

Patient-Specific Heartbeat Classification in Single-Lead ECG using Convolutional Neural Network

Elena Merdjanovska¹ and Aleksandra Rashkovska²

Abstract—For an expert cardiologist, any abnormality in the heart rhythm or electrocardiogram (ECG) shape can be easily detected as a sign of arrhythmia. However, this is a big challenge for a computer system. The need for automatic arrhythmia recognition comes from the development of many portable ECG measuring devices designed to function as a part of health monitoring platforms. These platforms, because of their wide availability, generate a lot of data and hence the need for algorithms to process this data. From the many methods for automatic heartbeat classification, convolutional neural networks (CNNs) are increasingly being applied in this ECG analysis task. The purpose of this paper is to develop arrhythmia classification model according to the standards defined by the Association for the Advancement of Medical Instruments (AAMI), using CNNs, on data from the publicly available MIT-BIH Arrhythmia database. We experiment with two types of heartbeat segmentation: static and dynamic. The ultimate goal is to implement an algorithm for long-term monitoring of a user's health, which is why we have focused on classification models from single-lead ECG, and, even more, on algorithms specifically designed for one person rather than general models. Therefore, we evaluate patient-specific CNN models also on measurements from a novel wireless single-lead ECG sensor.

I. INTRODUCTION

Electrocardiogram (ECG) is a recording of the electrical activity of the heart through electrodes usually placed on the skin. For an expert cardiologist, any irregularity in the cardiac rhythm or in the form of an ECG can be easily detected as an indication for arrhythmia. Arrhythmia is any change from the normal sequence of electrical impulses, which is accompanied by abnormal heart rhythm and can be a threat to the human life or require medical therapy to prevent further problems [1]. However, this is a big challenge for a computer system. The need for automatic recognition of arrhythmias comes from the development of many portable ECG measuring devices, such as various types of wireless ECG sensors [2], the purpose of which is to function as a part of health monitoring platforms. These systems, due to their wide availability, will generate a lot of data that will need suitable algorithms for processing it.

From the vast research field of ECG analysis, a large portion focuses on ECG arrhythmia classification. The research

on this topic has involved some standard methods in the past, such as frequency analysis [3], wavelet transform [4] and template matching [5], and have started to focus on deep learning more and more in recent years. Convolutional neural networks (CNNs), which extract knowledge from the input using the convolution operation, are one of the most widely used network architectures in deep learning. The most recent and advanced work in this area is the study described in [6]. The main advantage of this study is the collection of a new dataset from a wireless ECG sensor, and this data covers a lot of arrhythmia types. This research shows that a deep residual network works well on this type of a problem.

Most of the other works use the publicly available arrhythmia database – MIT-BIH. The study in [7] uses a multi-layer CNN, with includes signal denoising in the algorithm. The signal is divided into R-peak centered segments and data augmentation is performed to handle class imbalance. However, only the intra-patient evaluation paradigm [8] is used, which is a great drawback because it does not show how the algorithm would perform in a real-world scenario. In [9], on the other hand, the evaluation bias is considered by using the inter-patient paradigm [8] and a new type of loss function for the neural network is proposed. The deep learning architecture is similar to the one in [6]. The authors experiment with including neighbouring heartbeats in the classification decision, whereas for one heartbeat they consider R-peak centered, fixed-size window.

For more realistic evaluation of the results, other than the inter-patient paradigm, some works focus on patient-specific models [10], [11]. A more recent example of this can be found in [10]. They perform patient-specific arrhythmia detection using 1D CNN and include, in addition to the raw signal, fast Fourier transform (FFT) representation of the ECG in the network. They build a separate model for each patient using 5-minute ECG from each patient for training and a common dataset of hand-picked representative heartbeats. This makes their method a bit dependent on the database and on expert knowledge. All of the methods mentioned, other than the one in [6], which does not segment separate heartbeats, use some kind of fixed-length window for separating the heartbeats, i.e., static segmentation.

The final results of an ECG arrhythmia classification algorithm highly depend on the choice of the evaluation paradigm, and furthermore, on the choice in dividing the ECG signal into separate heartbeats, i.e., the type of segmentation. The aim of this study is to investigate which type of segmentation gives better classification results under the patient specific evaluation paradigm. We focus on single-lead

This work was supported by the Slovenian Research Agency under Grant P2-0095.

