



## COVID-19 Detection from Cough Sounds Using XGBoost and LSTM Networks

Elmoundher Hadjaidji\*<sup>ORCID</sup>, Mohamed Cherif Amara Korba<sup>ORCID</sup>, Khaled Khelil<sup>ORCID</sup>

Faculty of Science and Technology, LEER Lab, Mohamed Cherif Messaadia University, Souk Ahras 41000, Algeria

Corresponding Author Email: [h.elmoundher@univ-soukahras.dz](mailto:h.elmoundher@univ-soukahras.dz)

Copyright: ©2024 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ts.410234>

### ABSTRACT

**Received:** 25 April 2023  
**Revised:** 21 October 2023  
**Accepted:** 17 February 2024  
**Available online:** 30 April 2024

#### Keywords:

*COVID-19 detection, cough detection, deep learning, long short-term memory (LSTM), XGBoost, acoustic features, digital health*

COVID-19, a contagious respiratory virus with symptoms like a dry cough, prompted intensive diagnostic efforts. Current standards fall short in controlling transmission, driving researchers to explore automated identification methods. In this work, using artificial intelligence and audio signal processing techniques, an automated system is developed. After extracting cough segments from audio recordings through the eXtreme Gradient Boost algorithm, the system attempts to detect COVID-19 employing a deep learning-based approach known as Long Short-Term Memory. In particular, the XGBoost model identifies the cough segment, and the LSTM-based model conduct binary classification on it to establish whether a person is positive or negative for COVID-19. To assess the proposed detection scheme, several experiments were conducted with the use of two publicly available cough sound datasets, namely COUGHVID and VIRUFY, which were collected from coronavirus-infected and non-infected persons through a large-scale crowdsourced campaign. The suggested system results were validated through comparisons with prior studies, demonstrating its strong performance even in noisy environments. Additionally, the obtained results indicate that the proposed method for detecting COVID-19 performs admirably under ideal conditions, achieving approximately 97% accuracy on the VIRUFY dataset and an impressive classification rate of nearly 88% on the COUGHVID dataset.

## 1. INTRODUCTION

The SARS-CoV-2 (COVID-19) virus, identified in December 2019, has affected hundreds of millions of individuals worldwide, resulting in a global-scale mortality. According to medical references [1-3], COVID-19 infection is characterized by a variety of symptoms, including fever, fatigue, and a persistent cough [4, 5]. Understanding these symptoms and detecting them early is crucial for accurate diagnosis. Generally, the COVID-19-associated cough is typically dry, without mucus, and can result in intense, recurring coughing episodes that may endure for an extended duration. Hospitals and labs use various testing methods, such as PCR, Chest CT imaging, and antigen tests, which vary in effectiveness, cost, and speed [6, 7]. Utilizing artificial intelligence (AI) for COVID-19 detection through cough analysis is a promising alternative [8]. AI techniques, primarily machine and deep learning-based, are user-friendly, non-invasive, offer rapid detection, and ensuring swift results. In this regard, we will introduce several studies from the literature that focus on the automatic detection of COVID-19 based on cough sounds.

A review article [9] discussed recent research on COVID-19 diagnosis using AI and respiratory sound analysis. The researchers [10] outlined AI-driven efforts for diagnosing COVID-19, offering a valuable automated system that utilizes non-invasive biological signals from both speech and non-speech audio. In their study, using cough signals, Pahar et al.

[11] employed two datasets, Coswara and Sarcos, derived from smartphone voice recordings. Employing seven machine learning classifiers with various features, their top-performing classifier, ResNet50, achieved an impressive area under the ROC Curve (AUC) of 0.98, while the LSTM-based classifier reached an AUC of 0.94. A study by Imran et al. [12] employed a deep transfer learning-based multi-class classifier to diagnose COVID-19 based on cough, achieving an accuracy of 92.64%. MFCC features were extracted from cough recordings and processed through a Convolutional Neural Network (CNN) architecture featuring three pre-trained ResNet50 models and a parallel Poisson biomarker [13]. The results exhibited a remarkable COVID-19 sensitivity of 98.5% with a specificity of 94.2% and an AUC of 0.97. For asymptomatic individuals, the sensitivity reached 100% with a specificity of 83.2%.

Ponomarchuk et al. [14] claim to introduce a new deep learning and signal processing-based approach to detect COVID-19, as well as providing denoising methods, cough detection and classification. A classification was made between a dry cough and wet cough by applying some modern and traditional machine learning methods to the COUGHVID database, where it has been shown that the decision tree models are superior to the rest of the models in terms of overall performance [15]. Based on logistic regression and support vector machines for acoustic data, and decision tree models for symptoms data, the researchers [16] proposed a multi-modal diagnostic for COVID-19. Using Coswara dataset, an AUC of

0.92 has been reached.

To the best of our knowledge, all the examined relevant COVID-19 diagnosis methods utilizing cough signals are vulnerable to environmental noise and undesirable signals, which could compromise the efficiency of AI-driven COVID-19 detection systems. To address this noise issue, this study suggests an automated system referred to as XGBoost-LSTM, which relies on XGBoost and LSTM algorithms for COVID-19 detection using cough audio signals. Specifically, the main contributions of this work are:

(1) Generally, datasets of cough sounds are often contaminated by environmental noise and undesirable parts (laughter, speech, music, and loud instruments), which may affect drastically AI system training. Thus, an XGBoost-based signal preprocessing system is used to accurately detect cough segments in recordings. Then, the scheme evaluates cough segment quality by estimating the signal-to-noise ratio (SNR) of the audio signal, and effectively preserves records that actually contain cough sounds.

(2) COVID-19 affects males and females differently [17-19], leading to a slight distinction in their cough signals. As a result, two separate classification models have been developed for each gender.

(3) To classify the detected cough segments extracted from the acquired audio recordings into COVID-19+ and COVID-19-, the LSTM neural network, capable of modeling time series variations, is employed along with MFCC features and log energy.

The remainder of the paper is laid out as follows. A general idea of the proposed framework is depicted in section 2. In particular, Cough detection and Cough Classification are described. In Section 3, the experimental results with a detailed discussion are reported. The conclusion is described in Section 4.

