Towards Understanding ECG Rhythm Classification Using Convolutional Neural Networks and Attention Mappings

Sebastian D. Goodfellow SEBASTIAN.GOODFELLOW@SICKKIDS.CA *Department of Critical Care Medicine*
The Hospital for Sick Children
Toronto, Ontario, Canada

Andrew Goodwin ANDREW.GOODWIN@SICKKIDS.CA *Department of Critical Care Medicine*
The Hospital for Sick Children
Toronto, Ontario, Canada

Robert Greer ROBERT.GREER@SICKKIDS.CA *Department of Critical Care Medicine*
The Hospital for Sick Children
Toronto, Ontario, Canada

Peter C. Laussen PETER.LAUSSEN@SICKKIDS.CA *Department of Critical Care Medicine*
The Hospital for Sick Children
Toronto, Ontario, Canada

Mjaye Mazwi MJAYE.MAZWI@SICKKIDS.CA *Department of Critical Care Medicine*
The Hospital for Sick Children
Toronto, Ontario, Canada

Danny Eytan DANNY.EYTAN@SICKKIDS.CA *Department of Critical Care Medicine*
The Hospital for Sick Children
Toronto, Ontario, Canada

Editor: Editor's name

Abstract

In this study, a deep convolutional neural network was trained to classify single lead ECG waveforms as either Normal Sinus Rhythm, Atrial Fibrillation, or Other Rhythm. The dataset consisted of 12,186 labeled waveforms donated by AliveCor® for use in the 2017 Physionet Challenge. The study was run in two phases, the first to generate a classifier that performed at a level comparable to the top submission of the 2017 Physionet Challenge, and the second to extract class activation mappings to help better understand which areas of the waveform the model was focusing on when making a classification. The convolutional neural network had 13 layers, including dilated convolutions, max pooling, ReLU activation, batch normalization, and dropout. Class activation maps were generated using a global average pooling layer before the softmax layer. The model generated the following average scores, across all rhythm classes, on the validation dataset: precision=0.84, recall=0.85, F_1=0.84, and accuracy=0.88. For the Normal Sinus Rhythm class activation maps, we observed roughly constant attention, while for the Other Rhythm class, we observed attention spikes associated with premature beats. The class activation maps would allow for some level of interpretability by clinicians, which will likely be important for the adoption of these techniques to augment diagnosis.

1. Introduction

Many researchers have used various machine learning techniques for heart rhythm classification/detection from ECG waveforms using one of two machine learning strategies: (1) Step-By-Step Machine Learning and (2) End-To-End Deep Learning. Step-By-Step Machine Learning is the most common approach and involves signal processing, hand engineering features, and classification whereas End-To-End Deep Learning uses Deep Neural Networks, which take raw ECG waveforms as input and output a classification, thereby removing the need to hand engineer features.

Table 1: Training dataset summary.

		Duration (Seconds)			
Label	**Count**	**Mean**	**Std**	**Max**	**Min**
Normal Sinus Rhythm	5154	31.9	10	61	9
Atrial Fibrillation	771	31.6	12.5	60	10
Other Rhythm	2557	34.1	11.8	60.9	9.1
Noisy	46	27.1	9	60	10.2
Total	8528	32.5	10.9	61	9

When using the traditional Step-By-Step approach (Kara and Okandan (2007); Asl et al. (2008); Asgari et al. (2015); Rodenas et al. (2015); Goodfellow et al. (2017); Datta et al. (2017); Goodfellow et al. (2018)), a wide range of features are extracted from an ECG waveform. These features include standard heart rate variability features (Goldberger et al. (2000)), entropy features (Richman and Moorman (2000); Alcaraz et al. (2010); Zhou et al. (2010); Lake and Moorman (2011); DeMazumder et al. (2013)), Poincaré plot features (Park et al. (2009)), and heart rate fragmentation features (Costa et al. (2017)). They also include features that describe the QRS-complex morphology and detect the presence of P-waves and T-waves. Following feature extraction, a classifier is then trained to learn a function that maps from the ECG features to a given rhythm class.

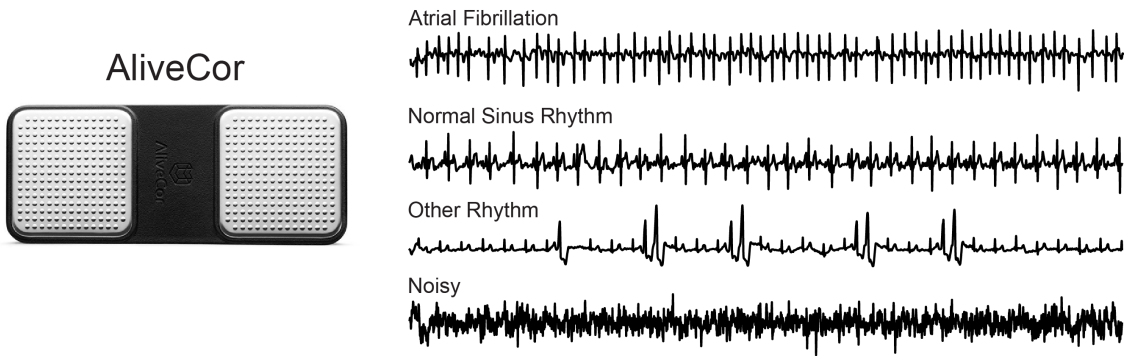

Figure 1: Left: AliveCor® hand held ECG acquisition device. Right: Examples of ECG recording for each rhythm class.

When using End-To-End deep learning for ECG rhythm classification, researchers have used convolutional neural networks (Pourbabaee et al. (2016); Rajpurkar et al. (2017); Acharya et al. (2017); Kamaleswaran et al. (2018)) and recurrent neural networks (Limam and Precioso (2017); Schwab et al. (2017)). A recent study by Rajpurkar et al. (2017) used a 33 layer residual neural network, a convolutional neural network architecture, to classify 14 different heart rhythms including Normal Sinus Rhythm, Ventricular Bigeminy, Atrial Fibrillation, Atrial Flutter, and others, and achieved an average F_1 score of 0.78.

In this study, we seek an improved understanding of the inner workings of a convolutional neural network ECG rhythm classifier. With a move towards understanding how a neural network comes to a rhythm classification decision, we may be able to build interpretability tools for clinicians and improve classification accuracy. Recent studies have focused on extracting attention mappings (class activation maps) from convolutional neural networks trained for image classification (Xu et al. (2015); Zagoruyko and Komodakis (2016); Zhou et al. (2016); Zhang et al. (2017)). Zhou et al. (2016) showed that class activation maps allow for predicted class scores to be visualized, demonstrating the implicit attention of a convolutional neural network on an image. The networks showed impressive localization ability despite having been trained for image classification. This study takes the formulations derived for convolutional neural network attention mapping of images and applies them to time series classification in order to visualize the region of an ECG waveform that receives the most attention during rhythm classification.

Technical Significance The main technical contribution of this study is the extension of Zhou et al. (2016)'s class activation maps from 2D image data to 1D time series data.

Clinical Relevance The clinical relevance of this study is that by using class activation maps for ECG rhythm classification, some level of interpretability is possible which will be important for the adoption of these techniques to augment decision making at the bedside. The class activation maps will potentially allow for clinicians to interpret and understand a model's classification decision.

