Revista de Informática Teórica e Aplicada - RITA - ISSN 2175-2745
Vol. 30, Num. 01 (2023) 44-52

RESEARCH ARTICLE

COVID-19 Detection Using Forced Cough Sounds and Medical Information

Detecção de COVID-19 Usando Sons de Tosse Forçada e Informações Médicas

Lucas A.M. de Souza[1], Heder S. Bernardino[1]*, Jairo F. de Souza[1], and Alex B. Vieira[1]

Abstract: The World Health Organization (WHO) has declared the novel coronavirus (COVID-19) outbreak a global pandemic in March 2020. Through a lot of cooperation and the effort of scientists, several vaccines have been created. However, there is no guarantee that the virus will shortly disappear, even if a large part of the population is vaccinated. Therefore, non-invasive methods, with low cost and real-time results, are important to detect infected individuals and enable earlier adequate treatment, in addition to preventing the spread of the virus. An alternative is using forced cough sounds and medical information to distinguish a healthy person from those infected with COVID-19 via artificial intelligence. An additional challenge is the unbalancing of these data, as there are more samples of healthy individuals than contaminated ones. We propose here a Deep Neural Network model to classify people as healthy or sick concerning COVID-19. We used here a model composed by an Convolutional Neural Network and two other Neural Networks with two full-connected layers, each one trained with different data from the same individual. To evaluate the performance of the proposed method, we combined two datasets from the literature: COUGHVID and Coswara. That dataset contains clinical information regarding previous respiratory conditions, symptoms (fever or muscle pain), and a cough record. The results show that our model is simpler (with fewer parameters) than those from the literature and generalizes better the prediction of infected individuals. The proposal presents an average Area Under the ROC Curve (AUC) equal to 0.885 with a confidence interval (0.881 - 0.888), while the literature reports 0.771 with (0.752 - 0.783).
Keywords: COVID-19 detection — Cough sounds — Deep Neural Networks

Resumo: A Organização Mundial da Saúde (OMS) declarou o surto do novo coronavírus (COVID-19) uma pandemia global em março de 2020. Por meio de muita cooperação e esforço dos cientistas, várias vacinas foram criadas. No entanto, não há garantia de que o vírus desapareça em breve, mesmo que grande parte da população seja vacinada. Portanto, métodos não invasivos, com baixo custo e resultados em tempo real, são importantes para detectar indivíduos infectados e possibilitar tratamento adequado precocemente, além de evitar a disseminação do vírus. Uma alternativa é usar sons de tosse forçados e informações médicas para distinguir uma pessoa saudável daquelas infectadas com COVID-19 por meio de inteligência artificial. Um desafio adicional é o desbalanceamento desses dados, pois há mais amostras de indivíduos saudáveis do que contaminados. Propomos aqui um modelo de Rede Neural Profunda para classificar pessoas como saudáveis ou doentes em relação ao COVID-19. Neste trabalho, usamos um modelo composto por uma Rede Neural Convolucional e duas outras Redes Neurais com duas camadas totalmente conectadas, cada uma treinada com dados diferentes do mesmo indivíduo. Para avaliar o desempenho do método proposto, combinamos dois conjuntos de dados da literatura: COUGHVID e Coswara. Esse conjunto de dados contém informações clínicas sobre condições respiratórias anteriores, sintomas (febre ou dor muscular) e um registro de tosse. Os resultados mostram que nosso modelo é mais simples (com menos parâmetros) do que aqueles da literatura e generaliza melhor a predição de indivíduos infectados. A proposta apresenta uma área média sob a curva ROC (AUC) igual a 0,885 com intervalo de confiança (0,881 - 0,888), enquanto a literatura relata 0,771 com (0,752 - 0,783).
Palavras-Chave: Detecção de COVID-19 — Sons de Tosse — Redes Neurais Profundas

[1] *Universidade Federal de Juiz de Fora (UFJF), Juiz de Fora - Minas Gerais, Brazil*
***Corresponding author**: heder@ice.ufjf.br
DOI: http://dx.doi.org/10.22456/2175-2745.126016 • **Received:** 20/07/2022 • **Accepted:** 04/01/2023

1. Introduction

On March 11, 2020, the World Health Organization (WHO) declared the COVID-19 pandemic caused by the new coronavirus (Sars-Cov-2) due to its fast geographic spread. On August 15, 2021, more than 243 million confirmed cases and more than 4.9 million deaths from the disease have been regis-

tered worldwide[1]. After much effort by scientists around the world, in December 2020, the population around the world started to get vaccinated against COVID-19. Nowadays, billions of doses have been administered in at least 201 countries[2]. Although, people can be infected even when they are vaccinated. Therefore, the best way to prevent the virus from spreading and ensure that people infected with COVID-19 have access to adequate treatment as soon as possible is through extensive testing and isolation of those infected ones. Currently, the Reverse Transcription Polymerase Chain Reaction (RT-PCR) test is considered a gold standard in the diagnosis of COVID-19. However, the report of this test can take days to be obtained. In addition, RT-PCR has not an accessible price and there is a limitation in the number of tests that can be performed daily, depending on the infrastructure and the amount of raw material available.

