Cardiac Abnormality Detection Based on an Ensemble Voting of Single-Lead Classifier Predictions

Pierre Aublin[1], Mouin Ben Ammar[2], Nina Achache[5], Mélina Benahmed[5], Aymane El Hichami[5], Michel Barret[4], Jérémy Fix[3], Julien Oster[1]

[1] IADI, U1254, Inserm , Université de Lorraine, Nancy, France
[2] ENSTA, Paris, France
[3] LORIA, CNRS, CentraleSupélec, Université Paris-Saclay, F-57000 Metz, France
[4] International Research Lab Georgia Tech - CNRS IRL 2958 CentraleSupélec
[5] CentraleSupélec, F-57070 Metz, France

Abstract

We developed a fully deep learning model to identify cardiac abnormalities from ECGs for the PhysioNet/CinC 2021 Challenge. Decision on different lead subsets was based as an average voting of all available single-lead predictions. ECG signals were bandpass filtered between 0.5 and 120 Hz, resampled at 250 Hz, cropped to 10 seconds and normalized (zero-mean, unit-variance). The neural network architecture consisted of fifteen blocks. Most blocks consisted in one-dimensional convolution followed by rectified linear unit activation, batch normalization, and dropout layers. Twelve blocks also contained a squeeze and excitation module. A global max pooling layer allowed for the extraction 512 features for each signal. Those features were inputted in fully connected MLP with two hidden layers with leaky rectified linear activation and linked to the outputs through a sigmoid activation. Our team (iadi-ecg) obtained scores of 0.48, 0.47, 0.47, 0.47, 0.46 on the twelve, six, four, three, two lead versions of the hidden challenge test set, resulting in final ranking between the 11th, and 12th out of 39 teams). The suggested approach had difficulties to generalize well on the hidden test set, and future works will aim at developing an hybrid model, as we assume that hand-crafted features might help for generalization purpose. The proposed technique demonstrated its ability to classify ECGs even when only two leads were available.

1. Introduction

Cardiovascular diseases are a leading cause of global mortality [1]. The Physionet/Computing in Cardiology Challenge 2020 aimed at classifying cardiac abnormalities from 12 standard lead electrocardiograms (ECG) [2]. Traditional automated ECG classification approaches rely on the use of handcrafted features extracted from the ECG signals and based on domain knowledge. These features are then fed into a classification stage [3].The domain knowledge based features used for rhythm classification can be divided into two categories: (i) temporal features, which depict the regularity of the heart rate and are extracted from the instantaneous heart rate (or RR intervals). These features are used to represent the level of predictability or order of these RR intervals. The presence of an arrhythmia comes with a specific signature, which can be identified using classical machine learning approaches. Such features include heart rate variability (HRV) characteristics [3], or predictability or irregularity of the RR intervals (Sample Entropy [4], or based on a Poincaré Plot representation). (ii) The second category of features consists in the analysis of the ECG morphology aiming at detecting pathological or abnormal electrical propagation (Premature Ventricular Contraction, presence or absence of P-wave (f-waves), prolonged QT interval or elevated ST segment).

More recently, deep learning approaches have also been proposed for the analysis of ECG signals, especially for rhythm classification [5]. Such solutions have been able to emerge thanks to the availability of large datasets of physiological signals (PTB [6], PTB-XL[7], Chapman-Shaoxing[8], CPSC 2017 [9], Ningbo [10], Georgia[2]). Several deep learning techniques have been suggested, starting from the use of convolutional neural networks (CNN) [11], to the use of recurrent neural networks (such as GRU or LTSMs) [12] in order to capture the temporal evolution of the signals, and finally to the use of attention-based mechanisms and more specifically Transformers [13] which revolutionized the field of Natural Language Processing.

Finally, as shown last year, the application of hybrid approaches based on a combination of deep learning and hand-crafted features allows for better classification per-

 ISSN: 2325-887X DOI: 10.22489/CinC.2021.046

formance, and seemingly better generalization on unseen datasets [14].

The Physionet/Computing in Cardiology Challenge 2021 [15] aims also at the classification of cardiac abnormalities, but from reduced-lead ECG signals.

2. Methods

For the classification task, we decided to develop a deep learning approach with a conventional CNN architecture for the automated extraction of features. In the following subsections, we will describe the data preparation, architecture of the network and how the network was optimized.

2.1. Dataset and Preprocessing

For training a cross-fold classification on five folds, each stratified by class was performed on the challenge training set [15]. We ensured that all the leads of the same patient were in the same split. Each ECG lead was seen as a one-dimensional signal and considered as an independent sample. However, to respect the maximum execution time allowed in the challenge, the submitted model did not use five folds, but only the one that gave the best challenge metric.