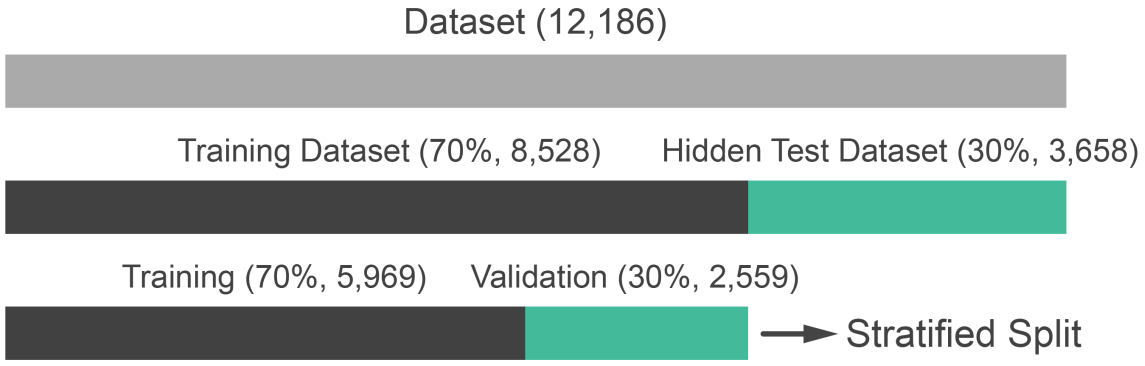

Figure 2: Dataset splitting strategy.

2. Dataset

The dataset used for this study is from the 2017 Physionet challenge where the objective was to build a model to classify a single lead ECG waveform as either Normal Sinus Rhythm, Atrial Fibrillation, Other Rhythm, or Noisy. The database consisted of 12,186 ECG waveforms that were donated by AliveCor®. Data were acquired by patients using AliveCor®'s single-channel ECG device and ranged in duration from 9 seconds to 61 seconds (Table 1). A measurement was taken by placing two fingers on each of the silver electrodes visible on the device in Figure 1. Data were stored as 300 Hz, 16-bit files with a ±5 mV dynamic range (Clifford et al. (2017)).

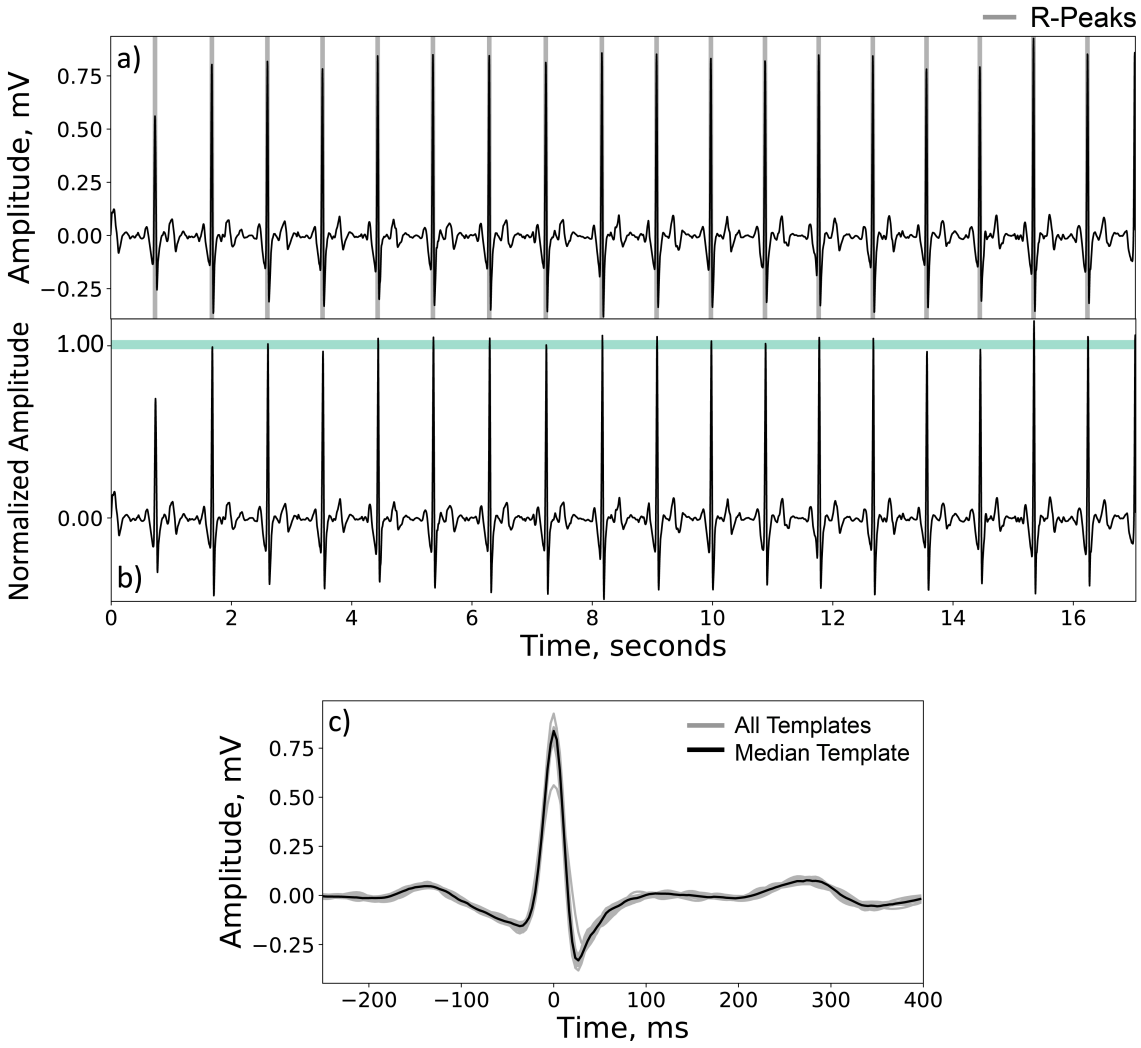

Figure 3: Example of (a) picked R-peaks from a raw waveform, (b) a normalized waveform and (c) extracted templates.

