed specifically for multi-class arrhythmia classification from single-lead ECG signals in data-scarce and low-separability conditions [8].

To address these challenges, we propose DynaECG-Net, a novel hybrid architecture that integrates a convolutional neural network (CNN) for local feature extraction with a channel attention mechanism to refine salient features. The model processes 253-sample heartbeat segments from single-lead ECG signals and performs three-class arrhythmia classification (N, S, and V). To enhance inter-class separability, DynaECG-Net employs a dynamic margin triplet loss, optimizing latent space embeddings for better discrimination, especially between morphologically similar classes like N and S. The results highlight the potential of DynaECG-Net for accurate, real-time arrhythmia detection using wearable devices in both clinical and home settings.

2. Materials and Methods

2.1. Data

Data consisted of the Massachusetts Institute of Technology-Beth Israel Hospital (MITBIH) arrhythmia database [9, 10]. It includes 48 two-channel ambulatory ECG records, each approximately 30 minutes long and digitized at a sampling rate of 360 Hz and gain of 200 analog-to-digital converter units per millivolt (adu/mV). The recordings were acquired from 47 subjects, 25 men aged 32—89 years and 22 women aged 23—89 years (record number 47 and 48 came from the same subject). Each record features simultaneous recordings from two leads, MLII and V5. This work tests the classification of three heartbeat classes, N, S, and V defined according to the AAMI/ANSI standard [11]. MITBIH provides annotations of associated with R-peak position of each heartbeat used for classification labels. Currently, we are using all the data available from the MITBIH database. No data exclusion based on outlier removal or feature reduction methods have been made.

2.2. Dynamic Margin Metric Learning

Dynamic triplet loss processes a set of three examples, referred to as a triplet, at a time to learn the desired embedding space where all possible triplets satisfy the triplet constraint. Here, it is utilized to train a hybrid neural network that combines CNN and BiLSTM with a layer of self-attention known as triplet at a time to learn the desired embedding representation as shown in Figure 1. The separation between N and S beats is especially difficult as the morphological and temporal similarities exist in both the heartbeats. We adjust the dynamic margin here to get the best separation between N and S beats especially. The representative embedding vector is then optimized for maximum separation between the arrhythmic classes that are already hard to separate. The margin is incrementally increased according to the feedback gradient. The next loop runs for a different margin. If the triplet loss for the current loop is lesser than the previous, the current setting is said to be optimized, otherwise the previous setting is carried as the optimum value. Similarly, the simulation goes on for 30 iterations to find the final perfect classification result for the currently simulated data.

Heartbeat input [12], defined as $x \in \mathrm{R}^{1\times L}$, where $L = 253$, is fed to the embedding network. The CNN layers extract local morphological patterns via two convolutional blocks with ReLU activation, batch normalization, and max-pooling, resulting in a downsampled sequence $x_c \in \mathrm{R}^{C\times T}$. The output is permuted to a temporal format $x_c \in \mathrm{R}^{T\times C}$ and fed into a bidirectional LSTM (Bi-LSTM) with hidden size h, producing a temporal embedding sequence $H = [h_1, h_2, \ldots, h_T] \in \mathrm{R}^{T\times 2h}$. We then apply an attention mechanism to compute a context vector $c \in \mathrm{R}^{2h}$, defined as:

$$\alpha_t = \frac{\exp(w^\top h_t)}{\sum_{j=1}^{T}\exp(w^\top h_j)}, \quad c = \sum_{t=1}^{T}\alpha_t h_t \tag{1}$$

where $w \in \mathrm{R}^{2h}$ is a learnable weight vector. This context vector is passed through a fully connected two-layer network and normalized to unit length, producing the final embedding $z \in \mathbb{R}$, where $\|z\|_2 = 1$.

To structure the embedding space so that similar heartbeats lie closer to dissimilar ones, we adopt a triplet loss with a dynamic margin m. A triplet consists of an anchor z^a, a positive z^p, and a negative z^n, where z^a, z^p belong to the same class, and z^n belongs to a different one. The loss is formulated as follows:

$$L_{triplet} = \max\left\{\|z^a - z^p\|_2^2 - \|z^a - z^n\|_2^2 + m,\ 0\right\} \tag{2}$$

In this work, we implement an adaptive strategy in which the triplet loss margin m is a dynamic variable that evolves during training. Initially set to a minimum value $m_{min} = 0.2$, it guides the model to learn an embedding space where the examples of each class are separated by at least this margin. As training progresses, the number of semi-hard triplets mined decreases. If this number falls below a predefined threshold for three consecutive iterations, m increases by 0.05. This cycle continues until it reaches a maximum value $m_{max} = 0.8$. The dynamic margin facilitates mining additional informative triplets and enhances inter-class separability. The minimum and maximum values of a were determined empirically.

