Identifying COVID-19 by Using Spectral analysis of Cough Recordings: A Distinctive Classification Study

Corresponding Author:

Negin MANSHOURI

Karadeniz Technical University

Faculty of Engineering

Department of Electrical and Electronics Engineering

61080 Trabzon/TURKEY

Phone (mobile): (+90) 5535791640

E-mail: negin.manshouri@ktu.edu.tr

Identifying COVID-19 by Using Spectral analysis of Cough Recordings: A Distinctive Classification Study

Negin MANSHOURI

Abstract

Sound signals from the respiratory system are largely the harbingers of human health. Early diagnosis of respiratory tract diseases is of great importance as it creates irreversible effects on human health when delayed. This diagnostic in the medical world has been made possible thanks to machine learning and signal processing analysis. The coronavirus epidemic, which is in question today and deeply shakes the whole world, has been revealed the importance of this issue even more. In terms of the coronavirus pandemic, it has become the focus of researchers to differentiate symptoms from similar diseases such as normal flu or influenza. Among these symptoms, the difference in cough sound has played a distinctive role in the proposed study. Several pioneering studies have proven that almost two-thirds of people who get corona have a dry cough. At this stage, the information of studies based on cough constitutes the main framework of our study. On the other hand, the basis of this study is based on machine learning algorithms. Clinical data collected under the supervision of doctors in a reliable environment was used as dataset. This dataset consists of 16 subjects suspected of the coronavirus with a specific patient demographic. In this study, using the polymerase chain reaction (PCR) test, suspected subjects were divided into two groups as negative and positive. The negative and positive labels represent the patient with non-COVID and with a COVID-19 cough respectively. Using the 3D plot or waterfall representation of the signal frequency spectrum, the salient features of the cough data are revealed. In this way, COVID-19 can be differentiated from other coughs by applying effective feature extraction and classification techniques. Power Spectral Density (PSD) based on Short Time Fourier Transform (STFT) and Mel Frequency Cepstral Coefficients (MFCC) were chosen as the efficient feature extraction method. Finally, among the classification techniques the Support Vector Machine (SVM) algorithm, was applied to the processed signals in order to identify and classify COVID-19 cough. In terms of results evaluation, the cough of subjects with COVID-19 has obtained with 95.86% classification accuracy thanks to the RBF kernel function of SVM and the MFCC method. In other words, the diagnosis of COVID-19 coughs was obtained with 98.6% and 91.7% sensitivity and specificity measures respectively.

Keywords: COVID-19, Cough, Signal processing, STFT, MFCC, SVM.

1. Introduction

Coronaviruses are a large family of viruses and a subset of coronavirus ranging from the common cold virus to the cause of more serious diseases such as SARS, MERS, and COVID-19. These were discovered in the 1960s and continued to be studied until the mid-1980s [1], [2]. This virus is common in mammals and birds in general, but seven coronaviruses of human origin have been identified so far [3]. The latest version of the virus, Coronavirus Acute Respiratory Syndrome 2, also known as SARS-

CoV-2, became widespread among people in Wuhan, China in December 2019 and unfortunately, it was transmitted to the whole world in a short time [4].

As of December 25, 2020, approximately 1,8 million people have lost their lives due to this disaster. Despite great efforts, the lack of a definitive solution for this disease shakes researchers and the medical world deeply. Several ways to keep people away from this pandemic have been proposed. These measures are limited to social distancing, frequent hand washing, and not touching the face as much as possible. Based on recent research the most common way this virus spreads is known by the droplets released into the air when the person with covid-19 coughs or sneezes [5], [6]. Since this virus behaves differently in subjects and has no fixed symptoms, it has been difficult for research studies to reach the main goal. In fact, common symptoms of COVID-19 include headache, fever, cough, and tiredness [7]. Other symptoms include shortness of breath, muscle and sore throat, sometimes diarrhea and vomiting [8], [9]. Among these symptoms, shortness of breath and dry cough caused by respiratory failure are the biggest causes of death. The painful side is that the coronavirus sometimes appears to affect other essential organs, namely the heart, kidney, brain and lungs [10]. By revealing all the information about this virus until now, there are still many uncertainties about the behavior of the virus, virus-patient interaction and pandemic process [11]. Based on the researches, the elderly and chronic patients are the most affected by this virus [12], [13], [14]. According to pioneering research, some COVID-19 patients are defined as virus carriers only. The data show that with 81 percent of COVID-19 patients, they are merely acting as carriers without very serious symptoms [11]. In the remaining percentage (19%), symptoms occur approximately 2 to 14 days after exposure to the virus. A study of 181 patients with Covid-19 found that the average incubation period was 5 days, and 97% of people had symptoms occurring on average 11.5 days after exposure to the virus [15], [16].

A detailed assessment of the situation of the COVID-19 infection, which was approximately 5 months old until May 2020, was carried out by Bchetnia et al. This study generally evaluated studies on the COVID-19 virus and played the role of information repository. Additionally, comparative analysis of SARS-CoV, MERS-CoV, and SARS-CoV-2 was prominently shown. Also the life cycle of the COVID-19 virus and how it affects infected cells was presented [17].

Clinical and radiological features of COVID-19 and its detailed effect on chronic kidney disease have been studied by medical researchers [13]. Based on the study, it has been observed that individuals with kidney disease are vulnerable to a severe COVID-19. In another medical study [10], damage to lung tissue biopsy samples of people who died due to COVID-19 were detected. Thus, it is hoped that when the study is expanded, it will shed light on the development of effective treatment methods. On the other hand, the behavior changes of this virus and its effects on the human body constitute the main target of many other medical studies [18], [19], [20], [21]. These studies were generally based on preliminary clinical data. The common point between them is that although they do not provide precise and clear information, they need more extensive and comprehensive research.

As is known, the general diagnostic method of this virus in the world is defined as the PCR test. But it seems that this way is not adequate to contain this global disease. The reasons for this are the limitation of the number of tests because of temporal factors, the low and cost of clinical tests despite the great demands and the fact that these tests are largely dependent on the hospital and clinical visit [22]. This mandatory visit can actually lead to regrets, as the COVID-19 virus is thought to last for hours and even days on different surfaces [23], [24]. Moreover, this Face-to-Face test endangers healthcare professionals and medical personnel and carries a serious risk. Since the appearance of this virus, test results were initially determined after as long as 10 days, so it was a huge waste of time. As time passed, faster types of tests evolved [25], but the decrease in accuracy raised different concerns [26]. After understanding the virus behavior a little, two different methods, namely X-ray and CT scan, were used for diagnosis [27], [28], [29], [30]. These methods have been seem to be as successful as the PCR test [31], [32], [22] but also require a clinic or hospital visit.

In the technology age, machine learning techniques can be used to control this pandemic. This technique actually develops a model by adapting to the environment and based on experience without explicit programming [33], [34]. The most important advantage of these techniques is to stay away from hospital or clinical visit-based COVID-19 test centers [35]. In short, it is aimed to make life easier within the scope of "digital health" thanks to computer science and engineering as well as the information obtained in medicine [7]. As another example, the spread prediction of this virus can be facilitated by the application of Artificial Intelligence (AI) [36], [37], [38]. In [7], the contribution of sound analysis to the diagnosis of COVID-19 virus was analysed and taken into consideration. Concentrating on this study, it is seen that computer audition (AC) technique is a ready and successful tool in diagnosis. In addition, these machine listening tools have just begun to be taken into account by researchers in the global epidemic process. Based on technologic information, it is predicted that the sensors of smart phones will also be useful in the early diagnosis of this virus [38]. In another cell phone-based research, a study that has the ability to make cheap, early and reliable diagnosis of COVID-19 has been conducted [39].

The work done in the field of COVID-19 machine learning is generally divided into two as image or sound analysis. Cough, one of the important symptoms of COVID-19, is the basis of sound-based studies [40], [7]. In fact, this symptom is an early harbinger of many diseases of the human body [41], [42], [43]. Cough samples containing COVID-19 can be classified with acceptable success using machine learning and signal processing techniques [9]. Another comprehensive study has been focused on different types of cough and distinguished patients with COVID-19 from asthma, bronchitis and healthy individuals with 96.83% success [44]. In cough samples recorded using a smartphone app, COVID-19 was classified with 97% AUC using the Resnet50 classifier due to the different pattern of cough [45].

It is an important and critical point that the dataset used to obtain strong information is safe [40], [33]. Using these datasets, all researchers around the world have focused on a single target. It is aimed to expand and verify the data within the scope of cooperation by using open source data in artificial intelligence-based research by carrying out different studies on engineering basis [33]. In fact, the possibility of easy access to open source data is the key step in machine learning [46]. These useful data lead to rich and scientific studies with high availability, accuracy and clarity in the world [47].

