

Multi-label Classification on 12, 6, 4, 3 and 2 Lead Electrocardiography Signals using Convolutional Recurrent Neural Networks

Niels Osnabrugge, Kata Keresztesi, Felix Rustemeyer, Christos Kaparakis, Francesca Battipaglia, Pietro Bonizzi, Joël Karel

Department of Data Science and Knowledge Engineering, Maastricht University, The Netherlands

Abstract

Automatic identification of cardiac abnormalities through the ECG with a reduced lead system (less than the standard 12-lead) can provide a valuable easy to use and lower cost diagnostic alternative to ordinary 12-lead ECG devices. This study investigates the use of Convolutional Recurrent Neural Networks (CRNNs) to identify cardiac abnormalities in 12, 6, 4, 3 and 2 lead ECG data. Multi-label classification with CRNNs relies on effective data pre-processing, model architecture and hyperparameter tuning. ECG signals were first pre-processed and then zero-padded or clipped to have an equal duration of 10 seconds). Additionally, a wavelet-based ECG segmentation algorithm was used to extract the characteristics and locations of the PQRST complexes (features), and both PQRST fiducial points and extracted features were used as inputs to two Convolutional Recurrent Neural Networks (CRNN), respectively, each one consisting of eight layers. The two CRNNs were subsequently concatenated. Preliminary results of the proposed method achieved an official score of 0.03 (team name: heartMAASters) and a cross-validation score of 0.46. The results of the enhanced model are notably worse due to an unknown failure. In the discussion we provide some theoretical considerations on why we would expect the proposed model to show a better performance.

1. Introduction

Wearable ECG devices are becoming increasingly relevant as a research and clinical tool for the identification of arrhythmia and for continuous monitoring of patients. In this respect, reduced lead ECG systems are a promising opportunity to diagnose cardiac abnormalities while being more convenient. Some cost-effective and wearable ECG devices are already available on the market, however, there is limited evidence about using reduced-lead ECGs to diagnose a wide range of abnormalities [1]. Several studies focused on proposing a subset of the 12-lead ECG

to identify cardiac pathologies [2–4], while other studies proposed online classification algorithms of cardiac abnormalities, which can be easily integrated in wearable devices and require a limited number of leads. [5, 6].

Several studies concluded that deep learning offers high performing methods for pattern recognition in ECG signals, with either twelve or less leads [7]. Convolutional Neural Networks (CNNs) are suitable to extract hierarchical patterns as well as local features from ECG signals [8, 9]. Recurrent neural networks (RNNs), on the other hand, capture temporal dependencies, handle inputs of various lengths, and can also be used to summarize local features to generate global features. Bidirectional Long Short Term Memory Networks (LSTMs), a variant of RNNs, improve performance and interpretability through their attention mechanism [10, 11]. Convolutional Recurrent Neural Networks (CRNNs) can handle long ECG signals of different lengths and multi-channel inputs. CRNNs have also been applied with attention mechanisms combined with expert features for disease detection [12, 13]. The state-of-the-art methods for signal denoising are autoencoders (AEs), which simultaneously conduct reconstruction and classification procedures for signal compression[14], and Generative Adversarial Networks (GANs), which allow to create synthetic datasets. To resolve the common problem of class imbalance, data augmentation has been performed with AEs and GANs to generate synthetic training sets [15]. Finally, various feature engineering techniques have been applied in previous studies to create features that may help classify ECG signals [16].

Although many of these algorithms are continually emerging and improving, achieving accurate classification of cardiac diseases remains a challenge. The objective of this study is to identify a high performing algorithm that can classify cardiac abnormalities from either 12-lead, 6-lead, 3-lead or 2-lead ECG signals, with a combination of the above state of the art methods, preprocessing, data augmentation, and feature selection. In particular, in this study, we investigated the use of CRNNs to identify cardiac abnormalities in 12, 6, 4, 3, and 2-lead ECG data.

2. Method

2.1. Data Preprocessing

Baseline wander was removed by means of a bandpass type II Chebyshev filter, with cut-off frequencies [0.5,100] Hz. All recordings were resized to 10 seconds by truncating it or by zero padding. Then, recordings were resampled to 500 HZ, yielding 5000 data points each.

