A CNN for COVID-19 Detection using ECG signals

Federico M. Muscato[1], Valentina D.A. Corino[1], Massimo W. Rivolta[2], Pietro Cerveri[1], Antonio Sanzo[3], Alessandro Vicentini[3], Roberto Sassi[2], Luca T. Mainardi[1]

1 Department of Electronics, Information and Bioengineering, Politecnico di Milano, Italy.
2 Department of Computer Science, Università degli Studi di Milano, Italy
3 Arrhythmia and Electrophysiology, and Coronary Care Unit,
Fondazione IRCCS Policlinico S. Matteo, Pavia, Italy

Abstract

We developed an end-to-end automatic algorithm for the detection of signs of COVID-19 virus infection in ECGs. We analyzed 12-lead ECGs from patients infected by COVID-19 (C-group) and from a control group (NC-group). The C-group (896 cases) included patients (age range [19-96] years) hospitalized at Ospedale San Matteo in Pavia (Italy) during the first 2020 pandemic outbreak. Infection was confirmed by nasal swab testing. The NC-group (also 896 cases) was built by collecting ECG in sinus rhythm from 3 datasets: Georgia ECG (USA), PTB-XL (Germany) and CPSC 2018 (China). Control ECGs were matched by gender, age and heart rate. An additional control group, only used for testing, was extracted from the Ningbo (China) database. A 4-layers convolutional neural network (CNN), with increasing filter size plus a final fully connected (FC) layer, was designed to classify C vs NC-group. The CNN was trained and k-fold cross validated (k=7) on 1536 ECGs (1316 for testing-220 for validation). Every fold model was used to classify the remaining, separate common test set of 256 ECGs. The accuracy was 0.86 ± 0.01 on validation, 0.86 ± 0.01 on the test set. The FPR on the NC-group was 0.14± 0.03 on validation, 0.13± 0.02 on test and 0.10± 0.01 on the Ningbo test set (p>0.05, ns) showing that no bias was induced by the selection of datasets.

1. Introduction

Coronavirus Disease 2019 (COVID-19) is a disease produced by Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) infection causing the current pandemic around the world. Due to the rapid propagation and the massive increase in the number of new infections of such a disease, a reliable and rapid identification of COVID-19 has become crucial to prevent its rapid spread. The quick antigenic swabs satisfy this requirement, nevertheless a patient needs to have the suspicion of being infected in order to proceed to testing, which is not possible for asymptomatic cases.

Previous studies have shown insight into the occurrence and implications surrounding electrocardiographic changes in the infected individuals and diverse patterns were observed in the electrocardiogram (ECG) of patients with COVID-19. In particular, various forms of cardiovascular variations such as prolongation in QT [1], arrhythmias [2], ST-segment modifications [3] and PR interval changes [4] have been observed in the ECG of COVID-19 cases. Such cardiovascular modifications and patterns [5] have promoted the study of ECG data as a new means of diagnosing the novel coronavirus.

Deep neural networks (DNNs) had a remarkable impact on different scientific fields [6] and showed huge potential in the medical domain creating automated diagnostic tools capable of analyzing medical data. Regarding COVID-19-induced pneumonia, several works reported that detection might be possible using DNNs on chest X-rays and CT scans [7-9]. Unrelated to COVID pneumonia, cardiac anomaly classification from ECG data through DNNs techniques has been developed in other studies demonstrating the effectiveness of an end-to-end approach both for single [10] and 12-lead signals [11]. On top of that, the PhysioNet/Computing in Cardiology Challenge in 2020 and 2021 [12], whose aim was classifying multi-type arrhythmia and cardiac abnormalities over annotated databases with thousands of 12-lead ECG recordings, has further stimulated the development of end-to-end ECG analysis tools. As an outcome, different models with increasing complexity and different approaches have been developed [13].

The promising performances achieved in this field suggest a potential application of DNNs for COVID diagnosis directly from the ECG traces. Therefore, in this work we have developed an automatic algorithm for the detection of signs of COVID-19 virus infection in 12-lead ECGs.