## 2. PROPOSED METHOD

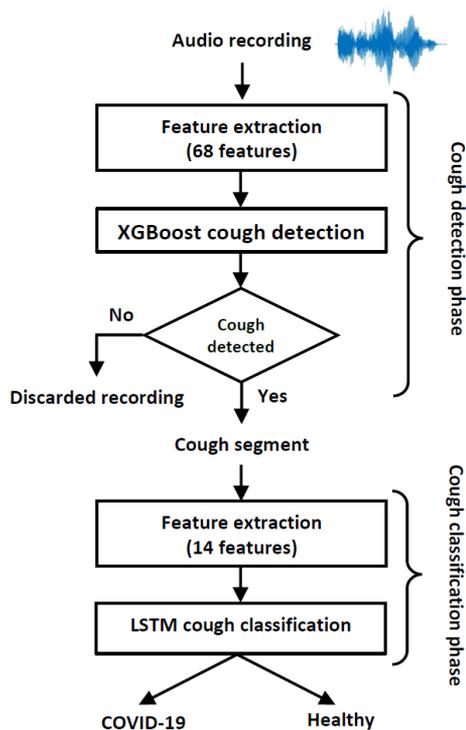


Figure 1. XGBoost-LSTM diagnostic system flowchart

In this section, the two main processing stages of the proposed system are described, as shown in Figure 1. The first stage detects cough segments in audio recordings and estimates their quality by computing the SNR. Then, the second part of the system performs a classification of previously detected cough segments to determine whether or not they come from a person with COVID-19.

### 2.1 Cough detection phase

Cough detection is an important step in any automatic COVID-19 detection systems. Commonly, a typical cough signal usually has three phases related to the person state of health: an early loud explosive first stage, an intermediate stage, and in most cases a later audible phase. Moreover, since the cough classification is highly dependent on identified cough segments and their assessed quality, the detection phase should take into account the cough signal temporal variations. Figure 2 depicts in detail the different processing steps to detect the cough portions in an audio file. In addition, the signal quality is measured using the signal to noise ratio (SNR).

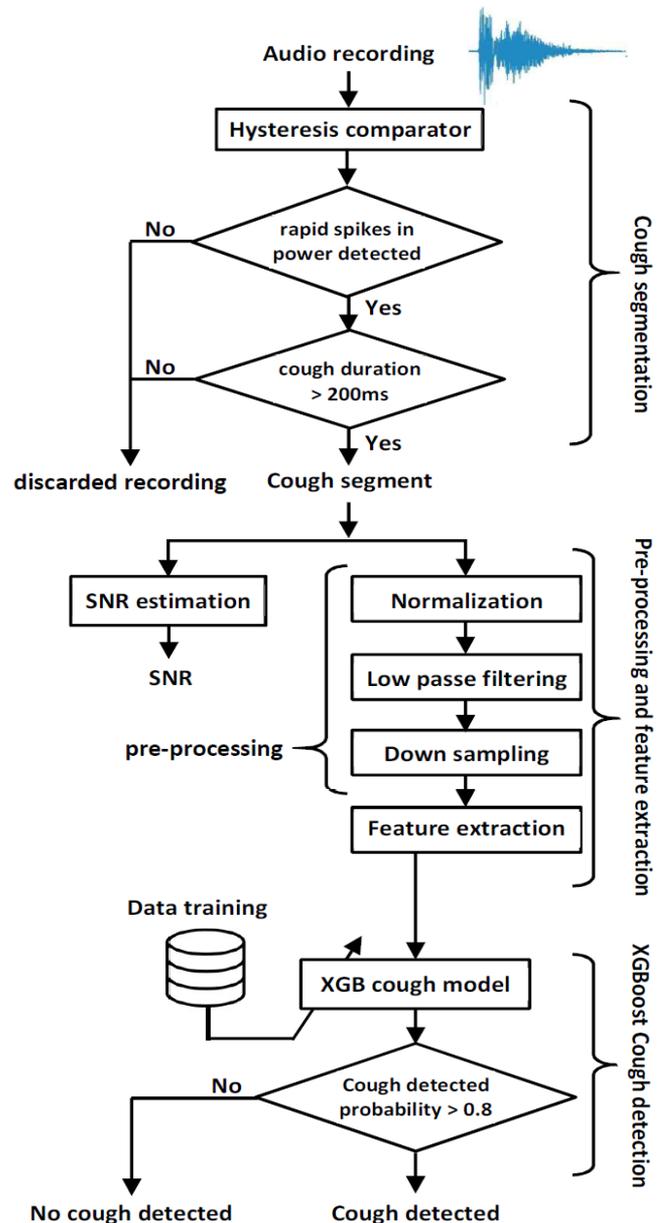


Figure 2. Cough detection system flowchart

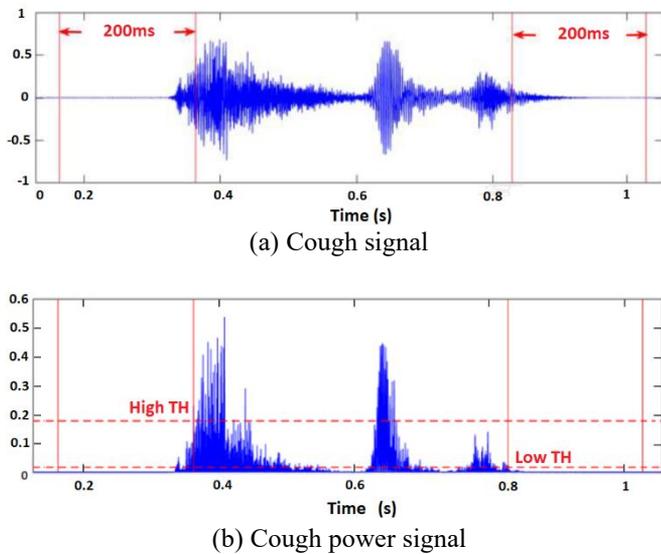
### 2.1.1 Cough segmentation phase

Using the signal power and a hysteresis comparator, cough segmentation is accomplished. The algorithm looks for regions of the signal with rapid power spikes, which are typical of cough sounds. It employs two thresholds: the high threshold  $TH_H$  for detecting the beginning of a cough and the low threshold  $TH_L$  for detecting the end of a cough segment. The two thresholds are defined by:

$$\begin{aligned} TH_H &= RMS_S \cdot M_H \\ TH_L &= RMS_S \cdot M_L \end{aligned} \quad (1)$$

where, the  $RMS_S$  represents the root mean square of the audio signal. The adopted values of  $M_H$  and  $M_L$  are chosen empirically as in the study of Orlandic et al. [20] and are equal to 2 and 0.2 respectively which allow efficient segmentation of the employed dataset.

