P-ISSN 1859-3585 | E-ISSN 2615-9619

DIAGNOSING PEOPLE INFECTED WITH COVID-19 BASED ON COUGH RECORDINGS

CHẨN ĐOÁN NGƯỜI NHIỄM COVID-19 DỰA TRÊN BẢN GHI TIẾNG HO

Do Manh Quang[1,*],
Vu Viet Thang[2], Ngo Thi Bich Thuy[2]

DOI: https://doi.org/10.57001/huih5804.2023.220

ABSTRACT

Rapid diagnosis of COVID-19 is crucial in preventing the spread of the disease. There are many solutions such as using early test kits or using new and modern technologies. One of the solutions, we aim to achieve is the application of artificial intelligence. There are many other types of research in this area but with the diversity and rapid growth of new strains, there is no such thing as a complete solution. For the same reason, we have joined the study to be able to diagnose people with COVID-19 early through cough recordings. In this paper, we present a CNN model and audio data augmentation in the COVID-19 diagnosis. The model uses features of MFCCs as input data. The test results of the models on the Virufy dataset, which are evaluated based on the ROC AUC results from 85.3% to 98.7%, the AUC of the public test dataset is 99.5% and the AUC of the private test dataset is 82.9%.

Keywords: *COVID-19, CNN, MFCCs, deep learning.*

TÓM TẮT

Chẩn đoán nhanh chóng COVID-19 là rất quan trọng trong việc ngăn chặn sự lây lan của bệnh. Có nhiều giải pháp như sử dụng bộ Kit xét nghiệm sớm hay sử dụng công nghệ mới, hiện đại. Một trong những giải pháp chúng tôi hướng tới là ứng dụng trí tuệ nhân tạo. Đã có nhiều nghiên cứu khác trong lĩnh vực này nhưng với sự đa dạng và tốc độ phát triển nhanh chóng của các chủng mới, không có giải pháp nào là hoàn chỉnh. Cũng vì lý do đó, chúng tôi đã tham gia nghiên cứu để có thể chẩn đoán sớm người mắc COVID-19 thông qua các bản ghi âm tiếng ho. Trong bài báo này, chúng tôi trình bày một mô hình CNN và kỹ thuật tăng cường dữ liệu âm thanh trong chẩn đoán COVID-19. Mô hình sử dụng các tính năng của MFCC làm dữ liệu đầu vào. Kết quả thử nghiệm của các mô hình trên bộ dữ liệu "Virufy" được đánh giá dựa trên kết quả ROC AUC từ 85,3% đến 98,7%, AUC của bộ dữ liệu thử nghiệm công khai là 99,5% và AUC của bộ dữ liệu thử nghiệm riêng tư là 82,9%.

Từ khóa: *COVID-19, CNN, MFCCs, học sâu.*

[1]R&D Team, Giao Hang Tiet Kiem Joint Stock Company, Vietnam
[2]Faculty of Information Technology, Hanoi University of Industry, Vietnam
[*]Email: quangdm2@ghtk.co
Received: 10/3/2023
Revised: 25/6/2023
Accepted: 25/11/2023

1. INTRODUCTION

Currently, until August 11, 2021, in the world, there are more than 204.000.000 people infected and more than 4.000.000 deaths because of COVID-19. In Vietnam, there are more than 225,000 cases and 3,757 deaths due to COVID-19. Rapid diagnosis of COVID-19 will be extremely useful for localizing and preventing the spread of the disease. Artificial intelligence is a good method to diagnose COVID-19 through cough recordings. A few recent types of research [1-5] have shown that the diagnosis of COVID-19 through cough sounds has high accuracy.

In the results of the research the authors Jordi Laguarta, Ferran Hueto, and Brian Subirana from MIT [1]. The diagnosis of COVID-19 used an AUC of 0.97 using the input MFCCs images for 3 ResNet50 models with parallel training and binary screening diagnostics. In the results of the research the authors Vipin Bansal, Gaurav Pahwa, and Nirmal Kannan from Altran [3]. They propose a basic CNN architectural model for COVID-19 diagnosis with input is MFCCs images, the proposed model results reached an accuracy of 0.7. In this research, We use the public cough dataset available on GitHub by the Virufy team [6]. This repository contains everything necessary to start writing a COVID-19 detection model. The goal of the model is to take audio files to predict whether they are infected with COVID-19 or not.