In this study, the diagnosis of COVID-19 was made within the scope of classification using the appropriate combination of machine learning and "GitHub" open source data sets [48]. In fact, "Virufy" is a voluntary trustworthy organization to identify patterns caused by COVID-19 coughing noises. It is worth noting that this organization offers free COVID-19 cough datasets as a pioneer in bringing innovation to humanity, health and the industry sector. The main goal of the study is to conduct an easy and early diagnosis study based on the classification technique for this global epidemic by focusing on cough sound samples. This model, based on computer algorithms and analysis of cough samples, makes predictions and decisions by taking into account the training data. Thus, it is predicted that early diagnosis of the virus will be possible by using effective signal processing methods in this dataset consisting of positive and negative samples of COVID-19. As a result, thanks to the collaboration of technology and engineering, people with suspected infection will be less likely to seek clinical treatment. It is useful to say that the proposed research is a basic study in terms of early diagnosis of COVID-19 using machine learning technique. This epidemic has the capacity to obtain encouraging and reliable information by making more and more detailed research in the academic and scientific field in a technologic sense. Considering the results of the study, it is seen that COVID-19 cough sound samples can be diagnosed with acceptable classification accuracy with software applications in homes and workplaces via smart tools [49].

2. Methods and test protocol

2.1. COVID-19 dataset and subjects

Dataset was collected at the hospital or clinic under the supervision of physicians or medical personnel. The cough sound data discussed were recorded with the "Virufy mobile app" on the request of Stanford University and made available on "GitHub". After the preprocessing steps were carefully done on the data, the results obtained from the PCR test were labeled as positive or negative [48]. In addition, subject demographics were presented in detail in Table 1. A data pool of 121 segmentations obtained from cough samples of sixteen subjects in total was created. As a result of this segmentation, there are a total of 73 non-COVID and 48 COVID-19 coughs. The segmentation preprocessing step actually plays an important role in the analysis by clarifying the importance and predominance infected regions considered [33], [50]. In this segmentation, the cough time of each subject was taken into account as 1640 milliseconds (ms). The sampling frequency in the study was determined as 48000.

Table 1. Demographics of subjects suspected of COVID-19 virus [48].

Age	Gender	Any chronic illness	Complaints and symptoms	PCR result
53	Male	None	None	Negative
50	Male	Congestive heart failure	Shortness of breath	Positive
43	Male	None	Sore throat	Negative
65	Male	Asthma or lung disease	Shortness of breath, worsening cough	Positive
40	Female	None	Sore throat, Loss of taste, Loss of smell	Positive
66	Female	Diabetes	None	Negative
20	Female	None	None	Negative
17	Female	None	Shortness of breath, Sore throat, Body aches	Negative
47	Male	None	Worsening cough	Negative
53	Male	None	Fever, chills, or sweating, Shortness of breath, worsening cough, Sore throat, Loss of taste, Loss of smell	Positive
24	Female	None	None	Positive
51	Male	Diabetes	Fever, chills, or sweating, worsening cough, Sore throat	Positive
53	Male	None	None	Negative
31	Male	None	Shortness of breath, worsening cough	Positive
37	Male	None	None	Negative
24	Female	None	Worsening cough	Negative

2.2. Data processing

The flow chart that highlights the stages of the study in detail was presented in Fig 1.

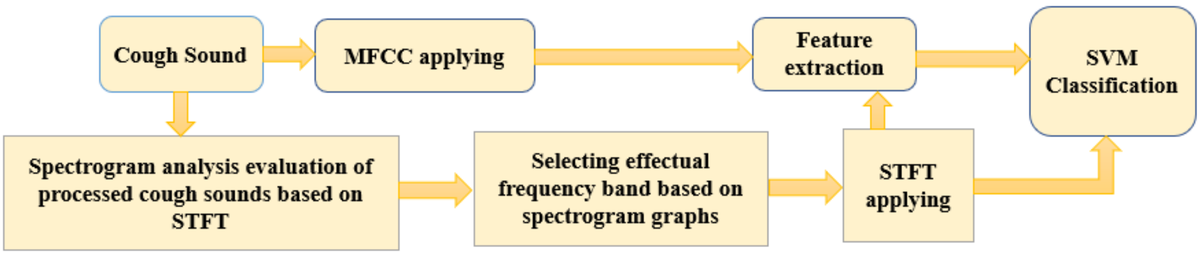


Fig. 1. Flow diagram of data processing.

2.2.1. Cough data STFT-based spectrogram analysis

One way of evaluating the frequency and phase behavior resulting from the change of signals over the time is the STFT [51]. In fact, this technique is to divide a long signal into short and equal lengths and apply the Fourier transform to each segment separately [52]. For a random signal $x(t)$ in time domain, STFT is shown by (1) [53], [54].

$$Y(t,f) = STFT\big(x(t)\big) = \int_{-\infty}^{+\infty} x(p)w(p-t)e^{-2\pi jfp}\,dp \tag{1}$$

In this equation, $x(t)$ and $w(t)$ are the continuous signal and windowing functions, respectively.

The visualization of this method, which provides a complete balance between time and frequency resolution [55], is defined as spectrogram or waterfall plot. In other words, it represents the square size of the coefficients of STFT [56].

Due to the unstable and unpredictable behavior of audio signals, the focus is on signal properties at a certain time by windowing process. Thanks to these overlapping windows, the behavior of the unstable signal will resemble a knowing and stationary signal. The length of the sliding window is an important factor in this transformation [57]. The purpose of the window is to actually select the time segment of the signal whose frequency properties are almost invariant [58]. Considering the properties of soft behaving windows, the Hamming window [59] was used in this study using trial and error method [60]. Also, the size and overlap of the window were selected as 2400 and 600, respectively.

In fact, spectrogram analysis was performed for each recording of COVID-19 and non-COVID cough sounds, then the average was calculated for each stage. Finally, taking into account the spectrograms based on STFT for COVID-19 and non-COVID cough sounds, the difference in PSD was calculated for analysis. When spectrogram graphics were interpreted, time and frequency slices that best reflect PSD were considered in terms of COVID-19 positive and negative samples classification. The dominant time and frequency intervals reflecting PSD were shown in Table 2. Thus, taking into account these time and frequency intervals, the basic step was taken for the feature extraction method.

Table 2. The dominant time and frequency intervals based on spectrogram.

Time (s)	**2**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Frequency intervals (Hz)	140-340, 460- 580	240-300	260- 620	340- 600	200-320, 480- 600	200-340	220-300, 460- 600	220-320, 440- 620	240- 340

2.3. STFT and MFCC feature extraction techniques

In the proposed study, two techniques were used as feature extraction methods, namely STFT and MFCC.

STFT is known as the feature extraction method widely used in the analysis of audio signals [55]. In the STFT method [61], [61] the features were chosen from thirteen effective frequency sub bands and their corresponding times, aiming at a reduced number of feature extraction destination. Using trapezoidal digital integration (Trapz) of these effective time and frequency intervals, a (121* 13) dimensional feature vector was obtained.

Another complex sound feature extraction technique to classify and distinguish differences between COVID-19 and non-COVID samples has been used in detail. MFCC technique is known to be compatible with the variation of the frequency band of the human ear [62]. In addition, MFCC is generally known as a technique with high success in recognizing sound systems [63], [64]. In a different study, MFCC has been shown to be a useful method in the differentiation of dry and wet cough [65]. The working principle of MFCC can be briefly summarized as signal windowing, discrete Fourier transform (DFT) application, calculating the logarithm of the coefficients magnitude, warping frequencies to a Mel scale and finally the application of discrete cosine transform (DCT) [66]. The detailed MFCC feature extraction flow chart is shown in Fig 2.

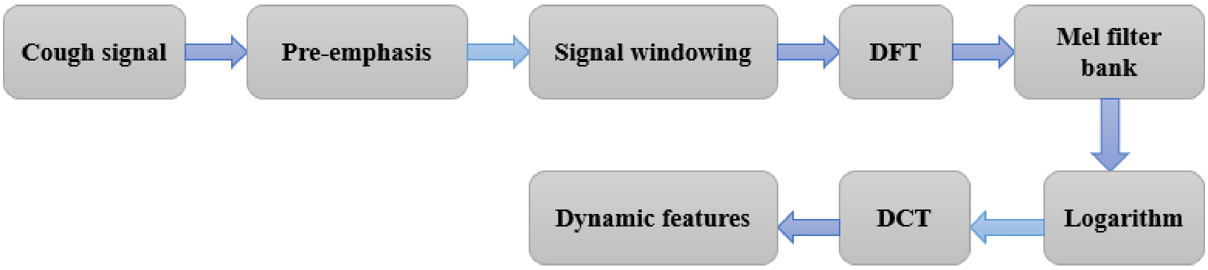


Fig. 2. MFCC feature extraction flowchart.

The Mel spectrum is calculated as a result of the signal obtained from the Fourier transform passing through a series of bandpass filters. Mel's approach from physical frequency (f) can be expressed as shown in (2).

$$f_{Mel} = 2595 log_{10}(1 + \frac{f}{700}) \quad (2)$$

In this technique, the Hamming window was chosen with 1024 window size and 512 overlap. Finally, the size of the feature vector obtained from MFCC was obtained as (121 * 13).

2.4. SVM classification algorithm

This supervised learning model is one of the powerful classification technique in class determination of two class problems. As a result of various scientific studies, it has been proven that this classifier is more successful than other classification methods [67], [68]. Among the advantages that distinguish SVM, it is its reliability, good generalization feature, strong theoretical basis and clear geometric lines in terms of classification [69], [70]. On the other hand, this classifier based on statistical functions minimizes structural risk possibilities [71]. The working principle of SVM in linear or nonlinear problems follows almost the same basis. This essential consists of discovering a hyperplane to distinguish between positive and negative samples [72].