¹Elena Merdjanovska is with the Department of Communication Systems, Jožef Stefan Institute, Ljubljana, Slovenia and the Jožef Stefan International Postgraduate School, Ljubljana, Slovenia elena.merdjanovska@ijs.si

²Aleksandra Rashkovska is with the Department of Communication Systems, Jožef Stefan Institute, Ljubljana, Slovenia aleksandra.rashkovska@ijs.si

ECG and propose a novel deep learning architecture for the task of heartbeat classification.

II. MATERIAL AND METHODS

A. Database Description

In this paper, the MIT-BIH Arrhythmia database was used to evaluate and develop the arrhythmia classification model [12]. This database is the most commonly used one in literature when developing ECG signal arrhythmia detection methods and is publicly available on PhysioNet [13]. It contains 48 excerpts from portable ECG recordings, each with a duration of 30 minutes. The recordings were digitized at 360 samples per second per channel with 11-bit resolution over a range of 10 mV. Each record is labeled by two cardiologists independently, resulting in a reference annotation for each heartbeat included in the database, approximately 110,000 annotations in total. Each entry in the MIT-BIH database contains two signals, one of which is always the modified limb lead II, while the second lead is one of the following: V1, V2, V4 or V5. Since the goal is to develop a model that can be applied to a single-lead ECG, this paper only uses one lead (MLII) from each of the records in the database to simulate such an environment.

Additionally to the MIT-BIH database, annotated measurements, obtained with a novel single-lead ECG wireless sensor – Savvy ECG [2], were also investigated. The database contains four 30-minutes long measurements, acquired during different activities, such as laying, sitting, walking. A moderate sampling rate of 125 Hz with 10 bit analogue/digital converter is used as an optimum between medical value and amount of generated data. The sensor was positioned to measure a single-lead ECG that corresponds to the difference in potential between the V1 and V2 electrodes from the standard 12-lead ECG.

B. CNN Architecture

A CNN with residual blocks was used for this classification problem. Residual networks introduce shortcut connections, which enable the network to be deeper by mitigating the problem of vanishing gradient. This type of neural network architecture was chosen because it has been shown to work very well on ECG without preprocessing of the signals. The size and parameters of the network are obtained by combining similar models for this problem used in the literature [14] [6] [9].

The network consists of several convolutional layers grouped into blocks, where each block has 2 layers. Each layer contains $32 * k$ filters, where k in the first convolutional layer has a value of 1 and increases by 1 after every block. In addition, the size of the input vector is reduced by a factor of 2 in each block. All blocks, except the first one, introduce additional shortcut connections and are referred to as residual blocks. A batch normalization layer, in combination with a ReLU activation layer, has been added to each convolutional layer in order to speed up training time. Dropout layers that drop 10% of the activations have been added between the convolutional layers to increase the generalization abilities of

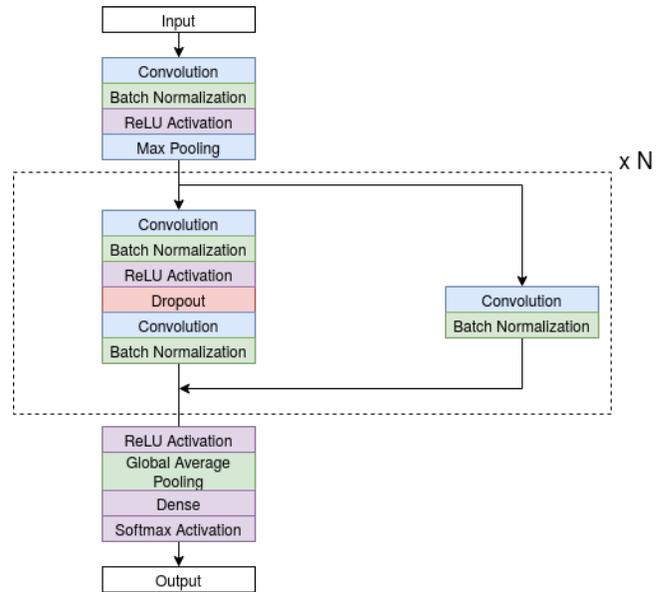


Fig. 1: Architecture of the used convolutional neural network

the model. Finally, there is one fully connected layer of 32 neurons, and one softmax fully connected layer that outputs the probabilities for the input vector to belong to a certain heartbeat class. The architecture of the network is presented in Fig. 1.

