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

Pierre Aublin[1], Mouin Ben Ammar [2], Michel Barret [3], Jérémy Fix [3], Julien Oster[1]

[1] IADI, U1254, Inserm , Université de Lorraine, Nancy, France
[2] ENSTA, Paris, France
[3] CentraleSupélec, Metz, France

Abstract

To tackle the 2021 PhysioNet/Cinc challenge, we (iadi-ecg team) proposed a deep learning model for the classification of a single-lead ECG signal. Decision on the reduced lead-set was taken as an average voting of the available single lead-based predictions.

Single lead ECG signals were resampled at 250 Hz, bandpass filtered between 0.5 and 120Hz using a 3rd order Butterworth filter, and normalized (zero mean and unit variance). The neural network architecture consisted of 15 blocks, which include a one-dimensional convolutional layer, followed by rectified linear unit activation, batch normalization, and dropout layers. Between consecutive blocks, squeeze and excitation layers were introduced. A final global max pooling layer extracted 512 features for each signal, which were inputted in two fully connected layers with leaky rectified linear activation units. Training was performed by minimising a custom loss function, which combined a dice loss and binary cross entropy, using a Stochastic Gradient Descent with a cyclic update of the learning rate, with a Batch size of 64 over 50 epochs. A 5-fold cross-fold validation scheme was used for training and evaluating the model locally. Each fold contained recordings from all the databases, but were stratified by abnormality and subject (all leads from the same recording were included in the same fold).

Using the challenge metric, an average score of 0.657, 0.643, 0.642, 0.639, 0.629 was obtained for the 12, 6, 4, 3, 2 lead datasets respectively for the cross validation, with the corresponding submitted entry scores of 0.586, 0.577, 0.574, 0.572, and 0.563.

Final scores reflect a good level of performance for the detection of cardiac abnormalities, which could be further improved with optimised hyperparameters (in the loss) or by incorporating hand-crafted features or pre-training with a representation learning approach.

1. Introduction

Cardiovascular disease is 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], and several models have been suggested from simple Logistic Regression to more complex approaches such as Support-Vector-Machines, Random Forests or Boosting (XGBoost).

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 signal, 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 performance, and seemingly better generalisability 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 optimised.

2.1. Dataset and Preprocessing

For training a cross-fold classification on five folds, each stratified by class was performed. We also ensured that all the leads of the same patient were in the same split.

Within each fold, each ECG lead was considered as an independent sample. First 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. The signals were finally normalized to have zero mean and unit variance.

2.2. Architecture

The final architecture of the neural network is depicted in figure 1.

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, 5th, 7th, 11th blocks were followed by an average pooling with a stride of 2. From the 3rd 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 and 26, with leaky rectified linear units as activation. The overall network consisted of 10,683,196 optimizable parameters.

