Article citation info:
Benmalek E, El Mhamdi J, Jilbab A, Jbari A. A cough-based COVID-19 detection with gammatone and mel-frequency cepstral coefficients. Diagnostyka. 2023;24(2):2023214. https://doi.org/10.29354/diag/166330.

DIAGNOSTYKA, 2023, Vol. 24, No. 2

e-ISSN 2449-5220
DOI: 10.29354/diag/166330

A COUGH-BASED COVID-19 DETECTION WITH GAMMATONE AND MEL-FREQUENCY CEPSTRAL COEFFICIENTS

Elmehdi BENMALEK *, **Jamal EL MHAMDI**, **Abdelilah JILBAB**, **Atman JBARI**

E2SN, ENSAM de Rabat, Mohammed V University in Rabat, Morocco
* Corresponding author, e-mail: elmehdi.benmalek@um5s.net.ma

Abstract

Many countries have adopted a public health approach that aims to address the particular challenges faced during the pandemic Coronavirus disease 2019 (COVID-19). Researchers mobilized to manage and limit the spread of the virus, and multiple artificial intelligence-based systems are designed to automatically detect the disease. Among these systems, voice-based ones since the virus have a major impact on voice production due to the respiratory system's dysfunction. In this paper, we investigate and analyze the effectiveness of cough analysis to accurately detect COVID-19. To do so, we distinguished positive COVID patients from healthy controls. After the gammatone cepstral coefficients (GTCC) and the Mel-frequency cepstral coefficients (MFCC) extraction, we have done the feature selection (FS) and classification with multiple machine learning algorithms. By combining all features and the 3-nearest neighbor (3NN) classifier, we achieved the highest classification results. The model is able to detect COVID-19 patients with accuracy and an f1-score above 98 percent. When applying FS, the higher accuracy and F1-score were achieved by the same model and the ReliefF algorithm, we lose 1 percent of accuracy by mapping only 12 features instead of the original 53.

Keywords: COVID-19, cough recordings, machine learning, Mel-frequency cepstral coefficients, gammatone cepstral coefficients, feature selection.

1. INTRODUCTION

SARS-Cov-2 infection is the cause of the new COVID-19, which includes mild and severe forms. Patients infected with SARS-Cov-2 may present with a wide range of symptoms. The most common signs of the disease are fever (73% of cases) as well as symptoms of the flu-like syndrome, in association with respiratory signs such as cough (82%) and dyspnoea (31%). More rarely, anosmia, ageusia, or hemoptysis can be found. Intestinal symptoms were also evident in 10% of patients, such as vomiting, diarrhea, or abdominal pain [1-2].

In the absence of a specific therapy available to date, it is essential to be able to diagnose this disease as early as possible to isolate infected subjects and thus limit the spread of the epidemic. The reference diagnostic method is laboratory research for viral RNA by reverse transcriptase–polymerase chain reaction (RT-PCR) from nasopharyngeal swabs. However, obtaining the results takes several hours, and only certain laboratories have this test. Furthermore, although the specificity of the viral test is excellent, its sensitivity is imperfect (60 to 70%) because it depends on the quality of the sample and the rate of viral replication within the upper respiratory tract [3-4].

During the first wave of the COVID-19 pandemic, the influx of patients challenged healthcare facilities to quickly adapt healthcare systems and services. Although a considerable effort was made to adapt facilities, care protocols and modalities, and infection protocols. In the fight against COVID-19, organizations have quickly applied their machine learning expertise to reduce the likelihood and risk of COVID-19 spreading.

Recent years have seen a huge increase in the use of deep learning and machine learning in medical applications. Several studies have proposed systems based on deep learning for the diagnosis of COVID-19 utilizing medical imaging [5-7]. In addition to tailored networks [8-9], others are constructed using pre-trained models using transfer learning [10-11].

Recent studies have been examining the differences between respiratory sounds from healthy persons and patients who tested positive for COVID-19 (such as coughs, breathing, and voice) [12-13], an electronic stethoscope was used to capture and evaluate the respiratory sounds of ten individuals who had COVID-19 infections. All patients were found to have anomalous breath sounds, including cackles, unclassifiable murmurs, abnormal vesicular breath sounds, and augmented or weakened voice resonance.

Recent pathological analyses revealed that COVID-19 patients' lungs displayed varying degrees of consolidation [14]. According to the findings of the imaging examination, the predominant

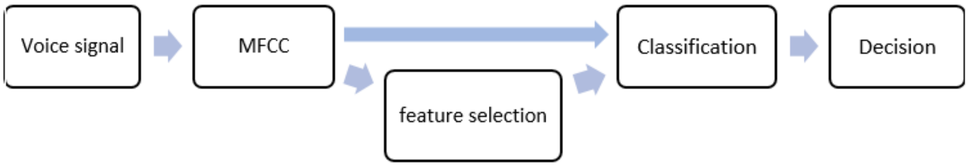


Fig. 1. COVID-19 diagnosis with cough recording diagram

symptoms of COVID-19 patients were many pulmonary plaques, ground glass shadows, infiltrating shadows, and in more severe cases, lung consolidation [15]. Other studies have suggested voice analysis to automatically detect patients suffering from COVID-19 [12, 16-17] since the voice and respiratory system are infected by the disease.