2. Materials and Methods

2.1. Data

We analyzed a case group (C: COVID) and a control (NC: non-COVID) group. The C-group included 12-lead ECG recordings from patients (age 66.93 ± 15.83, range [19-96] years) hospitalized at Ospedale San Matteo in Pavia (Italy) during the first 2020 pandemic outbreak. All the recordings had the same sampling rate (500 Hz) and Duration (10 seconds). All signals were stored with the standard set of 12 leads (the order was I, II, III, AVL, AVR, AVF, V1-V6). In all patients, infection was confirmed by nasal swab testing.

To address a possible bias of the network introduced by different data collection practices, the NC-group was built by selecting ECG in sinus rhythm from three public datasets collected in different countries and previously used in the Physionet Challenge 2021: Georgia ECG (USA), PTB-XL (Germany) and CPSC 2018 (China). Control ECGs were selected among those matching the C-group by gender, age (±3 year) and heart rate (±3 number of beats in the 10s segments). For each ECG in the C-group the order in which the three datasets were searched for a matching record in sinus rhythm was randomly changed to further avoid bias. Globally, the dataset consisted of 1792 signals balanced in respect of the two classes: 896 (C group) plus 896 (NC group). An additional NC-group of 790 recordings was extracted from the Ningbo (China) database, using the same criteria as before. Since this database did not match all the specific cases in the C-group, it was not possible to have the same number of recordings. In particular, there were no patients over 90 years of age in the Ningbo database. This final dataset was used to check model generalization capability on unseen data. A summary of the group characteristics is shown in Table 1.

Table 1. Features of the three different datasets. Age is expressed in years and Heart Rate in bpm.

Feature	C-group	NC-group	Ningbo
Age (mean±std)	66.9±15.8	66.1±15.3	64.4± 5.00
Age (min-max)	19 – 96	18 – 94	18 – 89
Gender (male)	566	566	513
Gender (female)	330	330	277
Heart Rate (mean±std)	76.68±15.2	74.16±12.6	75.66±11.0

2.2. Pre-Processing

A preprocessing step was applied in order to minimize the difference between the data from different sources. It consisted of a filtering step (3rd order Butterworth filter [0.5-45 Hz]) followed by Z-score normalization applied to each lead, separately.

2.3. Network architecture

The deep learning model was designed to expect in input a 10 s 12-lead ECG sampled at 500 Hz (assembled in a data matrix of 12x5000 samples). The architecture used in this work for classifying C vs NC group is a convolutional neural network (CNN 1D). This network aims at learning a compressed representation (encoding) of an input dataset with an approach similar to the biological sensorial processing of the visual cortex whose cells are sensitive to small sub regions of the visual field called receptive fields. In particular, the designed model consists of a network with 4 convolutional layers (see Figure 1) plus a fully connected (FC) layer for feature selection and then a final FC layer for classification with a softmax activation function. The filter size of the convolutional layers

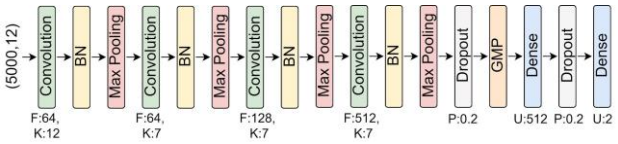

Figure 1. Architecture of the model used in this work. (F: Filters, K: Kernel, P: Probability, U: Units, BN: Batch Normalization, GMP: Global Max Pooling)

increases from 64 in the first two convolution layers, to 128 in the third and then up-to 512 in the last layer to capture as much information as possible in the different CNN filter banks. The model has a kernel size of 12 in the first convolutional layer and a kernel size equal to 7 in the last one. A previous work has shown that large kernels are more helpful for networks to learn meaningful features [14]. The ReLu activation function was used.

In our design of the network, we planned to have it as simple as possible. The choice of such a shallow network is motivated by the small number of available training samples and to have short training time. In order to avoid overfitting, a Global Max Pooling layer is inserted between the last convolutional layer and the first FC layer. This helped to improve (data not shown) model performance and to reduce the amount of model parameters. Furthermore, dropout layers were inserted before and after the first FC layer which are prone to overfit.

2.4. Network tuning and testing

The 1792 signals were initially split to create a hold-out test-set of 256 cases. The remaining recordings were split into training and validation. In particular, the model was k-fold cross validated (k=7) using 1316 ECGs for training and 220 for validation. Each signal entered in one set only (a single signal was not split into parts) and sets were balanced between C- and NC-class. Every fold model was then used to classify the common hold-out test set of 256 ECGs and the performance reported for this set. To check generalization capability on unseen data, another test set was created selecting 790 NC-signals from the Ningbo database.

