Non-architectural improvements for ECG classification using deep neural network

Conference Paper · December 2020

DOI: 10.1109/ISPRAS51486.2020.00016

Vladislav Ananev
Russian Academy of Sciences

Aram Avetisyan
Institute for System Programming, Russian Academy of Sciences

Vadim Gliner
Technion – Israel Institute of Technology

Non-architectural improvements for ECG classification using deep neural network

Pavel Andreev[1 6 7], Vladislav Ananev[1 2], Aram Avetisyan[1 8], Vladimir Makarov[1 2], Vadim Gliner[5], Assaf Schuster[5] and Evgeny Karpulevich[1 3 4]
[1]Ivannikov Institute for System Programming of the Russian Academy of Sciences, Moscow, Russia,
[2]Novgorod State University, Veliky Novgorod, Russia,
[3]National Research Center "Kurchatov Institute", Moscow, Russia,
[4]Genome Technologies and Bioinformatics Research Centre, Moscow Institute of Physics and Technology, Dolgoprudniy, Russia,
[5]Computer Science Department, Technion-IIT, Haifa, Israel,
[6]Skolkovo Institute of Science and Technology, Moscow, Russia,
[7]Higher School of Economics – National Research University, Moscow, Russia,
[8]Lomonosov Moscow State University, Moscow, Russia

***Abstract*—Due to the latest advances in machine learning algorithms new deep learning-based approaches to the interpretation of 12-lead electrocardiograms have been developed, demonstrating the quality of diagnostics comparable to the expert one. In this paper, we propose several techniques increasing the quality of ECG classification by a deep neural network. The techniques include patient metadata incorporation, signal denoising and self-adaptive model training. The experimental validation of the approaches was carried out on a novel dataset containing 64198 standard ECG recordings obtained during routine medical practice. The conducted experiments demonstrated statistically significant quality growth compared to the baseline, supporting the further application of our findings.**

***Index Terms*—ECG Classification, Convolutional Neural Network, Deep Learning, Denoising, Self-Adaptive Learning**

I. Introduction

Standard 12-lead electrocardiography is a common procedure for diagnosing cardiac abnormalities. Millions of people die from heart diseases annually, while the existing systems for automatic ECG interpretation are still subjected to a significant amount of errors [1] being only an auxiliary tool in clinical practice. The full automation of ECG analysis can significantly advance accessibility of quality medical care, provide clinicians with a second opinion and help patients to pre-monitor heart problems. Thus, improving the accuracy of predicting cardiac pathologies by ECG is an important research task.

Despite a large number of works devoted to the construction of effective algorithms for automatic classification of electrocardiograms, this problem is still of great relevance. The models based on hand-crafted feature extraction techniques might not take into account hidden dependencies in data and rely heavily on feature extraction algorithms which are often prone to errors, whereas deep neural network models don't require any explicit human knowledge about the domain they are applied to and can be trained in an end-to-end fashion unifying feature extractor and classification algorithms. The deep neural networks were first used to classify ECGs relatively recently, but they immediately showed promising results, demonstrating the recognition quality of some pathologies comparable human level. Nevertheless, in the studies reviewed, which use neural networks for ECG classification, an unprocessed signal is used as the input data of the models. In addition, patient metadata are ignored, and the neural network training process depends to a large extent on the quality of the markup, the receipt and validation of which is a significant problem.

In this paper we consider non-architectural ways of improving ECG classification performance using deep neural network. The techniques concerned include patients metadata incorporation, signal de-noising and self-adaptive model training. For the experiments a novel ECG dataset was collected and processed with the help of Russian cardiologists. An extensive experimental evaluation of the proposed techniques provides the evidence of their relevance and consistency.

II. Related work

In this section we briefly review ECG signal preprocessing techniques, as well as recently emerged algorithms for ECG classification.

A. Preprocessing

Most datasets with ECG signals contain various types of noise. These include a baseline wander due to the patient's movements and breathing, noises due to poor contact of the electrodes with the skin, and noises arising from the electrical activity of the patient's muscles [2]. Suppressing these noises is an important research direction as it allows to improve dataset consistency. The most common noise suppression methods known in the literature are digital filters, methods based on discrete wavelet transform and empirical mode decomposition [3].