Based on various research indicating that COVID-19 patients' voices are infected by the disease [16], we present in this study, a machine learning method based on gammatone cepstral coefficients (GTCC) and the Mel-frequency cepstral coefficients (MFCC) for COVID-19 diagnosis, extracted from cough recordings. A binary classification was performed to discriminate positive COVID patients from healthy controls. The records are collected from the Coswara Dataset, a crowdsourcing project from the Indian Institute of Science (IIS). After data collection, we extracted the GTCC and the MFCC from the cough records. These acoustic features are mapped directly or after selection to k-nearest neighbor (kNN) for k equals 3, 5, and 7, Decision Tree (DT), deep neural network (DNN), and support vector machine (SVM). The FS is done by minimum redundancy maximum relevance (mRMR), ReliefF, and analysis of variance (ANOVA). The model evaluation is performed by the confusion matrix and the metrics such as sensitivity, specificity, precision, accuracy, f1-score, and the Matthews Correlation Coefficient (MCC).

2. METHODS

2.1. Dataset

The data is collected from the Coswara Project at the Indian Institute of Science Bangalore [18]. The dataset consists of vowel sustained phonation (/a/, /e/, and /o/), a counting exercise, breathing sounds, and cough recordings used in this investigation. On April 13th, 2020, the collection of records began.

The primary purpose of the data collection strategy was to reach out to the worldwide human population. To accomplish this, a website application with a simple and interactive user interface was created. Start recording sound samples using the device's microphone by opening the application in a web browser on a computer or mobile device, and entering the necessary metadata anonymously. The application is utilized for 5 to 7 minutes on average. The user was told to use a personal device, to clean it with sanitizer before and after recording, and to keep it 10 cm away from their mouths. The sampling frequency used for the audio samples is 48 kHz. The annotator (human) listens to each sound clip and responds to a few questions.

The dataset is divided into 77 positive COVID-19 cases and 82 healthy controls reflecting the true negatives. Tables 1 and 2 list the demographic information, symptoms, and comorbidities for each class.

2.2 Acoustic features

GTCC have demonstrated superiority when modeling cough signals over the traditional MFCC with a comparable computing cost [19]. The diversity between the two methods leads to better performance when they rebuild a combined feature space [20]. Therefore, in this study we utilized both cepstral coefficients algorithms to analyze COVID-19 cough records.

We have extracted 13 GTCC and MFCC coefficients, with deltas, in addition to the pitch for each frame, per subject, resulting in 53 features. The concept behind employing delta (differential) coefficients is that understanding the dynamics of the power spectrum is necessary for better speech recognition. The delta coefficients are calculated by the formula below.

Table. 1. Sex and age group of the participants for each class

		Age (years old)				
	Sex	[20-29]	[30-39]	[40-49]	[50-59]	>=60
Positive	45 Males 32 Females	48	9	6	11	3
Negative	54 Males 28 Females	44	11	7	12	8

Table 2. Disease symptoms and comorbidities of the participants for each class

	asthma	cold	cough	loss_of_smell	diabetes	fever	pneumonia
Positive	2	13	19	5	1	14	1
Negative	0	0	0	0	2	0	0

$$d_t = \frac{\sum_{n=1}^{N} n(C_{t+n} - C_{t-n})}{2\sum_{n=1}^{N} n^2} \qquad (1)$$

where d_t is a delta coefficient estimated in terms of the static coefficients C_{t-n} to C_{t+n} from frame t. n is commonly assumed to be 2.

2.3. MFCC

The MFCC coefficients (Mel Frequency Cepstral Coefficients) are the most used parameters in speech recognition systems. MFCC analysis consists of exploiting the properties of the human auditory system by transforming the linear frequency scale into the Mel scale. This last scale is encoded through a bank of 15 to 24 triangular filters spaced linearly up to 1 kHz, then spaced logarithmically up to the high frequencies. The conversion from linear scale to Mel scale is given by:

$$mel = 2595 \log_{10}(1 + \frac{f_{Hz}}{700}) \qquad (2)$$

On a signal analysis frame, the MFCC coefficients are calculated from the energies of the bank of triangular filters in frequency scale Mel [21]. The first d cepstral coefficients (in general d is chosen between 10 and 15) C_k can be calculated directly by applying the discrete cosine transform to the logarithm of the energies E_i of a bank of M filters:

$$C_k = \sum_{i=1}^{M} \log(E_i) \,.\, \cos\left[\frac{\pi k}{M}\left(i - \frac{1}{2}\right)\right] k = 0, \ldots, d \leq M \ (3)$$

The discrete cosine transform provides uncorrelated coefficients. The coefficient C_0 represents the sum of the energies. Generally, this coefficient is not used. It is replaced by the logarithm of the total energy E_0 calculated and normalized on the analysis frame in the time domain. In this study, we have used 13 MFCC coefficients.

