Robust ECG Signal Classification for Detection of Atrial Fibrillation Using a Novel Neural Network

Zhaohan Xiong[1], Martin K. Stiles[2,3], Jichao Zhao[1]

[1]Auckland Bioengineering Institute, The University of Auckland, Auckland, New Zealand
[2]School of Medicine, The University of Auckland, Auckland, New Zealand
[3]Waikato Hospital, Hamilton, New Zealand

Abstract

Electrocardiograms (ECG) provide a non-invasive approach for clinical diagnosis in patients with cardiac problems, particularly atrial fibrillation (AF). Robust, automatic AF detection in clinics remains challenging. Deep learning has emerged as an effective tool for handling complex data analysis with minimal pre- and post- processing.

A 16-layer 1D Convolutional Neural Network (CNN) was designed to classify the ECGs including AF. One of the key advances of the proposed CNN was that skip connections were employed to enhance the rate of information transfer throughout the network by connecting layers earlier in the network with layers later in the network. Skip connections led to a significant increase in the feature learning capabilities of the CNN as well as speeding up the training time. For comparisons, we also have implemented recurrent neural networks (RNN) and spectrogram learning.

The CNN was trained on 8,528 ECGs and tested on 3,685 ECGs ranging from 9 to 60 seconds in length. The proposed 16-layer CNN outperformed RNNs and spectrogram learning. The training of the CNN took 2 hours on a Titan X Pascal GPU (NVidia) with 3840 cores. The testing accuracy for the CNN was 82% and the runtime was ~0.01 seconds for each signal classification. Particularly, the proposed CNN identified normal rhythm, AF and other rhythms with an accuracy of 90%, 82% and 75% respectively.

We have demonstrated a novel CNN with skip connections to perform efficient, automatic ECG signal classification that could potentially aid robust patient diagnosis in real time.

1. Introduction

Cardiovascular disease is one of the leading causes of death worldwide. Atrial fibrillation (AF) is the most commonly sustained arrhythmia and is associated with substantial morbidity and mortality [1-3]. Early AF detection may enhance the effectiveness of clinical treatment and prevention of its serious complications [4]. Electrocardiograms (ECG), used to obtain a record of a patient's heartbeat first by Muirhead in 1872 via wires attached to the patient's wrist, is a widely used non-invasive approach for clinical diagnosis in patients with AF, as well as other types of cardiac arrhythmia.

Due to the important role of ECG diagnosis, monitoring and interpretation in clinics, numerous automatic approaches were proposed in the past[5]. However, all these approaches for AF diagnosis from ECG recordings are not very effective.

Deep learning, in recent years, has emerged as an effective tool for data analysis. The use of artificial neural networks in deep learning have been responsible for drastically improving sequential data processing tasks such as speech recognition, language translation and text to speech software, due to the powerful feature learning abilities of neural networks for understanding complex datasets. Most state-of-the-art neural networks perform predictions from raw data inputs, taking efficiency in data analyses to a higher level and bypassing the need of expert knowledge. The purely data-driven nature of these algorithms allows their performance accuracy to increase accordingly with increasing amounts of data.
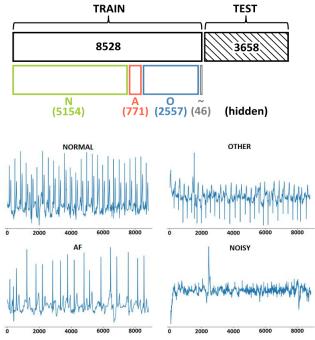
In this study, we have proposed a purely data-driven, deep learning pipeline, a 16-layer deep convolutional neural network (CNN), for the automatic classification of ECG signals from the Computing in Cardiology (CinC) Challenge 2017 into 4 distinct categories including AF. The large dataset of ECG data recorded from patients and associated labels provided by experts will provide an ideal framework for developing and validating an efficient approach for ECG diagnosis [5].

2. Methods

2.1. ECG Data

The ECG classification challenge was a sequential classification task where a single label was required for each individual input signal. The training dataset for the competition consisted of 8,528 single lead ECG recording ranging from 9 to 60 seconds in length with a sampling rate of 300 Hz [5]. Manual classification by a team of experts for the ECG data contained four classes: normal rhythm (N), AF rhythm (A), other rhythm (O) and noisy recordings (~). The online hidden testing set consisted of 3,658 samples similar to the training set (Figure 1). A

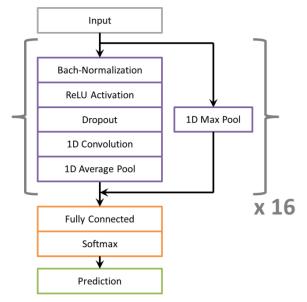typical waveform for each class is shown in the lower panel of Figure 1.

Figure 1: Top: The distribution of training and test set samples. **Bottom**: Typical recordings for each of the 4 classes in the data set.

2.2. Deep Learning Approaches

In this study, we have developed and compared three machine learning approaches: recurrent neural networks (RNN), spectrogram learning, and 16-layer CNN for the ECG classification task.