Adam optimizer was used as the algorithm for the iterative weight updates, with a learning rate of 0.0001 and batch size of 128. The training was limited to a maximum of 100 epochs, but to choose the best model, early stopping with patience of 8 epochs was used.

III. EXPERIMENTAL DESIGN

A. Data preprocessing

The Association for the Advancement of Medical Instrumentation (AAMI) provides guidelines for grouping heartbeats [15] into 5 classes: N (Nonectopic), S (Supraventricular ectopic beat), V (Ventricular ectopic beat), F (Fusion beat) and Q (Unknown beat). The annotations in the MIT-BIH database, in addition to class information, also contain the R-peak location of each heartbeat. The goal is to have a raw signal corresponding to one heartbeat as an input in the CNN. No signal processing is performed, nor additional attributes for the patient or ECG data are calculated. We experimented with two techniques for dividing the signal into separate beats, which are explained in the following subsections. This is a challenge because all beats are different in length, i.e. they do not last equally long, while, on the other hand, the input in the neural network is fixed in length.

1) *Static segmentation*: After examining the signals in the database and calculating the mean distance between two adjacent R-peaks, a width of 200 samples was chosen to represent a single heartbeat, in which the most important characteristics of the ECG signal would be included. Namely, the input to the CNN is a 200-element vector representing one heartbeat, which includes 100 samples of the original

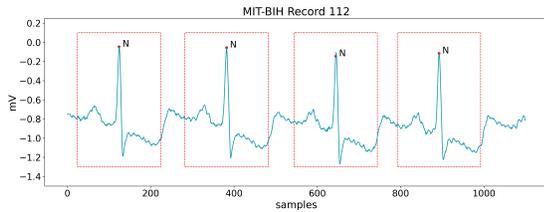


Fig. 2: Example of static segmentation of an ECG signal

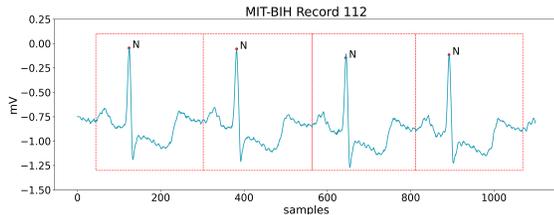


Fig. 3: Example of dynamic segmentation of an ECG signal

signal before the R-peak and 100 samples thereafter, in the case of MIT-BIH recordings. For the Savvy measurements, the same segmentation was applied, modified for the 125 sample rate (0.27 seconds before and after the R-peak) and the obtained heartbeat vector was resampled to 200 samples. Fig. 2 shows a few segmented beats of the signal from patient 112, with the red rectangles representing the part of the signal being considered as one heartbeat.

2) *Dynamic segmentation*: The second type of segmentation we examined was based on the knowledge that a heartbeat starts about 200 ms before the R-peak. In the case of the MIT-BIH database, this corresponds to about 80 samples. According to this, one heartbeat is the part of the ECG signal that starts 80 samples before the annotated R-peak and ends 80 samples before the next annotated R-peak. This way, heartbeats with different duration are obtained, but since the input to the convolutional network must have a fixed length, we have brought all heartbeats to 200 samples by down- or up-sampling. Fig. 3 shows several heartbeats segmented this way, with the first heartbeat lasting longer than the following ones, which necessitates re-sampling.

B. Creating a balanced dataset

Although the MIT-BIH Arrhythmia database contains several types of arrhythmias, even some that are rarely encountered, not all classes are equally represented. The difference in the number of samples from the different classes is not small at all: normal beats make up almost 90% of the data. This is a problem for any classification model, as it will be more difficult to learn to identify the under-represented arrhythmia classes.

In order to achieve a greater balance of classes so that underrepresented arrhythmias become more prominent, instead of including all heartbeat samples from the majority class during training of the classification model, only a fraction of them are included. Of all the heartbeats belonging to the N class, 10% are randomly selected, and the others are removed

during training. The rest of the classes are not downsized. The test set, for which the results are reported, has the original class distribution.