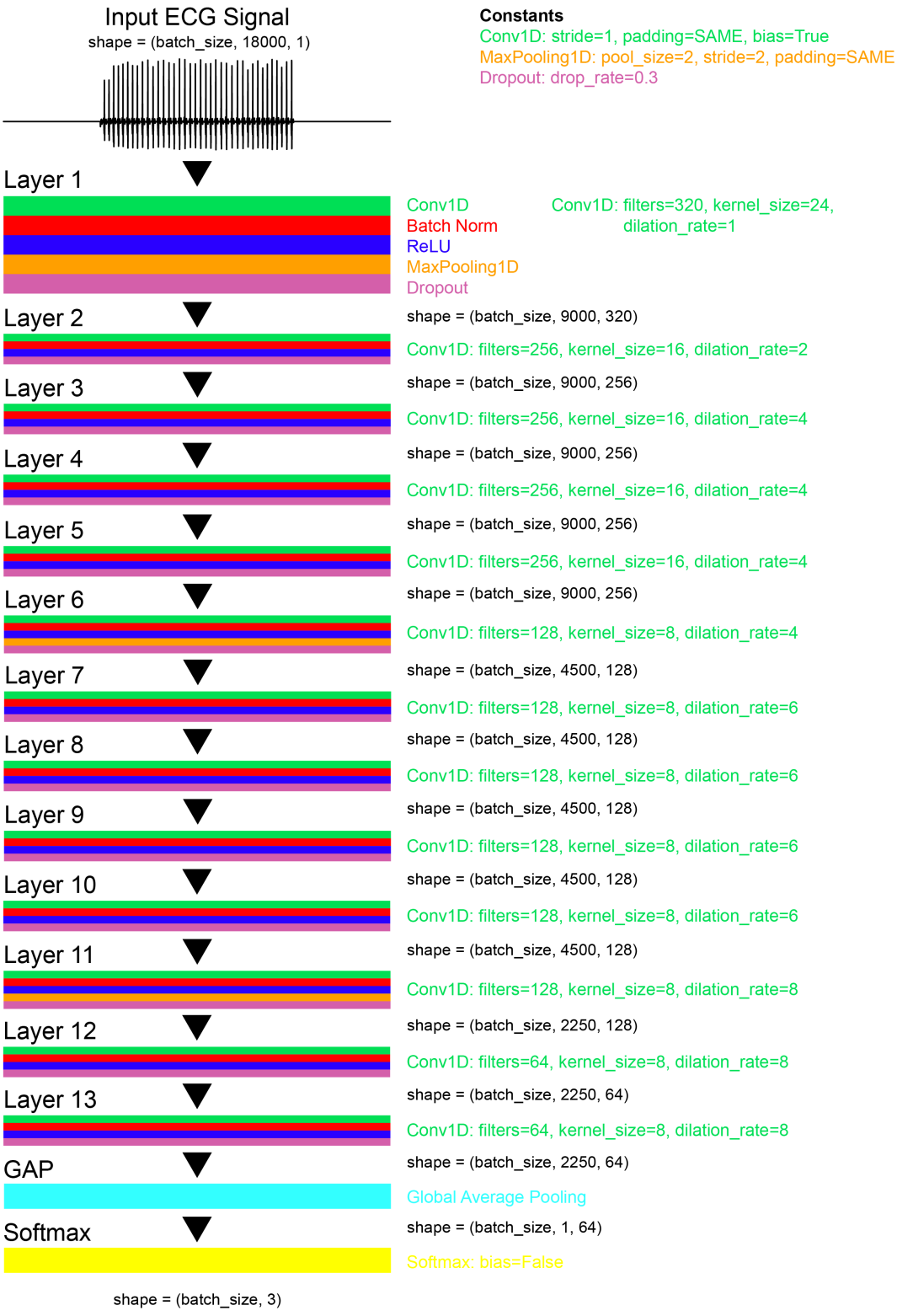

Figure 4: Deep convolutional neural network architecture.

The complete dataset consisting of 12,186 waveforms was first split into training (8,528) and hidden test (3,658) datasets following a 70%/30% split (Figure 2). Whenever the dataset was split, steps were taken to preserve the proportions of different rhythm classes to ensure each sub-dataset was representative of the original. The training dataset, consisting of 8,528 waveforms, was given to contestants for use in building a heart rhythm classifier. Contestants had a total of five months to build their model after which their model was uploaded to Physionet server where it was evaluated one time on the hidden test dataset. The top ten finalists produced F_1 scores between 0.81 and 0.83.

The dataset consisted of four rhythm classes; Normal Sinus Rhythm (NRS), Atrial Fibrillation (AF), Other Rhythm (OR), and Noisy (N) (Figure 1). Normal Sinus Rhythm is the default heart rhythm and Atrial fibrillation is the most common tachyarrhythmia of the heart in adults and is associated with an increased risk of stroke and heart failure. The Other Rhythm class consisted of many different heart arrhythmias all grouped into one class. This class could include arrhythmias such as Ventricular Tachycardia, Atrial Flutter, Ventricular Bigeminy, and Ventricular Trigeminy, however, this level of detail was not available to competitors. Lastly, the Noisy class consisted of real ECG waveforms where the labelers could not produce a confident heart rhythm classification. As presented in Table

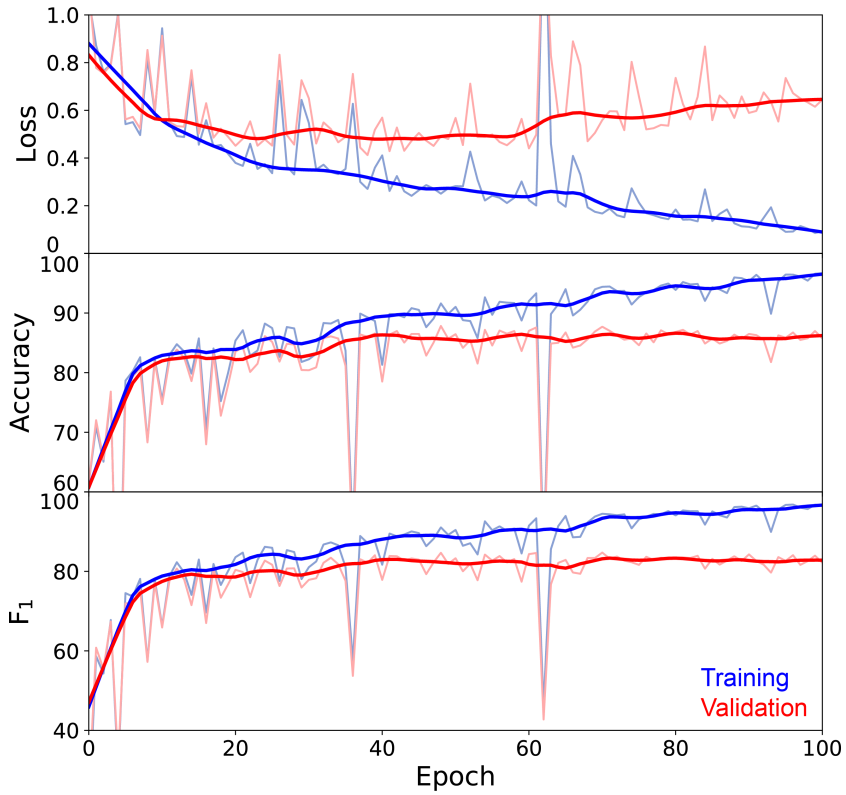

Figure 5: Training and validation dataset metrics as a function of training epoch (cross-entropy loss, accuracy, and F_1)

1, this dataset is imbalanced in order to reflect the true rates of occurrence in practice. For the purpose of this study, the Noisy class was removed from the dataset. See Clifford et al. (2017) for more information on how rhythm labels were generated.

3. Pre-Processing

Before being used for training and validation, each ECG waveform was pre-processed in the following ways. Full waveforms were first filtered using a finite impulse response (FIR) bandpass filter with a lower limit of 3 Hz and an upper limit of 45 Hz. With the ECG waveform filtered, the R-peaks were picked using the Hamilton–Tompkins algorithm (Hamilton and Tompkins (1986)) which returned an array of picked R-peak times as presented in Figure 3 (a) as vertical grey lines. Lastly, each ECG waveform was normalized to the median R-peak amplitude as shown in Figure 3 (b).

As shown in Table 1, the waveform durations ranged from 9 seconds to 61 seconds and a convolutional neural network needs a static waveform length for input. A waveform duration of 60 seconds (18,000 sample points) was chosen and any waveform with a duration of less than 60 seconds was zero padded as seen in Figure 4.