2.4. GTCC

A collection of filters for the cochlea simulation is called the gammatone filter bank. The magnitude parameters of a human auditory filter are quite similar to the impulse response of a gammatone filter. A gammatone filter bank can be used to represent the mobility of the basilar membrane. The impulse response of a gammatone filter is the product of a sinusoidal tone with a center frequency f_c and a Gamma distribution.

$$g(t) = Kt^{n-1}e^{-2\pi Bt}\cos(2\pi f_c t + \varphi) \quad (4)$$

where K amplitude gain, n filter order, B is the filter bandwidth, f_c center frequency (in Hertz), and φ is a phase shift. The function utilized to simulate the human auditory response is comparable to that of the fourth-order gammatone filter [22].

As the center frequency rises, the gammatone filters' bandwidth also rises. The bandwidth of a fourth-order gammatone filter can be calculated using the equations:

$$B = 1.019 \times ERB(f_c) \qquad (5)$$

Where ERB is the equivalent rectangular bandwidth abbreviation. The auditory filter width at each location along the cochlea is measured psychoacoustically using the ERB. Rectangular band-pass filter simplification is used to estimate the bandwidth in human hearing. The value of ERB centered at frequency f_c is the auditory filter's bandwidth. ERB can be modeled in many different ways from f_c. The link between ERB and f can roughly be described as [23]

$$ERBS(f) = 24.7 + 0.108f \qquad (6)$$

The number of $ERBs$ below each frequency is the definition of the ERB scale:

$$ERBS(f) = 21.4 \log_{10}(1 + 0.00437f) \qquad (7)$$

Every point in the ERB space should be covered by the filter bank to accurately mimic the human auditory frequency spectrum. According to Eq.6, the center frequencies of each gammatone filter are evenly spaced on the ERB scale. which is given by

$$f_{ci} = ERBS^{-1}(ERBS(f_{low}) + \frac{ERBS(f_{high}) - ERBS(f_{low})}{N} i) \qquad (8)$$

Where $ERBS^{-1}$ is the inverse of the $ERBS$, N is the number of gammatone filters, f_{low} is the lowest frequency taken into consideration set to 10Hz, f_{high} is the highest frequency set to 2kHz, and i is the filter index. GTCC calculation block diagram is illustrated in figure 2

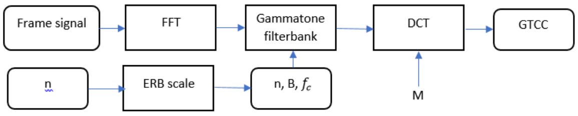


Fig. 2. GTCC calculation block diagram, where n is the number of filters in the filter bank, and M is the number of coefficients recovered by GTCC.

2.5. Classification algorithms

2.5.1. K-NN

The K-nearest neighbors (kNN) is an easy-to-implement supervised Machine Learning algorithm that can be used to solve classification and regression problems [24]. The intuition behind the K nearest neighbors algorithm is one of the simplest of all supervised machine learning algorithms:

Step 1: Select the number K of the neighbors

Step 2: Calculate the distance

Step 3: Take the K nearest neighbors according to the calculated distance.

Step 4: Among these K neighbors, count the number of points belonging to each category.

Step 5: Assign the new point to the most present category among these K neighbors.

2.5.2. SVM

Support vector machines (SVM) are supervised machine learning models focused on solving mathematical discrimination and regression problems. They were conceptualized in the 1990s from a statistical learning theory developed by Russian computer scientists Vladimir Vapnik and Alexey Chervonenkis: the Vapnik-Chervonenkis theory [25]. This model was quickly adopted due to its ability to work with high-dimensional data and the good results achieved in practice. Requiring a small number of parameters, SVMs are appreciated for their ease of use.

The principle of SVMs consists in reducing a classification or discrimination problem to a hyperplane (feature space) in which the data is separated into several classes whose boundary is as far as possible from the data points (or "maximum margin"). Hence the other name given to SVMs is wide-margin separators. The concept of boundary implies that the data is linearly separable. To achieve this, support vector machines use kernels, i.e. mathematical functions to project and separate data in vector space, "support vectors" being the data closest to the border. It is the furthest boundary of all the training points which is optimal, and which therefore presents the best capacity for generalization.

2.5.3. Decision tree

In graph theory, a tree is an undirected, acyclic, and connected graph. The set of nodes is divided into three categories:

The root node (access to the tree is through this node),

Internal nodes: nodes that have descendants (or children), which are in turn nodes,

Terminal nodes (or leaves): nodes that have no descendants.