2.2. Model Architecture

First a benchmark CNN-based model was implemented, which later was extended to a CRNN-based model. Both models were implemented in Python using the Keras library, and trained on a Volta V100-SXM2 GPU. Since we are dealing with a multi-label classification problem, binary cross entropy was chosen as loss function. The Adam optimizer was used because it is one of the most commonly used optimization algorithms, and ReLU as activation function for all the layers except the last, where a sigmoid function was used. The sigmoid function is selected as the last layer because it maps all the outputs in the 0-1 range, while ReLU was used in the previous layers due to its faster derivative computation time. Training was conducted with 10 epochs and using batches of size 128, which was determined empirically.

The benchmark method consists of one CNN with time domain ECG signals as input. We selected a CNN as a benchmark model because of their ability to handle signals along with a low training time compared to other Neural Networks [7]. Moreover, CNNs act as a foundation of several complex Neural Networks, like ResNets or CRNNs. Throughout this paper, the convolutional layers used are 1D convolutional layers. The reason for this is that there is a logical ordering over time, while over the leads convolution is not logical. Furthermore, max pooling is used to avoid overfitting on the noise and to have better performance on new data. Moreover, the networks apply batch normalization in order to improve time gain, performance and dropout. A dropout rate of 0.3 resulted in a model that could generalize best, while a higher dropout resulted in the loss of too much information. Moreover, a learning rate of 0.01 was used in this setting to speed up the learning process, and a clipping value of 0.5 was implemented to avoid exploding gradients, given the large amount of parameters.

The proposed CRNN-model consists of two neural networks running in parallel, prior to concatenation, with fully connected layers and a final sigmoid layer. As illustrated in Figure 1, the model is composed of two CRNNs, one taking the ECG multi-lead recording as an input and another taking the PQRST fiducial points as input. The idea to have fiducial points as an input is that the time

points of ECG waves may be important for recognizing arrhythmia. The motivation for choosing CRNNs was that the Long Short-term Memory (LSTM) [17] layer they include, allows to capture time-dependencies between data points. The concatenation, the last three fully connected layers and the sigmoid layer, ensures that the captured information from the two CRNNs are combined into one output, a prediction of which classes a specific recording belongs to. The sigmoid layer creates a probability per class; the classes that exceed the cutoff ratio are selected as predicted classes. In this paper a cutoff ratio of 0.6 was empirically determined.

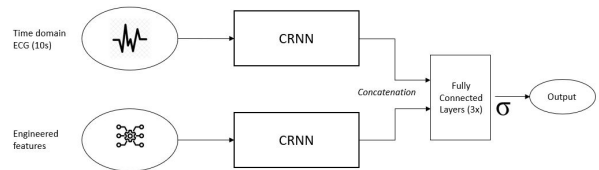


Figure 1. Scheme of the model

2.3. CRNN with ECG input

As introduced above, the first CRNN in the proposed model uses ECG signals as inputs, and it consists of an LSTM (20 hidden units), five convolutional layers with max pooling, and two dropout layers with a convolutional layer. At the end of the network there is a fully connected layer, as illustrated in Figure 2. In the first 4 convolutional blocks, the size of the convolutional kernels was set to 5×5 . The last two convolutional blocks, focusing on smaller segments of the ECG, included a convolutional kernel size of 4×4 .

The preprocessed ECG recordings are fed as input and have a size of 5000×12 (10 sec. sampled at 500 Hz, for each of the 12-leads).

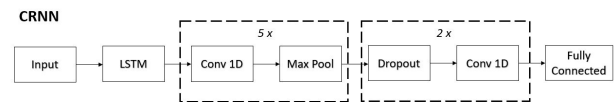


Figure 2. Scheme of the two CRNNs

2.4. CRNN with fiducial points input

The fiducial points CRNN was constructed to gain more specific information about the recordings. The QRS complex, P and T wave fiducial points were extracted by means of a wavelet-based ECG delineation algorithm using the

WTdelineator library in Python [18]. As illustrated in Figure 3, 13 fiducial points were extracted, including the location of onset and end of the P, Q, R, S, and T waves, the first and second peak of the P and T waves, and the peaks of the Q, R and S wave. Every feature is encoded to a specific number (for example: 1 for the onset of the P wave), and for every recording an array is built with the length of the recording (5000). When a data point at location i corresponds to one of the 13 fiducial points, the value of element i is set to the corresponding detected feature. Locations with no detected features have a value of zero.