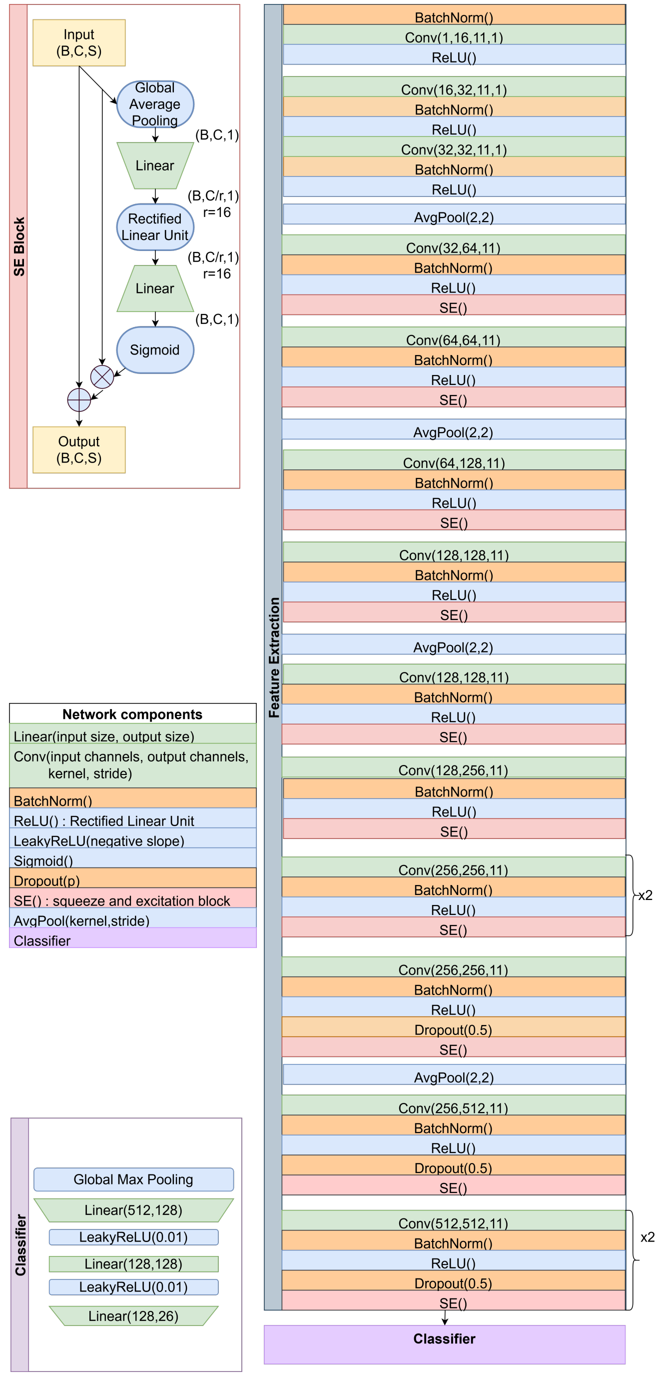

Figure 1. Left : Squeeze and Excitation (SE) block. Middle : global architecture

2.3. Loss and optimizer

The loss was a combination of a binary cross-entropy $bceloss$ and a soft-dice loss $diceloss$. For a given sample i, let x_i be the preprocessed ECG signal, F 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 activation function:

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

To deal with class imbalance, the weights of positive samples were adjusted for each class as

$$p_c = \frac{1 - d(c)}{d(c)} \quad (2)$$

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 for 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, i.e. $d(y_i^c = 1 | x_i \in database) = 0$. A more important contribution to the loss was given to the sample i if it only had a few positive outputs. Hence

$$\omega_i^c = \frac{db_i^c}{max(n_{act}, 1)} \quad (3)$$

with n_{act} the number of positive labels in the vector y_i, and $db_i^c = 1$ if the class is present in the database where the training sample i was drawn, or $db_i^c = 0$ otherwise. The final weighted binary cross entropy is given by

$$bceloss(F, x_i, y_i) = \sum_c (bce(F_c, x_i, y_i)) \quad (4)$$

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

$$dcloss(F, x_i, y_i) = 1 - 2\frac{1 + \sum_{c=1}^{26} F_c(x_i).y_i^c}{1 + \sum_{c=1}^{26} F_c(x_i) + \sum_{c=1}^{26} y_i^c} \quad (5)$$

The final loss combining (4) and (5) was given by

$$loss(F, x_i, y_i) = bceloss(F, x_i, y_i) + dcloss(F, x_i, y_i). \quad (6)$$

Different optimizers were tested and finally a stochastic gradient descent was chosen. 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 Cross-fold validation scores assembled in table 1 with an 0.05 improvement in the Challenge metric compared with a weighted BCE.

Criterion	Score
$bceloss$	0.609 ± 0.002
$loss$	0.658 ± 0.003

Table 1. Challenge metric score on cross fold validation for the different losses (on training dataset)

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

Leads	Training	Validation	Test	Ranking
12	0.657	0.586	???	???
6	0.643	0.577	???	???
4	0.642	0.574	???	???
3	0.639	0.572	???	???
2	0.629	0.563	???	???

Table 2. Challenge scores for our selected entry (team iadi-ecg) obtained on the hidden validation set. Hidden test set results and final ranking are not yet available.

4. Discussion

During this competition, several avenues of research have been explored in order to optimise the performance of the deep learning models. The initial model was inspired from a previous work [20], and as seen during the optimisation process the use of SE has been significant. The other main improvement factors came from the use of the combined DICE and BCE loss, and the use of the SGD optimiser with a cycling rate.

There is probably 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.

Other avenues of research that remain to be explored include for example the use of fully attention-based mechanisms (or Transformers-based architecture). Several tests were conducted in order to replace the global max pooling layer with a Transformer network, but those tests did not lead to significant performance improvements.

The use of a representation learning approach in order to initialise 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 [21]. 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 hand-crafted 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 P, et al. Classification of 12-lead ECGs: the PhysioNet/computing in cardiology challenge 2020. Physiological Measurements 2020;.

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

[4] Richman JS, 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 2017 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 2017 Computing in cardiology (CINC). IEEE, 2017; 1–4.

[13] Vaswani A, et al. Attention is all you need. In Advances in neural information processing systems. 2017; 5998–6008.

[14] Natarajan A, et al. A wide and deep transformer neural network for 12-lead ECG classification. In 2020 Computing in Cardiology. IEEE, 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. Identification of 27 abnormalities from multi-lead ecg signals: an ensembled se_resnet framework with sign loss function. Physiological Measurement 2021;.

[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 2017 IEEE winter conference on applications of computer vision (WACV). IEEE, 2017; 464–472.

[20] Vogt N. CNNs, LSTMs, and attention networks for pathology detection in medical data. arXiv preprint arXiv191200852 2019;.

[21] Kiyasseh D, et al. Clocs: Contrastive learning of cardiac signals across space, time, and patients. In International Conference on Machine Learning. PMLR, 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