2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

On the impact of ECG data quality for arrhythmia detection using convolutional neural networks and wearable devices

1st Juan Manuel Sancho
sanchojuanma@gmail.com

2nd Jônata Tyska Carvalho
Federal University of Santa Catarina (UFSC)
jonata.tyska@ufsc.br

***Abstract*—Cardiovascular diseases are the leading cause of death in the world, with arrhythmias being a significant symptom and risk factor. Advancements in technologies such as low-cost and low-power wearable devices, and machine learning techniques for analyzing big volumes of data offer opportunities to address this issue. However, low-cost devices may have limitations, including reduced data quality due to lower sampling rates, bit depth, and the number of leads recorded. These limitations might produce a significant decrease in machine learning models' performance in detecting arrhythmias. This study investigates the impact of data quality reduction on arrhythmia classification using deep neural network models. High-quality ECG data with 12 leads, 500Hz sampling rate, and 32-bits resolution were transformed into low-quality versions with varying leads (from one to six), 100Hz sampling rate, and 8-bits resolution. Training a state-of-the-art deep learning arrhythmia detection model on both high-quality and low-quality datasets revealed a decrease in performance from 95.3% to 93.9% in the worst case, which is concerning given the critical nature of the domain. To mitigate this performance loss, we propose an ensembling method that compensates for 42% of the loss, achieving an accuracy of 94.5% even with the low-quality dataset. The analysis also identifies the leads with the most promising classification performances. These results can aid in making better design decisions when creating cost-effective wearable ECG devices.**

***Index Terms*—E-Health, Deep Learning, Remote Sensing**

I. Introduction

According to the World Health Organization, cardiovascular diseases are the leading cause of death worldwide. Currently, these diseases account for a 16% of deaths globally [19]. Within the spectrum of possible cardiac issues, arrhythmias contribute as a risk factor [12]. The most common method for analyzing the heart's electrical and rhythmic patterns, in order to detect possible anomalies, is through an Electrocardiogram (ECG). A traditional ECG has 12 cardiac leads which measure the heart's activity from different planes by placing several electrodes on the chest and limbs and measuring the electrical current between them [4].

The traditional ECG exam, performed and analyzed by a physician in a medical facility, does not scale when the objective is to monitor entire populations of people at risk. In this context, telemedicine and wearable devices can aid in monitoring cardiac patients, allowing for a more extensive medical follow-up in a less invasive way, and even allowing better health services access for people who live far away from health facilities [18]. Wearable devices for these patients have the potential to improve their diagnostic and follow-up [10], especially when combining them with techniques for automatic anomaly detection. Since telemedicine devices are designed to report data over more extended periods of time than a standard screening, it becomes unfeasible for physicians to manually analyze a patient's complete long-form records, which could span days or weeks of monitoring. This makes the deployment of automatic detection techniques important for such devices.

Regarding automatic arrhythmia detection, multiple articles have been published about utilizing Machine Learning techniques to work on 12-lead, 2-lead and even 1-lead ECG signals [1], [11], [13], [20], [21]. As we can see, there is a gap of studies investigating the effectiveness of using a number of leads in between two and twelve, which is important when we think about the design and development of wearable devices.

The efficiency, accuracy and complexity of machine learning models depend highly on the quality of the training data utilized [8]. On the other hand, wearable devices are usually designed considering their cost, portability, energy consumption, and data transfer and storage costs [9], [17], which leads to lower-quality signals being captured by them. From this conflict emerges the need to investigate the concrete impact of data quality reduction on machine learning models' performance, as well as ways to compensate for it. This balance of cost and performance is dramatically important in applications related to medicine, where even small performance gains result in potentially more patients receiving adequate care and treatment, which in practice may determine life or death outcomes.