2. RELATED WORK

The recent works widely use the MFCC features and convolutional neural networks (CNNs) in the model. In the results of the research the authors Jordi Laguarta, Ferran Hueto, and Brian Subirana from MIT [1]. Cough recordings are transformed with Mel Frequency Cepstral Coefficient and inputted into a Convolutional Neural Network (CNN) based architecture made up of one Poisson biomarker layer and 3 pre-trained ResNet50's in parallel, outputting a binary pre-screening diagnostic. The CNN-based models have been trained on 4256 subjects and tested on the remaining 1064 subjects of the dataset. Transfer learning was used to learn biomarker features on a larger dataset, previously successfully tested in our Lab on Alzheimer's, which significantly improves the COVID-19 discrimination accuracy of architecture. When validated with subjects diagnosed using an official test, the model achieves COVID-19 sensitivity of 98.5% with a specificity of 94.2% (AUC: 0.97).

For asymptomatic subjects, it achieves a sensitivity of 100% with a specificity of 83.2%.

In the results of the research the authors Vipin Bansal, Gaurav Pahwa, and Nirmal Kannan from Altran [3]. They propose a CNN-based audio classifier using the open cough dataset. The dataset is labeled manually into cough categories with final labeling into Covid and Non-Covid classes. The two approaches proposed in this paper are based on mfcc features and spectrogram images as input to the CNN network. MFCC approach produced 70.58% test accuracy with 81% sensitivity and is better than the spectrogram-based approach. In the results of the research the authors Jiaxuan Guo, Shuo Xin from Applied Physics Stanford University [13]. They made several improvements to the original model of Virufy by processing mel-spectrogram with DenseNet, as well as other adjustments. Further, we augment the dataset by splitting individual coughs. The model improvement and data augmentation lead to better performance. The original Virufy model trained on the same dataset is also shown for comparison. Result got a ROC AUC of 0.88.

3. METHODS

3.1. COVID-19 Cough Dataset

In this paper, We use the Virufy team cough dataset to diagnose COVID-19. This cough audio dataset includes *.webm, *.ogg, *.mp3, *.m4a, *.wav formats. We only use the properties cough_detected, patient_id, audio_path, and pcr_test_result_inferred to extract features of the data. In this dataset, We found that the audio recordings included both cough and no cough recordings. To screen the dataset, We used the cough_detected attribute with an accuracy cough detection rate of 0.99 or higher. The pcr_test_result_inferred attribute includes positive, negative, untested, or pending. So We will only filter out the records that have positive and negative PCR test results. In Figure 1, We found that the number of positive samples was less than the number of negative samples. The Positive / Negative ratio is 0.25. That is why We will use the audiomentations library [7] sound library to augment the positive cough sample data. The following properties for enhancing audio data include:

- *Add Gaussian Noise*: Add gaussian noise to the samples.
- *Time Stretch*: Time stretches the signal without changing the pitch.
- *Pitch Shift*: Pitch shifts the sound up or down without changing the tempo.
- *Shift*: Shift the samples forwards or backward, with or without rollover.
- *Trim*: Trim leading and trailing silence from an audio signal.
- *Gain*: Multiply the audio by a random amplitude factor to reduce or increase the volume. This technique can help a model become somewhat invariant to the overall gain of the input audio.
- *Gain Transition*: Gradually change the volume up or down over a random period. Also known as fade in and fade out. The fade works on a logarithmic scale, which is natural to human hearing.
- *Polarity Inversion*: Flip the audio samples upside-down, reversing their polarity. In other words, multiply the waveform by -1, so negative values become positive, and vice versa. The result will sound the same compared to the original when played back in isolation. However, when mixed with other audio sources, the result may be different. This waveform inversion technique is sometimes used for audio cancellation or obtaining the difference between two waveforms. However, in the context of audio data augmentation, this transform can be useful when training phase-aware machine learning models.