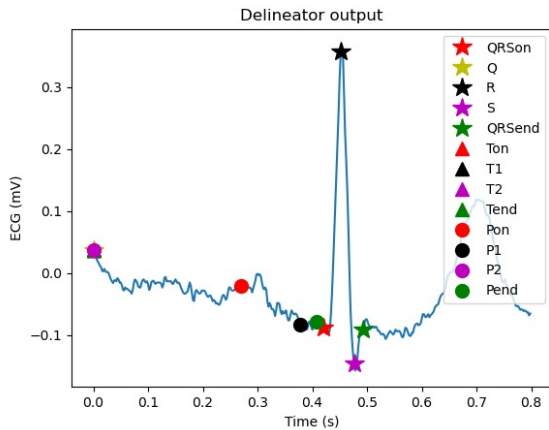


Figure 3. Fiducial points of the PQRST complex used as inputs to the second CRNN network.

The CRNN that uses the fiducial points as input consists of the same architecture of the CRNN with ECG signals as inputs, as illustrated in Figure 2.

3. Results

In order to evaluate the two models, we performed cross validation on the training set. The accuracy scores of the benchmark and proposed model, for 2, 3, 4, 6, and 12 leads are illustrated in Table 1.

Accuracy per model	2 leads	3 leads	4 leads	6 leads	12 leads
Benchmark model	0.30	0.42	0.42	0.30	0.46
Improved model	0.40	0.38	0.37	0.39	0.37

Table 1. Accuracy scores for the 2, 3, 4, 6 and 12 lead models of the benchmark and improved models.

As can be seen in Table 2, using the evaluation code provided by Physionet, several other metric scores were calculated, utilizing the scored labels only.

Scores	AUROC	AUPRC	Accuracy	F-measure	Challenge metric
Benchmark model	0.83	0.29	0.40	0.19	0.15
Improved model	0.50	0.06	0.06	0.00	-0.59

Table 2. Metric scores of the benchmark and improved models.

4. Discussion

4.1. Benchmark method performance

As showed in Table 2, the benchmark model predicts about 40% of the recordings correctly (40% accuracy). The $F1$ -measure, which represents the harmonic mean of precision and recall, is 0.19.

The area under the curve of the receiver characteristic operator (AUROC) for our multi-class classification problem was implemented in a way that, for each class, the score is generated as classifying the class against all other classes. The AUROC score obtained this way for all classes is 0.81. This metric provides an aggregate measure of performance across all possible classification thresholds and tells how much the model is capable of distinguishing between classes on a 0 to 1 scale.

The area under the precision-recall curve (AUPRC) score expresses how much the model is able to find all cardiac arrhythmia cases (which means high recall), without falsely marking negative examples (healthy patients) as positive for the arrhythmia (which would mean high precision). Since different classes have different fractions of positive examples, each class has a different baseline. The score of the model, 0.29, is the average of the AUPRC scores for each class.

4.2. Improved method performance

When looking at the performance of the benchmark and improved model (Table 1 and 2), it can be noted that the scores are significantly low. These low metric scores are likely caused by a failure in the code of the improved model; the origin of the failure is likely located in the CRNNs, because the benchmark model has notably better results.

5. Conclusion

Even though theoretically, the improved model is expected to show better performance through the use of engineered features, our results from the improved model are notably worse. This could be due to how the fiducial points are provided as input to the second CRNN. Currently, the fiducial points are stored in a single vector of length 5000, where an integer indicates if and what fiducial point is present at a specific location (corresponding to a specific discrete time index). Although the LSTM might

be able to handle such an input, the convolutional layers may not, since they smooth things out, and the resulting values may lose meaning for the convolutional part.

There are several ways by which performance of the current model could be improved. First, various experiments can be run to simplify the proposed architecture, as well as improve the training time. The proposed architecture is an ensemble of two CRNNs networks, which results in longer training times. This can be problematic when training time is limited. Secondly, as stated earlier, the current implementation of the engineered features decreases the $F1$, AUROC and AUPRC scores. This outcome is not aligned with our original hypothesis that providing more diverse information to the Neural Network should result in improving performance. However, the problem may be on how the information is provided to the model. Moreover, our model could be improved by adding an additional Neural Network with frequency domain information as input. This may provide auxiliary information that is not captured by the current CRNNs. Furthermore, hyper-parameter tuning can further enhance the model's performance.