State-of-the-art arrhythmia classification models usually work on high-quality ECG data and achieve maximum performance using only one or two leads of ECG data [2], [14], [15], [20]. However, in limited data quality scenarios, such as when building and using low-cost wearable devices for arrhythmia monitoring, measuring and using different leads for classification can be a promising direction. Given that every individual lead provides different information, it is also important to analyze the models' performances on each lead separately and their accuracy when detecting each specific arrhythmia class. This provides useful insight such as considering which leads to include when designing a new device, or how to best combine several available leads in order to increase classification performance.

This work presents a detailed analysis of the degradation of classification performance when using lower-quality signals, similar to those provided by wearable devices. The goal of this study is to better understand the necessary parameters and limitations for building more efficient monitoring devices. In order to achieve this, a state-of-the-art classifier based on the Convolutional Neural Network (CNN) architecture presented in [21] was trained to detect and classify cardiac arrhythmias of up to 4 classes. The dataset published by Zheng et al. [22] was used for training the model. The dataset contains over 10,000 12-lead ECG records of cardiac patients recorded at a 500 Hz sampling rate and a 32-bit resolution. Given that the intended use for this model is to utilize the data provided by wearable ECG devices, which usually do not provide the same data as a full traditional ECG recording, the model was also trained on downsampled and quantized signals to better approximate the data provided by such devices. As a reference device, the Galeno Sys [7] device as used, a 6-lead (leads I, II, III, aVR, aVF and aVL) wearable ECG device, which measures the heart's activity with a frequency of 100 Hz and a resolution of 8-bits. The results obtained from this experiment, including complete analysis on a per-lead and per-class basis, were then used to build a combinatory model of the 6 leads in order to reach a better classification efficiency. The results of our analysis show that the indeed data quality downgrade creates a significant impact on the model's accuracy, which decreases from 95.3% to 93.9% on the best-performing individual lead. However, by training a new classification model using the 6 downsampled leads, the loss of accuracy can be reduced by 42%, taking the detection accuracy on the low-quality data back up to 94.5%.

II. Methodology

The methodology used in this work has been structured as follows: first, the architecture and results of Yildirim et al. [21] were reproduced on single leads at the dataset's original 500Hz frequency and 32-bit resolution. Afterwards, since the objective of this work is mainly to study the potential performance loss with lower-quality samples, the original signals were downsampled and quantized and the same architecture was utilized in order to compare performances. The per-lead and per-class performances of both versions of the model were studied as well in order to determine if particular leads provide better results, or if any particular arrhythmia classes are harder to identify. Afterwards, in order to check whether the performance loss can be compensated, a multi-lead combinatory model was proposed and tested on the downgraded signals.

TABLE I
Reference architecture

Layer Type	Parameters	Output Shape
Conv1D	Filters=64, Size=21, Strides= 11	453 × 64
MaxPooling1D	Pool size=2	226 × 64
Batch Norm	-	226 × 64
LeakyReLU	Alpha=0.1	226 × 64
Dropout	Rate=0.3	226 × 64
Conv1D	Filters=64, Size=7, Strides= 1	220 × 64
MaxPooling1D	Pool size=2	110 × 64
Batch Norm	-	110 × 64
Conv1D	Filters=128, Size=5, Strides= 1	106 × 128
MaxPooling1D	Pool size=2	53 × 128
Conv1D	Filters=256, Size=13, Strides= 1	41 × 256
Conv1D	Filters=512, Size=7, Strides= 1	35 × 512
Dropout	Rate=0.3	35 × 512
Conv1D	Filters=256, Size=9, Strides= 1	27 × 256
MaxPooling1D	Pool size=2	13 × 256
LSTM	Unit=128, Return Sequences=True	13 × 128
Flatten	-	1664
Dense	Units=64, Activation=ReLU	64
Dense	Units=4, Activation=Softmax	4

A. Single original leads

Original signals were used to train the model presented by Yildirim et al. [21]. Table I presents the architecture as originally proposed by the authors.

Each lead's model (leads I, II, III, aVR, aVF and aVL) was trained 10 times in order to obtain average performances and reduce the impact of stochasticity in the results, varying the random split of train and validation sets (80% and 20% of the previously split training set, respectively). The resulting models were further evaluated on the separate test set.

