Transfer Learning for tuberculosis screening by single-channel ECG

V.N. Guryanova[1]

[1]Lomonosov Moscow State University, Leninskie Gory 1, Moscow, Russia, 119991

Abstract. Tuberculosis is one of the leading causes of death in the world. The majority of the population is not able to regularly conduct specific examinations, such as x-ray examinations, for the presence of tuberculosis. Currently, there are mobile devices for measuring ECG, which allow you to take measurements without leaving your home. This article explores the possibility of determining tuberculosis based on a single-channel mobile ECG. One of the general top-performance neural networks is used as a classifier. This article also explored the possibility of such a classification based not on raw data, but the generated image. The image allows you to interpret the prediction of the neural network and makes it possible for the doctor to understand the model's decision better. The article shows the promising quality and provides proof of concept of such screening. Different ratios of precision and recall are provided, which can be adjusted depending on the situation.

1. Introduction.

Tuberculosis (TB) is caused by bacteria (Mycobacterium tuberculosis) that most often affect the lungs. Tuberculosis is curable and preventable. TB is easily spread from person to person through the air. Tuberculosis (TB) is one of the 10 leading causes of death in the world [1]. It is very important to determine tuberculosis in time to prevent its spread and begin treatment of an ill person. It takes a long time for both the physician and the patient to carry out specific tests for TB (eg, such as X-rays or Xpert MTB/RIF) and as such, they cannot be conducted with relatively high regularity. So, it would be useful to have some pre-screening, which could be done quite often and which would allow identifying people with a high chance of having tuberculosis and conduct specific tests only for them.

ECG is a signal that displays the electronic activity of the heart. Each ECG recording shows the potential difference between two electrodes located on the surface of the body. Currently, on the market, there are mobile devices that are capable of reading an electrocardiogram (ECG) of a person, for example, AliveCor [2], CardioQvark [3]. Such devices read only one of the 12 leads. Such devices can be used as often as required. You can have just a few devices per institution or organization to conduct such pre-screenings every day. People who have a high chance of having a disease will need to see a TB specialist as soon as possible.

The motivation for this study was the work that shows the relationship between cardiological and tuberculosis diseases [4]. Besides, often the tuberculosis bacteria themselves affect the heart and thus affect its electrical activity. At the moment, this is the first work that is devoted to the problem of tuberculosis detection via single-channel mobile ECG.

Now there are plenty of works in which the ECG is analyzed using neural networks (NN), for example, [5], [6], [7]. Usually, such raw ECGs are analyzed using recurrent [8] or convolutional neural [6] networks with 1d convolutions. In this paper, it is proposed to analyze a visualization of an ECG and not a raw signal. There are some works in which neural networks analyze the image of an ECG signal [9] [10] and show good quality. However, in these articles, architectures were trained from scratch and were specially selected for the task. There are large number of pretrained architectures for images [11], [12], [13]. Since for many tasks pretrained architectures show significantly better quality than architectures trained from scratch [14], there is an assumption that they will show better quality on the paper's task. There is work [15] that shows that pretrained architectures can be used as a feature extractor for PCG signals, which also provides additional motivation for exploring the possibility of using pretrained architectures.

This paper aims to investigate the possibility of determining tuberculosis from ECG images, the possibility of using pretrained models for such tasks, and of the interpretation of such models.

The paper is organized as follows. Initially, the dataset and data preprocessing methods are described. Then, the process of generation of the ECG image and the architecture of the neural network are given. Then the results of experiments and interpretation of the neural network are reported. Finally, the limitations of the article are analyzed and the conclusion is given.

2. Dataset

All data used in this study was collected using the CardioQvark device [3]. The dataset consists of 1232 ECGs from 136 people with TB and 3609 ECGs from people without TB. Sex distribution among people with tuberculosis and people without tuberculosis is shown in figure 1 and figure 2. Age distribution among people with tuberculosis and people without tuberculosis is shown in figure 3 and figure 4.

