Convolution-Free Waveform Transformers for Multi-Lead ECG Classification

Annamalai Natarajan[1], Gregory Boverman[1], Yale Chang[1], Corneliu Antonescu[2], Jonathan Rubin[1]

[1]Philips Research North America, Cambridge MA, USA
[2]University of Arizona Banner Health, Phoenix AZ, USA

Abstract

*We present our entry to the 2021 PhysioNet/CinC challenge — a waveform transformer model to detect cardiac abnormalities from ECG recordings. We compare the performance of the waveform transformer model on different ECG-lead subsets using approximately 88,000 ECG recordings from six datasets. Our waveform transformer model achieves a challenge metric between 0.66 to 0.67 on 12, 6, 4, 3 and 2-lead sets on preliminary 10-fold nested cross validation experiments. Our entry (**team name: prna**) also achieves a challenge score of 0.58, 0.55, 0.55, 0.54 and 0.53, respectively on 12, 6, 4, 3 and 2-lead sets on the challenge validation set.*

1. Introduction

Cardiovascular diseases (CVDs) are the leading cause of death globally with an estimated 32% of deaths worldwide in 2019 [1]. It is important to detect cardiovascular diseases early so treatments can be provided to mitigate complications. The standard 12-lead ECG has been a popular choice in the diagnosis of various cardiac abnormalities, but, more recently subsets of ECG leads have been used due to their size, cost, performance and ease of use. In this challenge, we utilize ECG recordings from subsets of standard 12-lead ECG to evaluate the efficacy of detecting cardiac abnormalities. The subsets of ECG leads include 12, 6, 4, 3, or 2 leads respectively. The challenge provides ~88K ECG recordings from five training data sources assigned to belong to one or more of the 30 cardiac abnormalities. More details about this challenge can be found in [2,3].

Prior work has shown deep neural networks to be successful in detecting cardiac abnormalities from 12-lead ECG signals [4]. In this work, we experiment with a waveform transformer model which differentially weights different parts of the inputs using an attention mechanism. This feature is a natural fit to this problem since cardiac abnormalities tend to be transient in ECG recordings. Our waveform transformer model is similar to our entry in the 2020 PhysioNet challenge [4] with some modifications:

1. We remove the embedding network that applied a series of convolution operations to the input ECG recording and instead rely on a convolution-free approach where the input signal is split into smaller segments and fed directly to a transformer model.

2. We seed the network with pretrained vision transformer network weights that have been trained on two-dimensional still images. We tailor our architecture to allow single-dimensional waveform inputs to make use of these pretrained weights.

We continue to train wide and deep networks that include 22 static ECG and demographic features, which are concatenated to the learned deep features of the network. These 22 features are similar to the ones listed in Table 1 in [4].

2. Methods

2.1. Pre-processing

Recordings from all databases provided by the challenge organizers [2, 3], including CPSC [6], INCART [7], PTB [8], PTB-XL [9], Chapman-Shaoxing [10] and Ningbo [11], were used for model training. We first split the data into 10 folds utilizing multi-label stratification [12]. Each recording was standardized to a sampling rate of 500Hz. We apply a finite impulse response bandpass filter with bandwidth between 3 - 45 Hz. Each recording is also normalized so that each channels' signal lies within the range of -1 to +1. We extract random fixed width windows from each recording across the subset of leads. We set the fixed width to be 7680 samples ($T = 15.36$ seconds), which allows the signal to be split into divisible segments sizes. We apply zero-padding to the recordings at the end when the sequence length is less than T seconds.