Decision trees [26] are a category of trees used in data mining and business intelligence. They employ a hierarchical representation of the data structure in the form of sequences of decisions (tests) to predict an outcome or a class. Each individual (or observation), which must be assigned to a class, is described by a set of variables that are tested in the nodes of the tree. Testing is done in internal nodes and decisions are made in leaf nodes.

2.5.4. DNN

A set of interconnected formal neurons allows the resolution of complex problems, thanks to the adjustment of the weighting coefficients in a learning phase. A neural network is inspired by the functioning of biological neurons and takes shape in a computer in the form of an algorithm. The neural network can modify itself according to the results of its actions, which allows learning and problem-solving without algorithms, therefore without classical programming [27].

Unlike ordinary programs made to perform a given action, the neural network uses an algorithm to learn new data from previously recorded examples that it analyzes rigorously. This is a true model of learning by experience. It is also endowed with generalization and classification capabilities which allow it to carry out very sophisticated statistical operations. After standardizing the data, we set the first layer at 25 neurons the second and the third at 10 neurons each with ReLu activation function, and 100 iterations.

2.6. Feature selection

Feature selection is a dimensionality reduction method used in machine learning and data processing. It consists, in a high-dimensional space, in finding a subset of relevant variables. That is to say that one seeks to minimize the loss of information coming from the suppression of all the other variables. FS techniques are used for four reasons:

Simplify the models to facilitate their interpretation by researchers/users,

Reduce learning time,

Avoid the scourge of dimension,

Improve generalization by reducing overfitting.

2.6.1. mRMR

“Min-Redundancy, Max-relevance” (mRMR) is a filtering method for FS proposed by [28]. This method is based on classical statistical measures such as mutual information, correlation, etc. The basic idea is to take advantage of these measures to try to minimize the redundancy (mR) between the variables and maximize relevance (MR). The authors use mutual information to calculate the two factors mR and MR. The calculation of the redundancy and the relevance of a variable is given by the following equation:

$$Redondance(i) = \frac{1}{|F|^2}\sum_{i,j\in F} I(i,j) \qquad (9)$$

$$Pertinence(i) = \frac{1}{|F|^2}\sum_{i,j\in F} I(i,Y) \qquad (10)$$

– $|F|$: represents the size of the set of variables.

– $I(i,j)$: is the mutual information between the i^{th} and the j^{th} variable.

– $I(i,Y)$: is the mutual information between the i^{th} variable and the set of labels of class Y. The score of a variable is the combination of these two factors such as:

$$Score\ (i) = Relevance\ (i) - Redundancy\ (i) \qquad (11)$$

2.6.2. ReliefF

This algorithm, introduced under the name of ReliefF [29], does not content itself with eliminating redundancy but defines a criterion of relevance. This criterion measures the ability of each feature to group data with the same label and discriminates between data with different labels. The algorithm is described below.

1: Initialize the weights
2: Randomly draw a data X_i
3: Find the K nearest neighbors of X_i having the same tags (hits),

4: Find the K nearest neighbors of X_i having a label different from the class of Xi (misses)

5: For each characteristic update the weights

$$W_d = w_d - \sum_{j=1}^{K} \frac{diff(x_i,d_i,hits_i)}{m*k} + \sum_{c\neq class(x_i)} \left(\frac{p(c)}{1-p(class(x_i))}\right) \sum_{j=1}^{K} \frac{diff(x_i,d_i,hits_i)}{m*k} \quad (12)$$

6: The distance used is defined by:

$$diff(x_i, d_i, hits_i) = \frac{|x_i d - x_j d|}{\max(d) - \min\,(d)} \quad (13)$$

$Max(d)$ (resp. $min(d)$) designates the maximum (resp. minimum) value that the characteristic designated by the index d can take, on the set of data. $x_i d$ is the value of the d^{th} characteristic of x_i.

2.6.3. ANOVA

In ANOVA, we study a quantitative variable to which we assign one or two qualitative variables: categorical variables. These categorical variables are called “factors” or “variability factors”. If the analysis of variance focuses on a single factor, then it is called one-way analysis or One-way ANOVA. If several factors enter into the analytical test, then we speak of two-factor analysis, multifactorial or MANOVA for Multivariate Analysis Of Variance.

ANOVA is used concretely to highlight the existence of an interaction between these variability factors and the main quantitative variable studied, generally, a population divided into groups.

We use ANOVA to understand how the different groups respond to the statistical test. If there is a statistically significant result, i.e. the means of the different groups are equal on the factors studied, this means that the two population groups are similar.

Like other types of statistical tests, ANOVA compares the means of different groups and demonstrates the existence of statistical differences between the means. This statistical method is part of the omnibus tests. This means that it identifies a difference but does not tell which specific groups are statistically different from each other.