For preprocessing of the signals, a 3rd order Butterworth bandpass filter [0.5 and 120Hz] was applied, followed by two notch filters at 50 and 60Hz to remove the potential powerline interference. The signals were then resampled at 250Hz. Only 10 seconds for each ECG signal were kept: centered in the middle of the recording for signals longer than 10 seconds, or zero-padded on each side for signals shorter than 10 seconds. Each lead signal was normalized to have zero-mean and unit variance.

2.2. Architecture

Figure 1 shows our neural network architecture. For feature extraction, 15 convolution blocks were used; each block composed of a convolution layer, followed by a batch normalization, and a rectified linear unit. The 3rd, 4th, 5th, 7th, 12th blocks were followed by an average pooling with a stride of 2. From the 4th convolution layer on, each block finished with squeeze and excitation (SE) modules [16] [17]. A global max pooling was finally applied in order to extract 512 features from each 10-second ECG lead. Classification consisted in two fully connected layers of output dimensions 128, with leaky rectified linear units as activation. A final fully connected layer with output dimension 26 was used for classification. The network output consisted of the logits, as the sigmoid activation corresponding to the classifier's output probabilities and prediction was incorporated within the loss function.

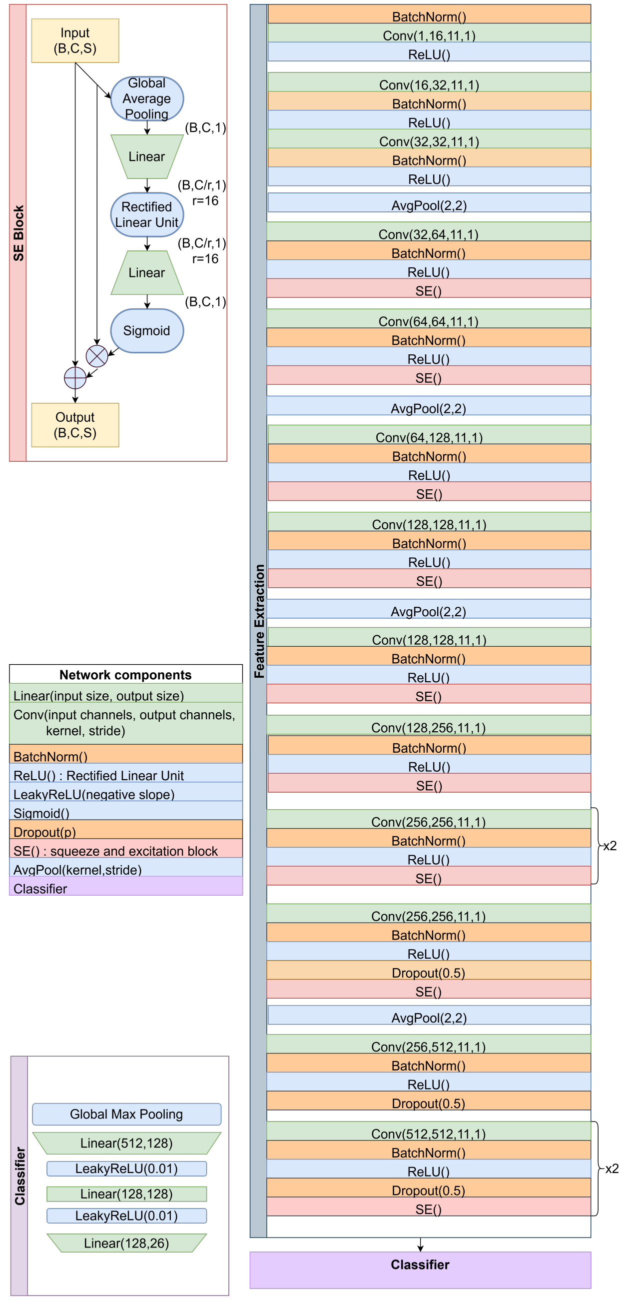

Figure 1. Left: Squeeze and Excitation (SE) block. Right: global architecture. To describe tensor shapes in SE block, B, C, N stand for Batch, Channels, number of Samples.

2.3. Loss and optimizer

The loss was a combination of a binary cross-entropy $\mathrm{bceloss}$ and a soft-dice loss dcloss. For a given sample i, let x_i be the preprocessed ECG signal, $F = (F_c)_c$ the network output, F_c being the network output (logit) for class c. Let y_i^c be the label of the sample i on class c, and $\hat{y}_i^c$ the predicted label. With σ the sigmoid function:

$$\mathrm{bce}(F_c, x_i, y_i^c) = -\omega_i^c[p_c.y_i^c \ln(\sigma(F_c(x_i)))+ (1-y_i^c)\ln(1-\sigma(F_c(x_i)))] \quad (1)$$

To deal with class imbalance, the weights of positive samples were adjusted for each class as $p_c = \frac{1-d(c)}{d(c)}$ with $d(c)$ the rate of occurrence for class c on the whole training dataset.