Table 2: Summary of model scores from the validation dataset.

Label	Precision	Recall	F_1	Accuracy	Size
Normal Sinus Rhythm	0.92	0.92	0.92	-	1523
Atrial Fibrillation	0.81	0.81	0.81	-	227
Other Rhythm	0.80	0.81	0.80	-	725
Average / Total	0.84	0.85	0.84	0.88	2475

4. Architecture

The objective of this study was to build an convolution neural network ECG rhythm classifier that performed at a level comparable to the top submission of the 2017 Physionet Challenge (F_1 = 0.81 - 0.83), and contained a global average pooling layer before the softmax classifier, allowing for class activation mappings to be extracted. The class activation mappings allowed for visualization of areas of the waveform the model was focusing on when making a classification decision. Our starting point was the 13 layer architecture of Kamaleswaran et al. (2018) that generated an F_1 score of 0.83 on the same dataset. Their model had 13 identical layers comprised of a 1D convolution, batch normalization, ReLU activation, max pooling, and dropout, which takes a (batch_size, 18275, 1) tensor as input and outputs a (batch_size, 1, 64) tensor for classification by a softmax layer. The temporal dimension of the waveform is reduced from 18275 to 1 by 13 max pooling operations with a pool size of 2. This architecture worked well for classification but was not ideal for extracting class activation mapping given the very low temporal resolution before the softmax layer. In order to improve the temporal resolution of the output from the final layer, 10 of the 13 max-pooling layers were removed, resulting in an output shape of (batch_size, 2250, 64) as opposed to (batch_size, 1, 64). As shown in Figure 4, each of the 13 layers contained a 1D convolution (bias term included), batch normalization, ReLU activation and dropout, and

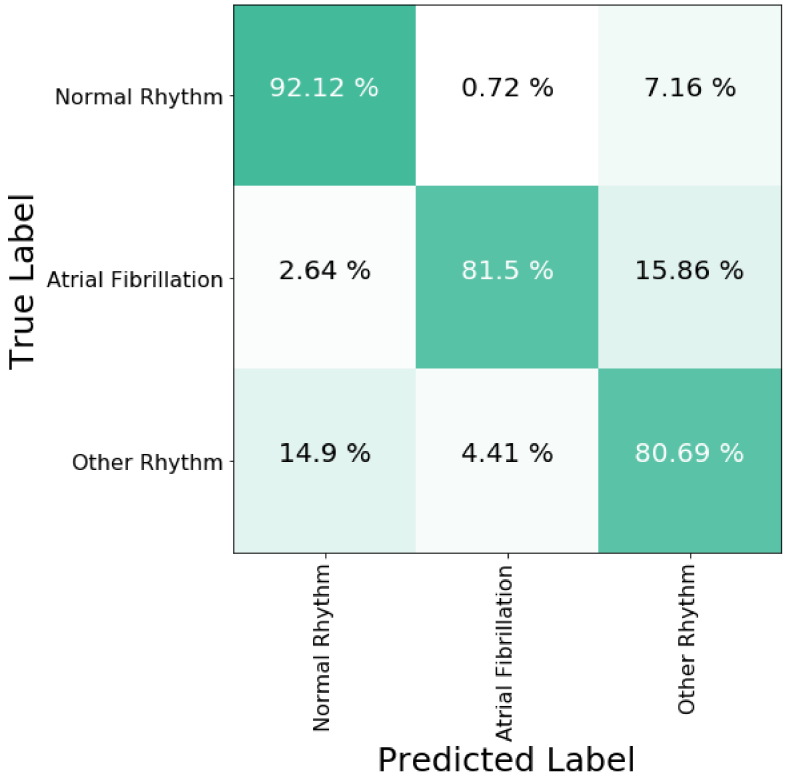

Figure 6: Confusion matrix for the validation dataset.

layers 1, 6, and 11 contained a max pooling layer between the ReLU and Dropout. Figure 4 contains details about the stride, pool size, filter count, padding, kernel size, and dropout rate for each layer. Dilation rates were increased from 1 to 8 for the 1D convolutions in order to increase the networks receptive fields. Lastly, a global average pooling (GAP) layer was used to compute the temporal average of the feature map of each unit from the last convolutional layer (layer 13). The output from the GAP layer had the shape (batch_size, 1, 64), which was then fed into a softmax classification layer, with the bias term excluded.

5. Training

For model training, the dataset given to contestant, consisting of 8,528 waveforms, was further split into a training and validation dataset following a 70/30 split (Figure 2). The splitting was stratified such to preserve the proportions of different rhythm classes and ensure that the training and validation datasets were representatives of the original. The model was trained for 100 epochs with a batch size of 10, resulting in 578 steps per epoch, and a constant learning rate of 1E-3. Optimization was carried out using ADAM (Kingma and Ba (2015)) and the cross-entropy loss function. The loss function was weighted by class to account for the class imbalance in the dataset (Table 1).