2.2.1 Recurrent Neural Networks

A 3 layer RNN was designed to extract temporal features from the raw waveform [6]. The RNN processed the input signal values sequentially with a feedback mechanism for any particular value by gaining the information learnt from the prior input values. This then propagated to the next temporal location where the same operations occurred. This allowed the raw ECG signal vector to be fed in one value at a time for the entire signal.

2.2.2 Spectrogram Learning

Spectrograms provided a representation of ECGs in the frequency domain using the Fourier transform when temporal domain representations alone were not enough. The converted 2D frequency-time matrix would then be processed like an image. The GoogleNet, winner of the 2014 ImageNet Large Scale Visual Recognition Challenge, was then used to process the resultant spectrograms [7].

2.2.3 A 16-layer convolutional neural network

Figure 2: The proposed neural network architecture with repeated 16 1D convolutions with skip connections.

A 16-layer CNN was developed for the ECG classification task (**Figure 2**). The design consisted of 16 sequential skip connections [8], as a means of increasing the efficiency of traditional CNNs.

In each block, the same operations were performed. During training, the data was fed into the CNN network in batches. To normalise the batch at each layer, batch-normalization [9] was performed to ensure the numerical values throughout the network were scaled to the same magnitude. Rectified linear activation (ReLU) units[10] were applied to speed up training by further normalising the values. Dropout[11] was then used to reduce overfitting of the CNN on the training data before the convolution layer.

The convolution layer was the major feature learning component of the CNN and involved a 15x1 filter with trainable weights that slid across the signals to extract features from the waveform. When significant features were detected, the filters activated by changings its weights to a more significant numerical value. By providing the CNN with labelled data, the model was able to learn the significant features that characterise different classes.

Pooling layers were added to down sample the signal by taking every two values in a vector and reducing it to 1 value by either averaging the two (average pool) or taking the max of the two (max pool). This forced the CNN to keep only the most relevant features and also decreased the memory burden of such a large network. Pooling

layers were also used on the skip connections to maintain dimensional consistency when the two separate paths joined back together at each block.

To produce a prediction, a fully connected layer was used which transformed the outputs from the convolution and pooling layers to a 4x1 vector of numerical values corresponding to the outputs for each class (N, A, O or ~). A softmax function was used to represent these values as a probability by normalising them between 0 and 1.

The network took 5 seconds long segments as input, and produced a prediction for each segment. The model output was a probability for each class, and the predicted class was the one with the highest probability. The overall classification for an entire ECG signal was the mean of the individual segment-wise predictions. The training set was randomly split into 80% training and 20% validation for 5-fold cross validation. The adaptive momentum estimation (Adam) optimizer, with a learning rate of 0.001, was used to optimize the network parameters.
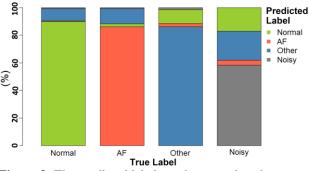
3. Results

	N	A	O	Overall
RNN	0.80	0.72	0.64	0.72
Spectrogram Learning	0.87	0.80	0.68	0.78
Proposed CNN	0.90	0.82	0.75	0.82

Table 1: The individual and overall F1 scores comparisons for the three machine learning methods.
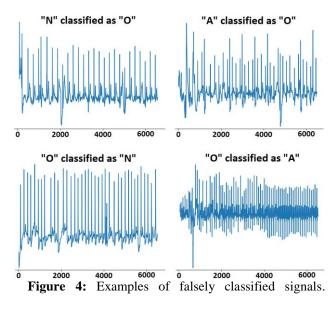
The performance in each class and the overall F1 score of the three investigated models on the test set are shown in **Table 1**. The proposed CNN approach outperformed the two traditional techniques by a large margin.

Figure 3: The predicted labels vs the ground truth.

Further analysis was performed on a randomly generated validation set from the training set and presented as a bar plot (**Figure 3**). The "N" subgroup had the highest classification accuracy (0.90). The "A" subgroup classification was the second best (0.86), followed by the "O" subgroup (0.85) and lastly the "~" subgroup (0.60). As seen from the figure, the CNN model had difficulties distinguishing normal sinus rhythm from other rhythm, as 8.8% of "N" signals were mistaken for

"O" and 10% of "O" signals were mistaken for "N". Many "A" were also classified as "O" (9%). Examples of signals from the most frequent false classifications are shown in **Figure 4**.

Figure 4: Examples of falsely classified signals.

4. Discussion

4.1 Challenge Dataset

The challenges of this competition data came from 3 main factors: relatively small overall dataset, major class imbalance across the 4 different classes, and varying signal lengths.

Firstly, despite the dataset being high quality in terms of the individual ECG recordings, overall, the dataset is still relatively small by today's standards in the modern world of deep learning. In the study by the Ng's group [12], 64,121 recordings with almost 2 million labels were used to develop a similar approach to classify ECGs. In contrast, the dataset used in this challenge had much less data, especially for the "A" and "O" subgroups.

