Automatic Concurrent Arrhythmia Classification Using Deep Residual Neural Networks

Deepankar Nankani, Pallabi Saikia, and Rashmi Dutta Baruah

Department of Computer Science and Engineering

Indian Institute of Technology Guwahati, Assam, India-781039

Abstract

This paper addresses the PhysioNet/Computing in Cardiology Challenge 2020. The challenge presents a problem to classify 26 types of arrhythmias and normal sinus rhythm using 12-lead electrocardiogram data. We were able to successfully perform the classification task using an eight layer deep residual neural network (ResNet). The skip connections present in the ResNet allowed the model to train faster and produce better challenge score. We also investigated sixteen other models that included convolution and recurrent neural network based models along with interpretability based attention mechanism as all of them are well suited for time series classification problems. The results depicted that the 8 layer ResNet model outperformed other models in terms of challenge score consuming significantly less time during the training phase. We preferred batch wise training to avoid having all the data in memory during training thereby alleviating the problem of memory choking. Our team, deepzx987, obtained a challenge score of 0.305 on validation data, −0.035 on the full test set, and ranked 35th in this year's challenge.

1. Introduction

Early diagnosis of concurrent cardiac arrhythmias using ECG signal helps in a timely treatment to reduce the mortality rate. The 12-lead ECG is a standard method that is representative of the heart's electrical activity. It assists cardiologists in screening and diagnosing cardiac abnormalities [1]. Due to the scarcity of expert cardiologists, the huge volume of data generated needs to be analysed automatically to assist these cardiologists. The PhysioNet/Computing in Cardiology Challenge (CinC) 2020 focuses on automatic approaches for classifying cardiac abnormalities using the standard 12-lead ECGs [2, 3].

In the past, the deep learning models have achieved remarkable performance in the domain of computer vision and natural language processing [4,5]. Deep learning models have achieved better performance for big and varied datasets than conventional machine learning models ac-companied with feature extraction methods for the task of ECG classification [6]. However, their usefulness in realistic clinical settings, using a standard 12-lead ECG still remains an open ended research problem. We explore this problem by the means of the PhysioNet/CinC challenge 2020 by investigating seventeen different models including convolution and recurrent neural network models along with an interpretability based attention mechanism. Our best approach for this challenge employs a segmentation method followed by an 8 layer residual neural network.

2. Methods

The workflow we followed can be categorized in three major parts: (i) data selection, (ii) segmentation and labelling, and (iii) record classification. The workflow is described in figure 1. All the six available databases consisting of standard 12-lead ECG provided in the challenge were used for experimentation. The details regarding the sampling rate, ADC gain, gender of patient, and diagnosis are present in the database [3]. The ECG signal length in the dataset vary from 2500 to 462600 samples. Since homogenization of signal length is important before feeding it to a neural network, segments of 2500 samples are extracted from each record. The segment length of 2500 samples was selected to utilise all the records from dataset as the length of the shortest segment available in the dataset was 2500 samples. The segments are then labelled according to the diagnosis present in each record. These segments are provided as input to the investigated models. The models classify the segments to one of the 26 types of arrhythmia or normal sinus rhythm. Although, the challenge specified three diagnosis pairs that are identically scored, namely, SVPB and PAC, PVC and VPB, CRBBB and RBBB. We did not merge these classes and approached the challenge with 27 classes.

We investigated five categories of models that include convolution neural networks (CNN), recurrent neural networks (RNN), combination of CNN and RNN, RNN with interpretability based attention mechanism [7], and lastly the combination of CNN, RNN, and Attention Mechanism. Figure 2 illustrates the schematic of different investigated
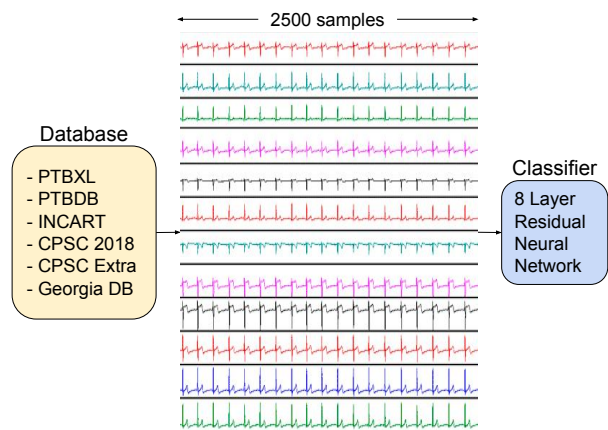
Figure 1. Illustration of the workflow. Segments are extracted from the 12-lead ECG and classified using ResNet.