In the proposed study, linear and nonlinear SVM model have been developed for cough-based COVID-19 diagnosis in the positively and negatively labeled problem. A brief description of linear and nonlinear SVM is explained below.

2.4.1. Linear SVM

The linear SVM algorithm will be effective when the model created by the training data set has a linear separable capability. The N-point training data set is defined as $(x_1, y_1), \ldots, (x_n, y_n)$, where y is the class label and x is the pattern to be classified. Now the main problem is to build a decision function that can properly classify any x input. Ultimately, to solve the problem, the decision function $g(x)$ is linearly defined in (3).

$$g(x) = w^T x + b \tag{3}$$

where w represent the hyperplane normal vector and b is a scalar. [71], [70]. Ultimately, in a two-class problem, the training examples are separated by the hyperplane $w^T x + b = 0$.

SVM chooses the plane with the maximum margin between the two classes from different hyperplanes for the training set [73]. Mathematical analysis of optimum hyperplane calculation is explained in detail in [70].

2.4.2. Nonlinear SVM and kernel functions

By using a nonlinear operator in the decision function, the linear SVM can be extended to a nonlinear classifier [73], [70], [74]. In addition, this operator allows the analysis of data to be transferred to the multidimensional property area that is also known as the "kernel trick" [75].

Due to the different kernel functions, it is extremely important to choose the appropriate function [76]. The most common kernel function in SVM classification method is defined as "Radial Basis Function" (RBF). In this research, the two most commonly used kernel types in SVM studies were exploited: namely polynomial and RBF [77]. These kernels were identified as follows in (4) and (5).

$$K_{poynomial}(x, y) = (x^T y + 1)^p \tag{4}$$

$$K_{RBF}(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right) \tag{5}$$

$p, \sigma > 0$ are a constant that defines the kernel order and width respectively. In this study the parameter p was chosen as 2. Hyperparameters were automatically optimized using "fitcsvm" [78].

2.5. Classifier training

To make a skill assessment of the proposed model, leave-one-out (LOO) and hold-out (HO) cross-validation processes of the statistical techniques were used. In the cross-validation technique [79], the dataset is divided into approximately equal sized subsets or folds named k [80]. If these subsets are shown as $N_1, N_2, \ldots., N_K$, the classifier learning algorithm is then applied k times for, $j = 1$ to k, each time using the combination of all subsets except N_j as the training set and N_j as the test set.

The LOO method, which is closely related to the jack-knife method [81], is actually a special form of cross validation. In this cross-validation method, where the number of folds is equal to the number of samples in the data set, a single sample is selected as the test set and all other samples are selected as the training set, and this situation is used once for each time [82], [83].

The hold out technique [84] is known in a simple and comprehensive way among cross validation methods [85], [86]. In this technique, the dataset is randomly divided into two, the training and the test set. After creating a model by training the training set, the proposed model will be evaluated in test validation data. General separating for training and testing is known as 80% and 20%, respectively. In this study, training and test splitting were done as 50.41% and 49.58%. But in the COVID-19 positive and negative two-class training data, two different separating were taken into account. This cross-validation process was repeated 1000 times.

2.4.3. Evaluating the performance of classification

A specific table called a confusion matrix is a useful tool in order to evaluate the classification performance. The parameters obtained from the elements of the complex matrix are named accuracy, sensitivity, and specificity [57]. These parameters were used for detailed performance evaluation of the SVM classification technique. The mathematical equation of these parameters is presented in (6), (7), (8). In the study, the non-COVID class was defined as the positive samples and the COVID class as the negative samples. This matrix is defined in Table 3. In the table below P = Positive, N = Negative, TP = True Positive, FP = False Positive, TN = True Negative and FN = False Negative .

Table. 3. Confusion matrix.

		Predicted class	
		Non-COVID	COVID
True class	Non-COVID (P)	TP	FN
	COVID (N)	FP	TN

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (7)$$

$$Specificity = \frac{TN}{TN+FP} \quad (8)$$

3. Results

The spectrogram graph showing the difference in PSD of cough sounds with COVID-19 positive and negative labels from the time and frequency perspective are shown in Fig, 3 and 4 respectively. The three-dimensional (3D) PSD graph reflecting this PSD is presented in Fig. 5. In short, this graph represents the PSD difference of the average cough sounds from COVID-19 and non-COVID. Focusing on the graph, the PSD difference in this cough sound analysis is evident up to 0.65 seconds. This distinction is useful for us in terms of the classification of COVID-19 positive and negative labelled cough sounds. Based on the information obtained from these graphs, effective time-frequency intervals were taken into account for feature extraction. SVM classification results using STFT feature extraction technique for LOO and HO cross validation are presented in Table 4 and 5 respectively. On the other hand, the same tables are also calculated in the MFCC feature extraction method, as Tables 6 and 7. Looking at the results in general, the classification of COVID-19 cough sounds seems successful with the suggested methods. Considering the results, it is taken into account that the MFCC technique gives a better percentage of success compared to STFT feature extraction, and it seems that RBF is more successful than Linear and Polynomial among SVM types. In addition, the classification performance analysis parameters largely agreed with each other. In the MFCC feature extraction method, in terms of Mel coefficients, the issue of how many overlapping windows will bring the highest success to the sum of 13 Mel coefficients was considered. To be clear, this description is given in detail in Fig. 6. The best accuracy achievement was obtained for 11 overlapping windows for almost three SVM classifier types in Fig.6 interpretation. Confusion matrix including sensitivity, specificity, precision, negative predictive value for the successful feature extraction method determined in Table 6 is shown in Figure 7. In Fig. 7, labels 1 and 0 indicate COVID-19 and non-COVID respectively.

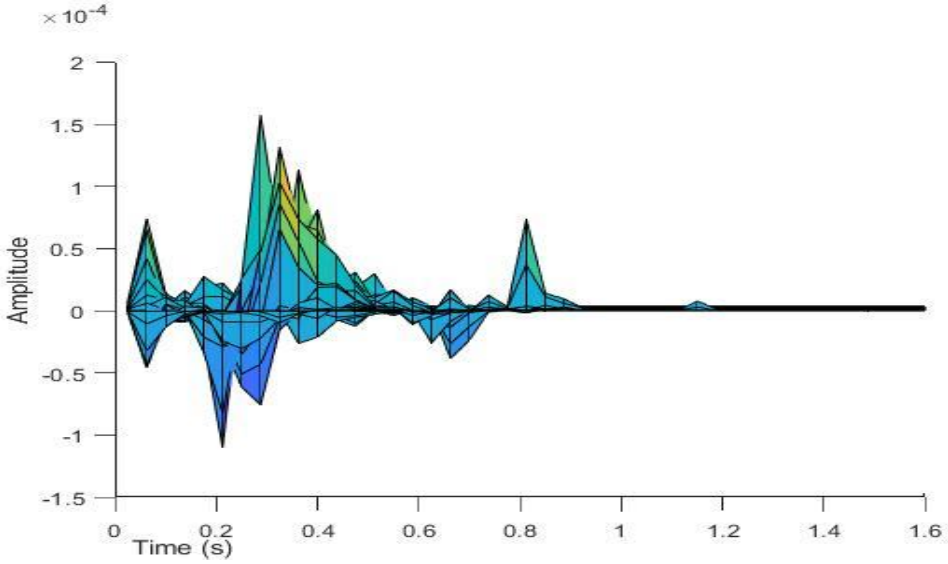


Fig. 3. Spectrogram graph from time perspective.

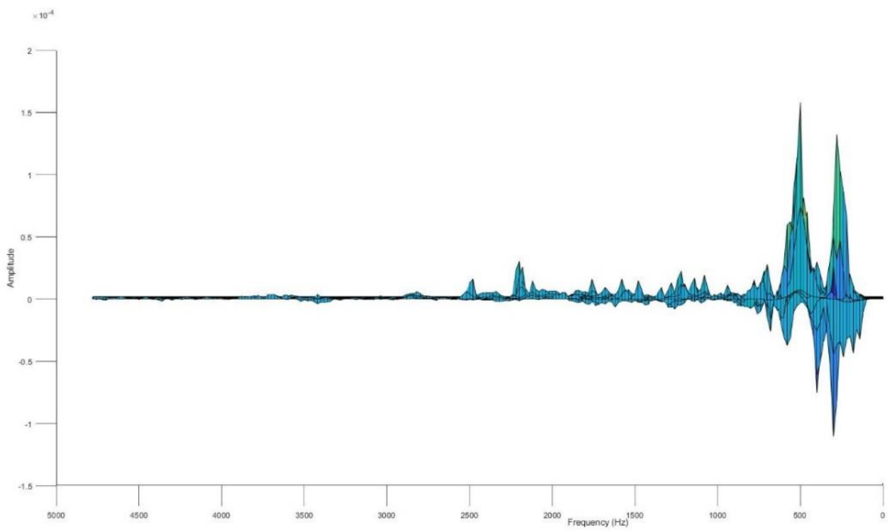


Fig. 4. Spectrogram graph from frequency perspective.