After enhancing the audio data in Figure 2, the Positive / Negative ratio is 0.76. With the above ratio, it will help train the deep learning model more accurately when the positive and negative samples are not too different.

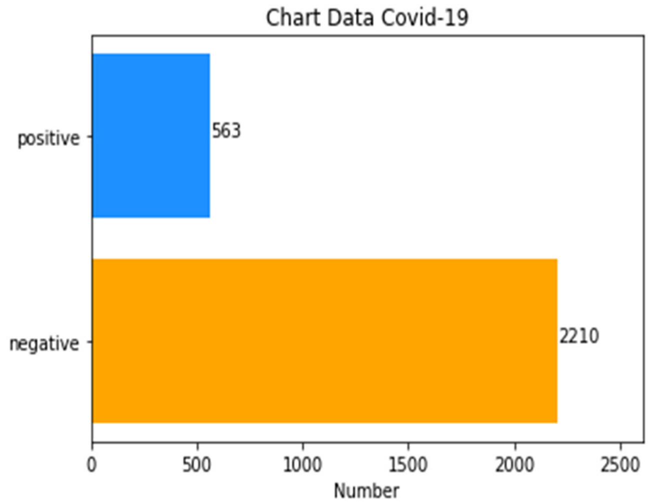

Figure 1. Initial cough data

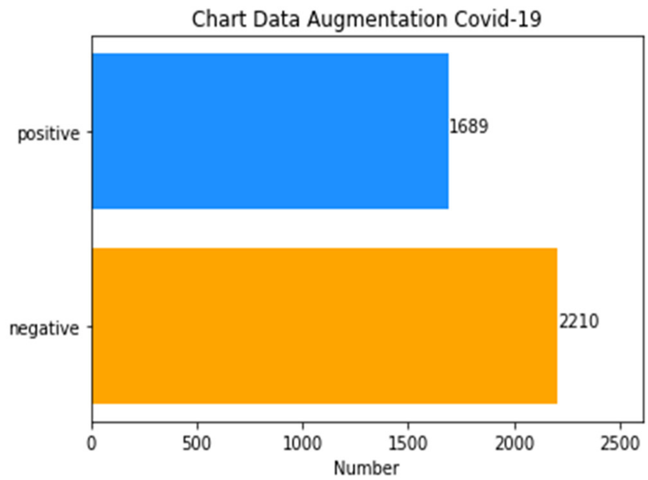

Figure 2. Cough data after augmentation

In the cough recording, the recording time fluctuates between 5 and 7 seconds. So We choose a size of 154350 which equates to 7 seconds when sampled at 22050 samples/sec as a constant amount for all input data. For feature extraction of audio recordings, We use MFCCs. MFCCs [8] is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear Mel scale of frequency. Mel-frequency cepstral coefficients (MFCCs) are coefficients that collectively make up an MFC. We use the librosa library [9] to extract 13 features per audio frequency band. The 7-second audio samples are transformed into a matrix of 13x702 MFCCs. The MFCCs features will be the input data and processed to produce the probability score of the classification as the output.

3.2. COVID-19 Experimental Model Architecture

We tested models such as SVM, RandomForest, CNN, and CNN with Data Augmentation on the private test dataset during the research process. The obtained results show that CNN combined with Data Aug has better diagnostic performance than the other models, the results are in Table 1. From the obtained results, we propose the CNN with Data Augmentation. The cough audio recordings are converted to a matrix of MFCCs used as input data. We use this input data for the CNN model in Figure 3 to extract features and use the sigmoid classifier to predict the probability of COVID-19 infection of the cough sample. The function classifier using activation sigmoid will use Equation 1. The basic CNN architecture in Table 2 with 3 convolutional layers, 2 pooling layers, and 3 fully connected layers will be used to calculate the probability of COVID-19 infection from the MFCCs matrix. To minimize the linearity of the data, the relu function is activated at each convolutional layer to help the feature extraction achieve better results.