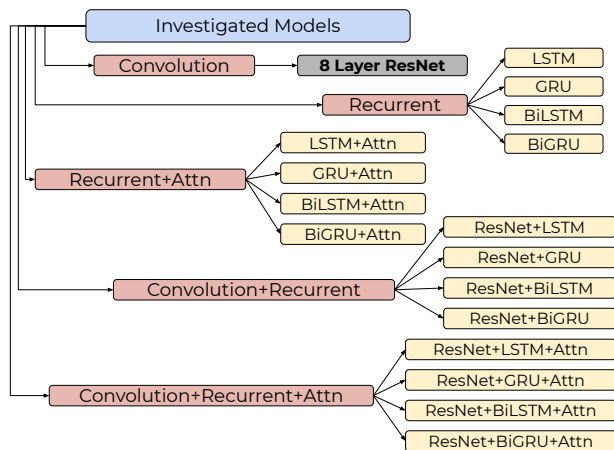models. The details of each category of models is provided in subsequent sections.

Figure 2. A schematic of different investigated models.

2.1. Convolution Based ResNet

The first category employs the convolution based Residual neural network (ResNet) [8]. The convolution filters present in the CNN makes the need of peak detection, feature extraction, ranking, and selection insignificant as the model recognises patterns of each rhythm during training as uses the learned representation during inference. The architecture of the proposed 8-layer ResNet is illustrated in figure 3. The model consists of a variety of layers including one dimensional convolution layer (1-D Conv), batch normalization (BN), dropout, max pooling, and fully connected layer or dense layer. Regularization of the model is attained using batch normalization and dropout to avoid-

ing overfitting on the training data. Batch normalization [9] normalizes the inputs of preceding layer. Dropout [10] drops random convolution filters by temporarily removing their contribution during training with dropout probability p and vice versa during the testing phase. The 1-D Conv accounts for temporal relations present in samples of ECG signal. The skip connections skip over the non-contributing layers, allowing the gradient to backpropagate to the initial layers, mitigating the problem of vanishing and exploding gradient. This allows the model to train faster with better accuracy. The non-linearity is introduced using Rectified Linear Unit (ReLU) activation in the model. The dense layer with 27 neurons is employed to provide more abstraction to the model representation. The softmax activation function assigns a probability to the final model prediction. The convolution filter weights were initialised using Xavier initialization [11]. 64 filters were used with the size of 16×1 and a stride of 1. Stride of Max Pooling is kept at 2 and dropout probability of 0.5.
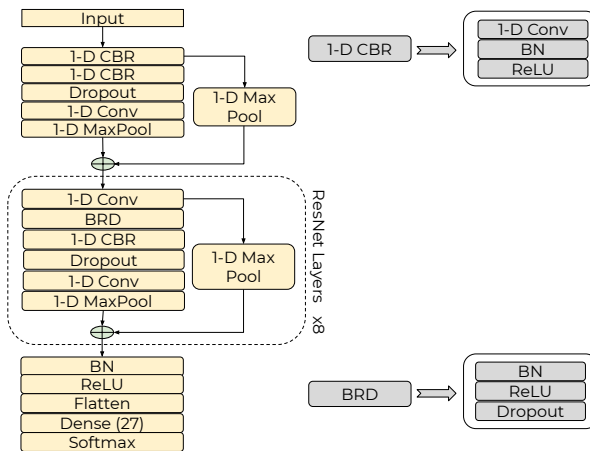

Figure 3. Proposed architecture of the 8-layer ResNet.