2.3. Classification using MLP

Once the embedding network is trained, we freeze its weights and extract embeddings $z \in \mathrm{R}^d$ for all samples. These are fed into a lightweight multi-layer perceptron (MLP) for classification:

$$\hat{y} = softmax(W_2 \cdot ReLU(W_1 z + b_1) + b_2) \tag{3}$$

where $W_1 \in \mathrm{R}^{h'\times d}$, $W_2 \in \mathrm{R}^{K\times h'}$, and K is the number of heartbeat classes. The MLP is trained using the cross-entropy loss:

$$L_{CE} = -\sum_{i=1}^{K} y_i \log \hat{y}_i \tag{4}$$

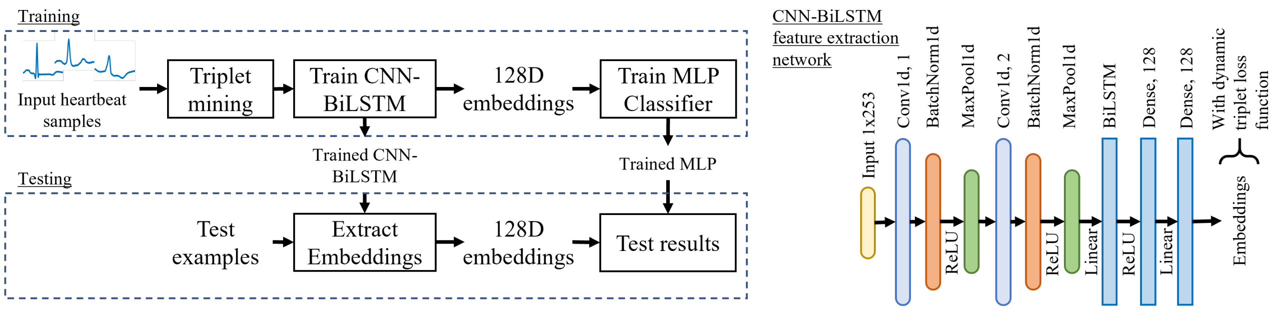

Figure 1. Overall diagram of the proposed model (left panel), and CNN-BiLSTM feature extraction and embedding optimization network (right panel).

where y_i is the true label (one-hot encoded) and $\hat{y}_i$ is the predicted probability.

2.4. Evaluation criteria

The performance of the model is evaluated and reported using both classwise and overall (macro-averaged) metrics of accuracy (Acc), sensitivity (Sen), precision (Pre) and F1-score (F1) using Equations 5, 6, 7, and 8, where $K = 3$ is the number of classes.

$$Acc = \frac{\sum_{k=1}^{K} TP_k}{N} \tag{5}$$

$$Sen_k = \frac{TP_k}{TP_k + FN_k},\ Sen_{mac} = \frac{1}{K}\sum_{k=1}^{K} Sen_k \tag{6}$$

$$Pre_k = \frac{TP_k}{TP_k + FP_k},\ Pre_{mac} = \frac{1}{K}\sum_{k=1}^{K} Pre_k \tag{7}$$

$$F1_k = \frac{2 \cdot Pre_k \cdot Sen_k}{Pre_k + Sen_k},\ F1_{mac} = \frac{1}{K}\sum_{k=1}^{K} F1_k \tag{8}$$

3. Results and Discussion

The effectiveness of this approach is reflected in the classification metrics. An overall accuracy of 98.97%, sensitivity of 97.76%, and F1-score of 96.82% indicate the model's strong ability to generalize, even when trained on only 50% of the available data. Notably, the classwise performance underscores the model's strength in distinguishing between morphologically similar classes. For instance, the sensitivity of 95.23% and F1-score of 93.36% for the class S is particularly impressive given its visual resemblance to class N. These results highlight that the dynamic margin contributes not only to better separation between majority and minority classes but also enhances the decision boundary between classes with overlapping morphologies. The t-SNE visualizations as shown in Figure 2 further validate this behavior by revealing well-separated clusters in the learned embedding space. Unlike conventional classification loss functions that do not explicitly enforce class separability, the dynamic triplet loss ensures that embeddings of the same class are pulled together, while embeddings of different classes are pushed apart based on the relative difficulty of each triplet. This dynamic adjustment avoids the pitfalls of a fixed margin approach, which may overfit or underfit depending on intra-class variability.

Table 1. Classification results for N, S and V classes on MITBIH dataset.