C. Evaluation schemes

The classification of ECG heartbeats can be considered from a few different test settings. i.e. evaluation schemes. From the AAMI-defined evaluation paradigms for ECG analysis, we focus on the patient-specific one, and also investigate the performance according to the intra-patient paradigm.

1) *Intra-patient scheme*: With this scheme, different heartbeats from the same person can be found in both the training set and the test set. All segmented heartbeats from all the patients are viewed as one data set. They are then shuffled and randomly divided into 10 subsets (each containing 10% of the total samples and having a similar distribution of classes as the complete dataset). According to this division, 10 models are trained so that each of the 10 subsets is a test set only once (and the remaining 90% are the training set). This technique is called cross-validation. The intra-patient scheme is only used for comparison with similar models in the literature, but has no application in the real world, because the model will almost always need to successfully classify arrhythmias in people whose data were not represented in the training process at all (inter-patient scheme) or be tailored to one person (patient-specific scheme, addressed in this paper).

2) *Patient-specific scheme*: The second type of model evaluation experiments are at the patient level. This means creating a separate model for each patient, using at most 5 minutes of the patient-specific recording for training and the remaining 25 minutes for testing [16]. The MIT-BIH database is divided into two parts, based on whether the patients have a normal rhythm (records number 100-124) or arrhythmias of some kind are prevalent (records number 200-234). The first group (100-124) is used to create a general set of heartbeats. This set is common, that is, it is included in training of each patient-specific model from the other group (200-234). The patient-specific training sets are constructed from this common dataset together with the first 2.5 minutes of the particular patient. The next 2.5 minutes (between minute 2.5 and minute 5) are kept as a validation set and the parameters for the training of the models are tuned based on the results obtained on this validation set. As mentioned above, the remaining 25 minutes are set aside only for testing of the models. This simulates a real-world application of such a model: annotated ECG recordings lasting several minutes from each user of a wireless sensor would be available and further arrhythmias in the same user would be successfully classified.

IV. RESULTS AND DISCUSSION

The first task in the experiments was selecting the best parameters for each evaluation scheme and for the two types of segmentation. The implementation of these experiments and models was written in the Python programming language (version 3.8), using several libraries including: Keras with

Tensorflow 2 (for neural networks), scikit-learn (for the performance evaluation) and matplotlib (for visualization of results).

For each scenario, the following CNN parameters were chosen:

- intra-patient: (1) static segmentation: 2 blocks, 32 filters in the first layer, 15 filter length; (2) dynamic segmentation: 1 block, 32 filters in the first layer, 15 filter length;
- patient-specific: (1) static segmentation: 1 block, 32 filters in the first layer, 15 filter length; (2) dynamic segmentation: 1 block, 64 filters in the first layer, 15 filter length.

The results show that a generally small network has been selected. This is due to the relatively small size of the dataset, for which many layers are not required to encompass all the data.

For performance evaluation of the classification models, the most commonly used metrics, proposed also by AAMI, are reported: accuracy (*Acc*), sensitivity or true positive rate (*TPR*), specificity or true negative rate (*TNR*), positive predictive value (*PPV*) and F1 score (*F1*). In addition, these metrics are calculated separately for the classes S and V, as these types of arrhythmias are of particular importance and, according to AAMI, it is recommended that they are evaluated also separately. We report the metrics obtained both on the validation and the test sets, generated as described in Section III-C.

Table I shows the performance metrics for the model trained using the intra-patient evaluation scheme (over all 10 cross-validation runs), for both types of segmentation. It can be noted that the dynamic and static segmentation yield very similar results by all metrics, and therefore none of these methods can be evaluated as better performing in the intra-patient scheme. Furthermore, it can be seen that for the class V, the scores are higher than for the S class, which can be explained by its greater representation in the database.

The second evaluation scheme consists of training a specific model for each patient – first for 24 patients from the MIT-BIH database. In doing so, 25 minutes from each patient’s recording are left aside for testing. The results of the classification of the heartbeats from these patient-specific test sets are summarized in Table II. The obtained scores are lower than those in the intra-patient scheme, but as mentioned, this evaluation scheme has the potential to be applied in a real-world scenario. Just like above, we can notice that *TPR* and *PPV* are weaker than the other metrics. But it is precisely in them that we can see a positive improvement with dynamic segmentation on the test set, that is, for the two most significant classes (S and V) they are higher than with static ECG segmentation. The largest difference can be seen in the *PPV* for the S class. This means that the number of S-class beats misclassified as N is much smaller using dynamic segmentation, which is of great importance in the problem of arrhythmia recognition. The other measures do not show such a significant difference for the different segmentation as in the case of *PPV*, but there is still an improvement. Therefore, it can be concluded

that dynamic segmentation is better for classification models of patient-specific arrhythmias.