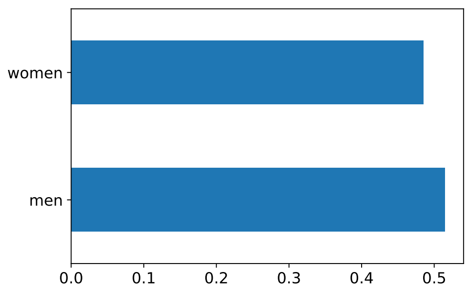

Figure 1. Sex distribution among people with TB.

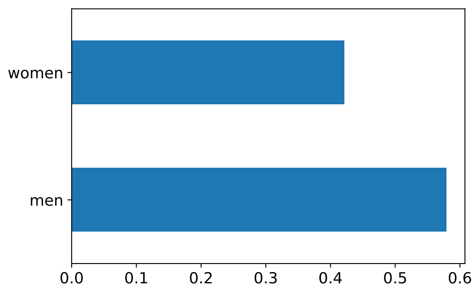

Figure 2. Sex distribution among people without TB.

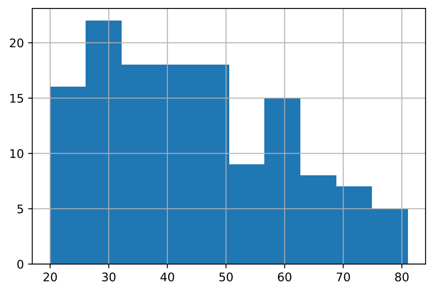

Figure 3. Age distribution among people with TB.

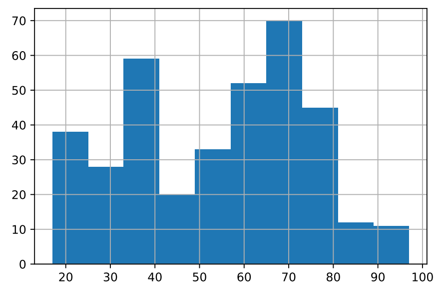

Figure 4. Age distribution among people without TB.

ECG recordings were sampled as 1000 Hz, and they have been filtered by the CardioQvark device. The length of ECG recordings varies from 30.5 seconds to 900 seconds. For trend subtraction, the median filter with a kernel size of 187 was used.

3. Image generation process and neural network architecture

To reduce the image size, signals were resampled to a frequency of 200 Hz, and also signals longer than 300 seconds were divided into sub-signals. To receive the sub-signals, the signal is divided sequentially into sections of 300 seconds, if the length of the last section is less than 30 seconds, then it was not added to the sample set.

Since the signals were taken from a device that can have noise, and the signal could disappear for various reasons, it is necessary to remove signals that contain a lot of noise or do not contain any useful information. R-peaks, that represent heartbeat, were calculated using the Pan–Tompkins algorithm [16]. Then the distances between the two peaks were calculated. It was measured heuristically that the signal section almost certainly does not contain useful information if the distance between the two R-peaks is more than 2.6s or less than 0.3s. The signal was removed from the dataset if there were more than 55 percent of such distances in it.

The image was generated using the matplotlib plot function [17]. The whole ECG signal was divided into 7 parts of 2857 points each, with dots per inch of 70. The size of the figure was 17 by 17. Since the matplotlib generates an image in the RGBA format, and most neural networks use RGB format, only the 4th channel was used. This channel was multiplied to create 3 channels. The example of the generated image is given in figure 5.

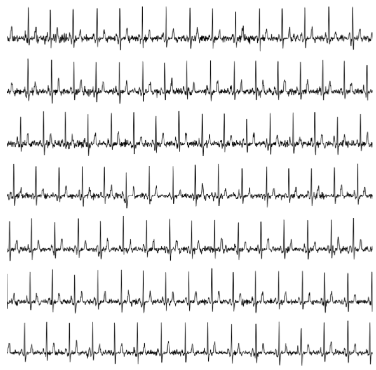

Figure 5. Generated ECG image.

