Adaptive filtering combined with deep ensembles for better arrhythmia detection

Jani Bizjak, Matjaž Gams
Institute Jožef Stefan, Slovenia
name.surname@ijs.si

Abstract

Heart rate diseases are leading cause for death, accounting for 1 in 4 deaths in USA today. It is thus important to detect and treat heart problems as soon as possible, sadly it is not feasible for everyone to visit the doctor and check their heart regularly. To alleviate the problem a reliable automated solution for arrhythmia detection is required, to allow each individual to check their heart at home and only visit the doctor if there is a problem. In this work we extend and improve previously developed methods so they achieve better accuracy and are able to classify more arrhythmias. We test the new architecture on MIT-BIH database and comment on the results.

1 Architecture

In our previous work [Bizjak, 2015] we used convolutional neural networks (CNN) for QRS peak detection and to classify 3 types of arrhythmias: Left Bundle Branch Block (LBBB), Right Bundle Branch Block (RBBB) and Premature Ventricular Contraction (PVC). The reason we chose to detect these arrhythmias was their distinguished shape of the signal. While we could not detect certain arrhythmias, for example the ones that happen out of the rhythm (Premature Atrial Contraction (PAC)), for normal, LBBB, RBBB and PVC peaks we achieved high classification accuracy of over 99%.

In order to improve the results, a different architecture was required, one that can detect time dependencies in the data.

1.1 Adaptive filtering for QRS detection

In [Dohare *et al.*, 2014] authors created filtering procedure that not only smooths the signal but also detect QRS peaks. When we tried to apply it to our dataset it did not perform at desired accuracy, mainly because of the noisy data. To solve this problem we filtered the signal with low pass filter ($\alpha = 0.2$) in order to reduce the high frequency noise. We then centred the signal by subtracting median from the signal, with this we removed signal drift. The median was calculated with sliding window containing 300 samples (approximately 1 second).

In the next step we calculated derivative of the signal, using sliding window of 5. This produced distinguishing artefacts where R peaks were located due to the fast changes in amplitude, while the rest of the signal approached 0. In order to easily find the peaks we detected an envelope around extremes as seen in Figure 1. The actual location of the R peak was then determined by checking when the signal breaches 80% average. Analogously we were able to detect deep valleys that are characteristic for PVC.

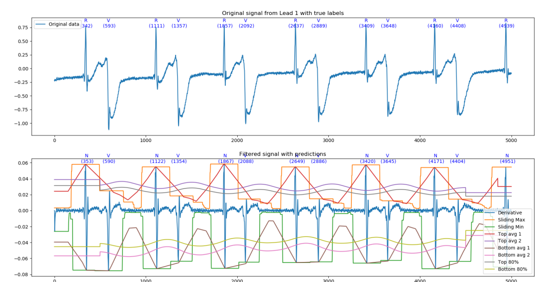

Figure 1: Different stages of filtering in order to detect QRS peaks.

1.2 Deep ensembles for arrhythmia classification

The reliable QRS peak detection is the first step in ECG analysis. While the methods described work well for detection of simple patterns, like QRS, they are unable to distinguish malformed shapes of the signal that are present with certain heart problems. In order to detect them, we combined multiple deep architectures, together with statistical model to capture time dependencies and improve overall accuracy.

While CNN is good at recognising visual clues it performs poorly when data is time dependent. We tried to solve this problem by introducing LSTM layers, we experimented with different combinations of layers (LSTM, Convolutional, Dense) and different sizes but all performed worse than a simple statistical model, that analyses frequency of peaks. The model checks whether the time between peaks is shorter than $(average - 2.2 * STD)$ (where 2.2 was determined experimentally). This approach produced good accuracy for both PAC and PVC detection but could not distinguish between the two.