References

- [1] PhysioNet. Will two do? varying dimensions in electrocardiography: The physionet/computing in cardiology challenge 2021, 2021. URL <https://physionetchallenges.org/2021/>.
- [2] Drew B, Pelter M, Brodnick D, Yadav A, Dempel D, Adams M. Comparison of a new reduced lead set ecg with the standard ecg for diagnosing cardiac arrhythmias and myocardial ischemia. *Journal of Electrocardiology* 2002;35:13–21.
- [3] Green M, Ohlsson M, Forberg JL, Björk J, Edenbrandt L, Ekelund U. Best leads in the standard electrocardiogram for the emergency detection of acute coronary syndrome. *Journal of Electrocardiology* 2007;40(3):251–256.
- [4] Aldrich HR, Hindman NB, Hinohara T, Jones MG, Boswick J, Lee KL, Bride W, Califf RM, Wagner GS. Identification of the optimal electrocardiographic leads for detecting acute epicardial injury in acute myocardial infarction. *The American journal of cardiology* 1987;59:230–23.
- [5] Alfaras M, Soriano MC, Ortín S. A fast machine learning model for ecg-based heartbeat classification and arrhythmia detection. *Frontiers in Physics* 2019;7:103.
- [6] Okada Y, Yoto TY, Suzuki T, Sakuragawa S, Suigiura T. Wearable ecg recorder with acceleration sensors for monitoring daily stress: Office work simulation study. 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society EMBC 2013;4718–4721.
- [7] Hong S, Zhou Y, Shang J, Xiao C, Sun J. Opportunities and challenges of deep learning methods for electrocardiogram data: A systematic review. *Computers in Biology and Medicine* 2020;122:103801.
- [8] Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, NG A. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nature Medicine* Jan. 2019; 25:65–69.
- [9] Weifang S, Nianyin Z, Yuchao H. Morphological arrhythmia automated diagnosis method using gray-level co-occurrence matrix enhanced convolutional neural network. *IEEE Access* 2019;7:67123–67129.
- [10] Saeed S, Mohammadhosein O, Matin H. Lstm-based ecg classification for continuous monitoring on personal wearable devices. *IEEE Journal of Biomedical and Health Informatics* 2020;24(2):515–523.
- [11] Li R, Zhang X, Dai H, Zhou B, Wang Z. Interpretability analysis of heartbeat classification based on heartbeat activity's global sequence features and bilstm-attention neural network. *IEEE Access* 2019;7:109870–109883.
- [12] Shashikumar SP, Shah AJ, Clifford GD, Nemati S. Detection of paroxysmal atrial fibrillation using attention-based bidirectional recurrent neural networks, 2018.
- [13] He R, Liu Y, Wang K, Zhao N, Yuan Y, Li Q, Zhang H. Automatic cardiac arrhythmia classification using combination of deep residual network and bidirectional lstm. *IEEE Access* 2019;7:102119–102135.
- [14] Xia Y, Gao HZLXZ, Zhang H, Liu H, Li S. An automatic cardiac arrhythmia classification system with wearable electrocardiogram. *IEEE Access* 2018;6:16529–16538.
- [15] Wang P, Hou B, Shao S, Yan R. Ecg arrhythmias detection using auxiliary classifier generative adversarial network and residual network. *IEEE Access* 2019;7:100910–100922.
- [16] Hong S, Wu M, Zhou Y, Wang Q, Shang J, Li H, Xie J. Encase: An ensemble classifier for ecg classification using expert features and deep neural networks. 2017 *Computing in Cardiology CinC* 2017;1–4.
- [17] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation* 1997;9(8):1735–1780.
- [18] Ledezma CA. Wtdelineator, 2021. URL <https://github.com/caledezma/WTdelineator>.

Address for correspondence:

Niels Osnabrugge
P.O. Box 616, 6200 MD, Maastricht, The Netherlands
n.osnabrugge@student.maastrichtuniversity.nl