Class	Sen	Pre	F1
N	99.00	100.00	99.00
S	95.00	91.00	93.00
V	99.00	98.00	98.00
Overall	97.66	96.33	96.66

Comparative analysis with existing methods also reveals the superiority of DynaECG-Net. Prior works have reported challenges in distinguishing between N and S classes, often leading to low sensitivity for the S class. The proposed method outperforms previous state-of-the-art models in this regard, particularly under data-scarce conditions, thereby demonstrating better generalizability and robustness. The high classification scores for both the majority (N) and minority (S, V) classes indicate that the model avoids the common issue of class imbalance bias. Another notable strength of this framework is its suitability for wearable ECG monitoring applications. The compact and discriminative nature of the learned embeddings suggests that they could be efficiently transferred to low-power edge devices for real-time classification, facilitating early detection of arrhythmic events in continuous monitoring scenarios. From a computational standpoint, although DynaECG-Net remains relatively lightweight compared to deeper networks, the triplet loss formulation requires careful triplet selection during training, which can increase training time and complexity. This has meaningful implications for personalized and remote healthcare.

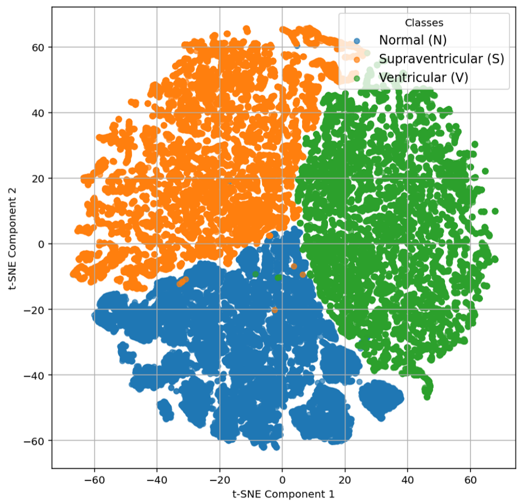

Figure 2. t-SNE visualization of embedding representation of training data in latent space.

While DynaECG-Net shows strong performance in classifying ECG beats under data-limited and low separability conditions, its current evaluation is limited to a disease-specific intra-patient setting and a focused three-class (N, S, V) task, which, although effective for analyzing hard-to-distinguish classes like N and S, does not represent the full range of arrhythmic conditions. Future work should explore domain adaptation techniques to enhance generalizability across cross-database scenarios by enabling the model to learn domain-invariant features, further strengthening the applicability of this model in clinical settings.

4. Conclusion

In conclusion, DynaECG-Net offers a robust and scalable solution for automated ECG classification, achieving high performance with limited data while ensuring interpretability and class separability.

Acknowledgments

This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) under the Innovative Human Resource Development for Local Intellectualization Program (IITP-2025-RS-2024-00439292).

References

[1] Şahin B, İlgün G. Risk factors of deaths related to cardiovascular diseases in world health organization (WHO) member countries. Health Social Care in the Community 2022;30(1):73–80.

[2] Kiranyaz S, Ince T, Gabbouj M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. IEEE Transactions on Biomedical Engineering 2015;63(3):664–675.

[3] Yin J, Hou B, Dai J, Zu Y. A CNN-BiLSTM method based on attention mechanism for class-imbalanced abnormal traffic detection. In Proceedings of the International Conference on Computer Vision and Deep Learning. 2024; 1–6.

[4] Mousavi S, Afghah F, Acharya UR. HAN-ECG: An interpretable atrial fibrillation detection model using hierarchical attention networks. Computers in biology and medicine 2020;127:104057.

[5] Zhang J, Liu A, Gao M, Chen X, Zhang X, Chen X. ECG-based multi-class arrhythmia detection using spatio-temporal attention-based convolutional recurrent neural network. Artificial Intelligence in Medicine 2020; 106:101856.

[6] Islam MR, Qaraqe M, Qaraqe K, Serpedin E. CAT-Net: Convolution, attention, and transformer based network for single-lead ECG arrhythmia classification. Biomedical Signal Processing and Control 2024;93:106211.

[7] Tao R, Wang L, Xiong Y, Zeng YR. IM-ECG: an interpretable framework for arrhythmia detection using multi-lead ECG. Expert Systems with Applications 2024; 237:121497.

[8] Thakur A, Thapar D, Rajan P, Nigam A. Deep metric learning for bioacoustic classification: Overcoming training data scarcity using dynamic triplet loss. The Journal of the Acoustical Society of America 2019;146(1):534–547.

[9] Moody GB, Mark RG. The impact of the MIT-BIH arrhythmia database. IEEE Engineering in Medicine and Biology Magazine 2001;20(3):45–50.

[10] Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. circulation 2000;101(23):e215–e220.

[11] Standard A. 'ANSI/AAMI ec57: Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms', 1998.

[12] Marinucci D, Sbrollini A, Marcantoni I, Morettini M, Swenne CA, Burattini L. Artificial neural network for atrial fibrillation identification in portable devices. Sensors 2020; 20(12):3570.

Address for correspondence:

Taesik Go, PhD
Jeonbuk National University
Department of Biomedical Engineering,
Jeonju, 54896, Republic of Korea
Email: tsgo@jbnu.ac.kr