Reading Between the Leads: Local Lead-Attention Based Classification of Electrocardiogram Signals

Gouthamaan Manimaran[1], Sadasivan Puthusserypady[1], Helena Dominguez[2], Jakob E Bardram[1]

[1] Technical University of Denmark, Copenhagen, Denmark
[2] Bisperbjerg-Frederiksberg Hospital, Copenhagen, Denmark

Abstract

Self-attention models have emerged as powerful tools in both computer vision and Natural Language Processing (NLP) domains. However, their application in time-domain Electrocardiogram (ECG) signal analysis has been limited, primarily due to the lesser need for global receptive fields. In this study, we present a novel approach utilizing local self-attention to address multi-class classification tasks using the PhysioNet/Computing in Cardiology Challenge 2021 dataset, encompassing 26 distinct classes across six different datasets. We introduce an innovative concept called "local lead-attention" to capture features within a single lead and across multiple configurable leads. The proposed architecture achieves an F1 score of 0.521 on the challenge's validation set, marking a 5.67% improvement over the winning solution. Remarkably, our model accomplishes this performance boost with only one-third of the total parameter size, amounting to 2.4 million parameters.

1. Introduction

Cardiovascular Diseases (CVD) continue to be the leading cause of mortality worldwide [1], and it is of utmost importance to provide timely and accurate diagnosis for effective intervention and patient care. ECG monitoring is the main component in the assessment of cardiac health, providing detailed insights into the activity of the heart. With advancements in computing technology and the growing availability of digital healthcare records, ECG data has become an essential resource for early detection and monitoring of cardiac conditions. Contemporary ECG algorithms heavily depend on deep learning technologies to deliver precise diagnostics, often in collaboration with human experts. The refinement of these algorithms holds paramount importance as they pave the way for advancing early prediction and treatment strategies for CVD.

This paper presents a solution to the PhysioNet/Computing in Cardiology Challenge 2021 [2]. The top contenders of this challenge [3–5] show impressive scores on the leaderboard of this challenge on both the validation and test sets. All these solutions use a convolution-only approach and for the right reason – ECG signal processing does not need the vanilla self-attention [6] module to achieve top scores. There has, however, been very little work done to investigate if a transformer network would perform better than a convolution network in ECG processing. This paper seeks to bridge this gap by presenting an approach that harnesses the power of attention in ECG signal classification. The motivation behind this approach arises from the recognition that the use of local self-attention mechanisms tailored to the unique characteristics of ECG signals could unlock new opportunities for improving classification accuracy and efficiency.

The paper introduces a novel method which is called "local lead-attention", which is designed explicitly for the ECG domain. This method allows our model to capture crucial features within a single lead and generalize its learning across multiple configurable leads. The findings presented in this paper show the untapped potential of self-attention models in medical signal analysis and demonstrate the significance of domain-specific adaptations to maximize their effectiveness.

2. Methods

This section explains our model architecture and our intuition behind this approach.

2.1. Architecture

To solve the problem of multi-class classification of arrhythmias with over 26 different classes in the PhysioNet Challenge 2021 [2, 7], we propose a novel architecture where the core processing module, the Local Self Attention, is based on the LongFormer [8] model from the NLP space. We adapt this model for time series analysis due to its smaller parameter size and its ability to give global context in every layer of the model.

The analysis of ECG signals does not require a global

context. Most arrhythmias can be determined from a single beat or the R-R intervals alone. Thus, a pure transformer network, which gives coarse local and fine-grained global interactions at every layer is not suitable for efficient signal classification, as seen from the top contenders of this challenge where they all use convolution-based models. In the proposed architecture, we mitigate this problem and use the established self-attention block to only process data points in a predefined kernel size. This would then work as a sliding window with a stride S to process all the features in a particular layer. This way, the network achieves a combination of the attention mechanism as well as the convolution architecture which is meaningful for a task like ECG signal processing. The global context is achieved in the later layers of the network where the features are shortened, similar to a convolution network as shown in Fig.1.