ResNext 101 [13] was used as the main model for the experiments. This model is one of the top 5 best models for classification on imagenet [18] and is easy to use and retrain on new data. This model has 4 blocks of convolutional layers. To explore the possibilities of transfer learning, the following neural network configurations were used: completely untrained network (UN), pretrained on imagenet network with no frozen layers (PN), pretrained on imagenet network with frozen layer 1 (or 2, 3, 4) and all layers before it (L1(2, 3, 4)). In case of freezing up to layer 4 only one fully connected classifier layer is trained.

4. Experiments, results and interpretation

To correctly evaluate the results, the entire dataset was divided into training, test and validation set. The training set was used to train the model. The validation set was used to select the hyperparameters and for early stopping. A test set was used to check the final quality. The full dataset was partitioned into three parts in a specific way, such that each of the three groups contained different patients. The train set contains 2648 ECGs from 247 patients, among them 69 patients with 689 ECGs have TB. The validation set contains 1062 ECGs from 112 patients, among them 30 patients with 255 ECG have TB. The test set contains 1130 ECGs from 144 patients, among them 37 patients with 288 ECGs have TB.

Due to limited GPU memory, the generated images were resized to 712 by 712. For reproducibility, all neural network configurations were trained with a random state 10. For all configurations, Adam was used as an optimizer and batch size was set to 16. The metric that was used for early stopping was the area under the precision-recall curve since classes are not balanced in samples. For the untrained network, the learning rate was set to 0.0005 and decreased after the 8th epoch with a coefficient of 0.001. For the pretrained network, the learning rate was set to 0.0005 and decreased after the 8th epoch with a coefficient of 0.001. For the pretrained network with unfrozen and frozen layers, the learning rate was set to 0.0001 and decreased after the 1st and 4th epoch with a coefficient of 0.001. All parameters that are not described were installed by default as in PyTorch library [19].

Table 1 shows the results from different network configurations on the area under the ROC Curve (ROC AUC) score and area under the Precision-Recall Curve (PR AUC) score.

Table 1. Experiment Results.

	UN	PN	L1	L2	L3	L4
ROC AUC	0.9079	0.93	**0.934**	0.9307	0.9138	0.8089
PR AUC	0.7956	0.83	**0.84**	0.8399	0.8081	0.615

As can be seen from table 1, the best quality is obtained with configurations L1 and L2. This table shows that transfer learning is useful for classifying ECG signals. This table also corresponds to the logic that the first layers of a neural network learn simple dependencies such as lines and simple shapes. If you freeze a neural network up to the last layers, then you can get worse quality than on a non-trained network, since the features on the last layers are not relevant for this specific task. Thus, based on metrics, L1 can be considered the best model.

In real life, the probability can not be used to make a certain decision. Specific labels for each patient are used. Labeling depends on the threshold. The threshold can be selected according to different metrics depending on the task. Since in this situation, the relationship between precision and recall is interesting, and the classes are not balanced, the threshold is selected by the f-beta score metric. The beta is in range $(0, +\infty)$. The greater beta is, the larger the weight the recall has. Beta can be selected differently based on the situation. If we want to find all TB-positive people and the cost of additional procedures is not that important than we give the recall the greater weight. If the cost of the additional procedures is crucial than the precision is much more important even at the cost of skipping some TB-positive people.

This article investigated the following beta threshold options: 0.25, 1, 2. The validation set was used for threshold calculation. With chosen thresholds, the following metrics were investigated: accuracy (A), precision (P), recall (R). Also, since each cardiogram was divided into several images, the following rule was used to determine the label for the entire cardiogram. If at least one cardiogram is labeled 1, then the entire cardiogram is labeled 1. The metrics that are determined for the entire ECG are called: (A1), (P1), (R1). Also as one person can have multiple ECG the patient score for cardiogram labels was calculated. The patient score equals to mean accuracy for person ECGs averaged through all people (PS). As shown in table 2 the results are quite promising and quite customizable depends on the situation.