In the literature, there are several studies on the detection of diseases using recordings of human sounds, such as speech, breathing, and coughing, through Artificial Intelligence (IA) techniques. In these studies, several acoustic analysis techniques are used, capable of extracting information from those audios, which will be used in the training of classification models to discriminate individuals who have the disease from those who do not. Next, studies that used Machine Learning (ML) algorithms in the detection of diseases such as Parkinson's, Amyotrophic Lateral Sclerosis, and tuberculosis will be presented.

A scheme for breath analysis to detect irregular patterns in respiratory cycles due to diseases is proposed in [1]. In that study, the data is obtained using smartphones, a Discrete Wavelet Transform (DWT) is adopted for reducing the noise, segments of the sounds are selected, the main features are extracted by the Bag-of-Features (BoF) technique [2], and the Support Vector Machine (SVM) is used as the classifier. The proposal was used to identify asthmatic inspiratory cycles and reached an accuracy of 75.21%.

Machine Learning (ML) models were trained in [3] to detect Parkinson's disease based on the analysis of speech recordings from patients. As it is a binary classification problem, with the class of each sample previously known, the models used in this task were: Random Forest (RF), Neural Networks (NN), and SVM. The data used consists of three databases with speech recordings. The first one is composed of 22 people who have Parkinson's disease, totaling 88.8 minutes of recorded speeches. Whereas, the second one is formed by 30 healthy people, totaling 28.2 minutes of audio. In addition, a third database available in the literature was used to validate the performance of the models used. The results reported show that the Random Forest obtained the best performance according to the metrics used, and obtained an accuracy of 99.94%.

In addition, Vashkevich et al. [4] trained a K-Nearest Neighbor (kNN) classifier in the detection of bulbar dysfunction in patients with Amyotrophic Lateral Sclerosis (ALS), which is a progressive neurodegenerative disease that affects the nervous system that causes the death of neurons responsible for the control of voluntary muscles. The database used consists of recordings of 54 people, 39 healthy ones, and 15 patients with ALS. All participants produced the sustained vowel $/a/$ in a comfortable tone and constant sound for as long as possible. This phonation was performed in one breath. Acoustic characteristics extracted from these recordings were used in training the model. The authors report that the best result obtained by the model has an accuracy of 90.7%, with a sensitivity of 86.7% and specificity of 92.2%. Sensitivity corresponds to the percentage of positive results obtained by the model among people with a certain disease, while specificity is the proportion of negative results of the method in individuals who do not have the disease.

The potential of using ML techniques for reporting symptoms of respiratory diseases and sleep disorders is investigated by Vhaduri [5]. Mel-frequency cepstral coefficient (MFCC) features and different classification techniques were applied to three datasets with combinations of nocturnal noises (such as sounds from air conditioners, dog barks, and sirens). In that work, an RF model found the best results and was able to identify respiratory diseases with an average accuracy of 96%, and AUC-ROC of 0.98.

Also, in [6] the authors have used forced coughing sounds and medical information to detect patients infected with tuberculosis, which is still one of the deadliest diseases in the world. The database used consisted of recordings of coughing from 38 individuals, 17 of whom were infected with the disease, and 21 were healthy. In addition, 5 clinical information were added: (i) arm circumference; (ii) temperature; (iii) body mass index; (iv) presence of pale conjunctiva and (v) heart rate. A Logistic Regression model was trained using these data, and obtained in the best case an Area Under the ROC Curve (AUC) of 0.95 with an accuracy of 78%.

These studies demonstrate the ability of AI algorithms to detect individuals infected with different diseases, based on acoustic analysis of sounds produced by human beings. There are promising studies on the use of AI methods to detect patients infected with COVID-19, using forced cough recordings. These works are presented in Section 2. Considering that these models can classify infected patients, they can be adopted on a large scale, as it would be possible to use cell phones and computers to record sounds that would be fed to the model. Thus, the objective of this work is to propose a model capable of differentiating individuals infected with COVID-19, which can be used to detect the disease, at a low cost, in a non-invasive way, and with the ability to generate results in real-time. It is noteworthy that these methods should not be used as a determining factor for the diagnosis of a patient, but as a form of pre-screening for performing medical tests, such as the RT-PCR test.

In this work, we proposed modifications in the topology

[1] ⟨https://covid19.who.int/⟩

[2] ⟨https://graphics.reuters.com/world-coronavirus-tracker-and-maps/vaccination-rollout-and-access/⟩

of the Deep Learning (DL) model proposed in [7] for the detection of individuals infected with COVID-19 through the analysis of acoustic characteristics of forced cough sounds and clinical information reported by them. In addition, we proposed a training protocol that includes an early stopping strategy. The results obtained were compared with those reported in the literature and prove that the proposed modifications increased the model's classification capacity. Thus, the contributions of this work are a simpler DL model than those available in the literature, and an approach capable of obtaining better results in the classification of people infected with COVID-19.

In the following sections we present the related work (Section 2), the datasets and methods used (Section 3), the computational experiments performed and the results achieved (Section 5). Finally, in Section 6 we present the conclusions obtained in this work, its limitations, and future work.