The network hyperparameters (kernel and filter size, first FC size and dropout probability) were determined with a Bayesian Optimization tuning that aimed to maximize the accuracy on the validation set. Bayesian Optimization was used because it does not sample hyperparameter combinations randomly but follows a probabilistic approach taking into account already tested combinations and uses this information to sample the next combination for a test. The model was compiled using an Adam optimizer and choosing categorical cross entropy as loss function. Moreover, a learning rate scheduler was used in order to prevent the training curves from diverging: if the validation loss was not decreasing in 20 epochs, the learning rate was reduced by a 0.1 factor.

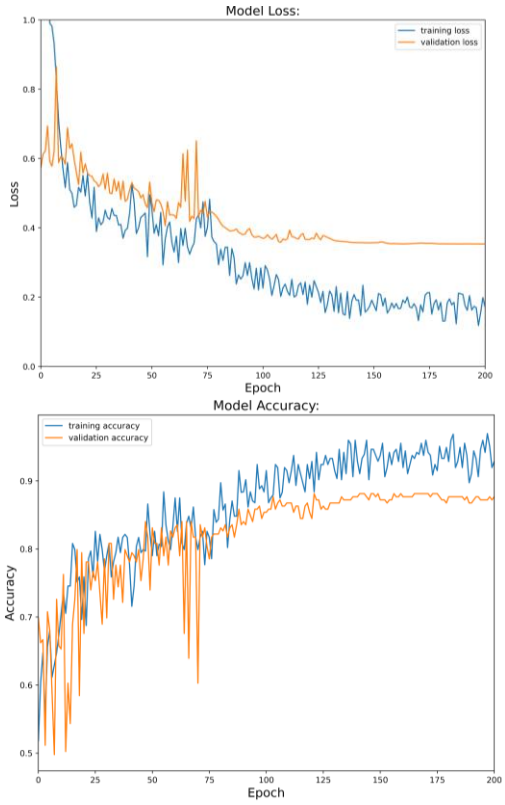

Figure 2. Trend of Accuracy and Loss in one fold.

Table 3. Confusion matrix showing the results on the hold test set of the model trained at fold #3.

	Predicted COVID	Predicted Control
Actual COVID	112	16
Actual Control	14	114

2.4. Metrics

The performance metrics for diagnosis of COVID-19 were reported as Accuracy, False Positive Rate (FPR) and True Positive Rate (TPR). TPR is calculated as the proportion of COVID patients classified correctly, while FPR is the proportion of non-COVID patients classified as infected. It is worth mentioning that since the dataset is balanced, accuracy provided an unbiased measure of the network performances.

3. Results

To show the learning curve of the network, in Fig. 2 the trends of accuracy and loss are reported for fold #3. These curves have oscillations at the beginning that decrease with the progression of the training when the learning rate decreases.

The resulting confusion matrix on the hold test set for this fold is represented in Table 3. A significant capability to discriminate between C and NC can be observed.

Overall, the accuracy was 0.85 ± 0.01 (mean ± std on the different folds) on validation, 0.86 ± 0.01 on the common test set and 0.89 ± 0.01 on the Ningbo test set. The performance summary is shown in Table 2. Data are reported as mean ± std.

Table 2. Performance summary on the different evaluated sets with the number of cases in brackets.

Group type	Validation		Test	
	TPR	FPR	TPR	FPR
C-group	0.85± 0.02 (110)	-	0.85± 0.01 (128)	-
NC Group	-	0.14± 0.03 (110)	-	0.13± 0.02 (128)
NC Ningbo	-	-	-	0.10± 0.01 (790)

For comparison, a t-test was also performed between the TPR or the FPR obtained on the validation and the test sets to prove no differences exist in the performance. Moreover, the FPR computed on the test set of the NC-group and the NC-Ningbo group was performed to assess that no bias was introduced by the arbitrary selection of the NC-group. All the tests had $p>0.05$ (ns).

4. Discussion

We found that patients infected with COVID-19 present electrocardiographic changes that could be identified by deep learning models. We chose to develop a network with a limited number of layers as it neither demanded prolonged training duration nor large number of training samples. On our patients, the proposed network showed very good discrimination capability.