The onset cough instant is detected when the signal power is above the high threshold, while its end is determined by the signal power falling below the lower threshold for a continuous duration of 0.01 second, ensuring the accurate detection of the end of the cough signal. Note that cough sounds lasting less than 0.2 second are discarded. However, to guarantee that the detected cough is not cut short, the algorithm considers a duration of 0.2 second before and after the signal as a part of the cough. Figure 3 depicts the process of detecting the onset and end times of the cough segment.



**Figure 3.** Detection of the onset and end times of the cough segment

### 2.1.2 Pre-processing and feature extraction phase

This phase includes three distinct stages: pre-processing, SNR estimation, and feature extraction. In this section, we will provide a detailed description of each step.

#### Pre-processing stage

This phase aims to minimize variations due to different recording conditions. Each recording is first normalized to the  $[-1, 1]$  range, and then filtered by a 4th order low-pass Butterworth filter, with a cut-off frequency of 6kHz. Subsequently, all recordings are downsampled to a 12kHz frequency since the relevant cough-related information is

practically present below 4kHz [20].

#### SNR estimation stage

Crowdsourced cough recordings are highly affected by background environmental noises such as: laughter, speech, music, phone ringtones, and other types of noise, affecting considerably the performance of the classifier, hence an estimation of the noise is essential. Hence, the SNR is estimated by calculating the ratio of the power of the cough portions to the rest of the signal samples power, suspected to be background noise. Note that a recording may contain one or more portions of cough, so the SNR can be expressed as:

$$SNR = 20 \cdot \log_{10} \left( \frac{\sqrt{\sum_{c=1}^C \left( \frac{1}{N_c} \sum_{k=1}^{N_c} x_c(k)^2 \right)}}{\sqrt{\frac{1}{N_n} \sum_{k=1}^{N_n} x_n(k)^2}} \right) \quad (2)$$

where,

- $x_c(k)$ : Cough portion samples
- $c$ : Cough portion index
- $C$ : Number of cough portions
- $N_c$ : Number of samples within a cough portion
- $x_n(k)$ : Noise samples
- $N_n$ : Number of noise samples

#### Feature extraction stage

The audio features used are mainly derived from different studies related to the analysis, segmentation, and classification of cough signals. In this work, 68 audio features, reported in Table 1, are extracted for each cough segment. The temporal features used in the study of Chatzarrin et al. [21] represent the 19 peaks detected in the energy envelope (EPPD) of the cough signal, with the aim of differentiating the sounds of dry cough from wet cough, because in an early stage COVID-19 can cause a dry cough in most cases. The spectral features outlined in the study of Monge-Álvarez et al. [22] depict the power spectral density within 8 specific energy bands, which have been chosen to ensure robust cough segmentation in various noisy environments. MFCC features are widely used with deep learning algorithms for cough detection and classification [23-26], the rest of the features are mainly used in audio signal analysis and in different natural language processing domains.

### 2.1.3 XGBoost cough detection model

Based on the audio features mentioned earlier, the cough detection is performed using the XGBoost model and COUGHVID dataset [26], which contains audio recordings of coughs, and was developed for COVID-19-related research. XGBoost provides the probability that a particular recording contains cough noises: if the probability is greater than 0.8, the recording is assumed to have at least one cough. To train this model, a set of 215 audio files from dataset are, first, randomly selected. Then, each audio file is classified as a cough sound if it contains at least one cough, otherwise it is considered as a non-cough sound. This procedure yielded a roughly balanced sample of 121 cough sounds and 94 non-cough sounds such as speaking, laughing, and silence. The resulting model is 95.5% accurate, indicating that it successfully removes the vast majority of data that do not contain cough while retaining the records that do contain cough.

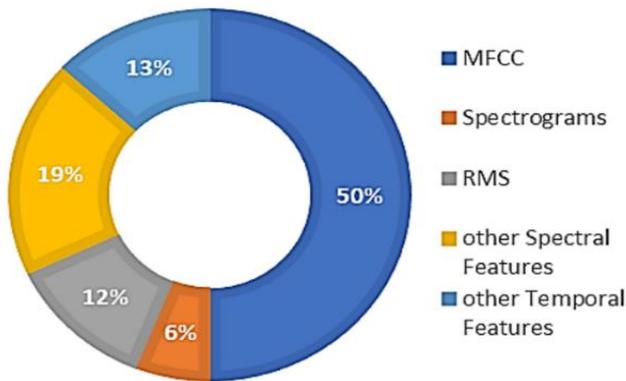
**Table 1.** Cough detection features

Features	Reference	Count	Computation Parameters
MFCC	[27]	26	Mean and St. dev of 13 MFCCs over time
EEPD	[21]	19	BPF intervals in 50-1000 Hz.
Power Spectral Density	[22]	8	Frequency bands (Hz): 0-200, 300-425, 500-650, 950-1150, 1400-1800, 2300-2400, 2850-2950, 3800-3900
RMS Power	[27]	1	None
Zero Crossing Rate	[27]	1	None
Crest Factor	[27]	1	None
Recording Length		1	None
Dominant Frequency	[27]	1	None
Spectral Centroid	[27, 28]	1	None
Spectral Rolloff	[27, 28]	1	None
Spectral Spread	[27, 28]	1	None
Spectral Skewness	[27, 28]	1	None
Spectral Kurtosis	[27, 28]	1	None
Spectral Bandwidth	[28]	1	None
Spectral Flatness	[27, 28]	1	None
Spectral Std Dev	[27, 28]	1	None
Spectral Slope	[27, 28]	1	None
Spectral Decrease	[27, 28]	1	None

**2.2 Cough classification phase**

**2.2.1 Feature extraction**

A crucial stage in any classification system is feature extraction. This study utilizes MFCC, the predominant feature for COVID-19 identification through cough sound (Figure 4 [29]). The cough segment is divided into 30 ms frames with a 13 ms overlap, and these frames are smoothed using a Hamming function. Subsequently, a Discrete Fourier Transform (DFT) computes the magnitude spectrum, which is further processed through Mel-scale triangular filters designed for auditory-guided analysis. These filters also include pre-emphasis to equalize high-frequency components. The resulting log energy from this process is then subjected to a Discrete Cosine Transform (DCT) for decorrelation. Each frame yields 13 MFCC coefficients along with log energy.