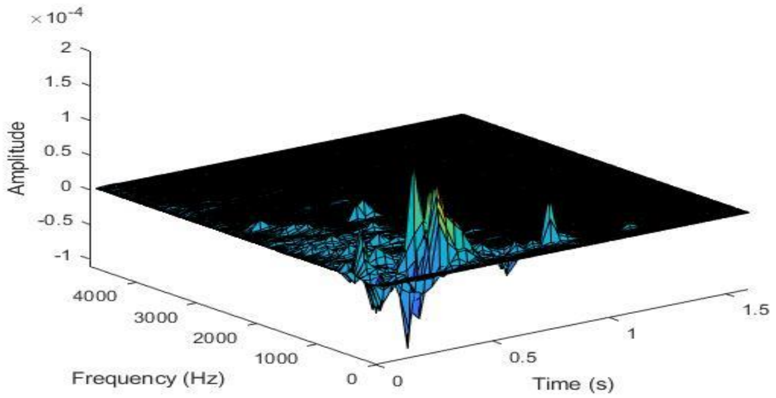


Fig. 5. 3D Spectrogram.

Table. 4. SVM classification results using STFT feature extraction technique for LOO cross validation.

	Accuracy	**Sensitivity**	**Specificity**
Linear	72.73	58.30	82.20
Polynomial	76.86	54.20	91.80
RBF	76.03	68.80	80.80

Table 5. SVM classification results using STFT feature extraction technique for two different training datasets splitting of HO cross validation.

	36 Non-COVID & 24 COVID for training			30 Non-COVID & 30 COVID for training		
	Accuracy	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
Linear	72.79	70.41	75.19	71.65	53.13	85.11
Polynomial	72.97	76.57	73.00	71.13	54.08	84.13
RBF	70.10	63.84	74.36	68.05	48.33	87.60

Table. 6. SVM classification results using MFCC feature extraction technique LOO cross validation.

	Accuracy	**Sensitivity**	**Specificity**
Linear	86.78	81.25	90.41
Polynomial	90.08	83.33	94.52
RBF	**94.21**	**89.58**	**97.26**

Table. 7. SVM classification results using MFCC feature extraction technique for two different training datasets splitting of HO cross validation.

	36 Non-COVID & 24 COVID for training			30 Non-COVID & 30 COVID for training		
	Accuracy	**Sensitivity**	**Specificity**	**Accuracy**	**Sensitivity**	**Specificity**
Linear	81.16	80.03	82.60	79.97	63.91	90.33
Polynomial	87.31	87.64	87.83	87.07	75.62	93.65
RBF	87.70	90.68	87.02	87.61	75.96	94.61

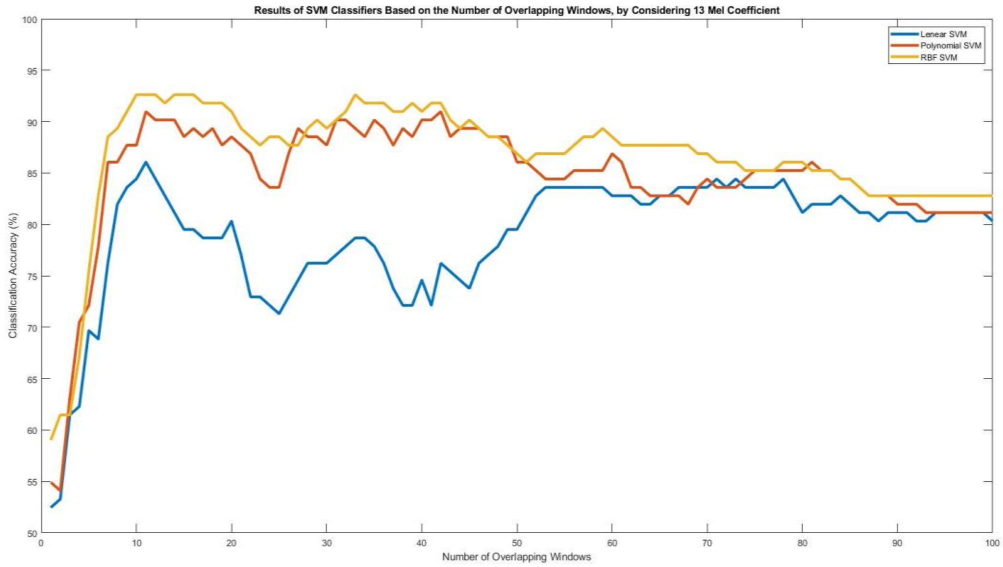
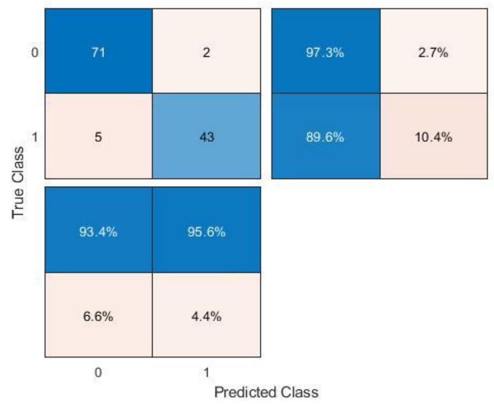
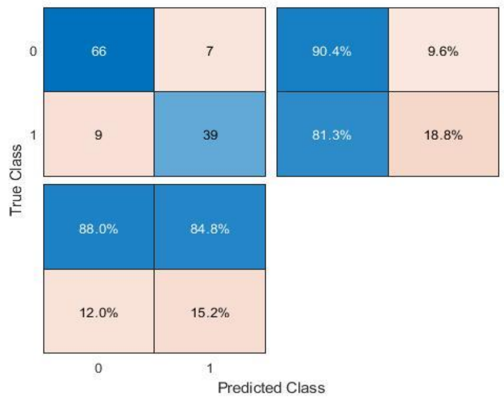


Fig. 6. Results of SVM based on number on overlapping windows by considering 13 Mel coefficients.

(a) Linear



(b) Polynomial



(c) RBF

Fig. 7. Confusion matrix for the best classification result.

3.1. Feature selection

Choosing the relevant features to choose the most suitable model is known as the critical and important step in machine learning. Feature selection technique [87] has many advantages, some of which can be summarized as simplifying the model for better understanding, reducing the size of the feature vector, and greatly reducing training time [88]. Since the signals received for analysis often carry unnecessary information, it will be useful to remove this data without losing too much information [89]. The presence of these irrelevant data deeply affects the accuracy of the model, while at the same time, it enables the model to be trained based on irrelevant features [90], [91], [40].

The Sequential Forward Search (SFS) method, which is frequently used in the scientific field [92], [93], was applied as an feature selection technique in this study. After performing SFS, a peak accuracy of 95.86% was observed on the dataset when using the best 9 features among the 13, as shown in Figure 8. Also, as a result of the application of the SFS technique, the success rate of RBF kernel SVM classification increased from 94.21% to 95.86% [94]. For this successful feature selection, the confusion matrix is shown in Fig. 9.

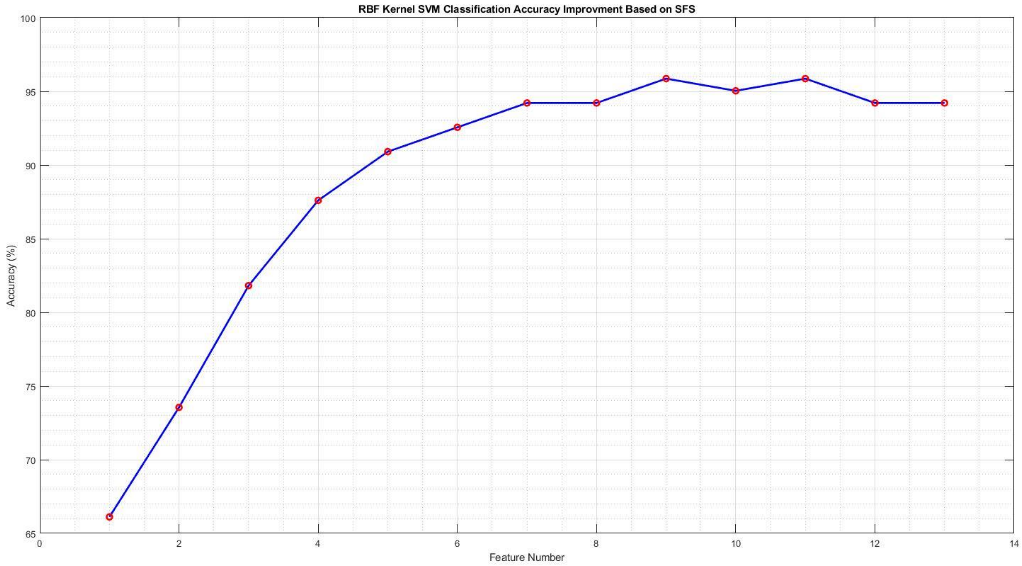


Fig. 8. RBF kernel SVM classification success percentage as a result of the SFS feature selection technique.