Digital filters are one of the simplest methods of noise suppression. They are based on filtering high-frequency and low-frequency signal components. The main disadvantage of these methods is that the frequency characteristics of the informative and noise components of the ECG signal can intersect, which can prevent effective filtering.

Another approach in removing noise in the ECG signal is filtering discrete wavelet transform coefficients [4]. The wavelet transform allows obtaining information not only about the frequency component of the signal, but also to take into

978-1-6654-1291-9/20/$31.00 ©2020 IEEE
DOI 10.1109/ISPRAS51486.2020.00016

account localization in time. The disadvantage of this method is the strong dependence of the obtained filtration on the mother wavelet, which leads to difficulties in applying the method to all twelve leads.

Empirical mode decomposition is also one of the methods to suppress noise in ECG [5]. Its main difference from the above techniques is that the basis decomposition functions are determined directly from the studied signal. This provides better adaptability of the method.

The final stage of preprocessing of the signal might be the isolation of R-peaks and the segmentation of the signal into individual heart beats. Pan-Tompkins algorithm is most often tool used for this purpose [6].

B. Feature Extraction-based Classification techniques

The classic and still widely used in practice approach to electrocardiogram classification is based on the selection of relevant features and the subsequent classification of the ECG using these features. An example of this method is the ECG analysis program of the University of Glasgow [7]. This program highlights clinically significant features, such as the duration of waves and intervals. The selected features are then used to classify the ECG using decision rules based on medical knowledge. The disadvantages of this approach are the strong dependence on the quality of identifying features and the limitations imposed by not always clear medical criteria for interpreting the ECG. The latter drawback, however, can be overcome by using machine learning models. Thus, in [8] 79 clinically significant signs were used to train the gradient boosting model, which allowed the authors to take first place in the PhysioNet competition in ECG classification.

In addition, features relevant for classification can be extracted from the ECG signal by methods that do not require any knowledge of the subject area. These methods include dimension reduction algorithms, signal decomposition in terms of basis (wavelet transform and others), calculation of high-order statistics, as well as combinations of these techniques. The extracted features are then used for machine learning algorithms. The most common among them are support vector machines and shallow neural networks [4].

C. Neural Network Classification Methods

The widespread use of digital health monitoring systems has led to the accumulation of large arrays of medical data and, as a result, to the possibility of improving existing systems for automatic diagnosis through the use of neural networks [9] [10].

To solve the problem of classifying electrocardiograms, various architectures of deep neural networks have been proposed. For instance, in [11], recurrent neural networks were used to recognize arrhythmias. The work compared architectures based on LSTM and GRU. The signal of only one derivation was used as input data. In addition, the authors limited themselves to a two-class classification problem. RNN with LSTM was also used as an algorithm for extracting global features in [8].

The most commonly used deep neural network architecture utilized to classify ECGs is the convolutional neural network. CNNs act simultaneously as a feature extraction algorithm and as a classifier. To detect atrial fibrillation Yong [12] use two-dimensional convolutional networks to classify electrocardiograms by spectrograms obtained using the windowed Fourier transform or discrete wavelet transform. Gliner [13] proposed two different CNN architectures, one trained using digital signals and one trained using images, to identify atrial fibrillation (AF). These methods have shown high results in atrial fibrillation detection.

In [14] and [15] CNNs were applied directly to the ECG signal. Datasets used in these papers were significantly larger than in all previously mentioned works, since the authors used data sets that are not publicly available, while most of the studies use data from the MIT BIH database [16], which contains only a small number of labeled ECG records.

Within the framework of the study [14], a dataset of one-lead ECG records was obtained. The collected data was annotated by certified ECG interpreters. Then unidimensional CNN with residual blocks was trained to predict the type of rhythm every 1.21 seconds from the corresponding ECG signal without preprocessing. Ribeiro [15] used data from the telediagnostic ECG monitoring system. The sample used consisted of 2,470,424 12-lead ECG recordings, corresponding to 1,676,384 patients. As in [14], the model is a unidimensional CNN with residual blocks. The trained model of the study achieved a classification quality superior to the human one for some of the concerned pathologies.