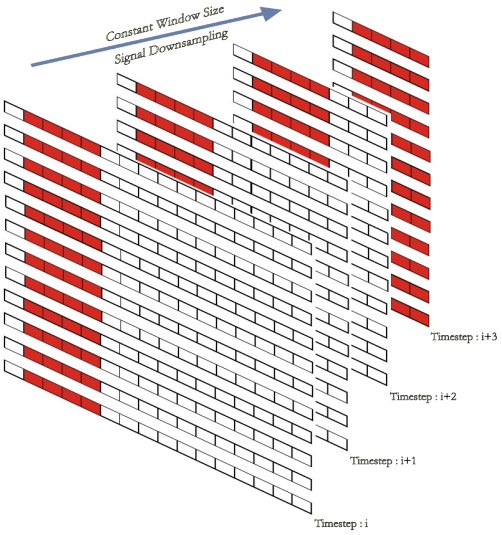

Figure 1. Local-Lead Attention with fixed window size

The work done by the winning solution of the ISIBrno Team [3] uses a 2D convolution algorithm for the various leads in the ECG signal. This model treats the individual leads as a separate entity till the final layer of the model after which the features from all the leads are aggregated. In our model, while we have a separate computation block with separate weights for each lead, we allow cross-talk between the leads inside the sliding window to learn features from the rest of the leads. In Fig. 1, all the highlighted boxes in a specific layer influence each other to learn lead-specific as well as lead-invariant features. This hypothesis is due to the fact that some arrhythmias are much more noticeable in some leads than other, thus we allow aggregation of features at every layer to learn these independencies as well as save on computation.

Local Lead-Attention: The core of a Transformer is self-attention. We briefly introduce the key idea to make the paper self-contained. Concretely, self-attention computes a weighted average of features with the weight proportional to a similarity score between pairs of input features. The Transformer network takes Z^0 as input. Given $Z^0 \in \mathbb{R}^{\mathbb{T}\times\mathbb{D}}$ with T time steps of D dimensional features, Z^0 is projected using $W_Q \in \mathbb{R}^{\mathbb{D}\times\mathbb{D}_{\shortparallel}}$, $W_K \in \mathbb{R}^{\mathbb{D}\times\mathbb{D}_{\urcorner}}$, and $W_V \in \mathbb{R}^{\mathbb{D}\times\mathbb{D}_{\gtrapprox}}$ to extract feature representations Q, K, and V, referred to as query, key, and value respectively with $D_k = D_q$. The outputs Q, K, and V are computed as:

$$\mathbf{Q} = \mathbf{Z^0 W_Q}, \quad \mathbf{K} = \mathbf{Z^0 W_K}, \quad \mathbf{V} = \mathbf{Z^0 W_V}. \tag{1}$$

The output of self-attention is given by:

$$\mathbf{S} = \text{softmax}\left(\frac{\mathbf{Q K^T}}{\sqrt{\mathbf{D_q}}}\right)\mathbf{V}, \tag{2}$$

where $S \in \mathbb{R}^{\mathbb{T}\times\mathbb{D}}$ and the softmax is performed row-wise. A multi-headed self-attention (MSA) further adds several self-attention operations in parallel. A main advantage of MSA is the ability to integrate temporal context across the full sequence. However, this comes at the cost of computation. A vanilla MSA has a complexity of $O(T^2D + D^2T)$ in both memory and time, making it inefficient for long videos. There have been several recent works on efficient self-attention [9]. In this paper, we adapt the local self-attention from [10] by limiting the attention within a local window. Our intuition is that the temporal context beyond a certain range is less helpful for ECG Classification. Such local self-attention significantly reduces the complexity to $O(W^2TD + D^2T)$, where W is the local window size ($\ll T$). Notably, local lead self-attention is used in conjunction with multiple leads $Z = \{L^1, L^2, \ldots, L^12\}$ with each lead feature contains multi-scale feature representation $L^1 = \{Z^1, Z^2, \ldots, Z^L\}\ldots$,$L^12 = \{Z^1, Z^2, \ldots, Z^L\}$, maintaining the same window size across each pyramid level. With this design, a small window size on a downsampled feature map covers a broad temporal range as shown in Fig. 1.

The final model architecture is shown in Fig.2. We initially project the N-Lead ECG signals using a shallow strided depthwise separable convolution block without overlap between the leads. This strided block is primarily used to shorten the ECG length before the attention module. The use of depthwise separable convolution [11] can be attributed to the parameter and computational efficiency, and more importantly, to reduce overfitting and introducing convolutional bias earlier on in the network.

The main block is the transformer encoder, which includes our novel local lead-attention block and also a strided version of this for down-sampling the signal. The final features are aggregated with a multi-head self-attention block with a global receptive field to aggregate attention across all the leads and time steps.