2. Related Work

Since COVID-19 was declared a pandemic in 2020, many groups [8, 9, 10, 11] started to collect data, mostly by crowdsourcing through mobile apps and websites on the internet, to set up systems capable of detecting individuals infected with COVID-19. The systems and algorithms these groups created can be used as pre-screening for conducting medical examinations, such as the RT-PCR test, by identifying individuals most likely to be infected. Another possibility of applying the algorithms is for contact tracing and preventive quarantine for individuals who had recent contact with a patient with a high probability of infection until the medical examination is performed. The data collected by these groups comprise human sounds (coughing, breathing, and/or speech), medical information, symptoms, and the patient's diagnosis.

The COUGHVID database [9] contains more than 20,000 cough recordings, collected between April 1, 2020, and September 10, 2020, through a website deployed on a private server. According to the authors, this database has a wide range of ages, genders, geographic areas, pre-existing respiratory conditions, and health status (infected or healthy), with the potential to allow computational models to obtain a good generalization. Despite the abundance of recordings available, only 1,010 of these were recorded by patients claiming to be infected with a COVID-19. A disadvantage observed in the crowd-sourced databases is the fact that the COVID-19 status of an individual (infected or healthy) can be a self-diagnosis, without necessarily having undergone any medical examination, thus, messing with the data and potentially confusing the model. To try to minimize this problem and legitimize that samples marked as COVID-19 positive came from infected individuals, Orlandinc et al. [9] analyzed the geographic location from which the samples came. The authors combined World Health Organization new case statistics[3] with a 2019 group database of United Nations [4], to determine the infection rate in the country of origin of the sample 14 days before it is standardized. According to the authors, this analysis revealed that 94.4% of recordings of infected patients originated from countries with more than 20 confirmed cases per million population.

To avoid samples with no coughs, the authors trained an eXtreme Gradient Boosting (XGB) classifier, based on 121 cough sounds and 94 non-cough sounds, to determine the probability that a given recording contains a coughing sound. This model obtained an AUC of 0.97 and was used to classify the available audios in the database. The output probabilities of the classifier were included in the metadata of each record under the label *cough_detected*.

The objective of Coswara [10] project, according to the authors, is to create a database with sound samples from healthy and contaminated individuals. Unlike the COUGHVID base, the Coswara project collected 9 different categories of sounds, namely: breathing (shallow and deep), coughing (shallow and heavy), sustained vowel phonation of three different types, and digit counting from 1 to 20 (normal and fast). In addition, metadata was collected with information on gender, age, location, health status (healthy, exposed, cured, or infected), and the presence of pre-existing medical conditions. As in the COUGHVID database, data collection took place online, through a website that can be accessed by computer or cell phone, where the individual must provide the aforementioned metadata, as well as the 9 categories of recordings. However, in this project, the process of cleaning and checking the quality of the samples is done manually and is still taking place. In the first version, published on August 7, 2020, in a GitHub repository[5] the project had data from 941 participants. The project contains 1575 samples, where there are 109 records of patients infected with COVID-19 labeled as *positive_asymp*, *positive_mild*, or *positive_moderate*, and 1476 COVID-19 negative records labeled *healthy*, *no_resp_illness_exposed*, *recovered_full*, or *resp_illness_not_identifier*.

In the literature, some studies show promising results in the detection of individuals infected with COVID-19 based on human sounds. For example, the contributions of a group of researchers from MIT [12], researchers from the Virufy group [7] and Brown et al. [8] stand out. All these studies used forced cough sounds in the model training, however, [7] added medical information of symptoms reported by the individuals in addition to coughing, while [8] also used breathing sounds. Regarding the extraction of characteristics of sound signals, the study of [12] used MFCC to train the model. Meanwhile, [7] besides the MFCC also used the Log Mel Spectrogram. On the other hand, [8] extracted several characteristics of the signal, some examples are: *Duration, Onset, Time, Period, RMS Energy, Spectral Centroid, etc*. In total, 477 characteristics of coughing and breathing sounds were extracted. About the databases used in the training of the model, all of them used different databases. [8] and [12] used their databases, collected online, while the first one is available through the

[3] ⟨https://data.humdata.org/dataset/coronavirus-covid-19-cases-and-deaths⟩
[4] ⟨https://population.un.org/wpp/Download/Standard/CSV/⟩
[5] ⟨https://coswara.iisc.ac.in/⟩

signing of an agreement, whereas to the best of our knowledge the second one was not made available. [7] combined data from the COUGHVID and Coswara databases, which are available for the entire scientific community.

DL models were used in [12, 7, 11], while other ML models were used in [8]. In [12], a model formed by 3 ResNet50 networks in parallel, each pre-trained on different datasets to identify different cough characteristics, was proposed. The output of these networks was concatenated and the method provided as output the binary classification of an individual's condition. This model achieved an AUC of 0.97 when validated with samples from individuals diagnosed with an official test, and sensitivity of 100% with a specificity of 83.2% for asymptomatic ones.