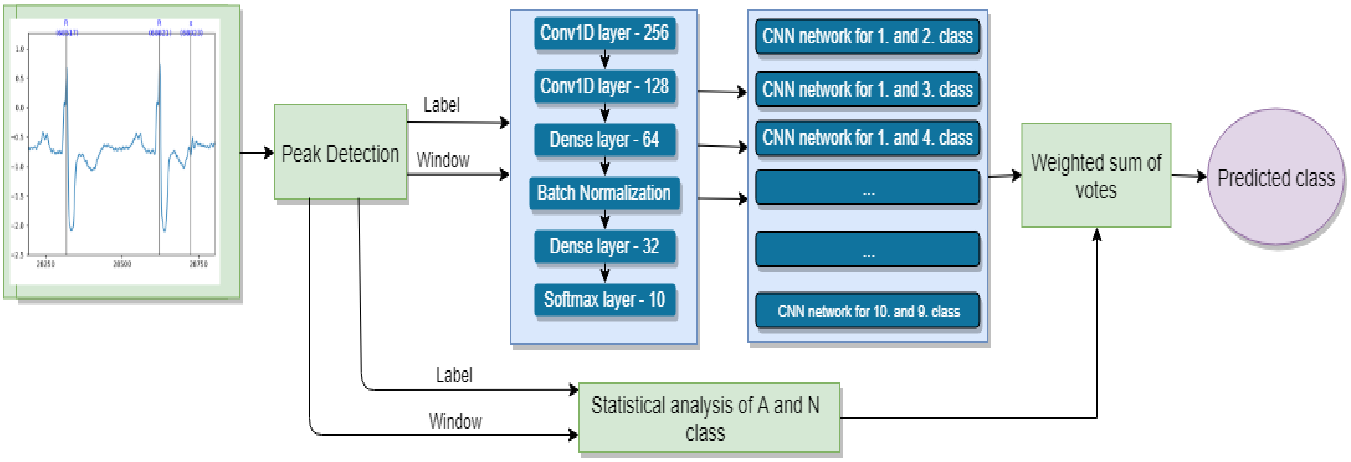

Figure 2: Architecture combines multiple models and performs weighted voting to produce prediction.

For QRS peak classification multiple architectures were tested. The combination of convolutional and fully connected layers proved to be the best. The architecture is presented in Figure 2.

The architecture performed well for most of the classes, but had problems distinguishing between certain pairs or triplets of classes. To solve this problem we added 45 smaller networks, consisting of convolution and two fully connected layers. If the larger network predicted a class with less than 80% certainty the ensemble of smaller networks were used, each classifying between 2 arrhythmias and others classes combined into one.

To combine the results of multiple models we used weighted voting.

2 Experiments

To test our model we used MIT-BIH [Goldberger *et al.*, 2000] dataset. The dataset consists of 48 recordings, each containing 2 leads that were marked by at least 3 medical professionals. In total there are over 11,000 labels, spread among 23 classes. Some of the labels denote noisy signal or other metadata. The classes are not balanced, for example some have less than 100 examples. In the end, we decided to use 10 most frequent classes. Each class we chose is present in at least 3 recordings and has more than 200 samples (except flutter (!)). In addition we combined classes for $+$ and $\sim$ into a single class representing noisy data.

First we tested how well the filtering stage performs for QRS peak detection. It is important that this step is as accurate as possible, since the rest of the model depends on accurately detected peaks for further analysis. Our model performed well, achieving combined accuracy of **99.99%**. For most of the recordings the model scored perfect, without misclassification, while on few the accuracy was lower with over 100 misclassified peaks. This happened on the recordings with extremely noisy data and labelling inconsistencies, the recordings where this is most prominent are: 203, 208, 210, 221, 223.

Secondly we tested how well the model performs at classification of the following classes: */, F, !, A, N, V, L, $\sim$, +, R*. Because classes are extremely imbalanced we randomly selected 500 representative QRS peaks for each class across all recordings, the rest were used for testing but no more than 500. While this stage of the testing was done separately of the first, if the first stage failed to identified QRS peak that was being tested in the second stage, we counted it as a miss, even if the second stage correctly classified it.

We have repeated the process several times to remove random factor from initial peak selection and achieved **91.05%** accuracy (*Note: Since the classes are balanced the majority class is 20% (combination of + and $\sim$)*. The ensembles architecture improved overall accuracy for over 6% compared to CNN architecture. For each individual class the average F1 score is: *+=96.4%, N=71.7%, A=83.2%, V=84.7%, $\sim$=98.5%, f=90.3%, F=88.1%, L=93.5%, R=94.5%, !=76.8* .

3 Conclusion

We presented a combination of different approaches and models for arrhythmia detection. The developed model was tested on a large dataset, where it not only achieved better accuracy, compared to the CNN model from our previous work, but was also able to classify higher number of arrhythmias.

References

[Bizjak, 2015] Jani Bizjak. Ecg signal analysis using deep neural networks. Master's thesis, Faculty of Mathematics and Physics, University of Ljubljana, Slovenia, 2015.

[Dohare *et al.*, 2014] Ashok Kumar Dohare, Vinod Kumar, and Ritesh Kumar. An efficient new method for the detection of qrs in electrocardiogram. *Computers & Electrical Engineering*, 40(5):1717–1730, 2014.

[Goldberger *et al.*, 2000] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet. *Circulation*, 101(23):e215–e220, 2000.