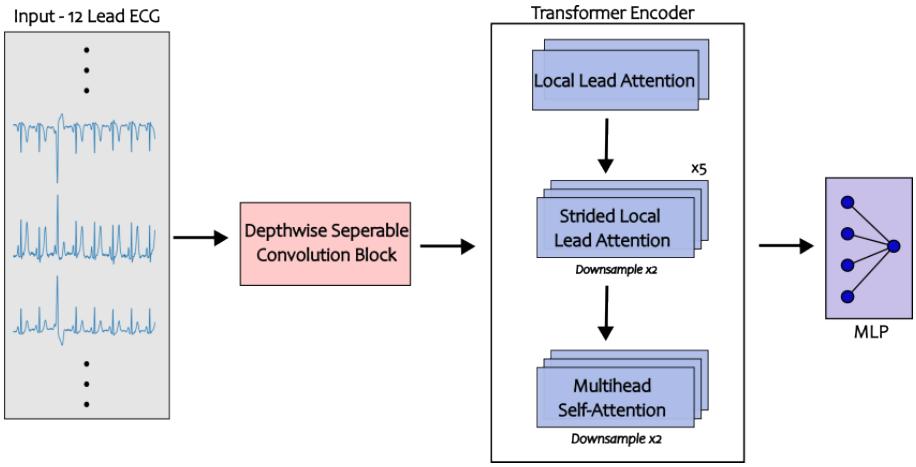

Figure 2. Proposed Model Architecture

2.2. Loss Functions

For the loss function, we used the constant weighted binary cross-entropy inspired via asymmetric loss (ASL) [12] similar to the work by *Han et al.* [4] that achieved good generalization ability. ASL is a method designed to tackle the inherent imbalance between positive and negative classes commonly encountered in multi-label classification tasks. It achieves this through the incorporation of asymmetric focusing and asymmetric probability shifting. The ASL formula is defined as follows:

$$ASL = \begin{cases} -(1-p)^{\gamma+}\log(p), & \text{if } y = 1 \\ -(p_m)^{\gamma-}\log(1-p_m), & \text{otherwise.} \end{cases} \quad (3)$$

Here, p represents the model's output probability, p_m signifies the shifted probability and $\gamma+$ and $\gamma-$ are positive and negative focusing parameters, respectively.

For practical implementation, we simplify the approach by setting the positive focusing parameter, $\gamma+$, to zero. We then conduct an exploration of a constant coefficient for the negative component, taking into account the optimal values of the negative focusing parameters, $\gamma-$, and the shifted probability, p_m. In our experimental setup, we assign a value of 0.1 to the negative coefficient, which approximately aligns with the positive-to-negative class ratio present throughout the dataset.

2.3. Model Training

The model is developed in PyTorch and has been trained for 100 epochs on an NVIDIA GeForce RTX 3090 GPU. The network is optimized through hyperparameter search and the values of the batch size are set to 16 with a learning rate of 0.0001 and a weight decay of 0.0001.

The data preprocessing pipeline involves several steps. Firstly, the provided data is expanded into a fixed 12-lead configuration, and any missing leads are padded with zeroes, resulting in a consistent matrix size of (12, time). Subsequently, resampling is applied to adjust the data to a uniform sampling frequency of 500 Hz, utilizing polyphase filtering when the original frequency is 1000 Hz or the FFT method for other cases. To enhance data quality, a zero-phase method with a 3rd-order Butterworth bandpass filter is employed, focusing on frequencies between 1 Hz and 47 Hz. Each ECG channel is subject to z-score normalized Data is zero-padded to a length of 8192 samples in the time domain, and if the signal exceeds this length, random sampling and cutting are performed to fit the required dimensions. Finally, during the training phase, a lead configuration is randomly selected (e.g., 12, 6, 4, 3, 2), with unused leads filled with zeros to augment the dataset for improved model generalization.

3. Results

We split the available data from the PhysioNet Challenge 2021 into a training and test dataset and performed our analysis here, due to the unavailability of the challenge test set. We set the random seed for the split to improve reproducibility and also have the same split for testing other solution winners. To have a reliable benchmark with the same data split, we retrain the winning solution of the challenge [3] with the same hyper-parameters. The results in Table 1 show the superior performance of our model, with a 5.3%, 6.61%, and 5.6% increase in AUPRC, AUROC, and F1 scores, respectively. This is achieved with the model size being 63% smaller than the previous state-of-the-art methods.