All the models were trained for 25 epochs using a batch size of 64 and Adam optimizer with a learning rate of 0.001. The loss function used was categorical cross-entropy.

Since there appears to be some small overfitting on 25 epochs of training as pointed out by Yildirim et al. [21], model checkpointing was set up in order to also preserve the model on the epoch that performs best on the validation set. Performance was evaluated on these models.

B. Single downsampled leads

After building and testing the classification model with high-quality data, the original signals were downsampled to 100 Hz and quantized to 8-bits, repeating the training of the original leads. The downsampling was performed through Fast Fourier Transform [5]. The same architecture was utilized, only modifying the strides of the first convolutional layer in order to preserve the shape of the feature maps propagated into the deeper layers, since the input dimensions are altered during downsampling. The performance of each lead and its downgraded counterpart were then compared to measure the impact of this modification.

On both cases (original and downsampled single leads), performance was also evaluated on a per-lead and per-class basis, in order to determine if specific leads classify different arrhythmias better than others, or if any lead in particular is generally better at classification.

C. Combined model

After defining the best performances on each lead, a multi-lead combinatory model (which combines the input of all 6 leads) was developed in order to determine if predictive performance can be improved. This model utilize the down-sampled and quantized versions of the data, and is built by

TABLE II
RHYTHM TYPES AND DISTRIBUTION IN THE DATASET

Acronym Name	Full Name	Frequency (%)
SB	Sinus Bradycardia	3,889 (36.53)
SR	Sinus Rhythm	1,826 (17.15)
AFIB	Atrial Fibrillation	1,780 (16.72)
ST	Sinus Tachycardia	1,568 (14.73)
AF	Atrial Flutte	445 (4.18)
SI	Sinus Irregularity	399 (3.75)
SVT	Supraventricular Tachycardia	587 (5.51)
AT	Atrial Tachycardia	121 (1.14)
AVNRT	Atrioventricular Node Reentrant Tachycardia	16 (0.15)
AVRT	Atrioventricular Reentrant Tachycardia	8 (0.07)
SAAWR	Sinus Atrium to Atrial Wandering Rhythm	7 (0.07)
All	All	10,646 (100)

removing the final dense and output layer from each lead's best models, and combining them with a new fully connected and output layer. This new model was re-trained on the training data, only updating the weights of the fully connected layers (leaving the original models intact, preserving their previously trained feature extraction and LSTM layers). Ten iterations were trained in order to reduce stochasticity and obtain an average performance. This training was performed over 10 epochs with a learning rate of 0.0001 as we saw that the combined model overfitted earlier than the original model.

D. Model evaluation

The comparison between models and leads was done through several metrics in terms of their predictive performance.

In addition to the general metrics of each model, performance on individual leads was analyzed in order to determine if there is any difference in their classification capabilities. Per-class performance is also presented to see if any particular class is more difficult to classify.

Finally, since the aim of this work is medical diagnosis, the kind of errors made by the classification process should also be considered. In medical diagnosis, there are two types of errors: Type I, where a patient is diagnosed with a disease and is healthy, and Type II, where a patient has a disease but is diagnosed as healthy. While both account for a misdiagnosis, Type II errors are widely considered more harmful, as the patient might miss the opportunity for a cure or treatment, endangering their health [6].

In this regard, a recall measure will be created in order to compare the correct classification as arrhythmia against normal rhythm. For this case, True Positives are the examples correctly labeled as any arrhythmia kind, while False Negatives are all the arrhythmia examples identified as normal. This measure will be higher as less examples are wrongly classified as "healthy", reaching 1 when no arrhythmia sample is classified as a normal rhythm.

E. Data

This work uses a dataset created under the auspices of Chapman University and Shaoxing People's Hospital [22]. It is composed of 10 second measurements of 12-lead ECG records of 10,646 patients. Each segment was manually labeled by a licensed physician in order to identify the presence of any arrhythmias or other cardiac issues, and then validated by a second physician.