The work [17] provides a number of benchmarking results for a recently published PTB-XL dataset, covering different machine ECG analysis tasks. Importantly, the authors concerned a comparison of different neural network architectures for ECG classification task and revealed a publicly available code for their testing. The current work is built upon their implementation of resnet1d50 neural network architecture.

III. Methodology

A. Model

A convolutional neural network with residual blocks resnet1d50 was used to classify electrocardiograms. The model was developed using the Pytorch Framework based on the implementation provided by [17]. The network architecture is shown in Fig. 1. The input data is a 12-lead ECG in the form of 768 x 12 matrix, as well as patients metadata (sex and age). 768 is the number of measurements which corresponds to 3.072 seconds recording at 250Hz frequency. To reduce the effect of incorrectly labeled samples on the learning process, a recently proposed self-adaptive training procedure [18] was elaborated.

The model was trained to predict the following classes: atrial fibrillation (AF), first-degree atrioventricular block (AVB), left bundle branch block (LBBB), right bundle branch block (RBBB), left atrial enlargement (LAE), left ventricular hypertrophy (LVH) and normal sinus rhythm (NSR).

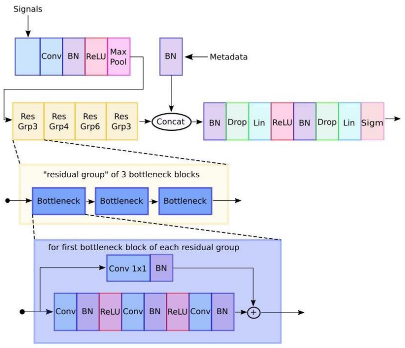

Fig. 1. The neural network architecture. Conv - unidimensional convolution operation, BN - batch normalization, ReLU - nonlinearity function, Drop - dropout operation, MaxPool - pooling by the maximum function, Lin - fully connected linear layer, Concat - vector concatenation, Sigm - element-wise sigmoid function.

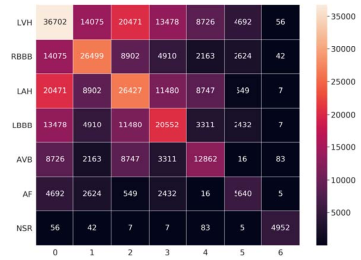

Fig. 3. Distribution of ECG samples among classes. The number at the intersection of the i-th row and the j-th column corresponds to the number of records in which the i-th and j-th pathologies are presented simultaneously, the i-th number on the diagonal denotes the total number of records corresponding to the i-th pathology.

B. Data

The dataset containing ECG records in 12 standard leads used within this work was obtained from several local medical centers in Veliky Novgorod, Russia. The duration of each recording is 4 seconds. Each recording was preprocessed in two steps. At first, LOWESS method was used to correct the isoline using locally weighted smoothing. Then high-frequency noise was removed by discrete wavelet transform.

The analysis of modified data showed that n recordings are still noisy, mostly at the edges of the ECG. Therefore, edges of all the ECG recording were cropped, the rest of the noisy samples were removed from the dataset.

The final set contains 64198 anonymized ECG records, corresponding to patients aged 13 to 95 years. Distribution of patients by age is shown in Figure 2.

The data was annotated by Russian cardiologists in the form of descriptions of the diagnoses of the respective patients during routine practice. From the textual descriptions of diagnoses by keywords, several classes of disorders, given above, were distinguished, as well as the case of the absence of significant deviations from the age norm. Distribution of patients by classes is shown in Figure 3.

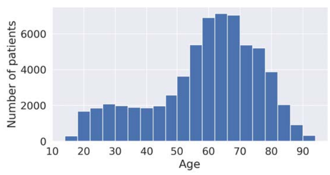

Fig. 2. Distribution of patients by age.

C. Experiments

In this section we provide experimental results indicating the relevance of the considered approaches to improvement of classification performance.