The sample weighting w_i^c parameters were also adjusted to take into account the heterogeneity between the databases (either due to a different patient population or a different annotation process) and to favor sparse model outputs. The contribution of a class c to the loss for a given sample was null if this class did not occur in the sub database where the training sample was from. A more important contribution to the loss was given to the sample i if it only had a few positive outputs. Hence $\omega_i^c = \frac{b_i^c}{\max(n_i,1)}$ with n_i the number of positive labels in the vector y_i, and $b_i^c = 1$ if the class is present in the database where the training sample i was drawn, or $b_i^c = 0$ otherwise. The final weighted binary cross entropy is given by

$$\mathrm{bceloss}(F, x_i, y_i) = \sum_c \mathrm{bce}(F_c, x_i, y_i^c) \quad (2)$$

A second term in the final loss was based on the dice coefficient [18] using a soft version as follows :

$$\mathrm{dcloss}(F, x_i, y_i) = 1 - 2\frac{1+\sum_c \sigma(F_c(x_i)).y_i^c}{1+\sum_c \sigma(F_c(x_i)) + \sum_c y_i^c} \quad (3)$$

The final loss combining (2) and (3) was given by

$$\mathrm{loss}(F, x_i, y_i) = \mathrm{bceloss}(F, x_i, y_i) + \mathrm{dcloss}(F, x_i, y_i). \quad (4)$$

Different optimization schemes were tested including Stochastic gradient descent (SGD) and Adam optimizers with a constant learning rate. Optimal learning rates were found via grid search between 2.10^{-5} and 1.10^{-3} and the best learning rate was at 1.10^{-4}.

A SGD with cyclic learning rate update was also tested and finally chosen for the official submission. The learning-rate was updated at every batch iteration using a cyclic update policy [19]. Cycles were triangular, varying between 2.10^{-5} and 1.10^{-3}, with a period of 2000 iterations. Batch size was set at 64, and the final model was trained over 50 epochs. The different models were trained on several workstations with different GPUs (Nvidia A100 and Titan XP). Training and testing were performed in docker containers with memory limited to 60 GB to replicate the Challenge server environment.

2.4. Postprocessing

After training of the network a calibration step was performed. The decision threshold for each class was adjusted in order to maximize the final challenge metric on the validation fold.

To deal with the different sets of reduced leads, a simple voting of the single lead-based outputs was performed. The output probabilities of the classifier were averaged over all the provided leads for each reduced set (2-, 3-, 4-, 6- or 12-leads). The final classification was obtained using the previously described decision thresholds.

3. Results

The additional value of the custom loss function was demonstrated by 5-fold cross-validation scores assembled in table 1 with a 0.03 improvement in the Challenge metric compared with a weighted BCE.

Models	Score
$bceloss$,SGD cyclic l_r	0.619 ± 0.004
$loss$, SGD constant $l_r = 10^{-4}$	0.643 ± 0.007
$loss$, Adam constant $l_r = 10^{-4}$	0.646 ± 0.001
$loss$, SGD cyclic l_r, without SE modules	**0.656 ± 0.001**
$loss$, SGD cyclic l_r *(submitted entry)*	0.652± 0.004

Table 1. Challenge scores during 5-fold cross-validation for different models on the training set.

Table 2 gives the scores obtained by our entry during Cross-Fold validation, and on the validation and test sets.

Leads	Training	Validation	Test	Ranking
12	0.66	0.59	0.48	12
6	0.64	0.58	0.47	11
4	0.64	0.57	0.47	11
3	0.64	0.57	0.47	11
2	0.63	0.56	0.46	11

Table 2. Challenge scores for our selected entry (team iadi-ecg) obtained on the validation and test sets.

4. Discussion

During this competition, several avenues of research have been explored to optimize the performance of the deep learning models. From Table 1 the main improvement factor came from the use of the combined DICE and BCE loss, and the use of the SGD optimizer with a cyclic learning rate.

The proposed model was trained on single lead signals and the multilead prediction was only an average voting, not considering the lead positions or inter-leads relationships. The average drop of 0.2 on validation and 0.1 on test between the lead subsets and the 2 lead score suggest the final classifier remained able to use only few available leads. This suggests that despite limitations a deep learning based classifier could still learn lead-independent features to predict abnormalities from ECG.