The dataset consists of 5,956 male patients and 4,690 female patients; among them, 17% present a normal cardiac rhythm (Sinus Rhythm), while the rest possesses at least one anomaly. Among these anomalies, 12 rhythms are recognized (1 normal and 11 anomalous). The distribution of these rhythms is presented in Table II.

Each record contains 10 seconds of every ECG lead, sampled at a rate of 500 Hz and a resolution of 32 bits. There are two versions of each record: one is the original recording with noise, while the other is a version de-noised through a Butterworth low pass filter, LOESS curve fitting and non local means [22]; the latter is utilized in this work.

Since the reference point of this work the ECG data provided by the Galeno Sys device, which has a sampling rate of 100 Hz, the dataset needs to be downsampled from its original 500 Hz rate. While some there are works indicating that sampling rates as low as 62.5 Hz do not impact the ability to perform automatic ECG delineation [16], there is evidence supporting the fact that accuracy of such algorithms is unsatisfactory beyond 120 Hz [3]. Furthermore, the device has a resolution of 8 bits, which Ajdaraga and Gusev [3] conclude is insufficient, at least for some QRS detection algorithms. Since the device works at 100 Hz and 8-bits, these variables will be tested to measure their impact on performance.

As previously seen on Table II, some arrhythmia classes have very few examples, which could impact the model's ability to generalize their classification. For this reason, and as suggested by Zheng et al. [22], the classes are merged into 4 superclasses (AFIB,GSVT,SB,SR), following the recommendations of cardiologists. An additional 58 records are discarded because they are either empty or contain incomplete data. A test set of 20% of the records is separated in order to measure the models' ability to generalize outside of the training set. This test set is stratified in order to preserve the proportion of classes of the entire dataset. Table III presents the fused classes and their distribution.

III. RESULTS

A. Impact of data quality reduction on classification performance

Before detailing the per-lead and per-class performances on the original and downgraded signals, it is important to first address the general performance loss across all leads. An overview of the average performance across all leads on the original and downsampled signals can be seen on Table IV. As seen in these results, there is a nonnegligible performance loss

TABLE III
RESULTING CLASSES AFTER MERGING AND SPLITTING TRAIN-TEST SETS

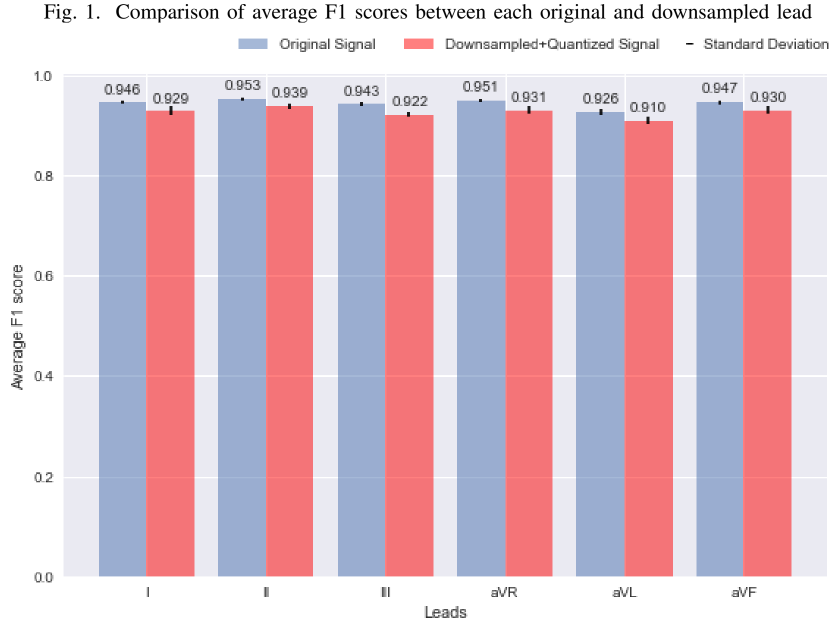
Original Classes	Merged Class	Total Examples (%)	Training set (%)	Test set (%)
AFIB, AF	AFIB	2218 (20.94)	1774 (20.94)	444 (20.94)
SVT, AT, SAAWR, ST, AVNRT, AVRT	GSVT	2260 (21.34)	1808 (21.34)	452 (21.34)
SB	SB	3888 (36.72)	3110 (36.72)	778 (36.72)
SR, SI	SR	2222 (20.98)	1778 (20.98)	444 (20.98)
All	All	10588 (100)	8470 (100)	2118 (100)