On the other hand, Chaudhari [7] developed a DL model, with 3 networks in parallel, being a Convolutional Neural Network (CNN) and two Neural Network (NN). The output of these networks is concatenated and used to feed a Fully-Connected Layer that will be responsible for classifying people infected by COVID-19. These networks receive as input the Log Mel Spectrogram and the MFCC extracted from cough recordings, as well as the medical information reported by the individuals. The model had a mean AUC of 0.771, with a 95% confidence interval of (0.752 - 0.783).

Logistic Regression, Gradient Boosting Trees and Support Vector Machines classifiers were used in [8]. These models were trained in three distinct tasks: (i) correctly classify healthy and infected individuals with COVID-19, (ii) distinguish an individual who tested positive for COVID-19 and has cough as a symptom from a healthy individual with cough, and (iii) differentiate the cough of people with COVID-19 from those with asthma. The researchers report that the model was able to achieve an AUC greater than 0.8 for each task. One has an extension of [8] in [11], where a CNN was adopted. That proposed CNN contains a key module called VGGish, which is a model pre-trained with an external massive general-audio dataset. An AUC-ROC of 0.71 with a 95% confidence interval of (0.65 – 0.77) was obtained in [11] in its more realistic situation. The dataset used in [11] is not publicly available.

In this work, the study developed by [7] was used as a basis, as the model was made available on Github[6] by the authors, as the databases used for training[7][8]. Modifications to the topology of the model and its training protocol were proposed, which resulted in a simpler model (i.e. with fewer parameters to be adjusted), and capable of obtaining better results, as shown by the computational experiments carried out.

3. Materials and Methods

In this study, cough sounds and medical information from individuals collected by the COUGHVID [9] and Coswara Project [10] groups were used. The first one obtained cough samples from individuals from different countries, in addition to age, gender, geographic location, and COVID-19 status. The collection was made through the website[9] and made available in September 2020 on the zenodo[10]. Whereas, the second one collected the data on the website[11] and made available on GitHub[12]. Coswara project requires participants to provide a recording of breath sounds, cough sounds, sustained vowel phonation, a counting exercise, and health conditions, as well as medical information.

3.1 Combined Dataset

Similarly to [7], in this work the datasets of the COUGHVID group and the Coswara project were combined. In this case, all 1,575 records from the Coswara database were selected, 109 from contaminated individuals and 1466 from healthy ones. Only short cough recordings were used. Regarding COUGHVID data, only samples with $cough\_detected \geq 0.9$ were selected. This process resulted in 441 samples from infected individuals and 5,651 healthy ones, which represents a high imbalance where only approximately 7.80% of the samples belong to the positive class. Thus, an undersampling was applied to the majority class, where 1,000 instances of the negative instances were randomly selected. In addition, all positive instances were considered. This process was performed as in the literature. As a result, 3,016 samples were available for training the model, with 550 data from the positive class (18.24% of the database) and 2,466 from the negative class (81.76% of the database). Although the final dataset is imbalanced, we respected the same process of creating the datasets performed by [7] to fairly compare the proposed modifications on the topology of the model and training protocol.

As there is no standard in the medical information collected, as each project collects different sets of data, it was necessary to define the information collected by both databases so that they could be aggregated. The common information that was used to train our model is the cough record available in both cases and two clinical information represented by logical values: (i) it indicates whether the subject has any pre-existing respiratory condition and (ii) it reveals whether the patient has the symptoms, fever or muscle pain.

The data preparation process started with the aggregation of metadata (cough audio path, symptoms, and individual contamination status) of the selected samples in a single database. Then, for each of the available recordings, the Log Mel Spectrogram was calculated, and the first λ Mel Frequency Cepstral Coefficients were extracted using the librosa [13] package. It is a package for audio and music analysis available for the Python language. The characteristics extracted from the cough recordings (Log Mel Spectrogram and MFCC), as well as the medical information, were used as input to the

[6] ⟨https://github.com/virufy/virufy-covid⟩
[7] ⟨https://zenodo.org/record/4048312⟩
[8] ⟨https://github.com/iiscleap/Coswara-Data⟩
[9] ⟨https://coughvid.epfl.ch/⟩
[10] ⟨https://zenodo.org/record/4048312⟩
[11] ⟨https://coswara.iisc.ac.in/⟩
[12] ⟨https://github.com/iiscleap/Coswara-Data⟩

classification model for the prediction of individuals infected with COVID-19.

4. The Proposed Model

We present here the proposed model, which uses three networks, an CNN and two other NNs with two full-connected layers, each one trained with different data from the same individual. The outputs of these three networks are inputs of a Fully-Connected (FC) layer, whose result is the probability of an individual being infected with COVID-19. The flowchart in Figure 1 presents our proposal and the most relevant information. One of the two NNs is trained with MFCCs extracted from the cough audio, while the other one receives the individuals' medical information as input. Finally, CNN is trained with the Log Mel Spectrogram images generated from the audio.

Figure 2 details the representation of the model proposed here. The two NNs are composed of two hidden layers with a ReLU activation function, followed by a dropout layer. The first NN receives the medical information of patients as input and contains the following FC layers: FC1 with 64 neurons and a dropout rate of 0.4, and FC2 with 32 nodes and a dropout rate of 0.2.