The results of Table 2 reflect the difficulties of the current approach to generalize on a different hidden test set, with an average drop of 0.10 on the challenge metric between the hidden validation set and the test set. There is still some room for improvement in the settings of some hyperparameters, for example in the weighting of the loss, either refining the weighting between DICE and BCE or better accounting for class imbalance and dataset heterogeneity in the DICE loss term. The use of a representation learning approach in order to initialize the network weights has also shown to yield better performance than using a random initialisation, and such technique has already been suggested for ECG analysis [20]. It would be interesting to assess in future works how such a technique would help. Finally, other techniques such as the use of an ensemble of models or the use of an hybrid model with added handcrafted features were also shown to be of added value, but such solutions were not investigated here either due to a lack of time or due to the limited available training time on the Challenge server.

Acknowledgements

PA and JO were supported by a grant from the ERA-CVD Joint Translational Call 2019, MEIDIC-VTACH (ANR-19-ECVD-0004). We also gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU partly used for this research.

References

[1] Virani SS, *et al.* Heart Disease and Stroke Statistics – 2021 Update: a Report from the American Heart Association. Circulation 2021;143(8):e254–e743.

[2] Alday EAP, *et al.* Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020. Physiological Measurement 2021;41(12).

[3] Clifford G, *et al.* Advanced Methods and Tools for ECG Data Analysis. 2007. ISBN 978-1-58053-966-1.

[4] Richman J, *et al.* Physiological Time-Series Analysis using Approximate Entropy and Sample Entropy. American Journal of Physiology Heart and Circulatory Physiology 2000; 278(6):H2039–H2049.

[5] Hannun AY, *et al.* Cardiologist-Level Arrhythmia Detection and Classification in Ambulatory Electrocardiograms Using a Deep Neural Network. Nature Medicine 2019;25(1):65–69.

[6] Bousseljot R, *et al.* Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. Biomedizinische Technik 1995;40(S1):317–318.

[7] Wagner P, *et al.* PTB-XL, a Large Publicly Available Electrocardiography Dataset. Scientific Data 2020;7(1):1–15.

[8] Zheng J, *et al.* A 12-lead Electrocardiogram Database for Arrhythmia Research Covering More Than 10,000 Patients. Scientific Data 2020;7(48):1–8.

[9] Liu F, *et al.* An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection. Journal of Medical Imaging and Health Informatics 2018;8(7):1368—1373.

[10] Zheng J, *et al.* Optimal Multi-Stage Arrhythmia Classification Approach. Scientific Data 2020;10(2898):1–17.

[11] Rubin J, *et al.* Densely Connected Convolutional Networks and Signal Quality Analysis to Detect Atrial Fibrillation Using Short Single-Lead ECG Recordings. In Computing in cardiology (CinC). IEEE, 2017; 1–4.

[12] Teijeiro T, *et al.* Arrhythmia Classification from the Abductive Interpretation of Short Single-Lead ECG Records. In Computing in Cardiology (CinC). 2017; 1–4.

[13] Vaswani A, *et al.* Attention is all you need. In Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R (eds.), Advances in Neural Information Processing Systems, volume 30. Curran Associates, Inc., 2017; .

[14] Natarajan A, *et al.* A Wide and Deep Transformer Neural Network for 12-Lead ECG Classification. In Computing in Cardiology (CinC). 2020; 1–4.

[15] Reyna MA, *et al.* Will Two Do? Varying Dimensions in Electrocardiography: the PhysioNet/Computing in Cardiology Challenge 2021. Computing in Cardiology 2021;48:1–4.

[16] Hu J, *et al.* Squeeze-and-Excitation Networks. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2018; 7132–7141.

[17] Zhu Z, *et al.* Classification of Cardiac Abnormalities From ECG Signals Using SE-ResNet. In Computing in Cardiology (CinC). 2020; 1–4.

[18] Deng R, *et al.* Learning to Predict Crisp Boundaries. In Proceedings of the European Conference on Computer Vision (ECCV). 2018; 562–578.

[19] Smith LN. Cyclical Learning Rates for Training Neural Networks. In IEEE Winter Conference on Applications of Computer Vision (WACV). 2017; 464–472.

[20] Kiyasseh D, *et al.* CLOCS: Contrastive Learning of Cardiac Signals Across Space, Time, and Patients. In Meila M, Zhang T (eds.), Proceedings of the 38th International Conference on Machine Learning, volume 139. PMLR, 18–24 Jul 2021; 5606–5615.

Address for correspondence:

Julien Oster, Laboratoire IADI (Inserm U1254)
Bâtiment Recherche, CHRU Nancy-Brabois, rue du Morvan
54500 Vandoeuvre, France
julien.oster@inserm.fr