The evaluation of the techniques was performed on 5-fold cross-validation. The 5 different cases which were taken into account for evaluation are as follows vanilla model (resnet1d50) with raw signals as inputs, vanilla model with preprocessed signals, model with incorporated metadata and preprocessed signals as inputs, self-adaptive training of vanilla model with preprocessed signals as inputs, and self-adaptive training of model with incorporated metadata and preprocessed signals as inputs. The mentioned cases are further denoted as "baseline", "PREP", "PREP + MET", "PREP + SELFAD" and "PREP + MET + SELFAD" correspondingly. Binary cross-entropy averaged over classes (BCE), area under ROC-curve (ROC-AUC), average precision (AP) and F1-Score were chosen as classification performance metrics. The results are shown in Figure 4 and Table I. The higher quality of the proposed model compared to the baseline suggests that the given non-architectural changes improve the quality of predictions. Interestingly, the results suggest that the quality gains from metadata incorporation and self-adaptive training are not cumulative.

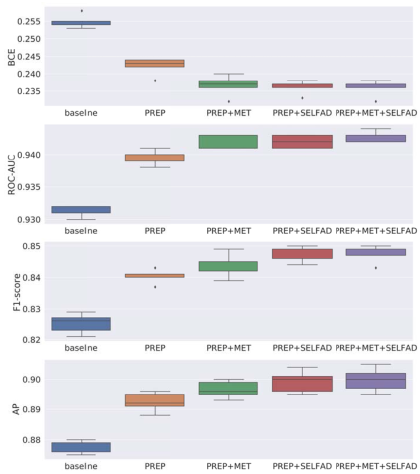

Fig. 4. Dependence of quality metrics (BCE, ROC-AUC, AP and F1-score) averaged over classes on the methodology used for ECG classification. baseline – vanilla model (resnet1d50) with raw signals as inputs, PREP – signal preprocessing, MET – incorporation of metadata into the model, SELFAD – self-adaptive training of the model.

TABLE I
95% CONFIDENCE INTERVALS (BASED ON STUDENT'S t-DISTRIBUTION) FOR BINARY CROSS-ENTROPY (BCE), AREA UNDER THE ROC-CURVE (ROC-AUC) AND AVERAGE PRECISION (AP), DEPENDING ON THE CONSIDERED CASES (SEE MAIN TEXT). 1-LEAD – DATA FROM ONLY ONE LEAD (II), 12-LEAD – DATA FROM ALL 12 STANDARD LEADS, PREP – PREPROCESSED SIGNALS AS INPUTS FOR THE NETWORK, MET – PATIENT'S METADATA INCORPORATION INTO THE MODEL, SELFAD – SELF-ADAPTIVE TRAINING OF THE MODEL.

Method	BCE	ROC-AUC	AP	F1-score
1-lead	0.348 ± 0.003	0.874 ± 0.001	0.763 ± 0.006	0.703 ± 0.008
12-lead(baseline)	0.255 ± 0.002	0.931 ± 0.001	0.878 ± 0.003	0.825 ± 0.004
12-lead+PREP	0.242 ± 0.003	0.939 ± 0.002	0.893 ± 0.004	0.840 ± 0.003
12-lead+PREP+MET	0.237 ± 0.004	0.942 ± 0.001	0.897 ± 0.003	0.844 ± 0.005
12-lead+PREP+SELFAD	0.236 ± 0.002	0.942 ± 0.001	0.899 ± 0.005	0.848 ± 0.003
12-lead+PREP+MET+SELFAD	0.236 ± 0.003	0.943 ± 0.001	0.810 ± 0.005	0.847 ± 0.003

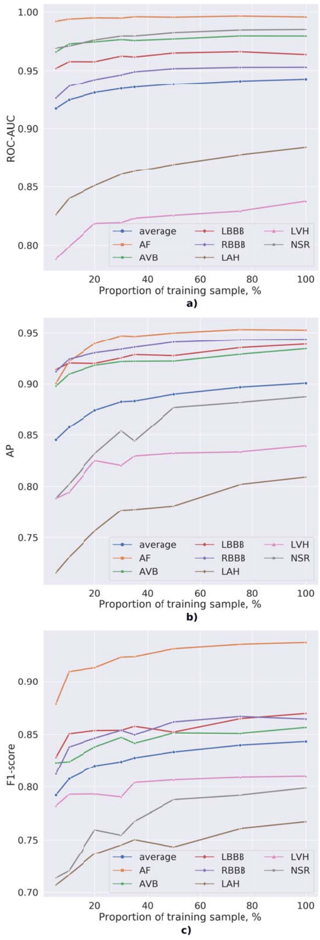

Fig. 5. Dependence of the quality metrics on the size of the training sample expressed in fractions of 48 thousand samples (the maximum available size of the training sample) for classification of the following conditions atrial fibrillation (AF), first-degree atrioventricular block (AVB), left bundle branch block (LBBB), right bundle branch block (RBBB), left atrial hypertrophy (LAE), left ventricular hypertrophy (LVH) and normal sinus (NSR).