The second network is a CNN with the Log Mel Spectrogram images with dimensions (64, 64, 1) as inputs. This is formed by three 2D Convolution Layers (CL) with kernels of size 3×3, where CL1 is formed by 32 kernels with a stride of size 2, and CL2 and CL3 consist of 64 kernels with a stride of 1. Each of these convolution layers is followed by a 2D average pooling layer with a kernel of size 2×2 and a stride of 2, followed by a layer of batch normalization and ReLU activation. The output of CL3 is flattened and used to feed the FC5 dense layer composed of 128 nodes with ReLU activation and a dropout rate of 0.5.

The last NN receives as input the first λ MFCCs extracted from the coughing sound and is formed by two dense layers similar to the first network, but the FC3 layer is formed by 64 neurons and a dropout rate of 0.4 and the FC4 layer by 32 and 0.2 dropout rate. Finally, the outputs of these networks are concatenated, and fed through the two dense layers FC6 with 64 nodes and FC7 with 32, both with ReLU activation, and combined into two nodes with softmax activation that predicts the probability of a subject being infected with COVID-19.

The DL model proposed here is based on that described in [7] and whose code is available at GitHub[13]. From this initial model, some changes were proposed in the structure of the network and in its training procedure to reduce overfitting, increase its prediction capability, and its performance in classifying infected patients. These changes were made to the number of neurons present in some of the FC layers, reducing the complexity of the network and preventing overfitting. With these modifications, the number of parameters in the model reduced from 279,826 (in the baseline model) to 154,370 in the proposal. This represents a reduction of about 45% in the number of parameters of the model.

As a form of regularization to avoid overfitting and degradation of the generalization of the model during the training, we proposed here the use of an Early Stopping (ES) policy. ES [14] is a technique used to monitor the model's performance through a validation dataset to prevent the deterioration of its generalization performance during the training. In this study, ES was implemented by monitoring the AUC of the validation data and the training stops when this value decreases by some epochs (a user-defined parameter).

5. Computational Experiments

Here we present the computational experiments performed, the metrics used to evaluate the model, and the results obtained. Furthermore, the results obtained by the model used in the literature base are presented [7], called here simply the Base Model (BM). For each proposed approach, the computational experiments were repeated 30 times, with different random divisions of data into training, validation, and testing sets. A division 70, 15, and 15 was used, as in the literature. The same proportion of the classes observed in the original dataset was kept in these sets. The number of samples for each class in these sets is shown in Table 1.

Table 1. Data distribution in training, validation and testing sets.

	Training	**Validation**	**Test**
COVID-19 positive	385	83	82
COVID-19 negative	1,726	370	370

The experiments performed used the dataset generated from the combination of COUGHVID and Coswara datasets. In this case, the first $\lambda = 39$ MFCC were extracted. Experiments were carried out to understand the impact of the inputs provided to the model (Log Mel Spectrogram, MFCC, and medical information). Initially, experiments were performed with the complete model (called MEL+MFCC+MI). Then, new experiments were executed with one of the inputs was disregarded. For example, the MEL+MI model is composed of networks that receive the Log Mel Spectrogram and the medical information as input. The MFCC+MI indicates that the DL model is fed with the symptoms and the MFCC. Whereas the MEL+MFCC model is composed of the networks with the MFCC and the Log Mel Spectrogram. Table 2 summarizes the composition information of the models.

Categorical cross-entropy as loss function and an Adam optimizer with a learning rate of 0.0001 was used in the training of the model. Data were split into batches of size 32 and the model was trained for 50 epochs. An ES policy was adopted, where AUC is monitored using a validation dataset and with a limit of 20 epochs.

The model and the data pre-processing were implemented using the Python programming language and the Keras pack-