2.7. Model evaluation

The dataset is split into two sections. For each label, 80% of the data is used for training, while the remaining 20% is used for testing. We have 63 positives and 66 negatives in the training set, and we used 14 positives and 16 negatives in the test set. We used all of the GTCC and MFCC gathered from cough records to increase the performance of our model. That is, we classified the signals based on their frames, and the frames were labeled based on the original data. As a result, there were 12684 observations, which were divided as follows:

- 7424 negatives (5940 training and 1484 test)
- 5260 positives (4208 training and 1052 test)

To avoid data leakage, the acoustical features of the test set were extracted independently from the training set, hence no test frame was used when training our models.

2.7.1. k-fold validation

To accomplish cross-validation, we employ the k-fold cross-validation approach. The input data is split into k subsets for k-fold cross-validation (in this work we set k to 5). The machine learning model is trained on all subsets except one (k-1) and then evaluated on the subset that was not utilized for training. This process is repeated k times, each time with a distinct subset saved for evaluation (and not used for training). We evaluated the models with the test set after verifying them with the training set using 5-fold cross-validation, first by frames, then by subjects.

2.7.2. Confusion matrix

A confusion matrix is a tool for measuring the performance of classification models with 2 or more classes. In the binary case (i.e. with two classes, the simplest case), the confusion matrix is a table with 4 values representing the different combinations of actual values and predicted. This matrix is essential to define the different classification metrics such as sensitivity, specificity, precision, accuracy, f1-score, and the MCC.

3. RESULTS

3.1. Dataset

In this section, we present the results achieved by applying the models to the test set (30 observations/2536 frames)

- 16 negatives (1484 frames)
- 14 positives (1052 frames)

3.2. All features

Table 3 illustrates the confusion matrix obtained using all features. The best classification results are obtained by the 3NN. This model was able to achieve:

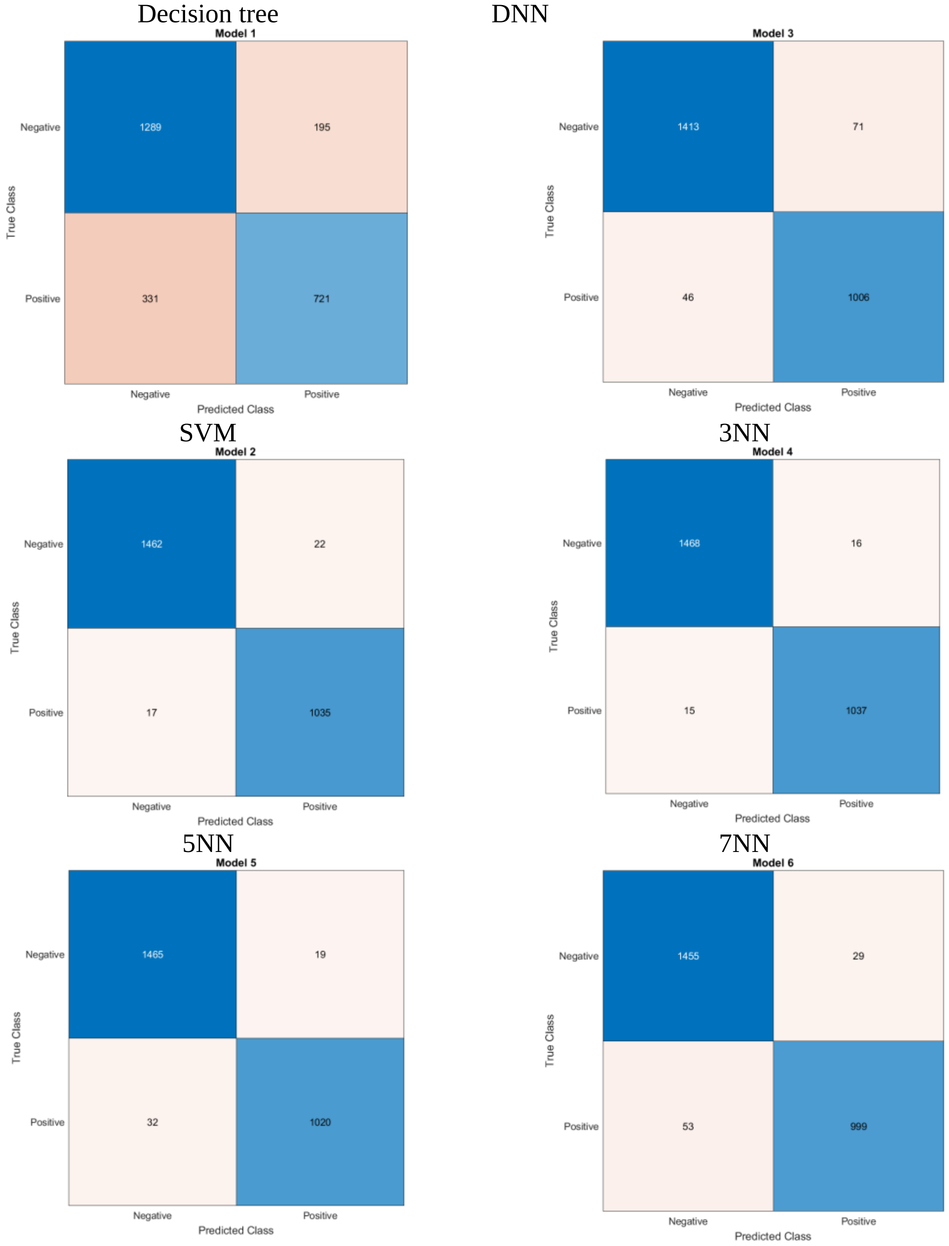
For the negatives, 1468 correctly classified observations, with only 16 false positives,