The results obtained with the proposed CNN, using the intra-patient and patient-specific paradigms, are also compared with results obtained with the best-performing state-of-the-art deep learning architectures for heartbeat classification on the same MIT-BIH Arrhythmia database, using the same evaluation schemes. It can be noted that the results are comparable.

The evaluation results using the patient-specific evaluation scheme on the Savvy recordings are given in Table III. It should be noted that these recordings only contain samples from the N and V classes, which is why they are not directly comparable to the results obtained on the MIT-BIH database. Nevertheless, it can be concluded that the performance of the patient-specific models is generally good, with *Acc* nearly 0.99. This is a very promising result for future expansion of the wireless ECG devices in patient-specific arrhythmia detection.

V. CONCLUSION

In this paper, a model for automatic arrhythmia recognition was proposed by using convolutional neural networks as a classification method. We used the standard ECG arrhythmias database – MIT-BIH, but instead of the full data available, we took only one channel to simulate the use of this model to data from single-lead ECG. Two types of separation of the signal into separate heartbeats were investigated. In doing so, the dynamic segmentation of the ECG signals proved to be a better solution for the classification model described in this paper. This is important because so far in literature, some type of static segmentation is commonly found. Furthermore, in terms of network architecture, although convolutional networks are most successful when they have multiple layers and process large amounts of data, in this case a network of 4 convolutional layers yields the best results. This can be explained by the relatively small amount of data, but in the future, with the increased use and popularity of wireless ECG sensors, larger data volumes will be collected.

The most important part of this paper is the development of patient-specific models, using a common dataset complemented by the first 5 minutes of the recording of the particular patient. This still requires the participation of a cardiologist when annotating the data for those 5 minutes, but further long-term arrhythmia detection has the potential to be automatic, as an accuracy of 0.98 has been achieved on the MIT-BIH Arrhythmia database, and nearly 0.99 on measurements acquired with a novel wireless single-lead ECG body sensor.

REFERENCES

- [1] A. Lyon, A. Mincholé, J. P. Martínez, P. Laguna, and B. Rodriguez, “Computational techniques for ECG analysis and interpretation in light of their contribution to medical advances,” *Journal of The Royal Society Interface*, vol. 15, no. 138, p. 20170821, 01 2018.
- [2] A. Rashkovska, M. Depolli, I. Tomašić, V. Avbelj, and R. Trobec, “Medical-Grade ECG Sensor for Long-Term Monitoring,” *Sensors*, vol. 20, 2020, 1695.

TABLE I: Evaluation metrics using the intra-patient evaluation scheme

		All					S					V						
		ACC	TNR	PPV	TPR	F1	ACC	TNR	PPV	TPR	F1	support	ACC	TNR	PPV	TPR	F1	support
val	static	0.9934	0.9928	0.8981	0.9409	0.9184	0.9912	0.9939	0.7862	0.8832	0.8319	3656	0.9951	0.9961	0.9487	0.9819	0.9651	10082
	dynamic	0.9912	0.9908	0.8784	0.9367	0.9057	0.9895	0.9927	0.764	0.8701	0.8136	3911	0.9914	0.9925	0.9019	0.9758	0.9374	9781
test	static	0.9934	0.9928	0.8948	0.939	0.9158	0.991	0.9938	0.7872	0.886	0.8337	2781	0.9949	0.996	0.9457	0.9794	0.9622	7235
	dynamic	0.9906	0.9898	0.8748	0.9221	0.8971	0.9882	0.9918	0.7302	0.8515	0.7862	2781	0.9908	0.9921	0.8971	0.9717	0.9329	7235
test	[9]	0.9979	0.9936	0.9771	0.9465	0.9613	0.9968	0.9993	0.9727	0.9083	0.9394	2781	0.9986	0.9992	0.9893	0.9906	0.9899	7007