Method	AUPRC	AUROC	F-Measure	Model Size
ISIBrno [3]	0.901	0.514	0.493	6.5M
Local Lead-Attention (this work)	**0.949**	**0.548**	**0.521**	**2.4M**

Table 1. Results of the Challenge

4. Conclusion

Our research aimed to address the challenge of multi-class classification in ECG signals, where the global context may not be as crucial as in other domains. We incorporated a local lead-attention block that processes data points within a predefined kernel size, effectively combining the benefits of attention mechanisms and convolutional architectures for efficient signal classification.

The concept of local self-attention, which allows the network to focus on relevant information within a limited context, aligns well with the nature of ECG signals. Most arrhythmias can indeed be determined from local patterns, such as individual beats or R-R intervals, making the efficient processing of these local contexts highly effective.

Our model achieved remarkable results, with an F1 score of 0.521 on the challenge's validation set, marking a 5.67% improvement over the winning solution. Notably, this superior performance was accomplished with only one-third of the total parameter size, highlighting the efficiency of our approach.

Acknowledgments

This research has been funded by the Innovation Fund Denmark as part of the CATCH project (Project No. #1061-00046B) and the Copenhagen Center for Health Technology.

List of Acronyms

CVD Cardiovascular Diseases
NLP Natural Language Processing
ECG Electrocardiogram

References

[1] Tsao C, Aday A, Almarzooq Z, Alonso A, Beaton A, Bittencourt M, Boehme A, Buxton A, Carson A, Commodore-Mensah Y, Elkind M, Evenson K, Eze-Nliam C, Ferguson J, Generoso G, Ho J, Kalani R, Khan S, Kissela B, Martin S. Heart disease and stroke statistics-2022 update: A report from the american heart association. Circulation 01 2022; 145:CIR0000000000001052.

[2] Reyna MA, Sadr N, Alday EAP, Gu A, Shah AJ, Robichaux C, Rad AB, Elola A, Seyedi S, Ansari S, Ghanbari H, Li Q, Sharma A, Clifford GD. Will two do? varying dimensions in electrocardiography: The physionet/computing in cardiology challenge 2021. In 2021 Computing in Cardiology (CinC), volume 48. 2021; 1–4.

[3] Nejedly P, Ivora A, Smisek R, Viscor I, Koscova Z, Jurak P, Plesinger F. Classification of ecg using ensemble of residual cnns with attention mechanism. In 2021 Computing in Cardiology (CinC), volume 48. 2021; 1–4.

[4] Han H, Park S, Min S, Choi HS, Kim E, Kim H, Park S, Kim J, Park J, An J, Lee K, Jeong W, Chon S, Ha K, Han M, Yoon S. Towards high generalization performance on electrocardiogram classification. In 2021 Computing in Cardiology (CinC), volume 48. 2021; 1–4.

[5] Wickramasinghe NL, Athif M. Multi-label cardiac abnormality classification from electrocardiogram using deep convolutional neural networks. In 2021 Computing in Cardiology (CinC), volume 48. 2021; 1–4.

[6] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Lu, Polosukhin I. Attention is all you need. In Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R (eds.), Advances in Neural Information Processing Systems, volume 30. Curran Associates, Inc., 2017; .

[7] Reyna MA, Alday EAP, Gu A, Liu C, Seyedi S, Rad AB, Elola A, Li Q, Sharma A, Clifford GD. Classification of 12-lead ecgs: the physionet/computing in cardiology challenge 2020. In 2020 Computing in Cardiology. 2020; 1–4.

[8] Beltagy I, Peters ME, Cohan A. Longformer: The long-document transformer. CoRR 2020;abs/2004.05150. URL `https://arxiv.org/abs/2004.05150`.

[9] Zhang CL, Wu J, Li Y. Actionformer: Localizing moments of actions with transformers. In European Conference on Computer Vision. 2022; .

[10] Choromanski K, Likhosherstov V, Dohan D, Song X, Gane A, Sarlos T, Hawkins P, Davis J, Mohiuddin A, Kaiser L, et al. Rethinking attention with performers. arXiv preprint arXiv200914794 2020;.

[11] Kaiser L, Gomez AN, Chollet F. Depthwise separable convolutions for neural machine translation. ArXiv 2017; abs/1706.03059.

[12] Ridnik T, Ben-Baruch E, Zamir N, Noy A, Friedman I, Protter M, Zelnik-Manor L. Asymmetric loss for multi-label classification. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021; 82–91.

Address for correspondence:

Gouthamaan Manimaran
gouma@dtu.dk
Department of Health Technology,
Technical University of Denmark, 2800 Kgs. Lyngby,
Copenhagen, Denmark