As for the class of the positives, 1037 observations have been detected, and 15 misclassified,

The ROC curve and AUC values for all classifiers are presented in table 4. To evaluate the performances of the models, we calculated the accuracy, sensitivity, specificity, precision, f1-score, and MCC for each classifier (Table 5). The k-NN classifiers have achieved very good performances, they obtained an accuracy of 98.78 percent for k=3, 98 percent for k=5, and 96.65 percent for k=7. The SVM was able to obtain the second-best accuracy result with 98.46 percent. The DNN has 95.31 percent accuracy, and lastly 79.26 percent for DT.

With 98.57 percent sensitivity, and 98.92 percent specificity, the 3NN is able to correctly identify patients with and without COVID-19 using cough records. Other models except for DT have achieved high percentages in both sensitivity and specificity exceeding 95 percent. The precision using 3NN is 98.48 percent, meaning 98 percent of COVID-19 patients labeled by the algorithm are truly suffering from the disease, which is very handy in a real-world application. The other models have comparable results; 98.17 percent for 5NN, 97.92 percent for SVM, 97 percent for 7NN, and 93.4 percent for DNN. As for the DT, the precision obtained was 78.71 percent. The F1-score values for each model were calculated to evaluate the overall performances. The 3NN has an f1-score of 98.53 percent, followed by the SVM at 98.15 percent, the 5NN at 97.56 percent, the 7NN at 95.68 percent, and the DNN at 94.42 percent. All these models are able to distinguish true positive, true negative, false positive, and false negative quite accurately. The DT comes last with a 73.27 percent f1-score. The MCC score of the 3NN is 0.97, 0.96 for the 5NN and SVM, 0.93 for the 7NN, and 0.9 obtained by the DNN. These models have an excellent correlation between prediction and observation, meaning highly precise COVID-19 diagnosis.

Table 3. Confusion matrices using all features



Decision tree

Model 1

True Class \ Predicted Class	Negative	Positive
Negative	1289	195
Positive	331	721