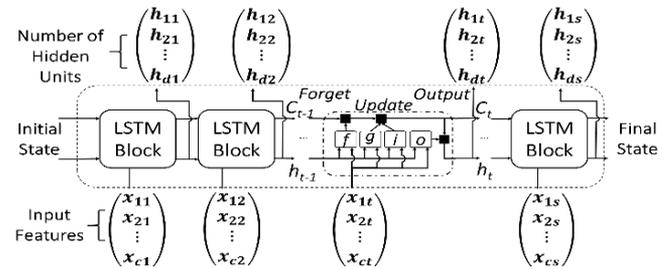
**Figure 4.** Acoustic features used for COVID-19 detection

**2.2.2 Description of the LSTM classifier**

In this work, LSTM networks, well suited for classification and time series prediction, are used to distinguish between COVID-19 cough and non-COVID-19 cough. It is a special type of recurrent neural network (RNN) known to have the ability to learn long-term sequences [30]. As previously mentioned, coughing generally comprises three distinct stages, which LSTM perceives as relatively stable time series with three states. The stability of these states depends on the specific respiratory tract condition causing irritation. The

subsequent section will provide a detailed description of the LSTM Layer Architecture.

The flow of a time series  $X$  with  $C$  features (channels) of length  $S$  through an LSTM layer is depicted in Figure 5. In this diagram,  $h_t$  and  $C_t$  denote the output (also known as the hidden state) and the cell state at time step  $t$ , respectively. In our work,  $C$  and  $S$  represent the dimensions of feature matrices extracted using MFCC and log energy.



**Figure 5.** LSTM Layer Architecture

The first LSTM block computes the first output and the updated cell state using the network initial state and the first-time step of the sequence. Then, the block computes the output and the updated cell state  $C_t$  at time step  $t$  using previous state of the network ( $C_{t-1}, h_{t-1}$ ) and the next time step of the sequence.

The LSTM layer output for this time step is stored in the hidden state at time step  $t$ , and the preceding time steps information is also saved in the cell state. At each time step, the layer modifies the cell state by adding or removing information, employing gates to control these updates.

Figure 5 illustrates the data flow at each the time step  $t$ . The block diagram shows how gates for a cell and hidden states are forgotten, updated, and ejected. In addition, Figure 5 illustrates the LSTM cell block diagram comprising four components responsible for controlling the layer cell state and hidden state: the input gate, the forget gate, the cell candidate, and the output gate, as defined by the following equations:

– Input gate ( $i$ ): Denotes the update level of the cell state

$$i_t = \sigma(W_i x_t + R_i h_{t-1} + b_i) \tag{3}$$

– Forget gate ( $f$ ): Represents the cell state reset control level (forget)

$$f_t = \sigma_g(W_f x_t + R_f h_{t-1} + b_f) \quad (4)$$

– Cell candidate ( $g$ ): Adds information to cell state

$$g_t = \sigma_c(W_g x_t + R_g h_{t-1} + b_g) \quad (5)$$

– Output gate ( $o$ ): Expresses the control level of the cell state added to the hidden state

$$o_t = \sigma_g(W_o x_t + R_o h_{t-1} + b_o) \quad (6)$$

$W$ ,  $R$ , and  $b$ , the learnable weights of the LSTM layer, are concatenations of the input weights, the recurrent weights, and bias, respectively:

$$W = \begin{bmatrix} W_i \\ W_f \\ W_g \\ W_o \end{bmatrix}, R = \begin{bmatrix} R_i \\ R_f \\ R_g \\ R_o \end{bmatrix}, b = \begin{bmatrix} b_i \\ b_f \\ b_g \\ b_o \end{bmatrix} \quad (7)$$

The cell state and the hidden state at the time step  $t$  are given by:

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (8)$$

$$h_t = o_t \odot \sigma_c(c_t) \quad (9)$$

where,  $\odot$  denotes the Hadamard product (element-wise multiplication of vectors). Also,  $\sigma_c$  and  $\sigma_g$  represent the state activation function and the gate activation function respectively. In this work, the hyperbolic tangent function ( $\tanh$ ) is used to compute the state activation function, whereas the sigmoid function given by  $\sigma(x) = (1 + e^{-x})^{-1}$  is employed to calculate the gate activation function.

### 3. EXPERIMENTAL RESULTS

#### 3.1 Datasets description

##### 3.1.1 COUGHVID dataset

Over 25,000 crowdsourced cough recordings are available in the COUGHVID collection, covering a wide range of participant ages, genders, geographic areas, and COVID-19 statuses. Between April 1st, 2020, and December 1st, 2020, all of the recordings were collected using a Web application hosted on a private server at the École Polytechnique Fédérale de Lausanne (EPFL) in Switzerland. There are roughly 35 hours of audio samples in the COUGHVID database corresponding to almost 37,000 segmented coughs with a sampling frequency of 48 kHz. More than 2,800 recordings were labeled by four qualified physicians in order to diagnose medical problems from the cough. All publicly available data records, as well as metadata, are stored in the Zenodo repository (<https://zenodo.org/record/4498364#.YgBQj-rMJPY>) [20], where we can distinguish three types of variables in metadata: (1) context information (timestamp and likelihood that the recording actually contains cough sounds), (2) user-reported information, and (3) labels provided by expert medical annotators about the cough recordings clinical

assessment.

##### 3.1.2 VIRUFY dataset

VIRUFY is an open source database (<https://github.com/virufy/virufy-data>) containing cough sounds for the two categories (COVID-19 positive and negative). This database was collected through the VIRUFY mobile data collection application as a resource of crowdsourcing data by smartphones public users. With the help of VIRUFY clinical researchers and medical advisors [31], the VIRUFY database was attached to information (symptoms, medical history, gender, and age) for all patients along with PCR test results and whether the patient was a smoker or not. VIRUFY contains 16 subjects (10 males and 6 women) containing more than 120 cough sound samples between positive and negative COVID-19 sounds sampled at 48kHz.

#### 3.2 Dataset cleaning

The utilization of crowdsourcing has paved the way for unconventional approaches in contrast to traditional data collection methods. This, in turn, has positioned crowdsourcing as a catalyst for research, given its capacity to facilitate cost-effective and expedited data access within the research community. Nonetheless, integrating crowdsourced data into artificial intelligence applications continues to encounter a range of challenges, with data quality, as highlighted in the study of Lease [32], standing out as a key concern. Crowdsourced data often includes samples unrelated to the content subject matter. Consequently, this study incorporates data cleaning processes to address this issue by using the metadata in order to choose the most reliable data in building the proposed automated system.