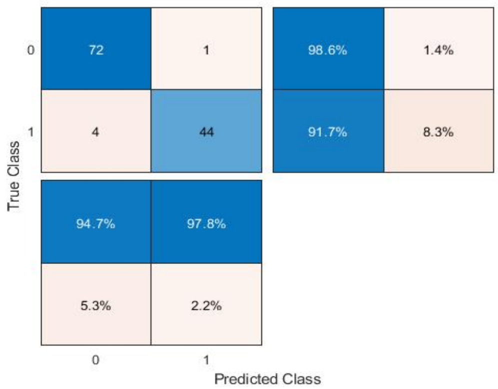


Fig.9. Confusion matrix after SFS applying for RBF kernel SVM.

4. Discussion

This study was conducted for the diagnosis of COVID-19, taking into account cough sounds, based on machine learning technology. In fact, the proposed research was defined as a machine learning study for the COVID-19 epidemic based on cough sounds without being in any hospital or clinical setting. This machine learning model provides great advantages by recording the cough sound of subjects suspected of COVID-19 based on the application of signal processing and audio recording of smartphones [35], [94], [45], [36]. One of the advantages of these studies is that they are performed anywhere, anytime and under easy conditions. These models are thought to be beneficial as a health aid tool when COVID-19 test kits are lacking. As stated from the beginning of the pandemic, the easiest way to protect from this dangerous virus is to stay at home and minimize interaction with people as much as possible. Taking this important issue into account, the proposed machine learning-based models protect healthcare professionals to a large extent [7], [35], [95], [96], [97]. In fact, computer-based ways to prevent the rapid and latent spread of corona virus are highly appreciated by researchers. Thanks to the application recommended in [35], the virus spread tracking has been easily possible. This study has developed a pre-diagnose model for COVID-19 from cough samples using artificial intelligence technology.

A small portion of the studies on COVID-19 has focused on machine learning-based cough sound classification researches. This topic has been the focus of attention for researchers recently. One reason for this is that there are a limited number of open access data sets based on cough sound. As these datasets are unveiled, they are quickly attracted by researchers. Because everyone's dream is to get rid of the COVID-19 epidemic as soon as possible and return to normal life.

As a result of the lack of a wide range of COVID-19 studies based on cough sound classification, we do not find very accurate the detailed comparison of our proposed classification study with a few classification studies [9], [94], [45], [98] because each study focuses on different datasets. The common goal is to be able to classify COVID-19 positive and negative labeled cough sounds with high success, using effective feature extraction and classification techniques using signal processing steps.

Using the STFT and MFCC feature extraction method, it has been observed that MFCC gives more successful results as expected in COVID-19 positive and negative cough sound classification process [99]. In order to perform the classification process of the study, the SVM classification method, which attracts great attention in this field [100], [101], [102] was selected and three different SVM types were considered as Linear, Polynomial, and RBF. Finally, it has been proven that feature selection is an important step in all aspects of signal processing. In short, it cannot be ignored that the classification performance increased by using the SFS feature selection technique in the proposed study [103].

5. Conclusion

The main reason for the spread of the coronavirus can be defined as the inadequacy of test kits, the cost and the loss of time in determining clinical test results. In this study, we tried to find a solution to the

COVID-19 epidemic surrounding the world using an open-access dataset, thanks to machine learning technology. Based on this technology, we have developed a model that can classify COVID-19 cough records from smartphones. We set off knowing that the dry cough symptom is an important factor in the diagnosis of COVID-19. Thus, we created the main framework of this research by taking cough-based studies into consideration. The most important goal of this study is to provide a virtual test opportunity away from the clinical and hospital environment based on machine learning technology. Cough records of 16 subjects with COVID-19 suspects were analyzed to create the dataset. The obtained cough sounds became a ready-made dataset by preprocessing and segmentation process. Considering the spectrogram graph based on STFT, we focused on effective features. Focusing on the PSD difference of negative and positively labeled cough sounds, strong features were prepared from dominant time and frequency ranges for the STFT feature extraction method. In addition, the MFCC feature extraction method was included in the analysis for the diagnosis of COVID-19 cough. Based on the work done for the diagnosis of COVID-19 cough sound, the MFCC feature extraction technique and RBF kernel SVM are selected as the method with the best percentage of success.

For the future of this study, a stronger cough-based COVID-19 diagnostic study can be performed by enlarging the dataset. A suggestion for future studies might be increasing the number of subjects and also the classification accuracy using different feature extraction and classification methods. Train and classification stages can be improved by using deep learning algorithms.

References

[1] D.A.J. Tyrrell, M.L. Bynoe, Cultivation of a Novel Type of Common-cold Virus in Organ Cultures, Br. Med. J. 1 (1965) 1467–1470. https://doi.org/10.1136/bmj.1.5448.1467.

[2] J.M. Visy, P. Le Coz, B. Chadefaux, C. Fressinaud, F. Woimant, J. Marquet, J. Zittoun, J.M. Visy, J.M. Vallat, M. Haguenau, Homocystinuria due to 5, 10–methylenetetra–hydrofolate reductase deficiency revealed by stroke in adult siblings, Neurology. 41 (1991) 1313–1315. https://doi.org/10.1212/wnl.41.8.1313.

[3] P. Zhou, X.-L. Yang, X.-G. Wang, B. Hu, L. Zhang, W. Zhang, H.-R. Si, Y. Zhu, B. Li, C.-L. Huang, H.-D. Chen, J. Chen, Y. Luo, H. Guo, R.-D. Jiang, M.-Q. Liu, Y. Chen, X.-R. Shen, X. Wang, X.-S. Zheng, K. Zhao, Q.-J. Chen, F. Deng, L.-L. Liu, B. Yan, F.-X. Zhan, Y.-Y. Wang, G.-F. Xiao, Z.-L. Shi, Addendum: A pneumonia outbreak associated with a new coronavirus of probable bat origin, Nature. 588 (2020). https://doi.org/10.1038/s41586-020-2951-z.

[4] N. Talaee, M. Varahram, H. Jamaati, A. Salimi, M. Attarchi, M. Kazempour dizaji, M. Sadr, S. Hassani, B. Farzanegan, F. Monjazebi, S.M. Seyedmehdi, Stress and burnout in health care workers during COVID-19 pandemic: validation of a questionnaire, J. Public Heal. (2020) 1. https://doi.org/10.1007/s10389-020-01313-z.

[5] What Is Coronavirus? | Johns Hopkins Medicine, (n.d.). https://www.hopkinsmedicine.org/health/conditions-and-diseases/coronavirus (accessed

December 25, 2020).

[6] Coronavirus, (n.d.). https://www.who.int/health-topics/coronavirus#tab=tab_1 (accessed December 25, 2020).

[7] B.W. Schuller, D.M. Schuller, K. Qian, J. Liu, H. Zheng, X. Li, COVID-19 and Computer Audition: An Overview on What Speech & Sound Analysis Could Contribute in the SARS-CoV-2 Corona Crisis, ArXiv. (2020). http://arxiv.org/abs/2003.11117 (accessed December 16, 2020).

[8] Clinical Manifestations of Patients with Coronavirus Disease 2019 (COVID-19) in a Referral Center in Iran - PubMed, (n.d.). https://pubmed.ncbi.nlm.nih.gov/33262799/ (accessed December 27, 2020).

[9] C. Brown, J. Chauhan, A. Grammenos, J. Han, A. Hasthanasombat, D. Spathis, T. Xia, P. Cicuta, C. Mascolo, Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data, Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. (2020) 3474–3484. http://arxiv.org/abs/2006.05919 (accessed December 16, 2020).

[10] F. Sadegh Beigee, M. Pourabdollah Toutkaboni, N. Khalili, S.A. Nadji, A. Dorudinia, M. Rezaei, E. Askari, B. Farzanegan, M. Marjani, A. Rafiezadeh, Diffuse alveolar damage and thrombotic microangiopathy are the main histopathological findings in lung tissue biopsy samples of COVID-19 patients, Pathol. Res. Pract. 216 (2020) 153228. https://doi.org/10.1016/j.prp.2020.153228.

[11] Features, Evaluation, and Treatment of Coronavirus - PubMed, (n.d.). https://pubmed.ncbi.nlm.nih.gov/32150360/ (accessed December 27, 2020).

[12] X. Yang, Y. Yu, J. Xu, H. Shu, J. Xia, H. Liu, Y. Wu, L. Zhang, Z. Yu, M. Fang, T. Yu, Y. Wang, S. Pan, X. Zou, S. Yuan, Y. Shang, Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study, Lancet Respir. Med. 8 (2020) 475–481. https://doi.org/10.1016/S2213-2600(20)30079-5.

[13] A. Abrishami, N. Khalili, N. Dalili, R.K. Tabari, R. Farjad, S. Samavat, A.N. Naghadehi, H. Haghighatkhah, M. Nafar, M. Sanei-Taheri, Clinical and radiologic characteristics of covid-19 in patients with ckd, Iran. J. Kidney Dis. 14 (2020). https://covid19.elsevierpure.com/en/publications/clinical-and-radiologic-characteristics-of-covid-19-in-patients-w (accessed December 27, 2020).

[14] Coronavirus disease 2019 (COVID-19) - Symptoms and causes - Mayo Clinic, (n.d.). https://www.mayoclinic.org/diseases-conditions/coronavirus/symptoms-causes/syc-20479963 (accessed December 27, 2020).