In addition, we assessed the prediction quality gain for each of the pathologies separately. The results are provided in Table II, which contains ROC-AUC and AP metrics for baseline and "PREP + MET + SELFAD" (combination of all proposed techniques) cases. The "PREP + MET + SELFAD" model have statistically significant quality improvement over baseline for all pathologies under consideration except for atrial fibrillation which is detected with very high quality by baseline.

Lastly, we provide experimental insights into the importance of a large training sample for good prediction performance. The dependence of the area under the ROC curve, average precision and F1-score on the size of the training sample expressed in fractions of 48 thousand samples (the maximum available size of the training sample for cross-validation) is presented in Figure 5. It can be noted that for some of the abnormalities concerned the training sample could be reduced up to 3 times with moderate quality degradation. Interestingly, one may consider the quality gain equivalent of our techniques in terms of training sample reduction. From this point of view, the application of concerned techniques accounts for reduction of training sample up to 5 times for some of the pathologies.

We demonstrated the effectiveness of the concerned techniques in terms of a number of quality metrics. The conducted experiments showed a statistically significant quality improvement over baseline for each of the considered pathologies as well as on average.

IV. CONCLUSION

In this work, we proposed several techniques for ECG classification performance improvement. The conducted experiments on a novel ECG dataset provided experimental evidence indicating their practical relevance. The techniques allowed to significantly increase the classification quality of the deep convolutional neural network model (resnet1d50) while not changing its architecture significantly. The latter means that the approaches concerned are compatible with other DNN architectures as well, supporting their future application in research and practice.

AUTHOR INFORMATION

Pavel Andreev, Vladislav Ananev, Aram Avetisyan contributed equally to this work.

CONTRIBUTIONS

P.A., V.A. and A.A. wrote the article. P.A. designed experiments and directed research. Data for experiments were provided by V.M. P.A. and V.A. performed research and analyzed data. E.K. supervised the project. V.G. and A.S. provided feedback on the text.

TABLE II

QUALITY METRICS OF THE MODEL PERFORMANCE FOR BASELINE AND THE CASE WHEN ALL TECHNIQUES PROPOSED IN THE CURRENT PAPER ARE EMPLOYED. THE FIGURES ARE REVEALED FOR EACH OF THE FOLLOWING CLASSES SEPARATELY AF – ATRIAL FIBRILLATION, AVB – 1ST DEGREE ATRIOVENTRICULAR BLOCK, LBBB – LEFT BUNDLE BRANCH BLOCK, RBBB – RIGHT BUNDLE OF HIS, LAE – LEFT ATRIAL ENLARGEMENT, LVH – LEFT VENTRICULAR HYPERTROPHY, NSR – WITHOUT SIGNIFICANT DEVIATIONS FROM THE AGE NORM. STATISTICALLY SIGNIFICANT IMPROVEMENTS ARE HIGHLIGHTED IN BOLD FONT.

Metric	ROC-AUC		Average precision	
	baseline	ours	baseline	ours
AF	0.994 ± 0.003	$\mathbf{0.995 \pm 0.003}$	0.948 ± 0.010	$\mathbf{0.960 \pm 0.005}$
AVB	0.975 ± 0.003	$\mathbf{0.980 \pm 0.002}$	0.924 ± 0.009	$\mathbf{0.935 \pm 0.009}$
LBBB	0.961 ± 0.003	$\mathbf{0.966 \pm 0.001}$	0.925 ± 0.004	$\mathbf{0.935 \pm 0.003}$
RBBB	0.946 ± 0.002	$\mathbf{0.956 \pm 0.004}$	0.935 ± 0.004	$\mathbf{0.947 \pm 0.005}$
LAE	0.851 ± 0.008	$\mathbf{0.879 \pm 0.007}$	0.755 ± 0.014	$\mathbf{0.798 \pm 0.017}$
LVH	0.813 ± 0.005	$\mathbf{0.835 \pm 0.004}$	0.818 ± 0.008	$\mathbf{0.837 \pm 0.012}$
NSR	0.979 ± 0.002	$\mathbf{0.986 \pm 0.001}$	0.840 ± 0.016	$\mathbf{0.885 \pm 0.014}$