To clean the data from the COUGHVID database, we selected records that have a probability greater than 0.8 by the XGBoost cough detection model described earlier which are assumed to contain cough. Then, we selected the subjects whose expert opinion agrees with the self-report of people with COVID-19 or healthy people only, while excluding subjects with other pathologies, since our main objective is the detection of people with COVID-19. After cleaning, the retained recording samples are reported in Table 2.

Concerning the VIRUFY database, depicted in Table 3, it does not require any cleaning due to its high-quality recording.

**Table 2.** Cough samples retained after cleaning the COUGHVID dataset

Gender	Healthy	Covid-19	Total
Male	236	241	477
Female	106	160	266
Total	342	401	743

**Table 3.** Cough samples retained after cleaning the VIRUFY dataset

Gender	Healthy	Covid-19	Total
Male	40	47	87
Female	26	12	38
Total	66	59	125

#### 3.3 Results and discussion

The LSTM model performs the classification of cough

audio segments transformed into MFCC supported by log energy features at its input. Using the Scikit-Learn random search algorithm with COUGHVID and VIRUFY datasets, the most effective obtained model configuration, adopted in this work, is presented in Table 4. It includes a sequence input layer for 14-dimensional sequences, a BiLSTM layer with 100 hidden units for capturing temporal dependencies, two fully connected layers for information transformation, a sigmoid activation function, and a classification output layer utilizing binary cross-entropy.

During experiments on the COUGHVID and VIRUFY datasets, two LSTM models, one for each sex, were utilized, since physiological differences in the phonatory apparatus has an impact on the two sexes in terms of infection and symptom variance, and therefore may affect the cough classification process [17-19]. Additionally, these two databases provide descriptive information on the sexual status of audio recordings.

In order to evaluate the performance of the proposed system, a set of metrics, namely specificity, recall (sensitivity), precision, accuracy and F1 score, were used:

- Specificity is a measure of how many healthy people can be correctly identified as healthy:

$$\text{Specificity} = \text{TN}/(\text{TN} + \text{FP}) \quad (10)$$

- Recall (sensitivity) indicates the proportion of genuinely unhealthy individuals correctly predicted:

$$\text{Recall} = \text{TP}/(\text{TP} + \text{FN}) \quad (11)$$

- Precision is, intuitively, the classifier capacity to avoid identifying a healthy sample as a positive covid-19 sample:

$$\text{Precision} = \text{TP}/(\text{TP} + \text{FP}) \quad (12)$$

- Accuracy is the proportion of correctly identified individuals to the total number of samples:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (13)$$

- F1-score is defined as an average of the model precision and recall:

$$\text{F1\_score} = \frac{2 \times (\text{Recall} \times \text{Precision})}{(\text{Recall} + \text{Precision})} \quad (14)$$

where,

- TP (True positive): The classifier detected covid-19 when covid-19 cough was present.

- TN (True negative): The classifier identified healthy cough when the healthy cough was present.

- FP (False positive): The classifier identified covid-19 in the presence of a healthy cough.

- FN (False negative): The classifier detected the healthy cough when covid-19 was present.

Tables 5-7 display the simulation results obtained for the XGBoost-LSTM diagnostic system using the COUGHVID and VIRUFY datasets with different LSTM minibatch sizes. These results were obtained through a 4-fold stratified cross-validation (CV) approach, where the data was divided into four subsets. Three folds were used for training, constituting 75% of the data, while one fold was reserved for testing, representing the remaining 25% of the data. This approach improved the reliability of model evaluation by systematically testing it on various data segments, thereby enhancing its overall performance.

Table 5 and Figure 6 present the results for females in the COUGHVID dataset, where it can be noted that the best results were obtained for SNR  $\geq 15$ dB with F1-score=92% and sensitivity=94.73%. It is noteworthy that for the case of better recording quality with SNR $\geq 20$ dB, a lower F1-score=90% was achieved which may be due to the low number of training cough samples, that may result in underfitting of the model.

**Table 4.** The LSTM configuration

Hyperparameter	Description
Input dimension	14
Number of classes	2
Number of layers	5
Minibatch size	~
Number of hidden units	100
Initial learn rate	1e-3
Optimizer	Adam
Number of epochs	100

**Table 5.** Female classification results for COUGHVID dataset (LSTM Minibach=30)

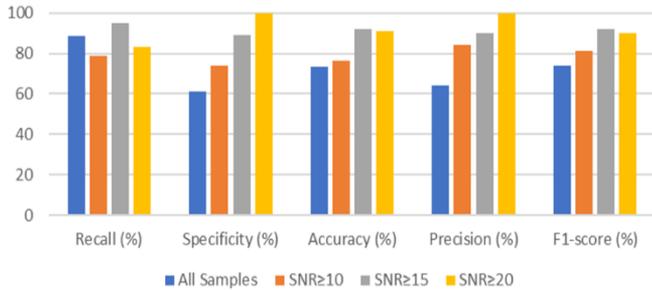
SNR Level (dB)	Number of Coughs Used for Training		Number of Coughs Used for Testing		Recall (%)	Specificity (%)	Accuracy (%)	Precision (%)	F1-score (%)
	Covid-19	Healthy	Covid-19	Healthy					
All Samples	112	75	48	31	88.57	61.36	73.41	64.00	74.00
SNR $\geq 10$	93	61	39	26	78.57	73.91	76.92	84.00	81.00
SNR $\geq 15$	47	42	20	17	<b>94.73</b>	88.88	91.90	90.00	<b>92.00</b>
SNR $\geq 20$	25	30	10	12	83.33	100.0	90.90	100.0	90.00