[15] S.A. Lauer, K.H. Grantz, Q. Bi, F.K. Jones, Q. Zheng, H.R. Meredith, A.S. Azman, N.G. Reich, J. Lessler, The incubation period of coronavirus disease 2019 (CoVID-19) from publicly reported confirmed cases: Estimation and application, Ann. Intern. Med. 172 (2020) 577–582. https://doi.org/10.7326/M20-0504.

[16] When is someone infectious? | FAQ, (n.d.). https://covid19.nj.gov/faqs/coronavirus-

information/about-the-virus/when-is-someone-infectious (accessed December 27, 2020).

[17] M. Bchetnia, C. Girard, C. Duchaine, C. Laprise, The outbreak of the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): A review of the current global status, J. Infect. Public Health. 13 (2020) 1601–1610. https://doi.org/10.1016/j.jiph.2020.07.011.

[18] N. Rajapakse, D. Dixit, Human and novel coronavirus infections in children: a review, Paediatr. Int. Child Health. (2020). https://doi.org/10.1080/20469047.2020.1781356.

[19] Y.R. Guo, Q.D. Cao, Z.S. Hong, Y.Y. Tan, S.D. Chen, H.J. Jin, K. Sen Tan, D.Y. Wang, Y. Yan, The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak- A n update on the status, Mil. Med. Res. 7 (2020). https://doi.org/10.1186/s40779-020-00240-0.

[20] E. Oliveros, Y. Brailovsky, P. Scully, E. Nikolou, R. Rajani, J. Grapsa, Coronavirus Disease 2019 and Heart Failure: A Multiparametric Approach, Card. Fail. Rev. 6 (2020). https://doi.org/10.15420/cfr.2020.09.

[21] R.W. Paterson, R.L. Brown, L. Benjamin, R. Nortley, S. Wiethoff, T. Bharucha, D.L. Jayaseelan, G. Kumar, R.E. Raftopoulos, L. Zambreanu, V. Vivekanandam, A. Khoo, R. Geraldes, K. Chinthapalli, E. Boyd, H. Tuzlali, G. Price, G. Christofi, J. Morrow, P. McNamara, B. McLoughlin, S.T. Lim, P.R. Mehta, V. Levee, S. Keddie, W. Yong, S.A. Trip, A.J.M. Foulkes, G. Hotton, T.D. Miller, A.D. Everitt, C. Carswell, N.W.S. Davies, M. Yoong, D. Attwell, J. Sreedharan, E. Silber, J.M. Schott, A. Chandratheva, R.J. Perry, R. Simister, A. Checkley, N. Longley, S.F. Farmer, F. Carletti, C. Houlihan, M. Thom, M.P. Lunn, J. Spillane, R. Howard, A. Vincent, D.J. Werring, C. Hoskote, H.R. Jäger, H. Manji, M.S. Zandi, The emerging spectrum of COVID-19 neurology: clinical, radiological and laboratory findings, Brain. 143 (2020) 3104–3120. https://doi.org/10.1093/brain/awaa240.

[22] B. Subirana, F. Hueto, P. Rajasekaran, J. Laguarta, S. Puig, J. Malvehy, O. Mitja, A. Trilla, C.I. Moreno, J.F.M. Valle, A.E.M. González, B. Vizmanos, S. Sarma, Hi Sigma, do I have the Coronavirus?: Call for a New Artificial Intelligence Approach to Support Health Care Professionals Dealing With The COVID-19 Pandemic, ArXiv. (2020). http://arxiv.org/abs/2004.06510 (accessed December 16, 2020).

[23] N. van Doremalen, T. Bushmaker, D.H. Morris, M.G. Holbrook, A. Gamble, B.N. Williamson, A. Tamin, J.L. Harcourt, N.J. Thornburg, S.I. Gerber, J.O. Lloyd-Smith, E. de Wit, V.J. Munster, Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1, N. Engl. J. Med. 382 (2020) 1564–1567. https://doi.org/10.1056/nejmc2004973.

[24] Opinion | Hospitals are overwhelmed because of the coronavirus. Here’s how to help. - The Washington Post, (n.d.). https://www.washingtonpost.com/opinions/2020/03/15/hospitals-are-overwhelmed-because-coronavirus-heres-how-help/ (accessed December 29, 2020).

[25] Detect COVID-19 in as Little as 5 Minutes | Abbott Newsroom, (n.d.). https://www.abbott.com/corpnewsroom/diagnostics-testing/detect-covid-19-in-as-little-as-5-

minutes.html (accessed December 29, 2020).
[26] FDA says Abbott's 5-minute Covid-19 test may miss infected patients, (n.d.). https://www.statnews.com/2020/05/15/fda-says-abbotts-5-minute-covid-19-test-may-miss-infected-patients/ (accessed December 29, 2020).
[27] J. Zhang, Y. Xie, G. Pang, Z. Liao, J. Verjans, W. Li, Z. Sun, J. He, Y. Li, C. Shen, Y. Xia, Viral Pneumonia Screening on Chest X-ray Images Using Confidence-Aware Anomaly Detection, IEEE Trans. Med. Imaging. (2020) 1–1. http://arxiv.org/abs/2003.12338 (accessed December 29, 2020).
[28] A. Narin, C. Kaya, Z. Pamuk, Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks, ArXiv. (2020). http://arxiv.org/abs/2003.10849 (accessed December 29, 2020).
[29] Y. Li, L. Xia, Coronavirus disease 2019 (COVID-19): Role of chest CT in diagnosis and management, Am. J. Roentgenol. 214 (2020) 1280–1286. https://doi.org/10.2214/AJR.20.22954.
[30] Z. Ye, Y. Zhang, Y. Wang, Z. Huang, B. Song, Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review, Eur. Radiol. 30 (2020) 4381–4389. https://doi.org/10.1007/s00330-020-06801-0.
[31] H.A. Gietema, N. Zelis, J.M. Nobel, L.J.G. Lambriksi, L.B.V. Alphen, A.M.L.O. Lashof, J.E. Wildberger, I.C. Nelissen, P.M. Stassen, CT in relation to rt-PCR in diagnosing covid-19 in the netherlands: A prospective study, PLoS One. 15 (2020). https://doi.org/10.1371/journal.pone.0235844.
[32] H.J.A. Adams, T.C. Kwee, R.M. Kwee, Coronavirus disease 2019 and chest CT: Do not put the sensitivity value in the isolation room and look beyond the numbers, Radiology. 297 (2020) E236. https://doi.org/10.1148/radiol.2020201709.
[33] J. Shuja, E. Alanazi, W. Alasmary, A. Alashaikh, COVID-19 open source data sets: a comprehensive survey, Appl. Intell. (2020) 1–30. https://doi.org/10.1007/s10489-020-01862-6.
[34] K. Lan, D. tong Wang, S. Fong, L. sheng Liu, K.K.L. Wong, N. Dey, A Survey of Data Mining and Deep Learning in Bioinformatics, J. Med. Syst. 42 (2018) 1–20. https://doi.org/10.1007/s10916-018-1003-9.
[35] A. Imran, I. Posokhova, H.N. Qureshi, U. Masood, M.S. Riaz, K. Ali, C.N. John, M.I. Hussain, M. Nabeel, AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app, Informatics Med. Unlocked. 20 (2020) 100378. https://doi.org/10.1016/j.imu.2020.100378.
[36] Z. Hu, Q. Ge, S. Li, L. Jin, M. Xiong, Artificial Intelligence Forecasting of Covid-19 in China, ArXiv. (2020). http://arxiv.org/abs/2002.07112 (accessed December 29, 2020).
[37] S. Wang, B. Kang, J. Ma, X. Zeng, M. Xiao, J. Guo, M. Cai, J. Yang, Y. Li, X. Meng, B. Xu, A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19), MedRxiv. (2020) 2020.02.14.20023028. https://doi.org/10.1101/2020.02.14.20023028.

[38] H.S. Maghded, K.Z. Ghafoor, A.S. Sadiq, K. Curran, D.B. Rawat, K. Rabie, A Novel AI-enabled Framework to Diagnose Coronavirus COVID-19 using Smartphone Embedded Sensors: Design Study, in: Proc. - 2020 IEEE 21st Int. Conf. Inf. Reuse Integr. Data Sci. IRI 2020, Institute of Electrical and Electronics Engineers Inc., 2020: pp. 180–187. https://doi.org/10.1109/IRI49571.2020.00033.

[39] P. Fozouni, S. Son, M. Díaz de León Derby, G.J. Knott, C.N. Gray, M. V. D'Ambrosio, C. Zhao, N.A. Switz, G.R. Kumar, S.I. Stephens, D. Boehm, C.L. Tsou, J. Shu, A. Bhuiya, M. Armstrong, A.R. Harris, P.Y. Chen, J.M. Osterloh, A. Meyer-Franke, B. Joehnk, K. Walcott, A. Sil, C. Langelier, K.S. Pollard, E.D. Crawford, A.S. Puschnik, M. Phelps, A. Kistler, J.L. DeRisi, J.A. Doudna, D.A. Fletcher, M. Ott, Amplification-free detection of SARS-CoV-2 with CRISPR-Cas13a and mobile phone microscopy, Cell. (2020). https://doi.org/10.1016/j.cell.2020.12.001.