DNN

Model 3

True Class \ Predicted Class	Negative	Positive
Negative	1413	71
Positive	46	1006

SVM

Model 2

True Class \ Predicted Class	Negative	Positive
Negative	1462	22
Positive	17	1035

3NN

Model 4

True Class \ Predicted Class	Negative	Positive
Negative	1468	16
Positive	15	1037

5NN

Model 5

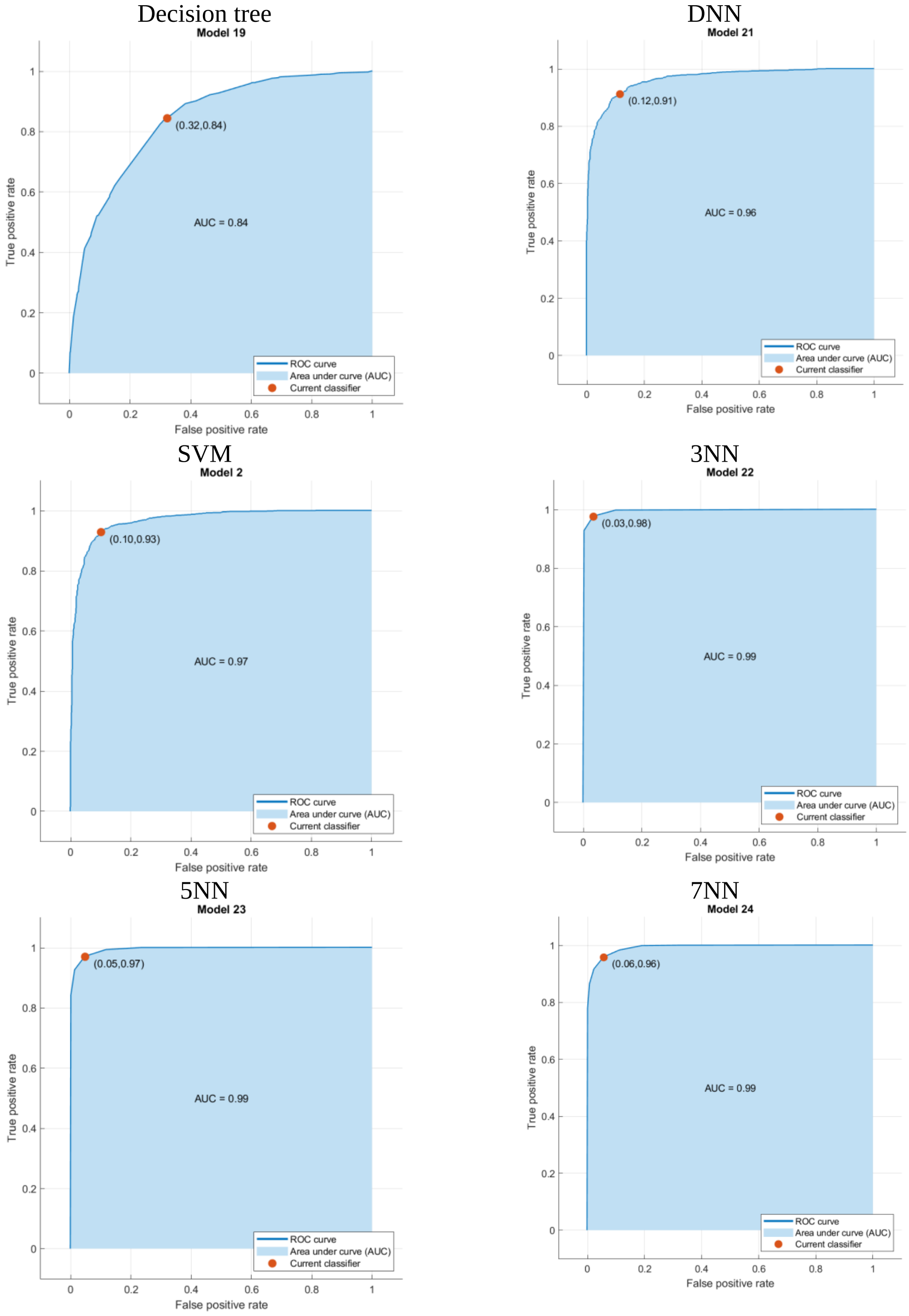
True Class \ Predicted Class	Negative	Positive
Negative	1465	19
Positive	32	1020

7NN

Model 6

True Class \ Predicted Class	Negative	Positive
Negative	1455	29
Positive	53	999

Table 4. ROC curves and AUC values using all features


Decision tree

DNN

SVM

3NN

5NN

7NN

Table 5. Evaluation metrics using all features

	Sensitivity %	Specificity %	Precision %	Accuracy %	F1-score %	MCC
DT	68.54	86.86	78.71	79.26	73.27	0.57
DNN	95.45	95.22	93.4	95.31	94.42	0.9
SVM	98.38	98.52	97.92	98.46	98.15	0.96
3NN	**98.57**	**98.92**	**98.48**	**98.78**	**98.53**	**0.97**
5NN	96.96	98.72	98.17	98	97.56	0.96
7NN	94.5	98	97	96.65	95.68	0.93

3.3. mRMR

We applied the mRMR to select the more pertinent features. Figure 3 shows the features ranked according to their importance score. In this case, the 12 features mapped to the classifiers are MFCC_coeff_9, GTCC_coeff_3, MFCC_coeff_12, MFCC_coeff_13, GTCC_coeff_5, MFCC_coeff_10, GTCC_coeff_6, GTCC_coeff_4, GTCC_coeff_13, GTCC_coeff_11, GTCC_coeff_12, and GTCC_coeff_7. We have selected the first 12 features since after the 12th feature the importance score becomes negligible. From figure 3, we noticed the first feature has high