TABLE IV
OVERVIEW OF PERFORMANCE LOSS ON DOWNSAMPLED LEADS

Model	Accuracy	Validation Accuracy	Test Accuracy	F1 Score (Test)
Original Signals	0.965 (0.008)	0.943 (0.011)	0.944 (0.01)	0.944 (0.01)
Downsampled Signals	0.941 (0.018)	0.923 (0.013)	0.927 (0.01)	0.927 (0.01)
Performance Loss	2.4%	2.0%	1.7%	1.7%

Note: Standard deviation in parentheses.

Fig. 1. Comparison of average F1 scores between each original and downsampled lead



when training the model on the 100 Hz and 8-bit signals that ranges from 1.7% to 2.4% depending on the metric analyzed. Further details can be found in the following subsections.

B. Single leads

1) Original signals: Each individual lead was trained 10 times on the original data at 500 Hz and 32-bit resolution with varying train/validation splits. Table V presents the overall average metrics.

As seen on Table V, every lead provides a good performance of over 0.92 F1 score, with lead aVL performing the worst.

Class Sinus Bradycardia (SB) performs the best in all leads, followed by Sinus Rhythm (SR), General Supraventricular Tachycardia (GSVT) and finally Atrial Fibrillation (AFIB). Lead II presents the best F1 score for AFIB, GSVT and SB, while lead aVR has the best score for SR. This indeed confirms that the combination of leads for improving classification is promising. Finally, the average Arrhythmia Recall metric is presented on Table VII.

TABLE V
AVERAGE MODEL RESULTS ON SINGLE ORIGINAL LEADS

Lead	Accuracy	Validation Accuracy	Test Accuracy	F1 Score (Test)
Lead I	0.96 (0.0108)	0.943 (0.0026)	0.947 (0.0033)	0.946 (0.0033)
Lead II	**0.975** (0.007)	0.951 (0.003)	**0.953** (0.0019)	**0.953** (0.0019)
Lead III	0.967 (0.0081)	0.943 (0.0038)	0.943 (0.004)	0.943 (0.004)
Lead aVR	0.971 (0.0067)	**0.953** (0.0033)	0.95 (0.0032)	0.951 (0.0032)
Lead aVL	**0.953** (0.0161)	0.923 (0.0062)	0.926 (0.0056)	0.926 (0.0058)
Lead aVF	0.962 (0.0067)	0.943 (0.0065)	0.947 (0.0045)	0.947 (0.0047)

Note: Standard deviation in parentheses.

TABLE VI
AVERAGE CLASS F1 SCORES ACROSS EACH ORIGINAL LEAD

Class	Lead I	Lead II	Lead III	Lead aVR	Lead aVL	Lead aVF
AFIB	0.902 (0.0094)	**0.92** (0.0037)	0.898 (0.0087)	0.911 (0.0065)	0.853 (0.0172)	0.903 (0.013)
GSVT	0.919 (0.0062)	**0.926** (0.0052)	0.918 (0.0062)	0.917 (0.0056)	0.906 (0.0058)	0.921 (0.0073)
SB	0.983 (0.0021)	**0.986** (0.0016)	0.984 (0.0028)	0.985 (0.0025)	0.978 (0.0027)	0.983 (0.0026)
SR	0.954 (0.0033)	0.957 (0.0028)	0.942 (0.0067)	**0.964** (0.0028)	0.928 (0.0061)	0.953 (0.0047)

Note: Standard deviation in parentheses.