To better understand performance, the model was evaluated by enriching the control population in the testing set with patients from different datasets thus varying the sources of data. Since these data were from different database sources, it was also a test of the generalization capability of the network.

While the results are promising, the study has some limitations. First, it was not possible to test the discrimination capability on a different population of COVID-19 subjects. Unlike the NC-group, where we could use the publicly available Ningbo database, no public, digitized 12 lead ECGs from COVID-19 patients were available at the time of this study. Secondly, the dataset used in the study is from confirmed, symptomatic hospitalized COVID-19 patients. Thus, the detection of asymptomatic infections in the general population may not achieve the same level of sensitivity. Finally, the NC-group was composed only by subjects in sinus rhythm. This might overestimate the TNR in the general population, where many other concurrent cardiac abnormalities are present.

In the future, we plan to validate these results prospectively, so as to permit the use of ECG as a screening test to exclude acute COVID-19 infection.

Acknowledgment

This project has been developed within the project "automatic COmputation of cardioVascular arrhythmIc risk from ECG data of COVID-19 patients" (COVIDSQUARED) which has received funding from the Italian Ministry of University and Research, call FISR 2020.

References

[1] F. Santoro *et al.*, 'QTc Interval Prolongation and Life-Threatening Arrhythmias During Hospitalization in Patients with Coronavirus Disease 2019 (COVID-19): Results from a Multicenter Prospective Registry', *Clin Infect Dis*, vol. 73, pp. e4031–e4038, 2021.

[2] A. Denegri *et al.*, 'Clinical and electrocardiographic characteristics at admission of COVID-19/SARS-CoV2 pneumonia infection', *Intern Emerg Med*, vol. 16, pp. 1451-1456, 2021.

[3] S. Bangalore *et al.*, 'ST-Segment Elevation in Patients with Covid-19 — A Case Series', *N Engl J Med*, vol. 382, pp. 2478-2480, 2020.

[4] B. B. Pavri *et al.*, 'Behavior of the PR interval with increasing heart rate in patients with COVID-19', *Heart Rhythm*, vol. 17, pp. 1434-1438,2020.

[5] B. Predabon *et al.*, 'The Electrocardiogram in the Differential Diagnosis of Cardiologic Conditions Related to the COVID-19 Pandemic', *J. Cardiac Arrhythmias*, vol. 33, pp. 133-141, 2020.

[6] A. Krizhevsky *et al.*, 'ImageNet classification with deep convolutional neural networks', *Commun. ACM*, vol. 60, pp. 84-90, 2017.

[7] M. Toğaçar *et al.*, 'COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches', *Comput Biol Med*, vol. 121, p. 103805, 2020.

[8] S. Toraman *et al.*, 'Convolutional CapsNet: A novel artificial neural network approach to detect COVID-19 disease from X-ray images using capsule networks', *Chaos Solitons Fractals*, vol. 140, p. 110122, 2020.

[9] A. M. Ismael *et al.*, 'Deep learning approaches for COVID-19 detection based on chest X-ray images', *Expert Syst Appl*, vol. 164, p. 114054, 2021.

[10] D. Li *et al.*, 'Classification of ECG signals based on 1D convolution neural network', in *2017 IEEE 19th Int. Conf. on e-Health Networking, Applications and Services (Healthcom)*, pp. 1-6, 2017.

[11] A. H. Ribeiro *et al.*, 'Automatic diagnosis of the 12-lead ECG using a deep neural network', *Nat Commun*, vol. 11, pp. 1-9, 2020.

[12] M. A. Reyna *et al.*, 'Issues in the automated classification of multilead ECGs using heterogeneous labels and populations', *Physiol. Meas.*, 2022.

[13] C.-Y. Chen *et al.*, 'Automated ECG classification based on 1D deep learning network', *Methods*, vol. 202, pp. 127-135, 2022.

[14] A. Y. Hannun *et al.*, 'Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network', *Nat Med*, vol. 25, pp. 65-69, 2019.

Address for correspondence:
Federico Maria Muscato
Department of Electronics, Information and Bioengineering
Via Camillo Golgi, 39, 20133 Milano, Italy
federicomaria.muscato@polimi.it