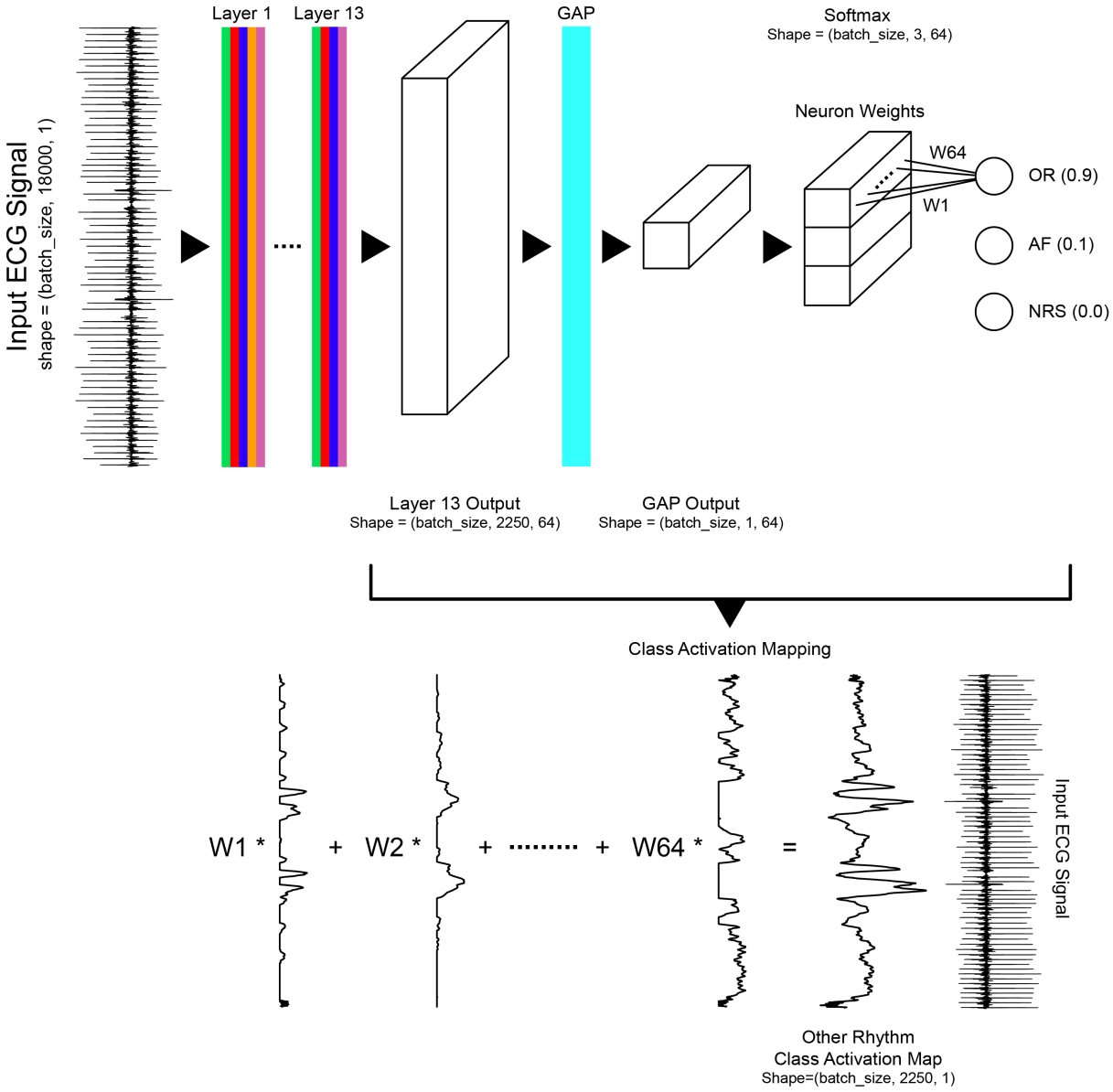

Figure 7: Class Activation Mappings (CAM) work flow.

6. Evaluation

Scoring for the 2017 Physionet Challenge used an F_1 measure, which is an average of three F_1 values for Normal Rhythm, Atrial Fibrillation, and Other Rhythm labels (Equation 1).

$$F_1 = \frac{F_{1,NRS} + F_{1,AF} + F_{1,OR}}{3} \tag{1}$$

The F_1 score is the harmonic mean of precision and recall where precision is defined as the number of true positives divided by the number of true positives plus the number of false

positives and recall is defined as the number of true positives divided by the number of true positives plus the number of false negatives.

The model was trained for 100 epoch (57,800 steps) and the cross-entropy loss, accuracy, and F_1 score, evaluated once per epoch on the training and validation datasets, can be seen in Figure 5. The final model weights used for the following class activations mapping analysis came from a checkpoint at epoch 48 where the validation F_1 score peaked at 0.847. Following epoch 48, the validation loss began to increase, suggesting the model may have started overfitting to the training dataset.

Table 2 presents a detailed overview of the model's performance (epoch 48 checkpoint) broken down by class rhythm. The model performs best for the Normal Sinus Rhythm class with an F_1 score of 0.92 and equally well for Atrial Fibrillation and Other Rhythm with F_1 scores of 0.81 and 0.80 respectively. Precision and recall were roughly equal for each rhythm class and the overall accuracy was 0.88.

Figure 6 displays a confusion matrix, for the validation dataset, which displays information about actual and predicted labels from the model. For example, in the validation dataset, there were 15,233 Normal Sinus Rhythm waveforms for which 92.12 % were correctly predicted as Normal Sinus Rhythm, 0.72 % were incorrectly predicted as Atrial Fibrillation, and 7.16 % were incorrectly predicted as Other Rhythm. The biggest errors made by the model were mislabelling Atrial Fibrillation as Other Rhythm and Other Rhythm as Normal Sinus Rhythm.

7. Class Activation Mappings

Zhou et al. (2016) demonstrate that convolutional neural networks trained for image classification appear to behave as object detectors despite information about the object's location not being part of the training labels (no bounding box annotations). Zhou et al. (2016) showed that their model could perform accurate object localization without any significant loss of classification accuracy. Object localization is enabled by class activation maps (CAM) and a global average pooling (GAP) layer before the softmax classifier allowing for the predicted class scores on any image to be visualized and for the discriminative object parts detected by the convolutional neural network to be highlighted.

Zhou et al. (2016)'s formulation was designed for analysis of images whereas our application is for time series. For our application, the class activation map for a particular rhythm class was used to indicate the discriminative temporal regions, not image regions, used by the convolutional neural network to identify that class rhythm. The following is a direct adaptation of Zhou et al. (2016)'s formulation for time series data.

For a given ECG time series, $f_k(t)$ represents the activation of filter unit k in the last convolutional layer (layer 13) at the temporal location t. From Figure 7, the activation map output from layer 13 has the shape (batch_size, 2250, 64), which corresponds to 64 filter units (k) with a temporal dimension of 2250 units. For filter unit k, the global average pooling, F^k, was calculated following $\sum_{t=1}^{2250} f_k(t)$, resulting in a reduction in the temporal dimension from 2250 to 1 (batch_size, 1, 64). The softmax logits for rhythm class c, S_c, were calculated following $\sum_{k=1}^{64} w_k^c F^k$, where w_k^c is the weights corresponding to class c for filter unit k, and the associated softmax class score, P_c, was calculated following $\frac{\exp(S_c)}{\sum_{c=1}^{3}\exp(S_c)}$. By substituting for F^k, we get equation 2.

$$S_c = \sum_{k=1}^{64} w_k^c F^k = \sum_{k=1}^{64} w_k^c \sum_{t=1}^{2250} f_k(t) = \sum_{t=1}^{2250} \sum_{k=1}^{64} w_k^c f_k(t) \tag{2}$$

The class activation map for class rhythm c, M_c, for a certain temporal unit, t, is given by

$$M_c(t) = \sum_{k=1}^{64} w_k^c f_k(t) \tag{3}$$

and thus, the softmax logits for class c are given by $S_c = \sum_{t=1}^{2250} M_c(t)$. This equations demonstrates that that activation map $M_c(t)$ represents the importance of activations for filter units $k = 1$ - 64 at temporal unit t leading to classification of an ECG time series as class c.

Figure 7 presents a visual representation of the prior formulation. The softmax layer contains three sets of weights (w_k^c), with dimension (batch_size, 1, 64), each corresponding to one rhythm class. In Figure 7, the high softmax score, $P_c = 0.9$, was associated with the Other Rhythm class. The Other Rhythm class activation map is computed by a weighted linear sum of the softmax wights (w_k^c) and the feature map output from layer 13 (f_k). The resulting class activation map had dimensions (batch_size, 2250, 1) and was later upsampled to the input temporal dimension of 18,000 using linear interpolation.

As presented in the previous section, this model performs at a level comparable to the top submission of the 2017 Physionet Challenge ($F_1 = 0.81$ - 0.83), although it is important to note that we did not evaluate the model on the hidden test dataset. This means that the addition of a global average pooling layer allowed for class activation maps to be extracted without sacrificing performance. However, the training time per epoch was roughly four times longer than that of Kamaleswaran et al. (2018) as a result of removing 10 of the 13 max-pooling layers in order to retain a high temporal resolution feature map at the final layer.

8. Results and Discussion

Class activation maps were extracted from each ECG waveform in the validation dataset and studies to understand the attention patterns associated with each rhythm class.