**Table 6.** Male classification results for COUGHVID dataset (LSTM Minibach=25)

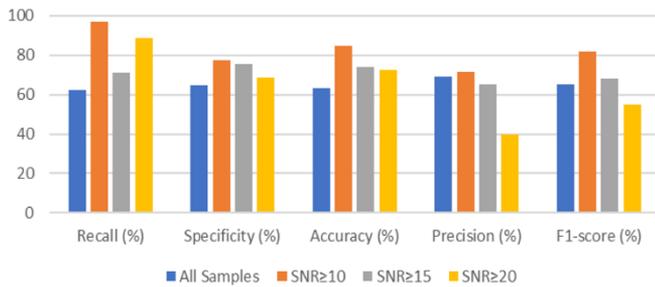
SNR Level (dB)	Number of Coughs Used for Training		Number of Coughs for Used Testing		Recall (%)	Specificity (%)	Accuracy (%)	Precision (%)	F1-score (%)
	Covid-19	Healthy	Covid-19	Healthy					
All Samples	169	166	72	70	62.50	64.51	63.38	69.00	65.00
SNR $\geq 10$	140	138	46	45	<b>97.05</b>	77.19	84.61	71.70	<b>82.00</b>
SNR $\geq 15$	79	105	26	35	70.83	75.67	73.77	65.00	68.00
SNR $\geq 20$	62	83	20	27	88.88	68.42	72.34	40.00	55.00

**Table 7.** Classification results for VIRUFY dataset (LSTM Minibatch=27)

Sex Type	Number of Coughs Used for Training		Number of Coughs Used for Testing		Recall (%)	Specificity (%)	Accuracy (%)	Precision (%)	F1-score (%)
	Covid-19	Healthy	Covid-19	Healthy					
Female	9	20	3	6	<b>100.0</b>	100.0	100.0	100.0	<b>100.0</b>
Male	36	30	11	10	<b>90.90</b>	100.0	95.23	100.0	<b>95.00</b>



**Figure 6.** Female classification results for COUGHVID dataset

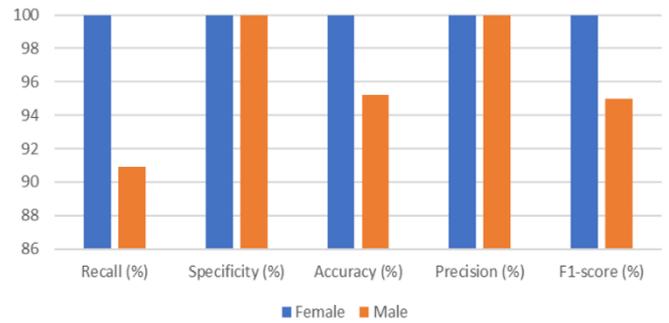


**Figure 7.** Male classification results for COUGHVID dataset

Table 6 and Figure 7 report the results for males in the COUGHVID dataset where the best results were achieved for SNR  $\geq 10$ dB with F1-score=82% and a sensitivity of about 97%. For the cases of SNR  $\geq 15$ dB where there is not enough number of samples to train the model properly, the classification system performs poorly with relatively low values of sensitivity and F1-score.

Table 7 and Figure 8 show the results for the VIRUFY dataset for both females and males. Because the recordings in this dataset are of good quality, the SNR level was not taken into account. Female samples reached a percentage of 100%

for all metrics, whereas male samples achieved a sensitivity (recall) of 90.9% and an F1-score of 95%. It is important to note that sensitivity and F1-score were the primary metrics used to assess the results because of their significance in evaluating the classifier ability to detect pathological samples. Meanwhile, specificity, accuracy, and precision play a critical role in evaluating the classifier performance in identifying healthy samples and determining the overall classification accuracy.



**Figure 8.** Classification results for VIRUFY dataset

Using the VIRUFY dataset, the suggested scheme shows high classification performance for both sexes, with a precision and specificity of 100%, and an average F1-score and accuracy of nearly 98%. It is noteworthy that all the data of this dataset are labeled with COVID-19 PCR test status and are quite accurate due to the fact that they were taken at a hospital under the observation of physicians using standard operating procedures. For the COUGHVID dataset, which is relatively less accurate due to self-reported information provided by the user and the lack of agreement among the experts about the patient health status, the suggested diagnosis system exhibits thoroughly good classification performance in terms of the SNR, and, specifically, a quite high sensitivity (recall) as reported in Tables 5 and 6.

**Table 8.** A comparative table of our classification results with relevant results in the literature

Dataset	Reference	Model	Features	Sex	Recall (%)	Specificity (%)	Precision (%)	Accuracy (%)	F1-score (%)
COUGHVID	[33]	LSTM	MFCC	-	60.00	62.00	-	62.00	-
	[34]	Multi-Branch Network Architecture	MFCC+	F	81.12	98.30	-	-	-
			clinical feature +	M	74.40	98.70	-	-	-
	[35]	LSTM	Mel-spectrograms	-	72.88	82.17	78.76	77.75	75.71
	Our approach	XGBoost-LSTM	13MFCC+log	F	94.73	100.0	90.00	91.89	92.00
			energy	M	97.05	77.19	71.7	84.61	82.00
[36]	CNN	Chromagram	-	-	-	87.60	92.90	93.40	
[36]	DNN	Chromagram	-	-	-	99.00	91.70	91.00	
VIRUFY	[37]	DNN	Frequency-domain feature vector	-	-	-	100.0	97.50	97.40
	Our approach	XGBoost-LSTM	13MFCC+log energy	F	100.0	100.0	100.0	100.0	100.0
				M	90.90	100.0	100.0	95.23	95.00

In order to evaluate the efficiency of the proposed COVID-19 detection technique, Table 8 gives a comparison of the classification achieved results with those reported in the literature based on the same two datasets. It can be considered that the proposed model outperforms the literature results for both datasets. First, from Table 8 it can be noticed that the suggested model performs better for females than for males, achieving an F1-score of 92% for COUGHVID and 100% for VIRUFY datasets. This difference may be attributed to the fact that the female model was trained on an imbalanced dataset with more COVID-19 cases than healthy cases, as shown in Table 5, while the male model was trained on a balanced dataset, as reported in Table 6.

For both datasets the suggested scheme exhibits relatively better sensitivity with 94.73% for females and 97.05% for males with COUGHVID dataset, and 100% for females and 90.9% for males with VIRUFY dataset. Furthermore, it also showcases superior accuracy, with results of 91.89% for females and 84.61% for males in the COUGHVID dataset, and 100% for females and 95.23% for males in the VIRUFY dataset. Additionally, the precision of our proposed scheme is noteworthy, attaining 90.0% for females and 71.7% for males in the COUGHVID dataset, and a 100% for both females and males in the VIRUFY dataset.