2.2. Recurrent Based Models

Recurrent Based Models consists a chain of repeating modules of neural network that learns about the 12-lead ECG signal and stores information in its repeating cells. Similar to the CNN, they also ingest a constant sized input. Inherently, RNNs work over sequences that makes it suitable for our application. However, RNNs struggle to remember information for longer period of time which led us to use long short term memory (LSTM), special class of RNN, designed to remember information for long periods of time without any struggle [12]. As arrhythmias occur arbitrarily in long term ECG, making long term information retention important, leading to the use of LSTM for this challenge. We also used gated recurrent unit (GRU), special class of RNN, that consists of update, reset, and

current memory gates and does not maintain any internal cell state and passes the information to next GRU [13]. Bidirectional LSTM and GRU were also employed as they focus on past and future samples during predictions. The models consists of 3 layers with 50 cells in each layer.

2.3. RNN with Attention Mechanism

The aspect of interpretability is also explored to explain the predictions of our models using attention mechanism along with the recurrent based networks [7]. It is better than conventional recurrent networks because in attention mechanism the entire input is passed along the network without creating any bottleneck situation, alleviating the learning problem of neural network. This allows recurrent networks to focus at specific morphologies present in ECG at different timestamps. Attention layer soft searches for parts of the signal that are relevant to predicting a specific class of arrhythmia. The model parameters for RNN based models were 2 layers with 50 cells in each layer and one attention layer with 10 cells.

2.4. Combination of Models

The combination of RNN variants along with a ResNet model was also performed to provide the models more flexibility to learn the data representation. The models used a 4-layer ResNet with 2 layers of RNN variants with 50 cells in each layer. The final category of models includes a combination of ResNet followed by RNN variants followed by an attention decoder. The model parameters were 5 layer ResNet followed by 2 layers of RNN with 50 cells each and one attention layer with 10 cells. All the models were trained locally using an Intel Xeon processor with 32 GB RAM and an NVIDIA GeForce Titan Xp graphics card with 12 GB GDDR6 VRAM. To avoid loading whole data in memory at once, we optimized the training procedure by employing the generator method available in keras library. Batch wise training was employed with a batch size of 16. Only the first label from each diagnosis was considered and similarly a single class label was predicted for each record in the test set. Only 27 classes were considered instead of 111 classes which lead to leaving out around 4000 records during model training. During the testing part we averaged the prediction of all segments of the record and choose that class of diagnosis that model predicted most number of times. We employed early stopping to avoid overfitting of the model. During the training, the model monitored the validation accuracy for 5 epochs and if the validation accuracy did not improve, the training was halted. The main data was split into 80% training and 20% validation data. Hyperparameters such as number of convolution & recurrent layers, size of kernels in CNNs and cells in RNNs were adjusted using the validation data.

3. Results

The challenge score [3] for all the models on 20% validation data extracted from the training data are described in figure 4. Resnet with LSTM and GRU with attention decoder produced scores of -0.128 and -0.418. Figure 5 describes the training time taken per epoch by each model on our system. Figure 7 describes the results over stratified 10 fold cross validation (CV) on the training data by the best performing 8 layer ResNet model. Figure 6 describes the convergence curves of accuracy and loss of ResNet model during training. At 29th epoch, the model training was halted as the validation accuracy did not improve after 24th epoch. Table 1 describes the final results for different evaluation metrics on validation and test data.
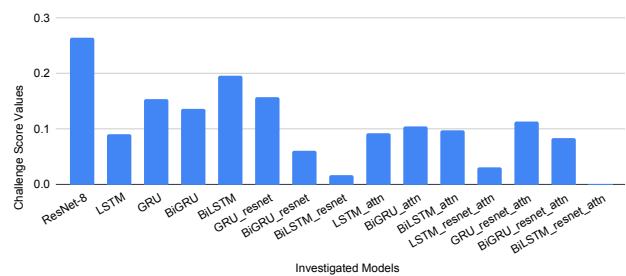

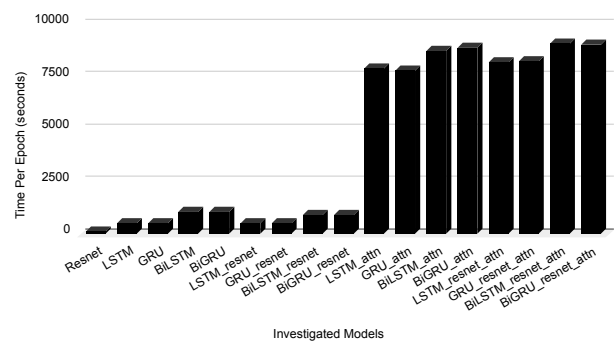Figure 4. Challenge Scores on train data for all models.