TABLE VII
AVERAGE ARRHYTHMIA RECALL ACROSS ORIGINAL LEADS

Lead	Arrhythmia Recall
I	0.989 (0.0017)
II	0.991 (0.0022)
III	0.989 (0.0032)
aVR	**0.994** (0.0023)
aVL	0.984 (0.0035)
aVF	0.99 (0.0036)

Note: Standard deviation in parentheses.

Lead aVR presents the smallest number of arrhythmias wrongly classified as a normal rhythm.

2) Downsampled signals: Next, new models were trained with data downsampled to 100 Hz and quantized to an 8-bit resolution. Each lead's models were once again trained 10 times to obtain an average performance. Table VIII presents the overall average metrics of the downsampled models, while Figure 1 presents the comparison of average F1 scores between the original and the downsampled signals.

As seen on Figure 1, every lead has worse performance when trained on the downsampled data. The average F1 performance loss across all leads is 1.74%.

Per-class average F1 scores are presented on Table IX. Finally, downsampled leads' average Arrhythmia Recall is presented on Table X. Overall, these results confirm that the models trained on the downsampled data perform considerably worse than those trained on the original signals.

TABLE X
AVERAGE ARRHYTHMIA RECALL ACROSS DOWNSAMPLED LEADS

Lead	Arrhythmia Recall
I	0.983 (0.0071)
II	**0.989** (0.0061)
III	0.982 (0.0066)
aVR	0.988 (0.0053)
aVL	0.979 (0.0031)
aVF	0.986 (0.0044)

Note: Standard deviation in parentheses.

C. Combined model

The combined model was created using the best models for each lead. It was trained 10 times to obtain average

TABLE VIII
AVERAGE MODEL RESULTS ON SINGLE DOWNSAMPLED LEADS

Lead	Accuracy	Validation Accuracy	Test Accuracy	F1 Score (Test)
Lead I	0.927 (0.0152)	0.926 (0.0074)	0.929 (0.0083)	0.929 (0.0086)
Lead II	0.953 (0.0099)	**0.94** (0.0051)	**0.939** (0.0046)	**0.939** (0.0047)
Lead III	0.938 (0.0189)	0.911 (0.0055)	0.922 (0.0046)	0.922 (0.0045)
Lead aVR	**0.958** (0.0182)	0.93 (0.0077)	0.931 (0.0064)	0.931 (0.0063)
Lead aVL	0.915 (0.0149)	0.904 (0.0037)	0.911 (0.006)	0.91 (0.0064)
Lead aVF	0.956 (0.0137)	0.927 (0.0046)	0.93 (0.0073)	0.93 (0.0073)

Note: Standard deviation in parentheses.

TABLE IX
AVERAGE CLASS F1 SCORES ACROSS EACH DOWNSAMPLE LEAD

Class	Lead I	Lead II	Lead III	Lead aVR	Lead aVL	Lead aVF
AFIB	0.855 (0.0233)	**0.883** (0.0119)	0.844 (0.0094)	0.868 (0.0145)	0.812 (0.0142)	0.864 (0.0177)
GSVT	0.91 (0.0065)	**0.914** (0.0087)	0.901 (0.0096)	0.903 (0.0119)	0.898 (0.006)	0.905 (0.0097)
SB	0.98 (0.003)	**0.981** (0.0024)	0.978 (0.0027)	0.979 (0.0036)	0.972 (0.0064)	0.978 (0.0025)
SR	0.932 (0.0112)	**0.945** (0.008)	0.924 (0.0086)	0.939 (0.0082)	0.913 (0.0086)	0.936 (0.0073)

Note: Standard deviation in parentheses.

performances. The results can be seen on Table XI.

TABLE XI
AVERAGE ACCURACY AND F1 SCORES OF THE COMBINED MODEL

Model	Accuracy (Test set)	F1 Score
Combined Model	0.946 (0.0052)	0.9456 (0.0054)
Downsampled Signals	0.926 (0.0053)	0.926 (0.0063)
Original Signals	0.944 (0.0037)	0.94 (0.0038)

Note: Standard deviation in parentheses.