#### 4. CONCLUSIONS

The present paper addresses the development of an automated system for COVID-19 detection based on cough sounds, aiming for swift and efficient identification of coronavirus patients. The proposed approach involves two steps: initially, an XGBoost model identifies the cough segment within an audio recording and assesses the signal-to-noise ratio (SNR) to evaluate background noise. Subsequently, LSTM-based models execute binary classification on the identified cough segments to determine COVID-19 positivity or negativity.

Crowdsourced cough sound data collected through mobile apps and websites often suffers from the interference of ambient noise, potentially compromising the effectiveness of automated COVID-19 detection systems based on cough sounds. In this study, we conducted experiments using two datasets: the COUGHVID dataset, contaminated by environmental noises, and the VIRUFY dataset, recorded in controlled acoustic conditions. According to the obtained results, the suggested XGBoost-LSTM based scheme for COVID-19 detection exhibited strong performance, achieving an average classification accuracy of approximately 88% for the noise-contaminated COUGHVID dataset and nearly 97% for the VIRUFY dataset. Furthermore, in comparison to other pertinent studies, our proposed method outperforms state-of-the-art algorithms, achieving an F1-score of 100% for females and 95% for males, along with an accuracy of 100% for females and roughly 95% for males. However, in noisy environments, while the suggested method still performs well with an F1-score of 92% for females and 82% for males, it shows a slight decrease in precision and specificity for males.

#### REFERENCES

- [1] Gavriatopoulou, M., Ntanasis-Stathopoulos, I., Korompoki, E., Fotiou, D., Migkou, M., Tzanninis, I.G., Psaltopoulou, T., Kastritis, E., Terpos, E., Dimopoulos, M.A. (2021). Emerging treatment strategies for COVID-19 infection. *Clinical and Experimental Medicine*, 21: 167-179. <https://doi.org/10.1007/s10238-020-00671-y>
- [2] Cipollaro, L., Giordano, L., Padulo, J., Oliva, F., Maffulli, N. (2020). Musculoskeletal symptoms in SARS-CoV-2 (COVID-19) patients. *Journal of Orthopaedic Surgery and Research*, 15: 1-7. <https://doi.org/10.1186/s13018-020-01702-w>
- [3] Lei, S., Jiang, F., Su, W., Chen, C., Chen, J., Mei, W., Zhan, L.Y., Jia, Y., Zhang, L., Liu, D., Xia, Z.Y., Xia, Z. (2020). Clinical characteristics and outcomes of patients undergoing surgeries during the incubation period of COVID-19 infection. *EClinicalMedicine*, 21. <https://doi.org/10.1016/j.eclinm.2020.100331>
- [4] Al-Ani, R.M., Rashid, R.A. (2021). Prevalence of dysphonia due to COVID-19 at Salahaddin General Hospital, Tikrit City, Iraq. *American Journal of Otolaryngology*, 42(5): 103157. <https://doi.org/10.1016/j.amjoto.2021.103157>
- [5] Pal, A., Sankarasubbu, M. (2021). Pay attention to the cough: Early diagnosis of COVID-19 using interpretable symptoms embeddings with cough sound signal processing. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, pp. 620-628. <https://doi.org/10.1145/3412841.3441943>
- [6] Vinh, D.B., Zhao, X., Kiong, K.L., Guo, T., Jozaghi, Y., Yao, C., Kelley, J.M., Hanna, E.Y. (2020). Overview of COVID-19 testing and implications for otolaryngologists. *Head & Neck*, 42(7): 1629-1633. <https://doi.org/10.1002/hed.26213>
- [7] Giri, B., Pandey, S., Shrestha, R., Pokharel, K., Ligler, F.S., Neupane, B.B. (2021). Review of analytical performance of COVID-19 detection methods. *Analytical and Bioanalytical Chemistry*, 413(1): 35-48. <https://doi.org/10.1007/s00216-020-02889-x>
- [8] Aly, M., Rahouma, K.H., Ramzy, S.M. (2022). Pay attention to the speech: COVID-19 diagnosis using machine learning and crowdsourced respiratory and speech recordings. *Alexandria Engineering Journal*, 61(5): 3487-3500. <https://doi.org/10.1016/j.aej.2021.08.070>
- [9] Lella, K.K., PJA, A. (2021). A literature review on COVID-19 disease diagnosis from respiratory sound data. *arXiv Preprint arXiv*: 2112.07670. <https://doi.org/10.3934/bioeng.2021013>
- [10] Deshpande, G., Batliner, A., Schuller, B.W. (2022). AI-Based human audio processing for COVID-19: A comprehensive overview. *Pattern Recognition*, 122: 108289. <https://doi.org/10.1016/j.patcog.2021.108289>
- [11] Pahar, M., Klopper, M., Warren, R., Niesler, T. (2021). COVID-19 cough classification using machine learning and global smartphone recordings. *Computers in Biology and Medicine*, 135: 104572. <https://doi.org/10.1016/j.compbiomed.2021.104572>
- [12] Imran, A., Posokhova, I., Qureshi, H.N., Masood, U., Riaz, M.S., Ali, K., John, C.N., Hussain, I., Nabeel, M. (2020). AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Informatics in Medicine Unlocked*, 20: 100378. <https://doi.org/10.1016/j.imu.2020.100378>
- [13] Laguarta, J., Hueto, F., Subirana, B. (2020). COVID-19 artificial intelligence diagnosis using only cough recordings. *IEEE Open Journal of Engineering in Medicine and Biology*, 1: 275-281. <https://doi.org/10.1109/OJEMB.2020.3026928>