$$S(x) = \frac{1}{1 + e^{-x}} \quad (1)$$

Table 1. Experimental results of diagnostic models

STT	Model	Results
1	SVM	80.9%
2	RandomForest	80.93%
3	CNN	81.2%
4	CNN with Data Augmentation	82.9%

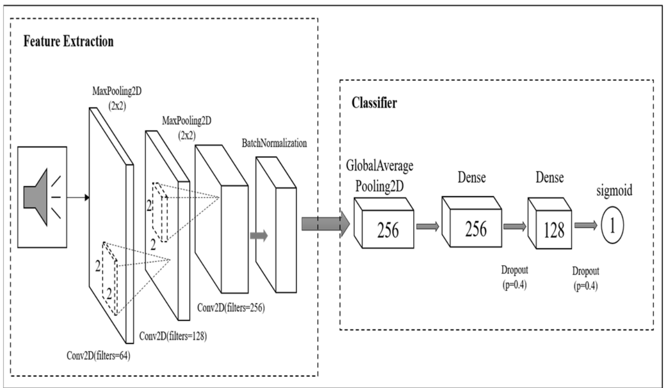

Figure 3. Proposed CNN model architecture

Table 2. The proposed theoretical architecture of the CNN model

Network type	Input value	Filters	Kernel	Activation
MFCCs Input	702x13x1			
Conv2D		64	2x2	relu
MaxPooling			2x2	
Conv2D		128	2x2	relu
MaxPooling			2x2	
Conv2D		256	2x2	relu
BatchNormalization				
GlobalAveragePooling2D	256			
Dense	256			relu
Dropout	0.4			
Dense	128			relu
Dropout	0.4			
Dense	1			sigmoid

We have tested using the Global Max Pooling2D layer but the diagnostic performance is not as good as the Global Average Pooling2D layer. With The function classifier using activation sigmoid, We use 1 BatchNormalization layer and 2 Dropout layers to reduce overfitting in the learning process of the model. It helps improve the performance of the model.

3.3. Results

We divide the random dataset into 2 parts: the *training dataset* and the *private test dataset* in a ratio of 90/10. In addition, We randomly generate 1 more public test dataset representing 20% of the training dataset. The purpose and parameters of the dataset will be described in Table 3.

Table 3. Description of training and evaluation dataset

Dataset	Target	Positive	Negative
Train	Train the proposed model	1521	1988
Public Test	Evaluate the learning ability of the model	304	398
Private Test	Evaluation of diagnostic performance capabilities with blinded data	168	222

During training, we use a Stratified K-folds cross-validation process. With K-fold 5 data will be divided into 5 parts. Provides train/test indices to split data into train/test sets. This cross-validation object is a variation of KFold that returns stratified folds. The folds are made by preserving the percentage of samples for each class. The model will be trained with 50 epochs, the training batch size is 64, and shuffle labels are True during training to minimize overfitting. We will train the model a total of 250 times. The public test dataset is selected from the training dataset to test the model's ability to learn and the private test dataset will not be trained.