[13] ⟨https://github.com/virufy/virufy-covid⟩

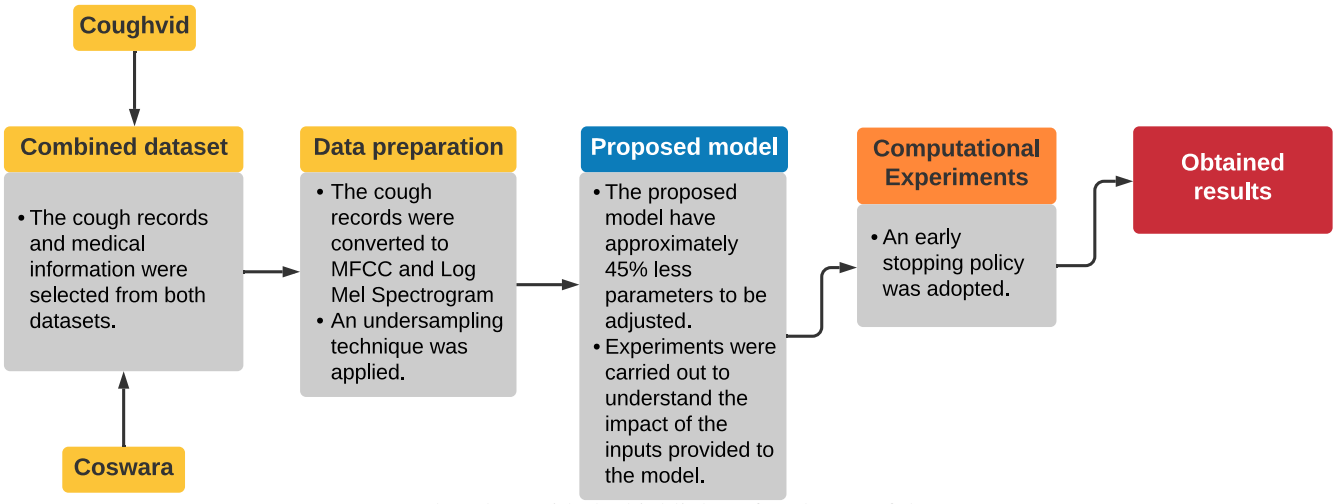

Figure 1. Flowchart with the highlights of each step of the process.

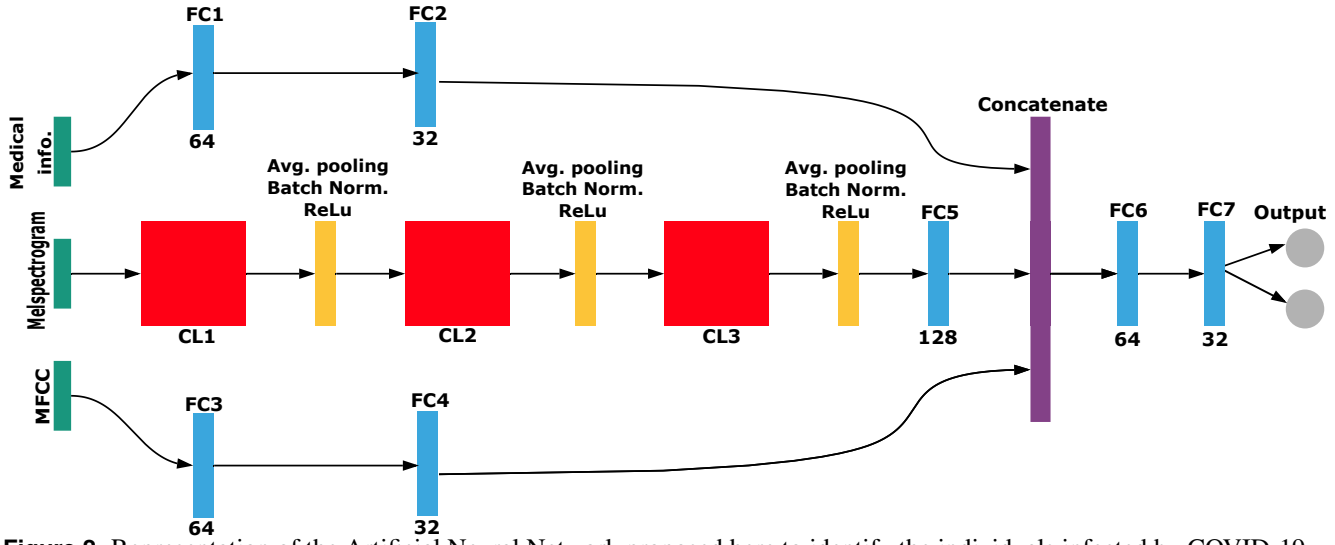

Figure 2. Representation of the Artificial Neural Network proposed here to identify the individuals infected by COVID-19. This model is based on that described in [7].

age, and the source code is public available[14]. The experiments were performed in the Google Colab development environment[15], and the model was trained with an Nvidia Tesla P100 graphics card.

In the study performed by [7], the only metric presented for the model is the mean AUC and the 95% confidence interval for 5 runs of the model. AUC is a metric based on the Receiver Operating Characteristic (ROC) Curve [15] that shows how good a model is to distinguish two classes and plots the True Positive rate against the False Positive rate. The AUC is a good way to summarize the ROC Curve in just one value that can assume a value in the interval [0, 1]. These metrics were calculated considering different random splits of the data in the training, validation, and test. In this study, in addition to the AUC, we present the model's Sensibility, Sensitivity, Positive Predictive Value (PPV), and Negative Predictive Value (NPV). Those metrics are calculated considering the number of True Positives (T_P), True Negative (T_N), False Positive (F_P), and False Negative (F_N). Sensibility (or sensitivity) is the probability of a positive prediction truly being positive, and can be calculated as $Sensitivity = T_P/(T_P + F_N)$. The Specificity is the probability of a negative test to truly be negative, and can be expressed as $Specificity = T_N/(T_N + F_P)$. Another metric, the PPV is the proportion of positive results that are true positive and can be computed as: $PPV = T_P/(T_P + F_P)$. Still, the NPV is the proportion of negative results that truly are negative: $NPV = T_N/(T_N + F_N)$. AUC is an important metric as it indicates the model's ability to differentiate between classes. Therefore, the higher the AUC, the better the method's ability to distinguish between infected and healthy individuals. In addition, PPV and NPV are important metrics to measure the model's performance in predicting correctly

[14] ⟨https://bit.ly/3LPvEwq⟩
[15] ⟨https://colab.research.google.com/⟩

Table 2. Summary of the models.