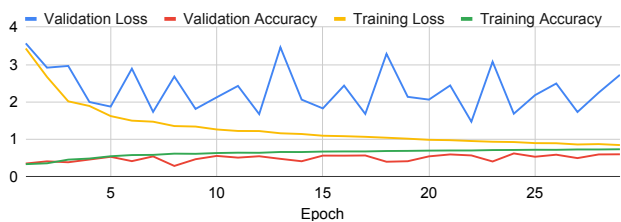Figure 5. Time to train per epoch for all the models.

Figure 6. ResNet convergence plots during the training.

Table 1. Classification results for the 8 layer ResNet model on validation and test data.

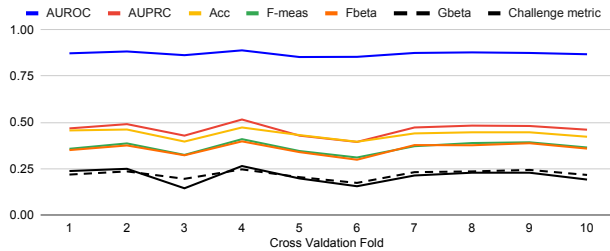| Metrics / Dataset	Validation Data	Test Data 1	Test Data 2	Test Data 3	Full Test set
AUROC	0.825	0.919	0.812	0.694	**0.742**
AUPRC	0.326	0.711	0.318	0.239	**0.229**
Accuracy	0.331	0.527	0.277	0.08	**0.181**
F-measure	0.286	0.228	0.259	0.118	**0.182**
Challenge Score	0.305	0.648	0.25	-0.287	**-0.035**

Figure 7. Stratified 10 fold CV results for ResNet model.

4. Discussion

The 8 layer ResNet outperformed other models in terms of challenge score and time complexity. Convolution layers in ResNet, LSTM, GRU layers, and cells in each layer along with the cells in attention layer were finalised empirically. The attention based models did not provide better results even after having a larger context vector. Training time per epoch of ResNet was around 110 sec. as compared to 9070 sec. by attention based models. The reason being the non optimised implementation for CUDA cores in graphics card. The main reason behind low performance of our models might be the selection of single label for each record. One interesting point we discovered was that even after providing 27 labels during the classification, our model predicted only 24 labels during inference.

5. Conclusions

Our team deepzx987 investigated five categories of models. The 8-layer ResNet outperformed all other models in terms of minimum execution time and challenge score. ResNet produced a score of 0.21 ± 0.04 during stratified 10 fold CV on training data, 0.305 on validation data, -0.035 on the full test set, and ranked 35th in this year's challenge. RNN and attention models took higher training time in addition to the poor performance as compared to ResNet.

References

[1] Kligfield P. The centennial of the Einthoven electrocardiogram. Journal of Electrocardiology 2002;35(4):123–129.

[2] 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.

[3] Perez Alday EA, Gu A, Shah A, Robichaux C, Wong AKI, Liu C, Liu F, Rad BA, Elola A, Seyedi S, Li Q, Sharma A, Clifford GD, Reyna MA. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. Physiological Measurement 2020; In Press.

[4] Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: A brief review. Computational Intelligence and Neuroscience 2018;2018.

[5] Young T, Hazarika D, Poria S, Cambria E. Recent trends in deep learning based natural language processing. IEEE Computational Intelligence Magazine 2018;13(3):55–75.

[6] Faust O, Hagiwara Y, Hong TJ, Lih OS, Acharya UR. Deep learning for healthcare applications based on physiological signals: A review. Computer Methods and Programs in Biomedicine 2018;161:1–13.

[7] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv14090473 2014;.

[8] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. corr abs/1512.03385 (2015), 2015.

[9] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv150203167 2015;.

[10] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research 2014;15(1):1929–1958.

[11] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. 2010; 249–256.

[12] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation 1997;9(8):1735–1780.

[13] Chung J, Gulcehre C, Cho K, Bengio Y. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv14123555 2014;.

Address for correspondence:

Deepankar Nankani
Department of Computer Science and Engineering
Indian Institute of Technology Guwahati, Assam, India
d.nankani@iitg.ac.in