Doctors often want to understand why a neural network made a certain decision. Therefore, the interpretation of the neural network is quite useful. Besides, interpretation can be used to identify previously unknown signs of the disease to think through treatment in the future.

Table 2. Experiment Results.

	A	P	R	A1	P1	R1	PS
0.25	0.8433	0.891	0.4714	0.8655	0.8688	0.556	0.8389
1	0.876	0.756	0.7911	0.8824	0.7264	0.8636	0.819
2	0.8492	0.665	0.8786	0.8388	0.621	0.937	0.763

Grad-CAM [20] shows quite good results in interpreting neural networks. The figures 6 and 7 show the interpretation results. The lighter the points on the image the more important these points are. Both images represent the ECG of TB-positive person. As can be seen from the images, the neural network draws attention to the PQST complexes, the distance between the peaks, and also the height of some peaks.

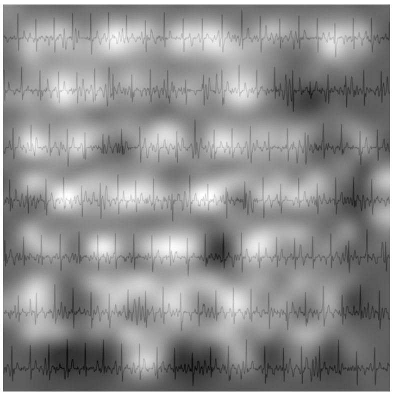

Figure 6. Interpretation of NN on TB-positive person.

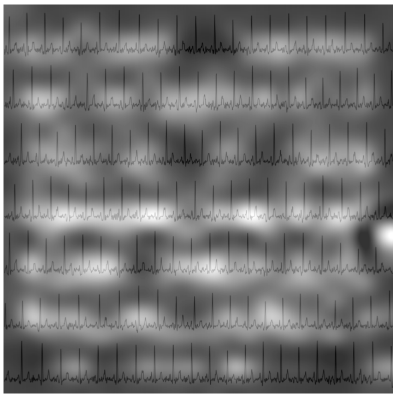

Figure 7. Interpretation of NN on TB-positive person.

5. Discussion

The reader of this article should know that this article has the following limitations: the sample size of TB-positive people is quite low, so for further investigation of quality the bigger sample should be used. It should also be noted that the sample contains patients who received treatment for tuberculosis. Due to the existing database structure, it is impossible to say whether the patient took the medicine, and if he did, how recent the use is. For more robust experiments, a new database must be compiled. But even in the current circumstances, this study is useful for the following reasons. Let us imagine an organization that employs a TB-positive person, who has already started taking the medicine and hides his condition, thus increasing the chance of infecting people around him. In such a situation it is still beneficial to screen people with the suggested algorithm. Besides, it is known that people taking medicine may experience an increase in QT-interval [21]. But in this dataset, the average value of the QT-interval showed an extremely low predictive quality for TB of about 0.51 ROC AUC. Based on this we can conclude that the neural network finds some other defining signs. Also with the help of the neural network interpretation in the future, the effect of TB-drugs on the heart can be explored.

6. Conclusion

In the course of this article, it was determined that the neural network can detect tuberculosis by a cardiogram with moderately high quality. It was found that transfer learning helps to improve the quality of the ECG image classification. ROC AUC score was increased from 0.9079 to 0.934 and PR AUC score was increased from 0.7956 to 0.84. So pretrained neural network should be used in different classification ECG cases. The method for interpretation of such models is described, which can be used further for drug and disease effect exploration purposes.

7. Acknowledgements

This article contains the results of a project carried out within the implementation of the Program of the Center for Competence of the National Technology Initiative “Center for Storage and Analysis of Big Data”, supported by the Ministry of Science and Higher Education of the Russian Federation under the Lomonosov Moscow State University Project Support Fund 13/1251/2018 from 11.12.2018.