This new model performs better than the best average performance on a single downsampled lead (Lead II, with an F1 of 0.939). Per-class average F1 scores of this model are presented on Table XII.

TABLE XII
PER CLASS AVERAGE F1 SCORES FOR THE COMBINED MODEL

Class	Combined Model	Downsampled Signals	Original Signals
AFIB	0.8983 (0.009)	0.854 (0.0015)	0.898 (0.0075)
GSVT	0.9268 (0.0055)	0.905 (0.0078)	0.918 (0.0061)
SB	0.9823 (0.0029)	0.978 (0.003)	0.983 (0.0024)
SR	0.9436 (0.007)	0.932 (0.0075)	0.95 (0.0044)

Note: Standard deviation in parentheses.

Finally, the average Arrhythmia Recall metric for the combined model is 0.990 (with a standard deviation of 0.0026).

IV. DISCUSSION AND CONCLUSIONS

In this work, we present a study investigating the impact of performing arrhythmia classification on lower-quality data similar to that provided by wearable devices. We presented to what extent the performance of state-of-the-art arrhythmia's classification models tested with lower-quality signal degrades. We present detailed results on how the degradation happens according to each class and ECG-lead used. We also proposed methods which attenuate the impact on data quality reduction. Using as a reference the ECG measurement parameters of the Galeno Sys [7] device (100 Hz sampling rate and 8-bits resolution), a CNN model was trained on a public dataset with over 10,000 ECG records in order to detect normal rhythms and 3 arrhythmia classes.

First of all, the results obtained confirmed that there is indeed a nonnegligible classification performance loss when utilizing the lower-quality signals, diminishing the maximum classification accuracy from 95.3% to 93.9%. In this scenario, it is no longer enough to only consider the classification provided by a single lead's signal in order to diagnose patients. It is noteworthy that while these differences in performance may seem small, given the application to the medical diagnosis field this difference could result in patients not receiving adequate care.

Secondly, it was observed that different leads provide different performances, in some cases with varying classification accuracies on different arrhythmia classes. This knowledge can help build better models by taking into consideration each lead's strengths and weaknesses.

Thirdly, by considering the previously mentioned results, leads can be combined in a way that allows to regain some of the performance lost due to the lower quality signals. It was shown that by training a new model combining all 6 leads, classification accuracy can be improved up to 94.5%, attenuating 42% of the performance loss when using lower-quality signals. Just to illustrate the impact of this result, in absolute numbers, the amount of patients correctly classified by this gain in comparison to the original network would represent 600 per 100.000 cardiac patients evaluated.

As future work, the techniques presented in this work could be tested with different datasets in order to determine if certain leads are universally better for classification on average, or if this difference is dataset and device-specific. Another possible direction for further research is to use patient's clinical information, such as gender and age, in order to improve the predictive capabilities of these models.

The findings discussed here can help low-cost, wearable ECG device designers decide which leads to track in order to improve performance of arrhythmia classification, whether it involves measuring single leads or choosing a specific combination for optimal performance.

ACKNOWLEDGEMENT

This work was partially funded by Ministério Público de Santa Catarina (MPSC) through the research Project Ceos: Data Intelligence for Society.

REFERENCES

[1] Acharya, U.R., Fujita, H., Oh, S.L., Raghavendra, U., Tan, J.H., Adam, M., Gertych, A., Hagiwara, Y.: Automated identification of shockable and non-shockable life-threatening ventricular arrhythmias using convolutional neural network. Future Generation Computer Systems **79**, 952–959 (2018)

[2] Acharya, U.R., Oh, S.L., Hagiwara, Y., Tan, J.H., Adam, M., Gertych, A., San Tan, R.: A deep convolutional neural network model to classify heartbeats. Computers in biology and medicine **89**, 389–396 (2017)