For Normal Sinus Rhythm (Figure 8 (a)), across hundreds of observations, the general class activation map pattern shows roughly constant attention (activation) for the duration of the waveform. Normal Sinus Rhythm is described by regular beat timing, as expressed by low variance entropy of interbeat interval times, as thus the observation of roughly constant attention fits the regular heart rhythm of Normal Sinus Rhythm. Normal Sinus rhythm is also characterized by the presence of a P-wave before every QRS-complex. To investigate the high-frequency attention trends, the class activation map was mapped to the templates as shown in Figure 8 (a). A template contains the P-wave, QRS-complex, and T-wave and was defined as 250 ms before the R-peak to 400 ms after the R-peak and therefore, each R-peak had an associated template (Figure 3 (c)). Figure 8 (a) shows that from template to template, there is a repeated class activation map shape, however, the shape does not intuitively map to a human understanding of Normal Sinus Rhythm template morphology.

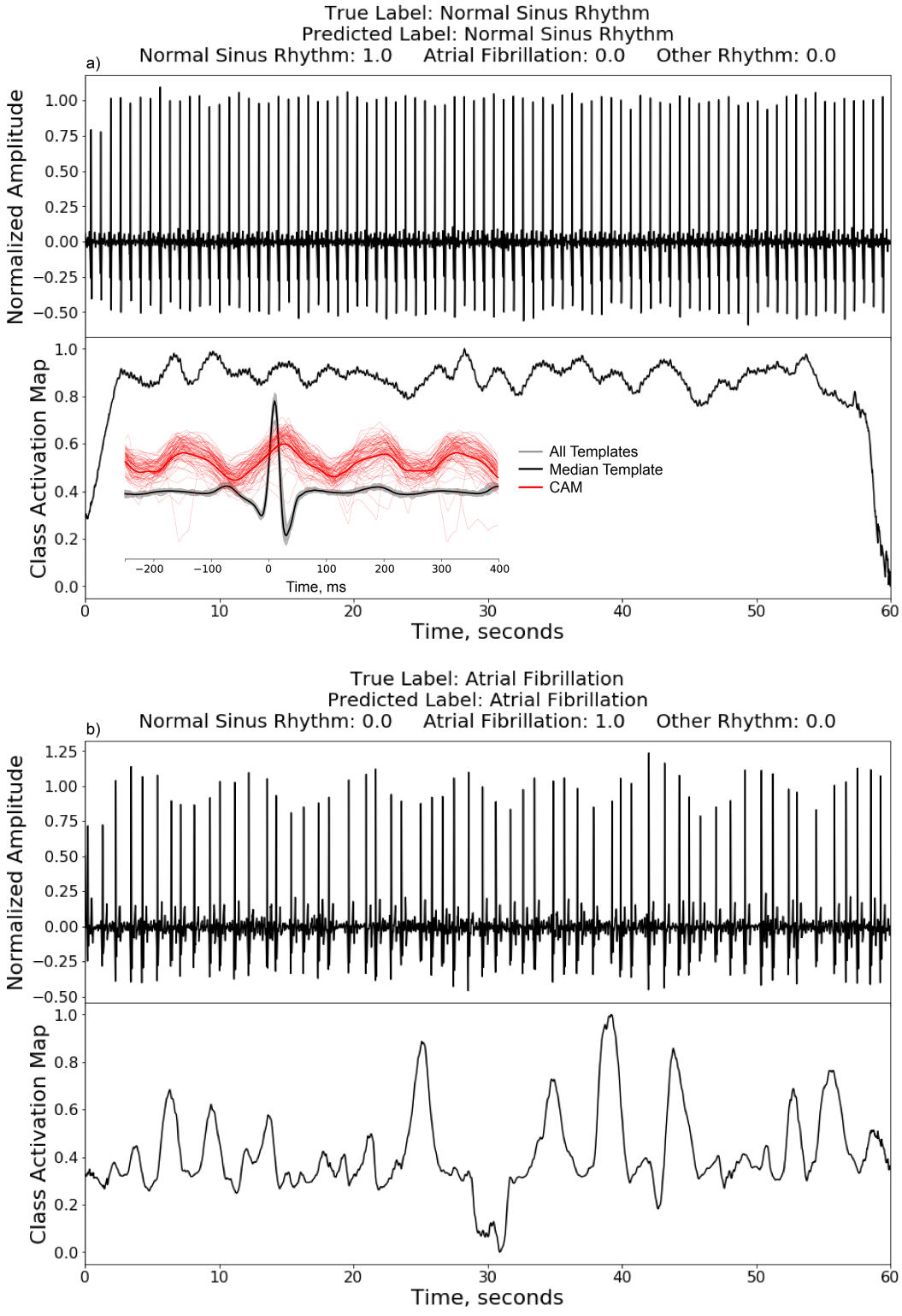

Figure 8: Examples of class activation maps for Normal Sinus Rhythm and Atrial Fibrillation.

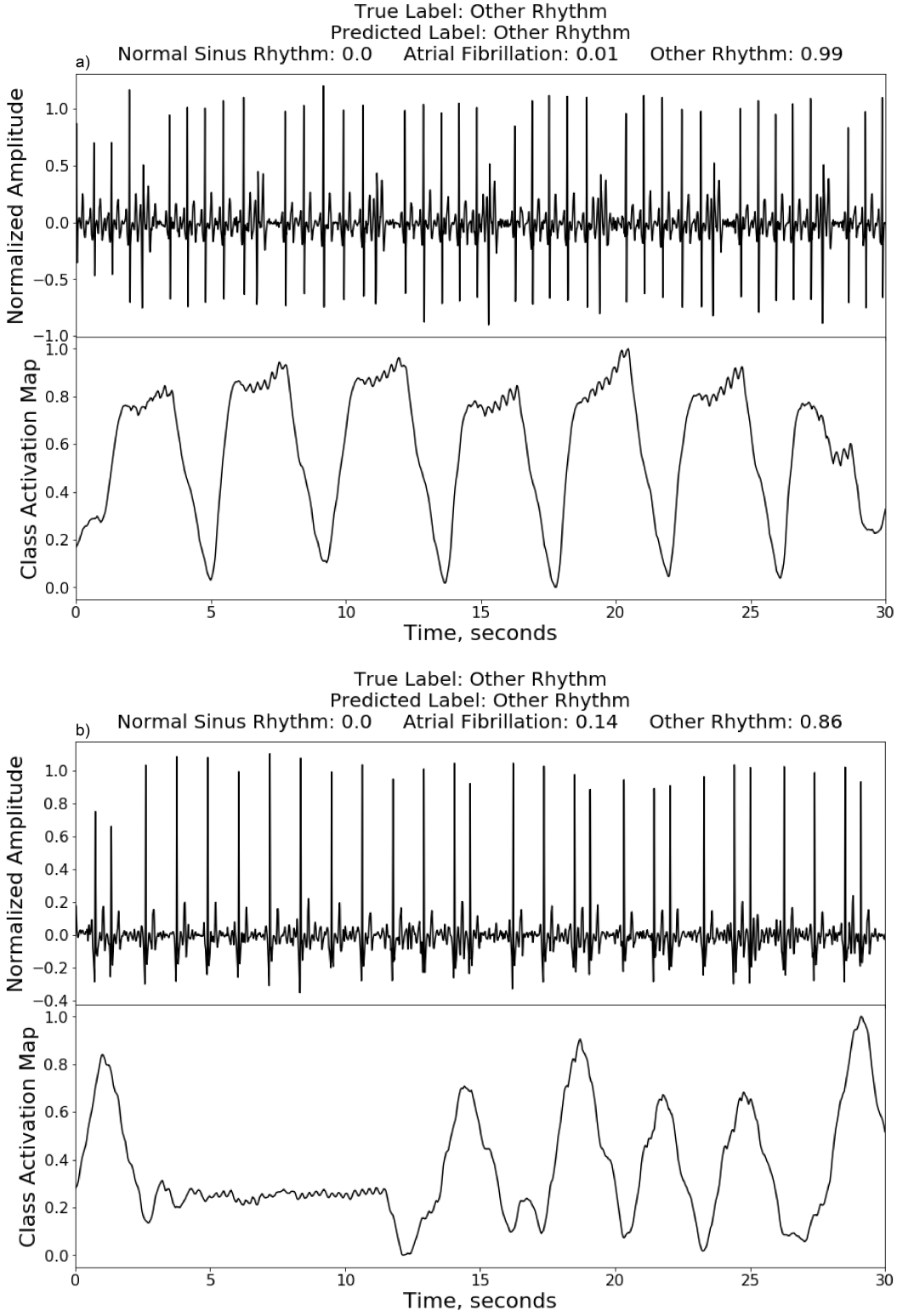

Figure 9: Examples of class activation maps for Other Rhythm.