Model Name	Description
MEL + MI + MFCC	Complete model.
MEL + MI	Model composed by the networks fed with the Log Mel Spectrogram and the medical information.
MI + MFCC	Model composed by the networks fed with the MFCCs and the medical information.
MEL + MFCC	Model composed by the networks fed with the Log Mel Spectrogram and the MFCCs.

samples in the positive and negative classes, respectively.

The results found by the complete version of the proposed model, as well as its variants, are presented in Table 3. This table contains values calculated using the test sets.

Considering the mean AUC and the confidence interval of 95%, one concludes that the proposals were able to generate models with greater capacity for classifying individuals infected by COVID-19. The results show that all proposed models obtained an average AUC larger than that of BM. This is also observed in the confidence interval: the worst case is still higher than the result from the literature (there are no overlapping values). Comparing the average AUC of the MEL+MI model with the result from the literature, it is possible to see that this model obtained a result about 16% better.

The AUC results indicate that the MEL+MI model, fed with the Log Mel Spectrogram and individuals' medical information, has the greatest ability to distinguish between an infected individual and a healthy one. Therefore, from the point of view of classification, this is the model with the best performance among those proposed. However, from a medical point of view, the MFCC+MI model is the one with the highest rates of correct classification of individuals in the positive class (infected) and in the negative class (healthy), as indicated by the VPP and VPN metrics. Therefore, this is the model with the highest success rates in the existing classes, which represents a lower rate of false negatives and false positives. The false negative rate is very important, as a sample from an infected individual classified as negative by the model could represent a person potentially spreading the virus.

Furthermore, it is possible to observe that MEL+MFCC reached the worst result in the classification of people infected with COVID-19. Thus, a patient's symptoms have a great contribution to the predictive capacity of the model and, when disregarded, the model's performance decreases. Figure 3 shows the ROC curve and the average AUC obtained by the methods.

6. Conclusion

Disease detection through forced cough sounds is a recent field of research, and it gained much prominence with the emergence of the new coronavirus pandemic in 2020. Here, we propose Deep Learning topologies and a training protocol with an Early Stopping (ES) policy to classify people as healthy or sick concerning COVID-19.

The proposals were evaluated using data from the COUGHVID and Coswara databases. The databases are formed of forced cough recordings and clinical information which were combined into a single dataset. To deal with the large unbalance of the data, with 81.76% belonging to the negative class, a data augmentation technique applied to the training set was proposed. Furthermore, to understand the contribution of each input provided to the proposed algorithm, the model performance was evaluated when one of the inputs was disregarded.

To evaluate the model, the average values and the confidence interval of 95% of AUC, Sensitivity, Specificity, Positive Predictive Value (VPP), and Negative Predictive Value (VPN) were presented. The results prove that the proposed model is superior to that from the literature. The proposals obtained an average AUC larger than that reached by the baseline in all experiments. Furthermore, it was possible to notice that the MEL+MI model considering only the Log Mel Spectrogram and the medical information of an individual obtained the best results. However, the model that obtained the highest success rates for each of the classes was the MFCC+MI model, fed with MFCC and the symptoms of a patient.

Despite the good results obtained in this work, further studies are needed for adopting this approach in real-world situations. There are still limitations in understanding whether the model can identify specific cough characteristics of COVID-19, or whether it only detects an anomaly. In the latter case, other diseases that have cough as a symptom would also be classified as COVID-19. In addition, it is necessary to understand how age, gender, place of birth, and other characteristics can influence the performance of the model.

As future work, new studies can be carried out on the increase in data applied to forced cough sounds, given that many of the available databases have unbalanced data. Another possibility is increasing the spectrogram data, as in Automatic Speech Recognition applications. To understand the capacity of the model to identify characteristics present in the cough of an individual with COVID-19, a study can be carried out using a database built with audios from individuals infected with COVID-19, as well as individuals who have other diseases whose cough is presented as a symptom. Thus, it will be possible to observe whether the model can extract specific characteristics for COVID-19 identification. Furthermore, evaluating the model in other databases is an interesting research avenue.

Table 3. The mean ($\bar{X}$) and the Confidence Interval (CI) of 95% obtained by the proposed models for the metrics used.