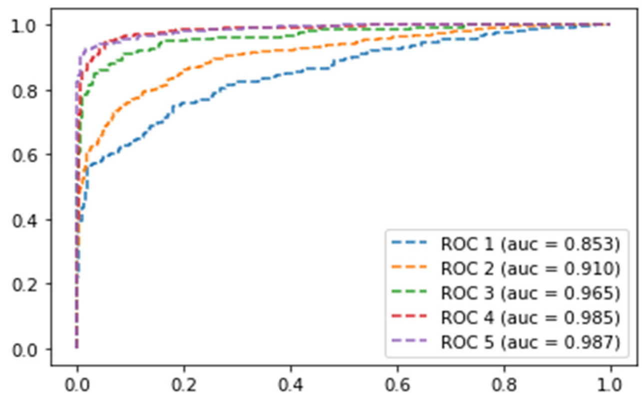

Figure 4. Training with a K-fold is 5

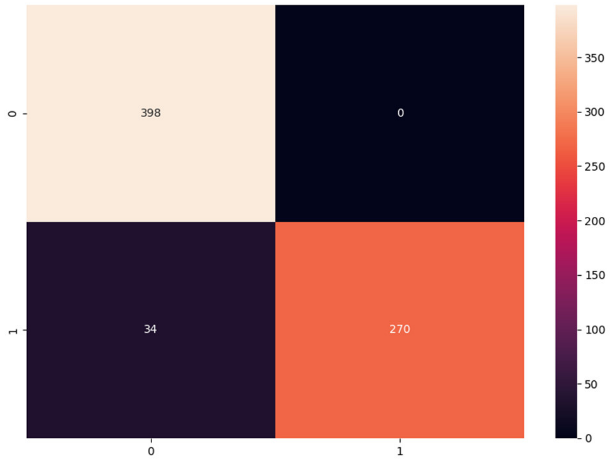

Figure 5. Confusion matrix of the public test dataset

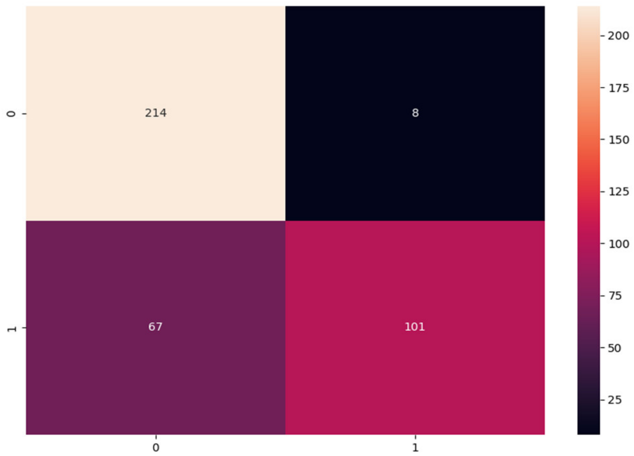

Figure 6. Confusion matrix of the private test dataset

After model training 5 times, We obtained ROC AUC from 85.3% to 98.7%, ROC AUC average is 94%. The average of 5 AUC training sessions was 94%. On each dataset, We use Precision Equation 2, Recall Equation 3, F1-score Equation 4, ROC AUC [10], Accuracy Equation 5, and Confusion matrix to evaluate the model. TN, FN, FP, and TP mean True Negatives, False Negatives, False Positives, and True Positives. We use the sklearn library [11] to describe the model evaluations.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (3)$$

$$\text{F}_1 = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \quad (4)$$

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (5)$$

With a public test dataset to evaluate the model's learning ability, the ROC AUC is 99.5%. Figure 5 shows that the learning model on the training dataset is good. The rate of recognizing negative samples is 0.89, positive samples are 1. Table 4 shows that the F1-scores are all over 90%, showing that the learning and diagnosis on the learned samples are very good.

Table 4. Results of indicators reported by public test dataset

	Precision	Recall	F_1-score
Class 0	0.92	1.00	0.96
Class 1	1.00	0.89	0.94
Accuray			0.95

With a private test dataset to evaluate the diagnosis ability with new data, the ROC AUC is 82.9%. Figure 6 shows a model that calculates the relative probability of COVID-19 infection. The rate of recognizing negative samples is 0.6, positive samples are 0.96. The rate of miscalculation in positive samples is quite high up to 0.4. Table 5 shows that the F1-scores are all over 70%, showing that the calculation of the possibility of COVID-19 infection is relative.

Table 5. Results of indicators reported by private test dataset

	Precision	Recall	F_1-score
Class 0	0.76	0.96	0.85
Class 1	0.93	0.60	0.73
Accuracy			0.81

Although our detection rate of infected people is not as high as that of the authors in [1] (98.5%) and [3] (94% with 13 features), however, when we experiment, we use a lot of control data. more experimental, and dependent cough-form data of different human races around the world.