Secondly, the uneven distribution of the different ECG classes introduced additional bias towards subgroups with more training data, creating more difficulties when classifying groups such as "A" where the number of training examples was small compared to "N". This made feature learning especially difficult for the "A" subgroup due to the limited examples provided. Furthermore, it was not entirely clear what other rhythms were in the "O". This provided further inconsistencies as the "O" class could possibly contain significantly different features within the same subgroup which would be difficult to group together by our CNN. This would also explain our observation where many "A" were classified as "O".

Lastly, the variation in the length of ECGs introduced further problems when developing a model. This was

especially problematic as in some scenarios, when an ECG recording, which was mostly "N", had a small episode of "A" or "O", it was very difficult to provide the correct classification based on a small signal segment. This resulted in many misclassifications between "O" and "N".

While the proposed CNN approach ignored the data inconsistencies, data augmentation could be investigated to ease the class imbalance as well as to increase the overall dataset size. Adaptive weighted averaging could be employed in future development to identify impactful segments in the ECGs for a more robust overall predict.

4.2 Performance of the Deep Learning Models

For the three methods investigated, the proposed CNN outperformed the others by a clear margin.

Although the RNN approach was also purely data driven, it was significantly limited by its learning capacity, while the capacity of the 16-layer CNN was further amplified by the skip connections introduced. A further limitation of RNN was their large computational costs. Due to the feedback mechanisms present, the number of computations increase exponentially with every increase in the length of the input as during each value input, the feedback loop had to be performed on all the previous input values.

The spectrogram method was fairly successful by adapting the state-of-the-art GoogleNet to the problem, however, transforming the raw wave form into the frequency domain was not as effective for learning. Furthermore, using an image classification CNN significantly constrained the input format into the network in terms of tuning the input patch size for optimal learning. Since small spectrograms did not provide enough information and large spectrograms contained sparse information which were both difficult to learn from. The proposed CNN, however, processed the raw waveform directly, and resulted in greater feature learning capabilities and a faster training process by the utilized skip connections between layers earlier in the network and layers later in the network. In comparison, traditional CNNs only contain layers one after the other, from input to output.

5. Conclusions

We have demonstrated a novel CNN to perform robust and efficient automatic ECG signal classification, particularly for AF, which could potentially be used to aid self-diagnosis for patients in a portable device.

Acknowledgements

References

[1] S. M. Narayan, M. Rodrigo, C. A. Kowalewski, F. Shenasa, G. L. Meckler, M. N. Vishwanathan, T. Baykaner, J. A. Zaman, and P. J. Wang, "Ablation of Focal Impulses and Rotational Sources: What Can Be Learned from Differing Procedural Outcomes?," *Current Cardiovascular Risk Reports,* vol. 11, no. 9, pp. 27, 2017.

[2] B. J. Hansen, J. Zhao, T. A. Csepe, B. T. Moore, N. Li, L. A. Jayne, A. Kalyanasundaram, P. Lim, A. Bratasz, K. A. Powell, O. P. Simonetti, R. S. Higgins, A. Kilic, P. J. Mohler, P. M. Janssen, R. Weiss, J. D. Hummel, and V. V. Fedorov, "Atrial fibrillation driven by micro-anatomic intramural re-entry revealed by simultaneous sub-epicardial and sub-endocardial optical mapping in explanted human hearts," *European heart journal,* vol. 36, no. 35, pp. 2390-2401, 2015.

[3] J. Zhao, B. J. Hansen, Y. Wang, T. A. Csepe, L. V. Sul, A. Tang, Y. Yuan, N. Li, A. Bratasz, K. A. Powell, A. Kilic, P. J. Mohler, P. M. Janssen, R. Weiss, O. P. Simonetti, J. D. Hummel, and V. V. Fedorov, "Three-dimensional Integrated Functional, Structural, and Computational Mapping to Define the Structural "Fingerprints" of Heart-Specific Atrial Fibrillation Drivers in Human Heart Ex Vivo," *Journal of the American Heart Association,* vol. 6, no. 8, pp. e005922, 2017.

[4] D. H. Lau, D. Linz, U. Schotten, R. Mahajan, P. Sanders, and J. M. Kalman, "Pathophysiology of Paroxysmal and Persistent Atrial Fibrillation: Rotors, Foci and Fibrosis," *Heart, Lung and Circulation*, 2017.

[5] C. L. Gari Clifford, Benjamin Moody, Ikaro Silva, Qiao Li, Alistair Johnson, Roger Mark, "AF Classification from a Short Single Lead ECG Recording: the PhysioNet Computing in Cardiology Challenge 2017," *Computing in Cardiology (Rennes: IEEE),* vol. 44, 2017.

[6] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks." *2013 IEEE Interational Conference.* pp. 6645-6649.

[7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions." *2013 IEEE Conference*. pp. 1-9.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition." *2015 IEEE Conference.* pp. 770-778.

[9] S. Ioffe, and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift." *2013 IEEE Conferenc* pp. 448-456.

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks." *Computer Vision ECCV 2016*. pp. 630 645.

[11] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of machine learning research,* vol. 15, no. 1, pp. 1929-1958, 2014.

[12] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks," *arXiv preprint arXiv:1707.01836*, 2017.