[3] Ajdaraga, E., Gusev, M.: Analysis of sampling frequency and resolution in ecg signals. In: 2017 25th Telecommunication Forum (TELFOR). pp. 1–4. IEEE (2017)

[4] Conover, M.B.: Understanding electrocardiography. Elsevier Health Sciences (2002)

[5] Cooley, J.W., Lewis, P.A., Welch, P.D.: The fast fourier transform and its applications. IEEE Transactions on Education **12**(1), 27–34 (1969)

[6] Cummins, R.O., Hazinski, M.F.: Guidelines based on fear of type ii (false-negative) errors: why we dropped the pulse check for lay rescuers. Circulation **102**(suppl_1), I–377 (2000)

[7] DataFlow: Galeno sys (2019), `https://galenosys.com.uy/`

[8] Gupta, N., Mujumdar, S., Patel, H., Masuda, S., Panwar, N., Bandyopadhyay, S., Mehta, S., Guttula, S., Afzal, S., Sharma Mittal, R., et al.: Data quality for machine learning tasks. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. pp. 4040–4041 (2021)

[9] Gupta, R., Gamad, R., Bansod, P.: Telemedicine: A brief analysis. Cogent Engineering **1**(1), 966459 (2014)

[10] Gusev, M., Stojmenski, A., Guseva, A.: Ecgalert: A heart attack alerting system. In: International Conference on ICT Innovations. pp. 27–36. Springer (2017)

[11] Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., Tison, G.H., Bourn, C., Turakhia, M.P., Ng, A.Y.: Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nature medicine **25**(1), 65–69 (2019)

[12] Khairy, P., Dore, A., Talajic, M., Dubuc, M., Poirier, N., Roy, D., Mercier, L.A.: Arrhythmias in adult congenital heart disease. Expert review of cardiovascular therapy **4**(1), 83–95 (2006)

[13] Li, Y., Pang, Y., Wang, J., Li, X.: Patient-specific ecg classification by deeper cnn from generic to dedicated. Neurocomputing **314**, 336–346 (2018)

[14] Li, Z., Zhou, D., Wan, L., Li, J., Mou, W.: Heartbeat classification using deep residual convolutional neural network from 2-lead electrocardiogram. Journal of electrocardiology **58**, 105–112 (2020)

[15] Mousavi, S., Fotoohinasab, A., Afghah, F.: Single-modal and multi-modal false arrhythmia alarm reduction using attention-based convolutional and recurrent neural networks. PloS one **15**(1), e0226990 (2020)

[16] Simon, F., Martinez, J.P., Laguna, P., van Grinsven, B., Rutten, C., Houben, R.: Impact of sampling rate reduction on automatic ecg delineation. In: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 2587–2590. IEEE (2007)

[17] Sodhro, A.H., Sangaiah, A.K., Sodhro, G.H., Lohano, S., Pirbhulal, S.: An energy-efficient algorithm for wearable electrocardiogram signal processing in ubiquitous healthcare applications. Sensors **18**(3), 923 (2018)

[18] Wootton, R., Craig, J., Patterson, V.: Introduction to telemedicine. CRC Press (2017)

[19] World Health Organization: Who reveals leading causes of death and disability worldwide: 2000-2019 (2020), `https://www.who.int`

[20] Yıldırım, Ö., Pławiak, P., Tan, R.S., Acharya, U.R.: Arrhythmia detection using deep convolutional neural network with long duration ecg signals. Computers in biology and medicine **102**, 411–420 (2018)

[21] Yildirim, O., Talo, M., Ciaccio, E.J., San Tan, R., Acharya, U.R.: Accurate deep neural network model to detect cardiac arrhythmia on more than 10,000 individual subject ecg records. Computer methods and programs in biomedicine **197**, 105740 (2020)

[22] Zheng, J., Zhang, J., Danioko, S., Yao, H., Guo, H., Rakovski, C.: A 12-lead electrocardiogram database for arrhythmia research covering more than 10,000 patients. Scientific Data **7**(1), 1–8 (2020)