4. CONCLUSIONS

With Virufy's dynamic public dataset, the challenges include a limited number of positive samples, different recording times and noisy cough recordings, etc. However, that is understandable as this is helping out the community in any way possible. Restricting positive cough samples makes it difficult to calculate COVID-19 infection rates on a private test dataset. But the proposed model based on the novel CNN architecture and audio data augmentation with the input of the MFCCs matrix still achieved the relative accuracy rate of AUC 82.9% and the F1-score over 70% for positive and negative samples other private test dataset (this is all source code of the project at Github [12]). Diagnosis of COVID-19 by cough is a rapid, non-contact screening diagnostic method. The method can be applied to smart mobile devices. This will reduce the burden on health systems around the world.

In the near future, We will continue to try to collect more cough recording samples from hospitals in Vietnam. Experiment to extract more Chroma Features, Spectral Contrast, RMS, Melspectrogram, etc to improve the model.

ACKNOWLEDGMENTS

In this research, we would like to thank the Virufy team for collecting cough samples from around the world and for building an API that makes it easy for us to access and process cough samples.

REFERENCES

[1]. Jordi Laguarta, Ferran Hueto, And Brian Subirana, 2020. *Covid-19 Artificial Intelligence Diagnosis Using Only Cough Recordings.* IEEE Open Journal Of Engineering In Medicine And Biology, Vol. 1.

[2]. Madhurananda Pahar, Marisa Klopper, Robin Warren, Thomas Niesler, 2021. *Covid-19 Cough Classification Using Machine Learning And Global Smartphone Recordings*. Comput Biol Med. 135:104572. doi: 10.1016/j.compbiomed.2021.104572.

[3]. Vipin Bansal, Gaurav Pahwa, Nirmal Kannan, 2020. *Cough Classification For Covid-19 Based On Audio Mfcc Features Using Convolutional Neural Networks*. 2020 IEEE International Conference On Computing, Power And Communication Technologies (Gucon) Galgotias University, Greater Noida, Up, India.

[3]. Rob Dunne, Tim Morris, Simon Harper, 2020. *High Accuracy Classification Of Covid-19 Coughs Using Mel-Frequency Cepstral Coefficients and A Convolutional Neural Network With A Use Case For Smart Home Devices*. https://doi.org/10.21203/rs.3.rs-63796/v1

[4]. Piyush Bagad, Aman Dalmia, Jigar Doshi, Arsha Nagrani, Parag Bhamare, Amrita Mahale, Saurabh Rane, Neeraj Agarwal, Rahul Panicker, 2020. *Cough Against Covid: Evidence Of Covid-19 Signature In Cough Sounds.* Arxiv Preprint Arxiv:2009.08790.

[5]. *Virufy Dataset Covid-19* (https://github.com/virufy/virufy-covid).

[6]. *Library Audiomentations* (https://github.com/iver56/audiomentations).

[7]. Min Xu, et al., 2004. *HMM-based audio keyword generation* (PDF). In Kiyoharu Aizawa; Yuichi Nakamura; Shin'ichi Satoh (eds.). Advances in Multimedia Information Processing - PCM 2004: 5th Pacific Rim Conference on Multimedia. Springer. ISBN 978-3-540-23985-7. Archived from the original (PDF) on 2007-05-10.

[8]. *Library Librosa* (https://librosa.org/doc/main/index.html).

[9]. *Area Under the Receiver Operating Characteristic curve.*

[10]. *Library Sklearn* (https://scikit-learn.org/).

[11]. *Source code*: https://github.com/DoManhQuang/diagnosis_covid_with_cough

[12]. *Jiaxuan Guo, Shuo Xin*. COVID-19 Cough Recognition. CS230. Stanford University.

THÔNG TIN TÁC GIẢ

Đỗ Mạnh Quang[1], Vũ Việt Thắng[2], Ngô Thị Bích Thúy[2]

[1]R&D Team, Công ty Cổ phần giao hàng tiết kiệm

[2]Khoa Công nghệ thông tin, Trường Đại học Công nghiệp Hà Nội