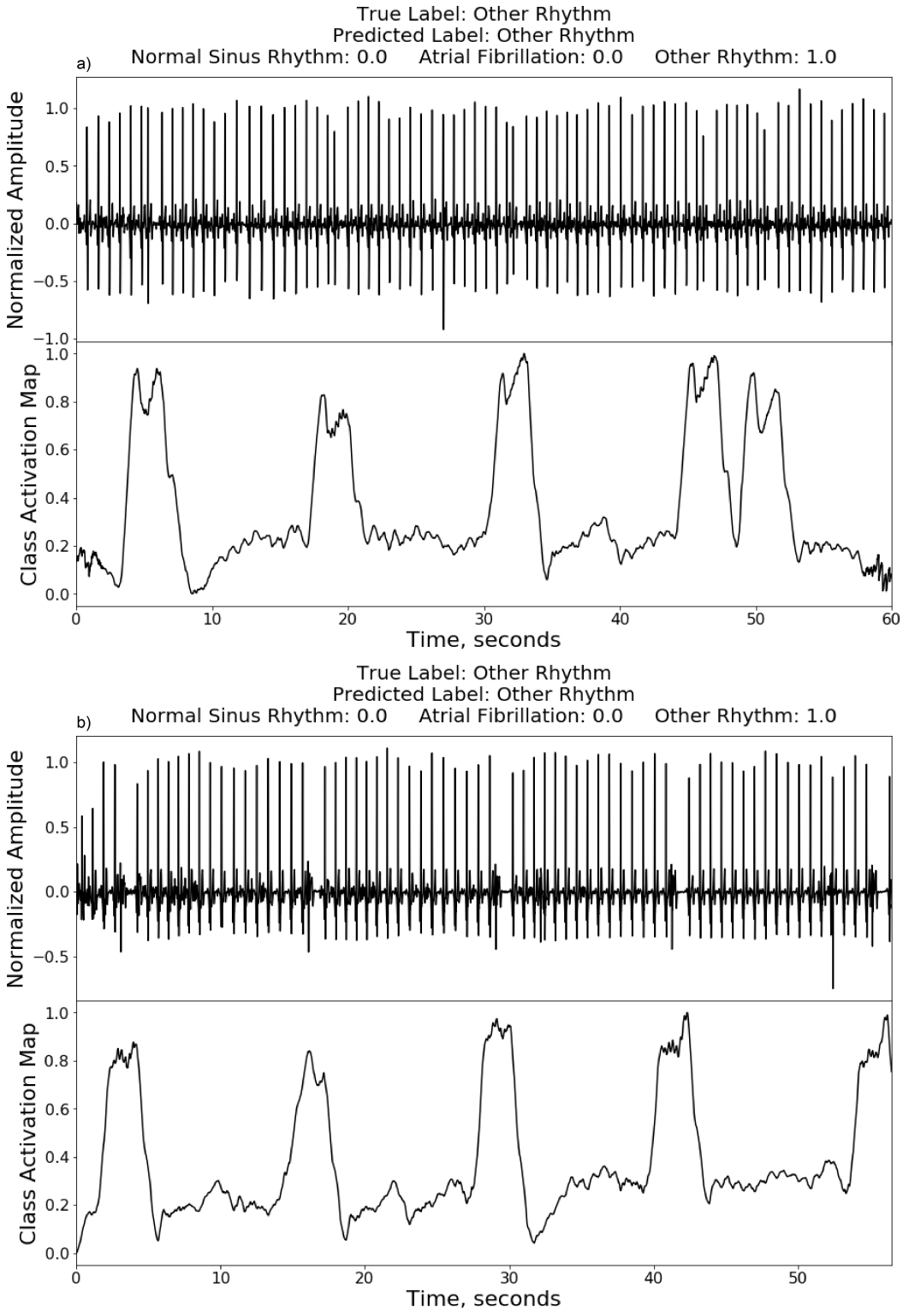

Figure 10: Examples of class activation maps for Other Rhythm.

For example, the neural network does not appear to be focusing its attention on the P-wave region of the template, as perhaps would have been expected.

For Atrial Fibrillation (Figure 8 (b)), the general class activation map pattern shows random attention fluctuations for the duration of the waveform. Atrial Fibrillation is described by an irregularly irregular beat rhythm and the lack of a P-wave before each QRS-complex. In Figure 8 (b), it appears that the activations are greatest where the longest gap between the T-wave and subsequent QRS-complex exists, which could be where the absence of a P-wave is most evident. The random attention fluctuations are consistent with an irregularly irregular beat rhythm, however, the class activation map may not intuitively map to a human understanding of this arrhythmia.

For Other Rhythm (Figures 9 and 10), the general class activation map pattern shows attention spikes associated with repeated patterns in the ECG waveform. The Other Rhythm class contains an unknown number of other heart arrhythmias potentially including Ventricular Tachycardia, Ventricular Bigeminy, Ventricular Trigeminy, Junctional Rhythm, and others. From our observation of the dataset, Other Rhythm waveforms typically contain premature beats that sometimes have different template morphologies from the rest. Figures 9 (a) and 10 (b) show regular premature beats with unique template morphologies. The class activation map for these waveform shows attention spikes associated with the premature beats. Figures 9 (b) and 10 (a) show irregular premature beats and again, attention spikes associated with them. The class activation maps for the Other Rhythm class are the most interpretable given they very easily map onto our clinical understanding of these arrhythmias.

Many researcher are working on the problem of automatic ECG rhythm classification given the clear clinical utility of such technology (Kara and Okandan (2007); Asl et al. (2008); Asgari et al. (2015); Rodenas et al. (2015); Pourbabaee et al. (2016); Rajpurkar et al. (2017); Acharya et al. (2017); Rajpurkar et al. (2017); Goodfellow et al. (2017); Datta et al. (2017); Kamaleswaran et al. (2018); Goodfellow et al. (2018)). The evaluation of ECG rhythm classification models typically focuses on their accuracy compared to health care professionals, however, one of their main advantages is that they can monitor continuously 24 hours per day. The problem with arrhythmias is that they can be hard to detect when occurring in short bursts and therefore can be overlooked. This does not change their significance, however, because short burst or sporadic arrhythmias may be a harbinger of more serious problems about to unfold.