		MEL+MFCC+MI	MEL+MI	MFCC+MI	MEL+MFCC	BM
AUC	$\bar{X}$	0.88	**0.89**	0.87	0.85	0.77
	CI	(0.88, 0.89)	(0.88, 0.89)	(0.87, 0.88)	(0.84, 0.85)	(0.74, 0.80)
Sensibility	$\bar{X}$	**0.82**	**0.82**	**0.82**	0.80	—
	CI	(0.81, 0.82)	(0.82, 0.83)	(0.82, 0.83)	(0.80. 0.81)	—
Specificity	$\bar{X}$	0.95	0.96	0.96	**0.97**	—
	CI	(0.94, 0.96)	(0.96, 0.97)	(0.95, 0.97)	(0.96, 0.98)	—
PPV	$\bar{X}$	**0.82**	**0.82**	**0.82**	0.80	—
	CI	(0.81, 0.82)	(0.82, 0.83)	(0.82, 0.83)	(0.80. 0.81)	—
NPV	$\bar{X}$	0.84	0.84	**0.85**	0.82	—
	CI	(0.84, 0.85)	(0.84, 0.85)	(0.84, 0.85)	(0.82, 0.82)	—

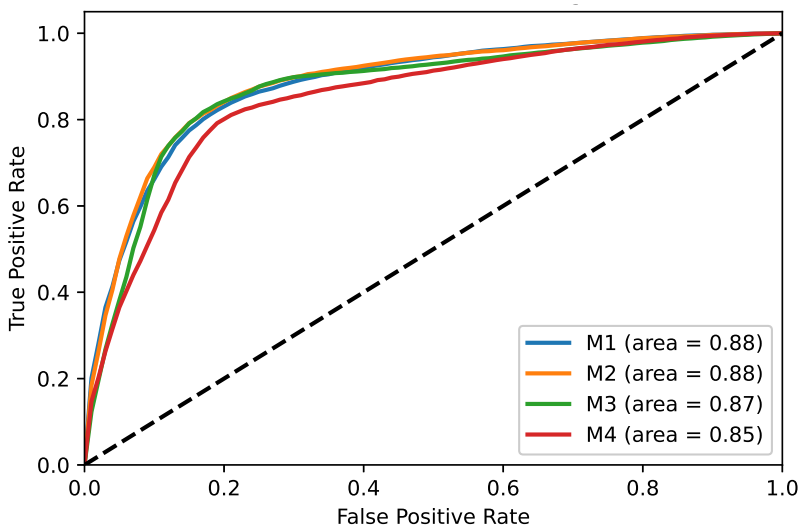

Figure 3. ROC curve and mean AUC obtained by the methods.

Acknowledgements

We thank the financial support provided by the funding agencies.

Author contributions

The authors contributed equally to this work.

References

[1] AZAM, M. A. et al. Smartphone based human breath analysis from respiratory sounds. In: *40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. USA: IEEE, 2018. p. 445–448.

[2] SIVIC, J.; ZISSERMAN, A. Efficient visual search of videos cast as text retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, USA, v. 31, n. 4, p. 591–606, April 2009.

[3] BRAGA, D. et al. Automatic detection of parkinson's disease based on acoustic analysis of speech. *Engineering Applications of Artificial Intelligence*, United Kingdom, v. 77, p. 148–158, January 2019.

[4] VASHKEVICH, M.; PETROVSKY, A.; RUSHKEVICH, Y. Bulbar als detection based on analysis of voice perturbation and vibrato. In: *2019 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*. USA: IEEE, 2019. p. 267–272.

[5] VHADURI, S. Nocturnal cough and snore detection using smartphones in presence of multiple background-noises. In: *Proc. of the SIGCAS Conference on Computing and Sustainable Societies*. New York, USA: ACM, 2020. p. 174–186.

[6] BOTHA, G. et al. Detection of tuberculosis by automatic cough sound analysis. *Physiological measurement*, United Kingdom, v. 39, n. 4, p. 045005, April 2018.

[7] CHAUDHARI, G. et al. Virufy: Global applicability of crowdsourced and clinical datasets for ai detection of

covid-19 from cough. *arXiv preprint arXiv:2011.13320*, New York, USA, November 2020.

[8] BROWN, C. et al. Exploring automatic diagnosis of covid-19 from crowdsourced respiratory sound data. In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. USA: ACM, 2020. p. 3474–3484.

[9] ORLANDIC, L.; TEIJEIRO, T.; ATIENZA, D. The coughvid crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms. *Scientific Data*, London, United Kingdom, v. 8, n. 1, p. 1–10, June 2021.

[10] SHARMA, N. et al. Coswara–a database of breathing, cough, and voice sounds for covid-19 diagnosis. *arXiv preprint arXiv:2005.10548*, New York, USA, 2020.

[11] HAN, J. et al. Sounds of covid-19: exploring realistic performance of audio-based digital testing. *npj Digital Medicine*, USA, v. 5, n. 1, p. 1–9, January 2022.

[12] LAGUARTA, J.; HUETO, F.; SUBIRANA, B. Covid-19 artificial intelligence diagnosis using only cough recordings. *IEEE Open Journal of Engineering in Medicine and Biology*, New York, USA, v. 1, p. 275–281, September 2020.

[13] MCFEE, B. et al. librosa: Audio and music signal analysis in python. In: *Proceedings of the 14th python in science conference*. USA: Citeseer, 2015. v. 8, p. 18–25.

[14] PRECHELT, L. Early stopping-but when? In: *Neural Networks: Tricks of the trade*. Berlin, Germany: Springer, 1998. p. 55–69.

[15] BRADLEY, A. P. The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, United Kingdom, v. 30, n. 7, p. 1145–1159, July 1997.