8. References

[1] World Health Organization Web Site [Electronic resource]. – Access mode: https:// www.who.int/news-room/fact-sheets/detail/tuberculosis (20.12.2019).
[2] AliveCor Web Site [Electronic resource]. – Access mode: https://www.alivecor.com (20.12.2019).
[3] CardioQvarkWeb Site [Electronic resource]. – Access mode: http://www.cardioqvark.ru.
[4] Huaman, M.A. Tuberculosis and cardiovascular disease: linking the epidemics // Tropical diseases, travel medicine and vaccines. – 2015. – Vol. 1(1). – P. 10.
[5] Kiranyaz, S. Real-time patient-specific ECG classification by 1-D convolutional neural networks / S. Kiranyaz, T. Ince, M. Gabbouj // IEEE Transactions on Biomedical Engineering. – 2015. – Vol. 63(3). – P. 664-675.
[6] Zihlmann, M. Convolutional recurrent neural networks for electrocardiogram classification / M. Zihlmann, D. Perekrestenko, M. Tschannen // Computing in Cardiology (CinC), 2017. – P. 1-4.
[7] Yao, Z. Atrial fibrillation detection by multi-scale convolutional neural networks / Z. Yao, Z. Zhu, Y. Chen // 20th International Conference on Information Fusion (Fusion), 2017. – P. 1-6.
[8] Banerjee, R. A Novel Recurrent Neural Network Architecture for Classification of Atrial Fibrillation Using Single-lead ECG / R. Banerjee, A. Ghose, S. Khandelwal // 27th European Signal Processing Conference (EUSIPCO), 2019. – P. 1-5.
[9] Parsons, S. Robust and fast heart rate variability analysis of long and noisy electrocardiograms using neural networks and images / S. Parsons, J. Huizinga // arXiv preprint arXiv: 1902.06151, 2019.
[10] Kim, M.G. A study on user recognition using 2D ECG based on ensemble of deep convolutional neural networks / M.G. Kim, H. Ko, S.B. Pan // Journal of Ambient Intelligence and Humanized Computing, 2019. – P. 1-9.
[11] He, K. Deep residual learning for image recognition // Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. – P. 770-778.
[12] Chollet, F. Xception: Deep learning with depthwise separable convolutions // Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. – P. 1251-1258.
[13] Xie, S. Aggregated residual transformations for deep neural networks // Proceedings of the IEEE conference on computer vision and pattern recognition, 2017. – P. 1492-1500.
[14] Simonyan, K. Very deep convolutional networks for large-scale image recognition / K. Simonyan, A. Zisserman // arXiv preprint arXiv: 1409.1556, 2014.
[15] Alaskar, H. The Implementation of Pretrained AlexNet on PCG Classification // International Conference on Intelligent Computing – Cham: Springer, 2019. – P. 784-794.
[16] Pan, J. A real-time QRS detection algorithm / J. Pan, W.J. Tompkins // IEEE Trans. Biomed. Eng. – 1985. – Vol. 32(3). – P. 230-236.
[17] Hunter, J.D. Matplotlib: A 2D graphics environment // Computing in science & engineering. – 2007. – Vol. 9(3). – P. 90.
[18] Mahajan, D. Exploring the limits of weakly supervised pretraining // Proceedings of the European Conference on Computer Vision (ECCV), 2018. – P. 181-196.
[19] Paszke, A. PyTorch: An imperative style, high-performance deep learning library // Advances in Neural Information Processing Systems, 2019. – P. 8024-8035.
[20] Selvaraju, R.R. Grad-cam: Visual explanations from deep networks via gradient-based localization // Proceedings of the IEEE International Conference on Computer Vision, 2017. – P. 618-626.
[21] Bykova, A.A. The prevalence of prolongation of the QT interval in patients receiving antituberculous chemotherapy // Cardiology and Cardiovascular Surgery. – 2019. – Vol. 12(2).