- [14] Ponomarchuk, A., Burenko, I., Malkin, E., Nazarov, I., Kokh, V., Avetisian, M., Zhukov, L. (2022). Project Achoo: A practical model and application for COVID-19 detection from recordings of breath, voice, and cough. *IEEE Journal of Selected Topics in Signal Processing*, 16(2): 175-187. <https://doi.org/10.1109/JSTSP.2022.3142514>
- [15] Leirgulen, J., Nuris-Souquet, M., Lévy-Fidel, C. (2021). Dry vs wet cough automatic classification using the COUGHVID Dataset. *Semantic Scholar*.
- [16] Chetupalli, S.R., Krishnan, P., Sharma, N., Muguli, A., Kumar, R., Nanda, V., Pinto, L.M., Ghosh, P.K., Ganapathy, S. (2023). Multi-modal point-of-care diagnostics for COVID-19 based on acoustics and symptoms. *IEEE Journal of Translational Engineering in Health and Medicine*, 11: 199-210. <https://doi.org/10.1109/JTEHM.2023.3250700>
- [17] Gebhard, C., Regitz-Zagrosek, V., Neuhauser, H.K., Morgan, R., Klein, S.L. (2020). Impact of sex and gender on COVID-19 outcomes in Europe. *Biology of Sex Differences*, 11: 1-13. <https://doi.org/10.1186/s13293-020-00304-9>
- [18] Agrawal, H., Das, N., Nathani, S., Saha, S., Saini, S., Kakar, S.S., Roy, P. (2021). An assessment on impact of COVID-19 infection in a gender specific manner. *Stem Cell Reviews and Reports*, 17: 94-112. <https://doi.org/10.1007/s12015-020-10048-z>
- [19] Biadsee, A., Biadsee, A., Kassem, F., Dagan, O., Masarwa, S., Ormianer, Z. (2020). Olfactory and oral manifestations of COVID-19: Sex-related symptoms-a potential pathway to early diagnosis. *Otolaryngology-Head and Neck Surgery*, 163(4): 722-728. <https://doi.org/10.1177/0194599820934380>
- [20] Orlandic, L., Teijeiro, T., Atienza, D. (2021). The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms. *Scientific Data*, 8(1): 156. <https://doi.org/10.1038/s41597-021-00937-4>
- [21] Chatzarrin, H., Arcelus, A., Goubran, R., Knoefel, F. (2011). Feature extraction for the differentiation of dry and wet cough sounds. In *2011 IEEE International Symposium on Medical Measurements and Applications*, Bari, Italy, pp. 162-166. <https://doi.org/10.1109/MeMeA.2011.5966670>
- [22] Monge-Álvarez, J., Hoyos-Barceló, C., San-José-Revuelta, L.M., Casaseca-de-la-Higuera, P. (2018). A machine hearing system for robust cough detection based on a high-level representation of band-specific audio features. *IEEE Transactions on Biomedical Engineering*, 66(8): 2319-2330. <https://doi.org/10.1109/TBME.2018.2888998>
- [23] Miranda, I.D., Diacon, A.H., Niesler, T.R. (2019). A comparative study of features for acoustic cough detection using deep architectures. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Berlin, Germany, pp. 2601-2605. <https://doi.org/10.1109/EMBC.2019.8856412>
- [24] Amoh, J., Odame, K. (2015). DeepCough: A deep convolutional neural network in a wearable cough detection system. In *2015 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, Atlanta, GA, USA, pp. 1-4. <https://doi.org/10.1109/BioCAS.2015.7348395>
- [25] Kadambi, P., Mohanty, A., Ren, H., Smith, J., McGuinness, K., Holt, K., Furtwaengler, A., Slepetyts, R., Yang, Z., Seo, J.S., Chae, J., Cao, Y., Berisha, V. (2018). Towards a wearable cough detector based on neural networks. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Calgary, AB, Canada, pp. 2161-2165. <https://doi.org/10.1109/ICASSP.2018.8461394>
- [26] Bansal, V., Pahwa, G., Kannan, N. (2020). Cough classification for covid-19 based on audio MFCC features using convolutional neural networks. In *2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)*, Greater Noida, India, pp. 604-608. <https://doi.org/10.1109/GUCON48875.2020.9231094>
- [27] Pramono, R.X.A., Imtiaz, S.A., Rodriguez-Villegas, E. (2016). A cough-based algorithm for automatic diagnosis of pertussis. *PloS One*, 11(9): e0162128. <https://doi.org/10.1371/journal.pone.0162128>
- [28] Sharma, G., Umopathy, K., Krishnan, S. (2020). Trends in audio signal feature extraction methods. *Applied Acoustics*, 158: 107020. <https://doi.org/10.1016/j.apacoust.2019.107020>
- [29] Deshpande, G., Schuller, B.W. (2020). Audio, speech, language, & signal processing for covid-19: A comprehensive overview. *arXiv Preprint arXiv: 2011.14445*. <https://doi.org/10.48550/arXiv.2011.14445>
- [30] Syed, S.A., Rashid, M., Hussain, S., Zahid, H. (2021). Comparative analysis of CNN and RNN for voice pathology detection. *BioMed Research International*, 2021: 1-8. <https://doi.org/10.1155/2021/6635964>
- [31] Chaudhari, G., Jiang, X., Fakhry, A., Han, A., Xiao, J., Shen, S., Khanzada, A. (2020). Virufy: Global applicability of crowdsourced and clinical datasets for AI detection of COVID-19 from cough. *arXiv Preprint arXiv: 2011.13320*. <https://doi.org/10.48550/arXiv.2011.13320>
- [32] Lease, M. (2011). On quality control and machine learning in crowdsourcing. In *Human Computation, Papers from the 2011 AAAI Workshop*, San Francisco, California, USA.
- [33] Son, M.J., Lee, S.P. (2022). COVID-19 diagnosis from crowdsourced cough sound data. *Applied Sciences*, 12(4): 1795. <https://doi.org/10.3390/app12041795>
- [34] Fakhry, A., Jiang, X., Xiao, J., Chaudhari, G., Han, A., Khanzada, A. (2021). Virufy: A multi-branch deep learning network for automated detection of COVID-19. *arXiv Preprint arXiv: 2103.01806*. <https://doi.org/10.48550/arXiv.2103.01806>
- [35] Hamdi, S., Oussalah, M., Moussaoui, A., Saidi, M. (2022). Attention-based hybrid CNN-LSTM and spectral data augmentation for COVID-19 diagnosis from cough sound. *Journal of Intelligent Information Systems*, 59(2): 367-389. <https://doi.org/10.1007/s10844-022-00707-7>
- [36] Islam, R., Abdel-Raheem, E., Tarique, M. (2021). Early detection of COVID-19 patients using chromagram features of cough sound recordings with machine learning algorithms. In *2021 International Conference on Microelectronics (ICM)*, New Cairo City, Egypt, pp. 82-85. <https://doi.org/10.1109/ICM52667.2021.9664931>
- [37] Islam, R., Abdel-Raheem, E., Tarique, M. (2022). A study of using cough sounds and deep neural networks for the early detection of COVID-19. *Biomedical Engineering Advances*, 3: 100025. <https://doi.org/10.1016/j.bea.2022.100025>