FUNDING

The reported study was funded by RFBR and MOST according to the research project 19-57-06004.

REFERENCES

[1] A. P. Shah and S. A. Rubin, "Errors in the computerized electrocardiogram interpretation of cardiac rhythm," *Journal of electrocardiology*, vol. 40, no. 5, pp. 385–390, 2007.

[2] A. Velayudhan and S. Peter, "Noise analysis and different denoising techniques of ecg signal-a survey," *IOSR journal of electronics and communication engineering*, vol. 3, pp. 641–644, 2016.

[3] E. H. Houssein, M. Kilany, and A. E. Hassanien, "Ecg signals classification: a review," *International Journal of Intelligent Engineering Informatics*, vol. 5, no. 4, pp. 376–396, 2017.

[4] M. Alfaouri and K. Daqrouq, "Ecg signal denoising by wavelet transform thresholding," *American Journal of applied sciences*, vol. 5, no. 3, pp. 276–281, 2008.

[5] S. Pal and M. Mitra, "Empirical mode decomposition based ecg enhancement and qrs detection," *Computers in biology and medicine*, vol. 42, no. 1, pp. 83–92, 2012.

[6] J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE transactions on biomedical engineering*, no. 3, pp. 230–236, 1985.

[7] P. Macfarlane, B. Devine, and E. Clark, "The university of glasgow (uni-g) ecg analysis program," in *Computers in Cardiology, 2005*. IEEE, 2005, pp. 451–454.

[8] T. Teijeiro, C. A. García, D. Castro, and P. Félix, "Arrhythmia classification from the abductive interpretation of short single-lead ecg records," in *2017 Computing in cardiology (cinc)*. IEEE, 2017, pp. 1–4.

[9] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *Jama*, vol. 316, no. 22, pp. 2402–2410, 2016.

[10] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[11] S. Singh, S. K. Pandey, U. Pawar, and R. R. Janghel, "Classification of ecg arrhythmia using recurrent neural networks," *Procedia computer science*, vol. 132, pp. 1290–1297, 2018.

[12] Y. Xia, N. Wulan, K. Wang, and H. Zhang, "Detecting atrial fibrillation by deep convolutional neural networks," *Computers in biology and medicine*, vol. 93, pp. 84–92, 2018.

[13] V. Gliner, N. Keidar, V. Makarov, A. I. Avetisyan, A. Schuster, and Y. Yaniv, "Automatic classification of healthy and disease conditions from images or digital standard 12-lead electrocardiograms," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.

[14] A. Y. Hannun, P. Rajpurkar, M. Haghpanahi, G. H. Tison, C. Bourn, M. P. Turakhia, and A. Y. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature medicine*, vol. 25, no. 1, p. 65, 2019.

[15] A. H. Ribeiro, M. H. Ribeiro, G. M. Paixão, D. M. Oliveira, P. R. Gomes, J. A. Canazart, M. P. Ferreira, C. R. Andersson, P. W. Macfarlane, M. Wagner Jr *et al.*, "Automatic diagnosis of the 12-lead ecg using a deep neural network," *Nature communications*, vol. 11, no. 1, pp. 1–9, 2020.

[16] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[17] N. Strodthoff, P. Wagner, T. Schaeffter, and W. Samek, "Deep learning for ecg analysis: Benchmarks and insights from ptb-xl," *arXiv preprint arXiv:2004.13701*, 2020.

[18] L. Huang, C. Zhang, and H. Zhang, "Self-adaptive training: beyond empirical risk minimization," *arXiv preprint arXiv:2002.10319*, 2020.