[40] A.N. Belkacem, S. Ouhbi, A. Lakas, E. Benkhelifa, C. Chen, End-to-End AI-Based Point-of-Care Diagnosis System for Classifying Respiratory Illnesses and Early Detection of COVID-19, ArXiv. (2020). http://arxiv.org/abs/2006.15469 (accessed December 30, 2020).

[41] A.B. Chang, G.J. Redding, M.L. Everard, Chronic wet cough: Protracted bronchitis, chronic suppurative lung disease and bronchiectasis, Pediatr. Pulmonol. 43 (2008) 519–531. https://doi.org/10.1002/ppul.20821.

[42] T. Higenbottam, Chronic cough and the cough reflex in common lung diseases, in: Pulm. Pharmacol. Ther., Academic Press, 2002: pp. 241–247. https://doi.org/10.1006/pupt.2002.0341.

[43] A. Windmon, M. Minakshi, P. Bharti, S. Chellappan, M. Johansson, B.A. Jenkins, P.R. Athilingam, TussisWatch: A Smart-Phone System to Identify Cough Episodes as Early Symptoms of Chronic Obstructive Pulmonary Disease and Congestive Heart Failure, IEEE J. Biomed. Heal. Informatics. 23 (2019) 1566–1573. https://doi.org/10.1109/JBHI.2018.2872038.

[44] A. Pal, M. Sankarasubbu, Pay Attention to the cough: Early Diagnosis of COVID-19 using Interpretable Symptoms Embeddings with Cough Sound Signal Processing, (2020) 1–15. http://arxiv.org/abs/2010.02417 (accessed December 30, 2020).

[45] J. Laguarta, F. Hueto, B. Subirana, COVID-19 Artificial Intelligence Diagnosis using only Cough Recordings, IEEE Open J. Eng. Med. Biol. (2020) 1–1. https://doi.org/10.1109/ojemb.2020.3026928.

[46] E.J. Yates, L.C. Yates, H. Harvey, Machine learning "red dot": open-source, cloud, deep convolutional neural networks in chest radiograph binary normality classification, Clin. Radiol. 73 (2018) 827–831. https://doi.org/10.1016/j.crad.2018.05.015.

[47] J.S. Frazer, A. Shard, J. Herdman, Involvement of the open-source community in combating the worldwide COVID-19 pandemic: a review, J. Med. Eng. Technol. (2020) 1–8. https://doi.org/10.1080/03091902.2020.1757772.

[48] GitHub - virufy/virufy_data, (n.d.). https://github.com/virufy/virufy_data (accessed December 30, 2020).

[49] R. Dunne, T. Morris, S. Harper, S. Report, High accuracy classication of COVID-19 coughs using Mel-frequency cepstral coecients and a Convolutional Neural Network with a use case for smart home devices, (2020). https://doi.org/10.21203/rs.3.rs-63796/v1.

[50] F. Shan, Y. Gao, J. Wang, W. Shi, N. Shi, M. Han, Z. Xue, D. Shen, Y. Shi, Lung infection quantification of COVID-19 in CT images with deep learning, ArXiv. (2020).

[51] E. Sejdić, I. Djurović, J. Jiang, Time-frequency feature representation using energy concentration: An overview of recent advances, Digit. Signal Process. A Rev. J. 19 (2009) 153–183. https://doi.org/10.1016/j.dsp.2007.12.004.

[52] N. Kehtarnavaz, Frequency Domain Processing, in: Digit. Signal Process. Syst. Des., Elsevier, 2008: pp. 175–196. https://doi.org/10.1016/b978-0-12-374490-6.00007-6.

[53] A. He, J. Yu, G. Wei, Y. Chen, H. Wu, Z. Tang, Short-time Fourier transform and decision tree-based pattern recognition for gas identification using temperature modulated microhotplate gas sensors, J. Sensors. 2016 (2016). https://doi.org/10.1155/2016/7603931.

[54] R. Bajric, N. Zuber, G.A. Skrimpas, N. Mijatovic, Feature Extraction Using Discrete Wavelet Transform for Gear Fault Diagnosis of Wind Turbine Gearbox, Shock Vib. 2016 (2016) 1–10. https://doi.org/10.1155/2016/6748469.

[55] D. Mitrović, M. Zeppelzauer, C. Breiteneder, Features for Content-Based Audio Retrieval, in: Elsevier, 2010: pp. 71–150. https://doi.org/10.1016/s0065-2458(10)78003-7.

[56] N.J. Sairamya, L. Susmitha, S. Thomas George, M.S.P. Subathra, Hybrid approach for classification of electroencephalographic signals using time-frequency images with wavelets and texture features, in: Intell. Data Anal. Biomed. Appl. Challenges Solut., Elsevier, 2019: pp. 253–273. https://doi.org/10.1016/B978-0-12-815553-0.00013-6.

[57] N. Manshouri, M. Maleki, T. Kayikcioglu, An EEG-based stereoscopic research of the PSD differences in pre and post 2D&3D movies watching, Biomed. Signal Process. Control. 55 (2020). https://doi.org/10.1016/j.bspc.2019.101642.

[58] B. Boashash, Heuristic Formulation of Time-Frequency Distributions, in: Time Freq. Anal. A Compr. Ref., Elsevier Ltd, 2003: pp. 29–57. https://doi.org/10.1016/B978-008044335-5/50023-1.

[59] F.J. HARRIS, Multirate FIR Filters for Interpolating and Desampling, in: Handb. Digit. Signal Process., Elsevier, 1987: pp. 173–287. https://doi.org/10.1016/b978-0-08-050780-4.50008-4.

[60] N. Manshouri, T. Kayikcioglu, A Comprehensive Analysis of 2D&3D Video Watching of EEG Signals by Increasing PLSR and SVM Classification Results, Comput. J. (2019). https://doi.org/10.1093/comjnl/bxz043.

[61] N. Mehala, R. Dahiya, A Comparative Study of FFT, STFT and Wavelet Techniques for Induction Machine Fault Diagnostic Analysis, n.d. https://pdfs.semanticscholar.org/3b7a/7c7a2bb4b8bccca373c1fd5c555d13a339cf.pdf (accessed May 23, 2019).

[62] M.A. Hossan, S. Memon, M.A. Gregory, A novel approach for MFCC feature extraction, in: 4th Int. Conf. Signal Process. Commun. Syst. ICSPCS'2010 - Proc., 2010. https://doi.org/10.1109/ICSPCS.2010.5709752.

[63] A. Winursito, R. Hidayat, A. Bejo, Improvement of MFCC feature extraction accuracy using PCA in Indonesian speech recognition, in: 2018 Int. Conf. Inf. Commun. Technol. ICOIACT 2018, Institute of Electrical and Electronics Engineers Inc., 2018: pp. 379–383. https://doi.org/10.1109/ICOIACT.2018.8350748.

[64] Y. Wang, B. Lawlor, Speaker recognition based on MFCC and BP neural networks, in: 2017 28th Irish Signals Syst. Conf. ISSC 2017, Institute of Electrical and Electronics Engineers Inc., 2017. https://doi.org/10.1109/ISSC.2017.7983644.

[65] H. Chatrzarrin, A. Arcelus, R. Goubran, F. Knoefel, Feature extraction for the differentiation of dry and wet cough sounds, in: MeMeA 2011 - 2011 IEEE Int. Symp. Med. Meas. Appl. Proc., IEEE Computer Society, 2011: pp. 162–166. https://doi.org/10.1109/MeMeA.2011.5966670.

[66] K.S. Rao, M. K E, Speech Recognition Using Articulatory and Excitation Source Features, Springer International Publishing, Cham, 2017. https://doi.org/10.1007/978-3-319-49220-9.

[67] C.W. Hsu, C.J. Lin, A comparison of methods for multiclass support vector machines, IEEE Trans. Neural Networks. 13 (2002) 415–425. https://doi.org/10.1109/72.991427.

[68] T. Joachims, Text categorization with Support Vector Machines: Learning with many relevant features, in: Springer, Berlin, Heidelberg, 1998: pp. 137–142. https://doi.org/10.1007/bfb0026683.

[69] M.E. Mavroforakis, S. Theodoridis, A geometric approach to support vector machine (SVM) classification, IEEE Trans. Neural Networks. 17 (2006) 671–682. https://doi.org/10.1109/TNN.2006.873281.

[70] I. El-Naqa, Y. Yang, M.N. Wernick, N.P. Galatsanos, R.M. Nishikawa, A support vector machine approach for detection of microcalcifications, IEEE Trans. Med. Imaging. 21 (2002) 1552–1563. https://doi.org/10.1109/TMI.2002.806569.

[71] R. Romero, E.L. Iglesias, L. Borrajo, A linear-RBF multikernel SVM to classify big text corpora, Biomed Res. Int. 2015 (2015). https://doi.org/10.1155/2015/878291.

[72] Y.-P.P. Lin, C.-H.H. Wang, T.-L.L. Wu, S.-K.K. Jeng, J.-H.H. Chen, Support vector machine for EEG signal classification during listening to emotional music, in: IEEE, 2008: pp. 127–130. https://doi.org/10.1109/MMSP.2008.4665061.