2.2. Waveform Transformer

An overview of the waveform transformer architecture is shown in Figure 1. Input to the network is a multi-lead ECG recording (e.g. 12, 6, 4, 3, or 2 leads). The input ECG recording is first broken up into smaller contiguous
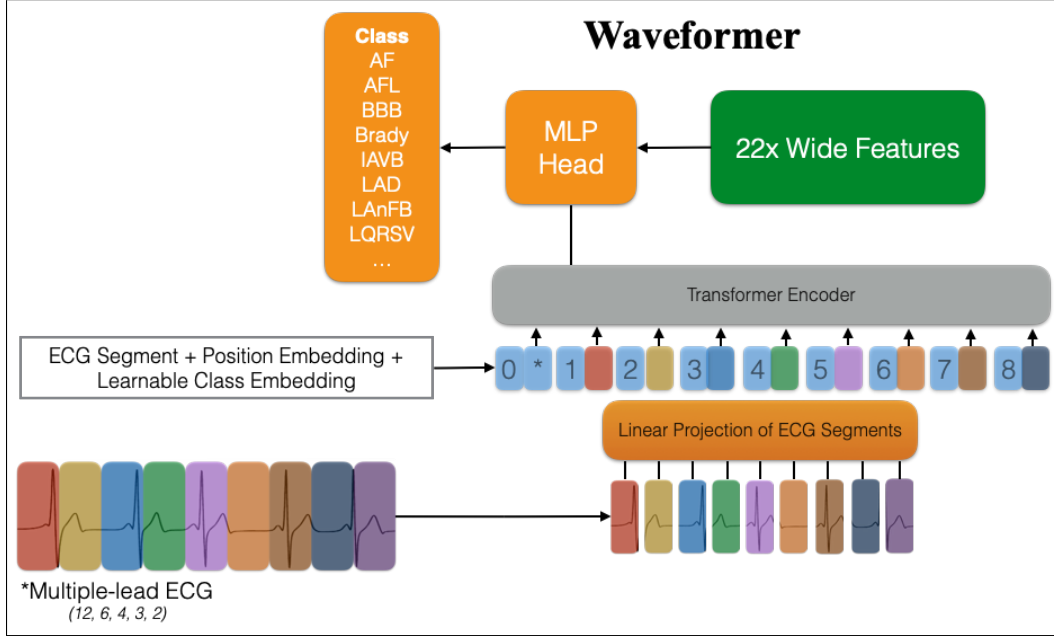
Figure 1. Architecture diagram illustrating the Waveform Transformer (Waveformer) for multi-lead, multi-label ECG classification. Linear projection is used to allow matching of pretrained weights. Wide (static) features can be directly fed as input to the MLP head. Based upon an original image in [5].

segments. Each segment undergoes a linear projection to embed it into a one dimensional vector that captures information for that time point in the overall recording. Linear projection of segments can be handled via a multi-layer perceptron (MLP) or convolution[1]. A sequence of embedded segments are then fed to the transformer encoder. Positional embedding is used to retain sequence order information. In addition, an extra learnable class token is fed to the transformer network that attends to all other tokens. The transformer model consists of 12 layers, using 12 attention heads and an embedding dimension of 768. We rely on pretrained weights of a vision transformer [5] trained on still image data, and as such, need to ensure the dimensions of the network match. We chose a base vision transformer model trained on 16x16 image patches with an embedding dimension of 768. To make a final prediction, the learnable class embedding is sent as input into an MLP head consisting of two linear layers. Static (wide) features are concatenated to the final linear layer of the network and a sigmoid operation is applied to make binary predictions about 26 classes[2]. Table 1 provides details on model architecture, settings and hyper-parameters used in our experiments. All models were trained using PyTorch using base models and pretrained weights from PyTorch Image Models [13]

Hyper-Parameter	Value
Global	
ECG window size (secs), T	15.36
Sampling frequency (Hz)	500
Batch size (train)	128
Batch size (validation)	64
Wide feature size, d_{wide}	22
Deep feature size, d_{deep}	64
Number of classes, d_{class}	26

Waveform Transformer	
ECG patch size, d_{patch}	64
Number of encoding layers	12
Embedding size, d_{model}	768
Number of heads	12
Dimension of feed forward layer	768
Dropout	0.1

Fully connected layers	
FC 1 size	64
FC 2 size	26
Dropout	0.2

Table 1. A listing of hyper-parameters selected to train the wide and deep neural network model for classification of cardiac abnormalities.

---

[1] Making the approach *almost* convolution-free.
[2] Equivalent classes are combined, reducing 30 classes to 26.

Fold	12-lead	6-lead	4-lead	3-lead	2-lead
1	0.635	0.608	0.630	0.619	0.626
2	0.642	0.632	0.644	0.622	0.634
3	0.660	0.644	0.645	0.645	0.649
4	0.674	0.645	0.660	0.667	0.658
5	0.677	0.657	0.667	0.666	0.656
6	0.689	0.676	0.686	0.670	0.677
7	0.694	0.678	0.690	0.673	0.676
8	0.699	0.683	0.687	0.683	0.680
9	0.691	0.684	0.697	0.693	0.674
10	0.692	0.678	0.689	0.689	0.675
Mean$\pm$s.d.	**0.675$\pm$0.023**	**0.658$\pm$0.026**	**0.669$\pm$0.024**	**0.663$\pm$0.026**	**0.661$\pm$0.019**

Table 2. 10-fold nested cross validation results showing challenge scores for different lead sets.