The motivation for this study came from a need to bridge the gap between machine learning research and delivering direct care to patients at the bedside. The class activation maps would allow for some level of interpretability by clinicians, which will likely be important for the adoption of these techniques to augment diagnosis. In addition to displaying a classification score on a bedside monitor (Atrial Fibrillation: 98%), the class activation maps could provide an added visual means for clinicians to interpret the model output.

9. Conclusions

In conclusion, we trained a deep convolutional neural network to classify single lead ECG waveforms as either Normal Sinus Rhythm, Atrial Fibrillation, or Other Rhythm. The model generated the following average scores, across all rhythm classes, on the valida-

tion dataset: precision=0.84, recall=0.85, F_1=0.84, and accuracy=0.88. A global average pooling layer was used to extract class activation maps without hurting the model's classification performance. The class activation mappings allowed for visualization of regions of the waveform the model was focusing on when making a classification decision. For Normal Sinus Rhythm, we observed roughly constant attention, which fits the regular heart rhythm of Normal Sinus Rhythm. The class activation maps for the Other Rhythm class, which showed attention spikes associated with premature beats, were the most interpretable given they fit our clinical understanding of these arrhythmias.

Acknowledgments

The authors would like to acknowledge support from the David and Stacey Cynamon Chair in Pediatric Critical Care at The Hospital for Sick Children and thank PhysioNet and AliveCor® for providing the dataset.

References

U. R. Acharya, H. Fujita, O. S. Lih, Y. Hagiwara, J. H. Tan, and M. Adam. Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. *Information Sciences*, 405:81—90, 2017.

R. Alcaraz, D. Abasolo, R. Hornero, and J. J. Rieta. Optimal parameters study for sample entropy-based atrial fibrillation organization analysis. *Comput Methods Programs Biomed*, 99:124—-132, 2010.

S. Asgari, A. Mehrnia, and M. Moussavi. Automatic detection of atrial fibrillation using stationary wavelet transform and support vector machine. *Computers in Biology and Medicine*, 60:132–142, 2015.

B. Mohammadzadeh Asl, S. Kamaledin Setarehdan, and M. Mohebbi. Support vector machine-based arrhythmia classification using reduced features of heart rate variability signal. *Artificial Intelligence in Medicine*, 44:51–64, 2008.

G. D. Clifford, C. Liu, B. Moody, L. H. Lehman, I. Silva, A. E. Johnson, and R. G. Mark. AF Classification from a short single lead ECG recording: the PhysioNet/Computing in Cardiology Challenge 2017. In *Proceedings of Computing in Cardiology 2017, Renness, France, September 24 - 27, 2017*, 2017.

M. D. Costa, R. B. Davis, and A. L. Goldberger. Heart Rate Fragmentation: A New Approach to the Analysis of Cardiac Interbeat Interval. *Front. Physiol.*, 8:255, 2017.

S. Datta, C. Puri, A. Mukherjee, R. Banerjee, A. Dutta Choudhury, R. Singh, A. Ukil, S. Bandyopadhyay, A. Pal, and S. Khandelwal. Identifying Normal, AF and other Abnormal ECG Rhythms using a Cascaded Binary Classifier. In *Proceedings of Computing in Cardiology 2017, Renness, France, September 24 - 27, 2017*, 2017.

D. DeMazumder, D. E. Lake, A. Cheng, T. J. Moss, E. Guallar, R. G. Weiss, S. R. Jones, G. F. Tomaselli, and J. R. Moorman. Dynamic analysis of cardiac rhythms for discriminating atrial fibrillation from lethal ventricular arrhythmias. *Am J Physiol Heart Circ Physiol*, 6:555—-561, 2013.

A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, and R. G. Mark. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation*, 101:215–220, 2000.

S. D. Goodfellow, A. Goodwin, R. Greer, P. C. Laussen, M. Mazwi, and D. Eytan. Classification of Atrial Fibrillation Using Multidisciplinary Features and Gradient Boosting. In *Proceedings of Computing in Cardiology 2017, Renness, France, September 24 - 27, 2017*, 2017.

S. D. Goodfellow, A. Goodwin, R. Greer, P. C. Laussen, M. Mazwi, and D. Eytan. Atrial fibrillation classification using step-by-step machine learning. *Biomedical Physics and Engineering Express*, 4, 2018.

P. S. Hamilton and W. J. Tompkins. Quantitative investigation of QRS detection rules using the MIT/BIH arrhythmia database. *IEEE Trans Eng Biomed Eng*, 33:1157—-1165, 1986.

R. Kamaleswaran, R. Mahajan, and O. Akbilgic. A robust deep convolutional neural network for the classification of abnormal cardiac rhythm using single lead electrocardiograms of variable length. *Physiological Measurement*, 39:035006, 2018.

S. Kara and M. Okandan. Atrial fibrillation classification with artificial neural networks. *Pattern Recognition*, 40:2967–2973, 2007.

D. P. Kingma and J. Ba. Adam: A Method for Stochastic Optimization. In *Proceedings of 3rd International Conference for Learning Representations, San Diego*, 2015.

D. E. Lake and J. R. Moorman. Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices. *Am J Physiol Heart Circ Physiol*, 300:319–325, 2011.

M. Limam and F. Precioso. Atrial Fibrillation Detection and ECG Classification based on Convolutional Recurrent Neural Network. In *Proceedings of Computing in Cardiology 2017, Renness, France, September 24 - 27, 2017*, 2017.

J. Park, S. Lee, and M. Jeon. Atrial fibrillation detection by heart rate variability in Poincare plot. *Biomed Eng Online*, 8:38, 2009.

B. Pourbabaee, M. J. Roshtkhari, and K. Khorasani. Feature leaning with deep Convolutional Neural Networks for screening patients with paroxysmal atrial fibrillation. In *Proceedings of the International Joint Conference on Neural Networks*, 2016.

P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng. Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv preprint arXiv170701836*, 2017.

J. S. Richman and J. R. Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.*, 278:2039–2049, 2000.

J. Rodenas, M. García, R. Alcaraz, and J. J. Rieta. Wavelet Entropy Automatically Detects Episodes of Atrial Fibrillation from Single-Lead Electrocardiograms. *Entropy*, 17:6179–6199, 2015.

P. Schwab, G. C. Scebba, J. Zhang, M. Delai, and W. Karlen. Beat by Beat: Classifying Cardiac Arrhythmias with Recurrent Neural Networks. In *Proceedings of Computing in Cardiology 2017, Renness, France, September 24 - 27, 2017*, 2017.

Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron C. Courville, Ruslan Salakhutdinov, Richard S. Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visualattention. *CoRR*, abs/1502.03044, 2015.

Sergey Zagoruyko and Nikos Komodakis. Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer. *CoRR*, abs/1612.03928, 2016.

Zizhao Zhang, Yuanpu Xie, Fuyong Xing, Mason McGough, and Lin Yang. Mdnet: A semantically and visually interpretable medical image diagnosis network. *CoRR*, abs/1707.02485, 2017.

B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. Learning Deep Features for Discriminative Localization. *CVPR*, 2016.

X. Zhou, H. Ding, B. Ung, E. Pickwell-MacPherson, and Y. T. Zhang. Automatic online detection of atrial fibrillation based on symbolic dynamics and Shannon entropy. *Biomed Eng Online*, 13:18, 2010.

Appendix A.

The Python scripts for this model can be downloaded from https://github.com/Seb-Good/deep_ecg.