[73] C.J.C. Burges, A tutorial on support vector machines for pattern recognition, Data Min. Knowl. Discov. 2 (1998) 121–167. https://doi.org/10.1023/A:1009715923555.

[74] V. Viitaniemi, M. Sjöberg, M. Koskela, S. Ishikawa, J. Laaksonen, Advances in visual concept detection: Ten years of TRECVID, in: Adv. Indep. Compon. Anal. Learn. Mach., Elsevier, 2015: pp. 249–278. https://doi.org/10.1016/B978-0-12-802806-3.00012-9.

[75] X. Li, X. Chen, Y. Yan, W. Wei, Z. Wang, X. Li, X. Chen, Y. Yan, W. Wei, Z.J. Wang,

Classification of EEG Signals Using a Multiple Kernel Learning Support Vector Machine, Sensors. 14 (2014) 12784–12802. https://doi.org/10.3390/s140712784.

[76] Y. Lei, Individual intelligent method-based fault diagnosis, in: Intell. Fault Diagnosis Remain. Useful Life Predict. Rotating Mach., Elsevier, 2017: pp. 67–174. https://doi.org/10.1016/b978-0-12-811534-3.00003-2.

[77] Advances in Kernel Methods | The MIT Press, (n.d.). https://mitpress.mit.edu/books/advances-kernel-methods (accessed January 6, 2021).

[78] Train support vector machine (SVM) classifier for one-class and binary classification - MATLAB fitcsvm, (n.d.). https://www.mathworks.com/help/stats/fitcsvm.html (accessed January 8, 2021).

[79] T.R. Shultz, S.E. Fahlman, S. Craw, P. Andritsos, P. Tsaparas, R. Silva, C. Drummond, C.X. Ling, V.S. Sheng, C. Drummond, P.L. Lanzi, J. Gama, R.P. Wiegand, P. Sen, G. Namata, M. Bilgic, L. Getoor, J. He, S. Jain, F. Stephan, S. Jain, F. Stephan, C. Sammut, M. Harries, C. Sammut, K.M. Ting, B. Pfahringer, J. Case, S. Jain, K.L. Wagstaff, S. Nijssen, A. Wirth, C.X. Ling, V.S. Sheng, X. Zhang, C. Sammut, N. Cancedda, J.-M. Renders, P. Michelucci, D. Oblinger, E. Keogh, A. Mueen, Cross-Validation, in: Encycl. Mach. Learn., Springer US, 2011: pp. 249–249. https://doi.org/10.1007/978-0-387-30164-8_190.

[80] G.I. Webb, C. Sammut, C. Perlich, T. Horváth, S. Wrobel, K.B. Korb, W.S. Noble, C. Leslie, M.G. Lagoudakis, N. Quadrianto, W.L. Buntine, N. Quadrianto, W.L. Buntine, L. Getoor, G. Namata, L. Getoor, X.J.J. Han, J.-A. Ting, S. Vijayakumar, S. Schaal, L. De Raedt, Leave-One-Out Cross-Validation, in: Encycl. Mach. Learn., Springer US, 2011: pp. 600–601. https://doi.org/10.1007/978-0-387-30164-8_469.

[81] B. Efron, The Jackknife, the Bootstrap and Other Resampling Plans, Society for Industrial and Applied Mathematics, 1982. https://doi.org/10.1137/1.9781611970319.

[82] J. Yuan, Y.M. Li, C.L. Liu, X.F. Zha, Leave-one-out cross-validation based model selection for manifold regularization, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Springer, Berlin, Heidelberg, 2010: pp. 457–464. https://doi.org/10.1007/978-3-642-13278-0_59.

[83] G.C. Cawley, Leave-one-out cross-validation based model selection criteria for weighted LS-SVMs, in: IEEE Int. Conf. Neural Networks - Conf. Proc., Institute of Electrical and Electronics Engineers Inc., 2006: pp. 1661–1668. https://doi.org/10.1109/ijcnn.2006.246634.

[84] P.G. Giresi, E.J. Stevenson, J. Theilhaber, A. Koncarevic, J. Parkington, R.A. Fielding, S.C. Kandarian, Identification of a molecular signature of sarcopenia, Physiol. Genomics. 21 (2005) 253–263. https://doi.org/10.1152/physiolgenomics.00249.2004.

[85] Understanding 8 types of Cross-Validation | by Satyam Kumar | Towards Data Science, (n.d.). https://towardsdatascience.com/understanding-8-types-of-cross-validation-80c935a4976d (accessed January 7, 2021).

[86] Hold-out vs. Cross-validation in Machine Learning | by Eijaz Allibhai | Medium, (n.d.). https://medium.com/@eijaz/holdout-vs-cross-validation-in-machine-learning-7637112d3f8f (accessed January 7, 2021).

[87] P. Pudil, J. Novovičová, J. Kittler, Floating search methods in feature selection, Pattern Recognit. Lett. 15 (1994) 1119–1125. https://doi.org/10.1016/0167-8655(94)90127-9.

[88] G. Chandrashekar, F. Sahin, A survey on feature selection methods, Comput. Electr. Eng. 40 (2014) 16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024.

[89] M.L. Bermingham, R. Pong-Wong, A. Spiliopoulou, C. Hayward, I. Rudan, H. Campbell, A.F. Wright, J.F. Wilson, F. Agakov, P. Navarro, C.S. Haley, Application of high-dimensional feature selection: Evaluation for genomic prediction in man, Sci. Rep. 5 (2015) 1–12. https://doi.org/10.1038/srep10312.

[90] P. Langley, Selection of Relevant Features in Machine Learning, 1994. www.aaai.org (accessed March 22, 2019).

[91] Z. Zhao, F. Morstatter, S. Sharma, S. Alelyani, A. Anand, H. Liu, Advancing feature selection research, (2010). https://eprints.kku.edu.sa/65/ (accessed March 22, 2019).

[92] B. Xue, M. Zhang, W.N. Browne, X. Yao, A Survey on Evolutionary Computation Approaches to Feature Selection, IEEE Trans. Evol. Comput. 20 (2016) 606–626. https://doi.org/10.1109/TEVC.2015.2504420.

[93] A.K. Jain, R.P.W. Duin, J. Mao, Statistical pattern recognition: A review, IEEE Trans. Pattern Anal. Mach. Intell. 22 (2000) 4–37. https://doi.org/10.1109/34.824819.

[94] M. Pahar, M. Klopper, R. Warren, T. Niesler, COVID-19 Cough Classification using Machine Learning and Global Smartphone Recordings, (2020). http://arxiv.org/abs/2012.01926 (accessed December 30, 2020).

[95] S. Lalmuanawma, J. Hussain, L. Chhakchhuak, Applications of machine learning and artificial intelligence for Covid-19 (SARS-CoV-2) pandemic: A review, Chaos, Solitons and Fractals. 139 (2020) 110059. https://doi.org/10.1016/j.chaos.2020.110059.

[96] P. Bachtiger, N.S. Peters, S.L. Walsh, Machine learning for COVID-19—asking the right questions, Lancet Digit. Heal. 2 (2020) e391–e392. https://doi.org/10.1016/S2589-7500(20)30162-X.

[97] M. van der Schaar, A.M. Alaa, A. Floto, A. Gimson, S. Scholtes, A. Wood, E. McKinney, D. Jarrett, P. Lio, A. Ercole, How artificial intelligence and machine learning can help healthcare systems respond to COVID-19, Mach. Learn. (2020) 1–14. https://doi.org/10.1007/s10994-020-05928-x.

[98] N. Sharma, P. Krishnan, R. Kumar, S. Ramoji, S.R. Chetupalli, N. R., P.K. Ghosh, S. Ganapathy, Coswara -- A Database of Breathing, Cough, and Voice Sounds for COVID-19 Diagnosis, ArXiv. (2020). http://arxiv.org/abs/2005.10548 (accessed January 10, 2021).

[99] M. Huzaifah, Comparison of Time-Frequency Representations for Environmental Sound

Classification using Convolutional Neural Networks, (2017). http://arxiv.org/abs/1706.07156 (accessed January 10, 2021).

[100] A.M. Ismael, A. Şengür, Deep learning approaches for COVID-19 detection based on chest X-ray images, Expert Syst. Appl. 164 (2021) 114054. https://doi.org/10.1016/j.eswa.2020.114054.

[101] V. Singh, R.C. Poonia, S. Kumar, P. Dass, P. Agarwal, V. Bhatnagar, L. Raja, Prediction of COVID-19 corona virus pandemic based on time series data using support vector machine, J. Discret. Math. Sci. Cryptogr. 23 (2020) 1583–1597. https://doi.org/10.1080/09720529.2020.1784535.

[102] M. Loey, G. Manogaran, M.H.N. Taha, N.E.M. Khalifa, A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic, Meas. J. Int. Meas. Confed. 167 (2021) 108288. https://doi.org/10.1016/j.measurement.2020.108288.

[103] H. Cen, R. Lu, Q. Zhu, F. Mendoza, Nondestructive detection of chilling injury in cucumber fruit using hyperspectral imaging with feature selection and supervised classification, Postharvest Biol. Technol. 111 (2016) 352–361. https://doi.org/10.1016/j.postharvbio.2015.09.027.