3. Results and Conclusions

In this section, we present preliminary results from our waveform transformer model. Our setup is a standard 10-fold nested cross validation. In each fold, we utilize data from the validation fold to learn probability thresholds and other hyper-parameters. In Table 2, we list the challenge metric on the held out test fold for different ECG lead subsets. We observe that overall, the results are very comparable across different ECG lead subsets are in the range of 0.66 to 0.67. As expected there is a monotonic improvement, albeit small, in the scores from the 2-lead to the 12-lead setup with an exception of 6-lead. This hints that the waveform transformer network is leveraging information from a small subset of ECG leads for the cardiac abnormalities of interest. Adding additional leads does not lead to significant improvement in performance in our preliminary experiments. However, there appears to be a significant improvement in performance when adding additional leads on the validation set. Our scores on the leader board for the 12, 6, 4, 3 and 2 leads are 0.578, 0.549, 0.550, 0.543, 0.532 respectively, which shows a 0.04 units improvement in the challenge metric between 2 and 12 leads.

We examined the AUROCs for each cardiac abnormality computed using the probabilities as output by the transformer model on a single fold. We observed similar trends in AUROC scores across different subsets of ECG leads. Our waveform transformer model achieved an AUROC of 0.85 and 0.77 on detecting low qrs voltage in 12-lead and 2-lead models respectively. We hypothesize that low qrs amplitude ECG recordings are very similar to normal sinus rhythms but with low amplitudes which makes detecting them challenging. We hypothesize that this poor performance on these cardiac abnormalities is that they often co-occur with other abnormalities that can be detected with high confidence. This low AUROC was followed by detection of T wave abnormal, Q wave abnormal and T wave inversion with AUROC's in the range of 0.88 to 0.90 respectively. All ECG lead subsets did exceptionally well in detecting pacing rhythm, left/right bundle block branch, tachycardia and bradycardia with AUROC's in the 0.98 to 0.99 range. This superior performance can be attributed to the unique physiological signatures embedded in ECG recordings and our inclusion of wide features such are heart rate. The AUROC for normal sinus rhythm is ~ 0.97 across all ECG lead subsets.

Lastly, we present the attention maps from the waveform transformer model in Figure 2. These attention maps illustrate which parts of the inputs are critical to making a final prediction much like a clinical expert manually scanning through ECG recordings. For each sample patient we show the ECG recording, attention maps along with ground truth (GT) and predicted (Pred.) labels. Both the x and y axis in the heatmaps represent time. Each cell represents a block of 64 samples (in time). Each sample (time point) can attend to all other time points. The high intensities (bright colors) along vertical time slices indicate that there are common time points that all time points are attending to. A potential shortcoming is highlighted for the bottom right patient which incorrectly predicts sinus rhythm, here the attention of all time points is on the length of the ECG recording which is non-informative.

References

[1] Cardiovascular Diseases. `https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)`. Accessed: 2021-08-10.

[2] Perez Alday EA, Gu A, Shah A, Robichaux C, Wong AKI, Liu C, et al. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. Physiological Measurement 2020;41.

[3] Reyna MA, Sadr N, Perez Alday EA, Gu A, Shah A, Robichaux C, et al. Will Two Do? Varying Dimensions in Electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021. Computing in Cardiology 2021;48:1–4.