TABLE II: Evaluation metrics using the patient-specific evaluation scheme

		All					S					V						
		ACC	TNR	PPV	TPR	F1	ACC	TNR	PPV	TPR	F1	support	ACC	TNR	PPV	TPR	F1	support
val	static	0.9889	0.9803	0.7253	0.7117	0.7168	0.9918	0.9953	0.8258	0.8583	0.8417	127	0.9845	0.9882	0.8849	0.9453	0.9141	439
	dynamic	0.9893	0.9801	0.7278	0.709	0.7176	0.9924	0.9971	0.8803	0.811	0.8443	127	0.9851	0.9891	0.8922	0.9431	0.9169	439
test	static	0.9652	0.9476	0.6057	0.6145	0.6067	0.9569	0.9815	0.5516	0.4611	0.5023	2340	0.9561	0.9608	0.714	0.9124	0.8011	4808
	dynamic	0.9766	0.9507	0.6746	0.6067	0.6269	0.9689	0.9959	0.8357	0.4239	0.5625	2340	0.9737	0.9813	0.838	0.9027	0.8691	4808
test	[10]	0.9514	0.9694	0.6573	0.665	0.6610	0.964	0.981	0.621	0.646	0.633	2340	0.986	0.9868	0.895	0.95	0.92	4808

TABLE III: Evaluation metrics using the patient-specific evaluation scheme on the Savvy recordings

		All				
		ACC	TNR	PPV	TPR	F1
val	static	0.9946	0.8979	0.785	0.898	0.832
	dynamic	0.9837	0.6938	0.5888	0.6939	0.6209
test	static	0.9887	0.5973	0.7163	0.5973	0.6335
	dynamic	0.9818	0.6459	0.6124	0.6458	0.6268

[3] K. Minami, H. Nakajima, and T. Toyoshima, "Real-time discrimination of ventricular tachyarrhythmia with fourier-transform neural network," *IEEE Transactions on Biomedical Engineering*, vol. 46, no. 2, pp. 179–185, 1999.

[4] O. T. Inan, L. Giovangrandi, and G. T. A. Kovacs, "Robust neural-network-based classification of premature ventricular contractions using wavelet transform and timing interval features," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 12, pp. 2507–2515, 2006.

[5] V. Krasteva and I. Jekova, "Qrs template matching for recognition of ventricular ectopic beats," *Annals of biomedical engineering*, vol. 35, pp. 2065–76, 01 2008.

[6] A. Hannun, P. Rajpurkar, M. Haghpanahi, G. Tison, C. Bourn, M. Turakhia, and A. Ng, "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Medicine*, vol. 25, 2019.

[7] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, M. Adam, A. Gertych, and R. S. Tan, "A deep convolutional neural network model to classify heartbeats," *Computers in Biology and Medicine*, vol. 89, pp. 389–396, 2017. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0010482517302810>

[8] E. Luz, W. Schwartz, G. Chávez, and D. Menotti, "ECG-based heartbeat classification for arrhythmia detection: A survey," *Computer Methods and Programs in Biomedicine*, vol. 127, pp. 144–164, 12 2015.

[9] A. Sellami and H. Hwang, "A robust deep convolutional neural network with batch-weighted loss for heartbeat classification," *Expert Systems with Applications*, vol. 122, pp. 75–84, 2019.

[10] S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 3, pp. 664–675, 2016.

[11] S. Raj and K. Ray, "A personalized arrhythmia monitoring platform," *Scientific Reports*, vol. 8, p. 11395, 12 2018.

[12] G. Moody and R. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 06 2001.

[13] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C.-K. Peng, and H. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet : Components of a New Research Resource for Complex Physiologic Signals," *Circulation*, vol. 101, no. 23, pp. E215–20, 07 2000.

[14] K. Weimann and T. O. F. Conrad, "Transfer learning for ECG classification," *Scientific Reports*, vol. 7, p. 6067, 03 2021.

[15] "Testing and reporting performance results of cardiac rhythm and ST segment measurement algorithms," Association for the Advancement of Medical Instrumentation, Standard, 2012.

[16] "Recommended practice for testing and reporting performance results of ventricular arrhythmia detection algorithm," Association for the Advancement of Medical Instrumentation, Standard, 1987.