[4] Natarajan A, Chang Y, Mariani S, Rahman A, Boverman G, Vij S, et al. A wide and deep transformer neural net-
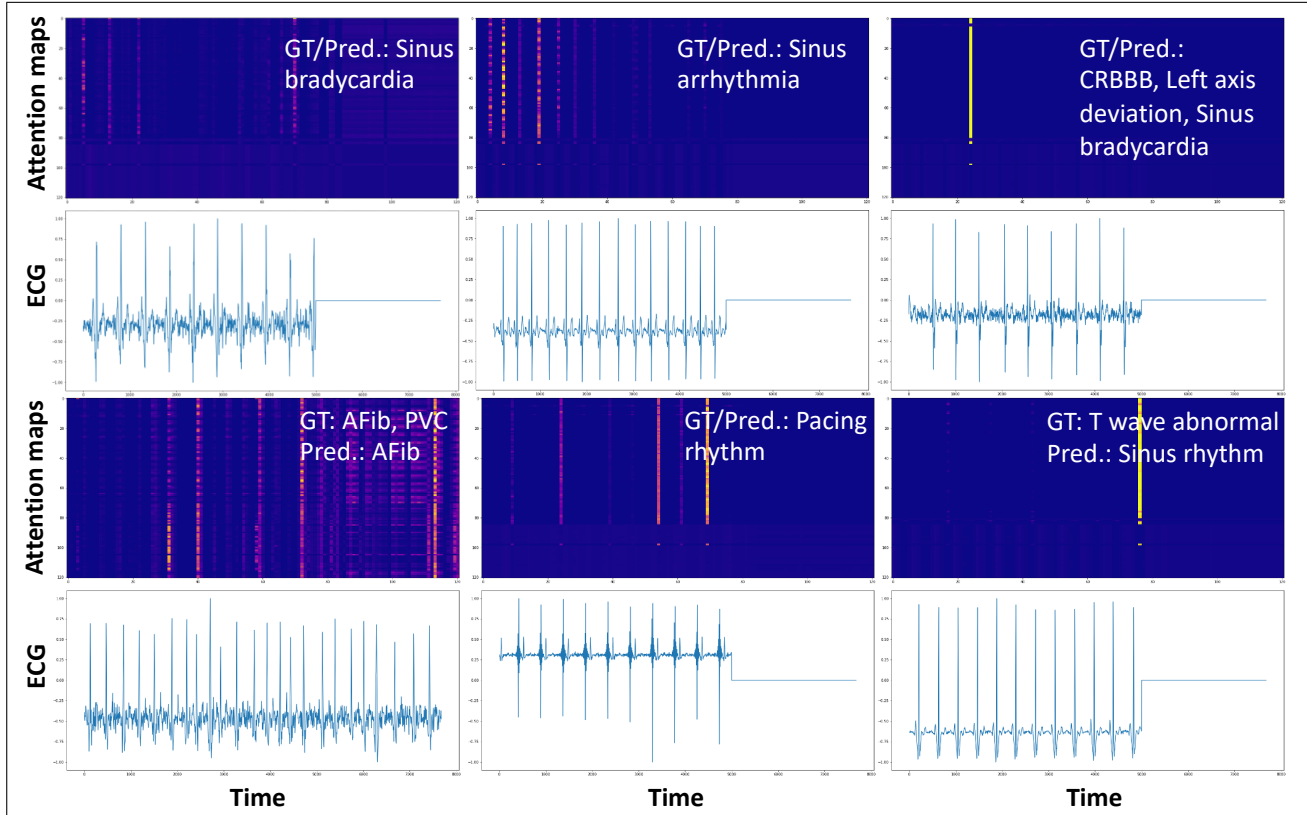
Figure 2. Six randomly selected attention maps that show where attention is being paid in the corresponding ECG recordings (shown below). Recordings consist of 7680 samples that are encoded via a waveform transformer network into $N = 120$ sequential embeddings. Attention heat maps are shown for lead II in the 12-lead setup. Also shown are the ground truth (GT) and predicted (Pred.) class labels.

work for 12-lead ecg classification. In 2020 Computing in Cardiology. IEEE, 2020; 1–4.

[5] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv201011929 2020;.

[6] Liu F, Liu C, Zhao L, Zhang X, Wu X, Xu X, et al. An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection. Journal of Medical Imaging and Health Informatics 2018;8(7):1368—1373.

[7] Tihonenko V, Khaustov A, Ivanov S, Rivin A, Yakushenko E. St Petersburg INCART 12-lead Arrhythmia Database. PhysioBank PhysioToolkit and PhysioNet 2008;Doi: 10.13026/C2V88N.

[8] Bousseljot R, Kreiseler D, Schnabel A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. Biomedizinische Technik 1995;40(S1):317–318.

[9] Wagner P, Strodthoff N, Bousseljot RD, Kreiseler D, Lunze FI, Samek W, et al. PTB-XL, a Large Publicly Available Electrocardiography Dataset. Scientific Data 2020;7(1):1–15.

[10] Zheng J, Zhang J, Danioko S, Yao H, Guo H, Rakovski C. A 12-lead Electrocardiogram Database for Arrhythmia Research Covering More Than 10,000 Patients. Scientific Data 2020;7(48):1–8.

[11] Zheng J, Cui H, Struppa D, Zhang J, Yacoub SM, El-Askary H, et al. Optimal Multi-Stage Arrhythmia Classification Approach. Scientific Data 2020;10(2898):1–17.

[12] Szymański P, Kajdanowicz T. A scikit-based Python environment for performing multi-label classification. ArXiv e prints February 2017;.

[13] Wightman R. Pytorch image models. https://github.com/rwightman/pytorch-image-models, 2019.

Address for correspondence:

Jonathan Rubin
222 Jacobs St Cambridge, MA 02141, United States
jonathan.rubin@philips.com