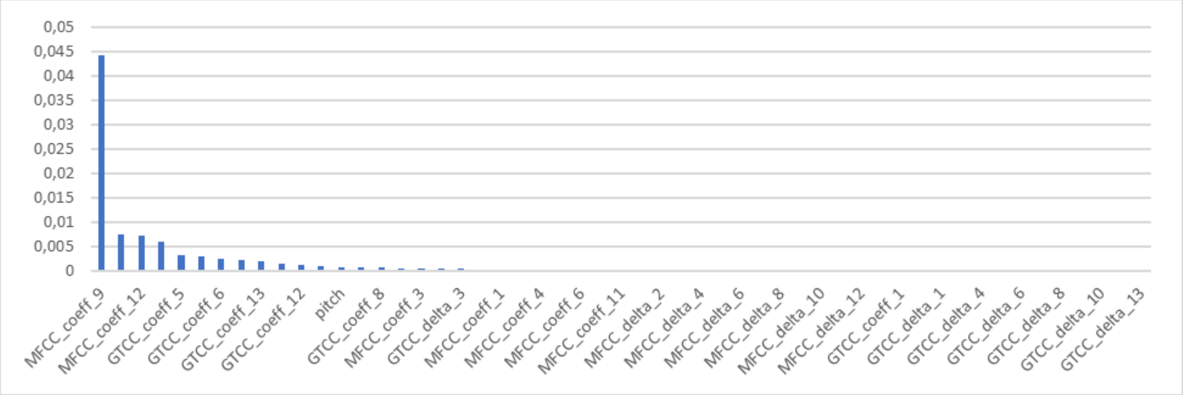
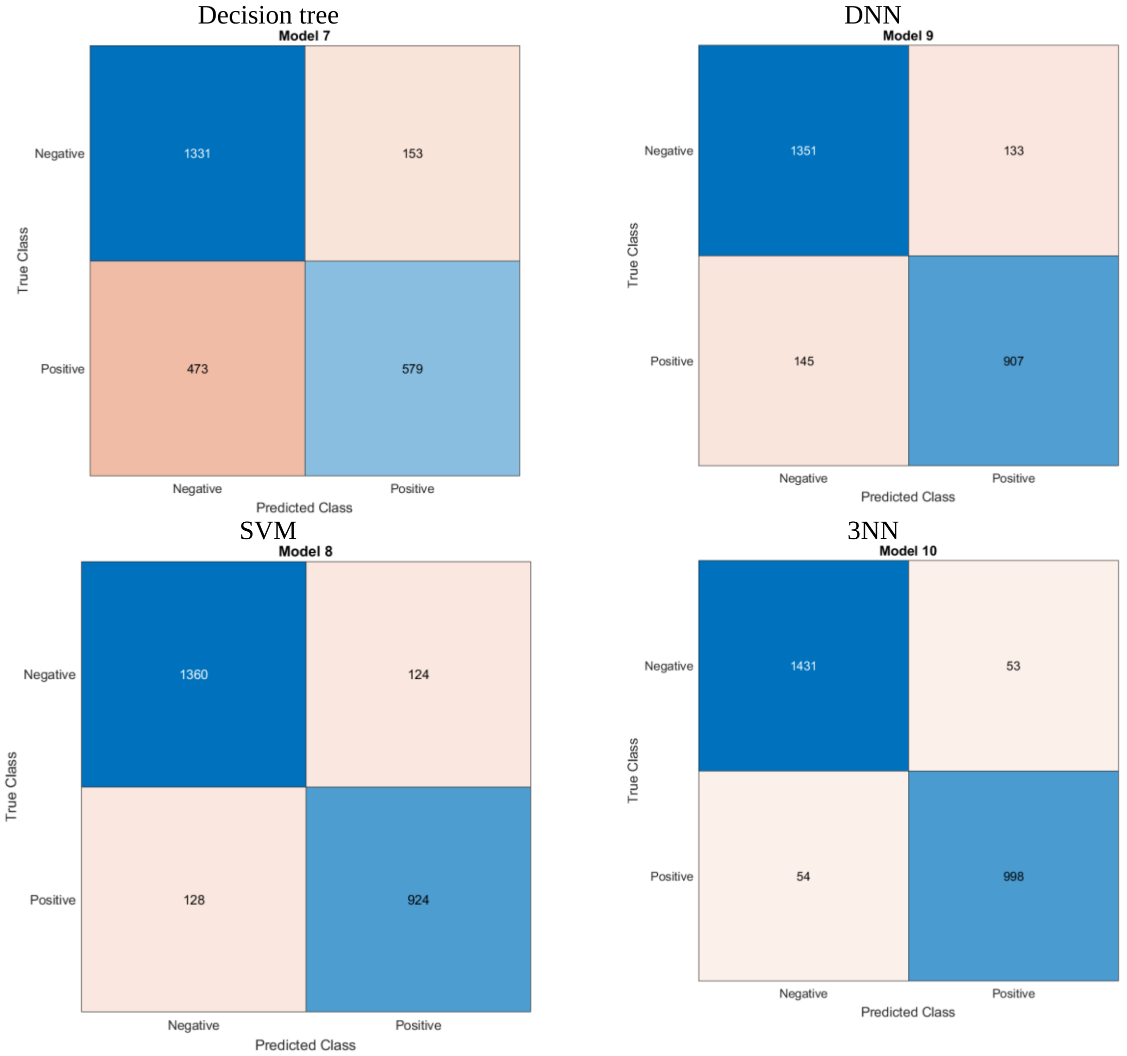


Fig. 3. Feature importance score sorted using the mRMR

Table 6. Confusion matrices using the mRMR



Decision tree

Model 7

True Class \ Predicted Class	Negative	Positive
Negative	1331	153
Positive	473	579

DNN

Model 9

True Class \ Predicted Class	Negative	Positive
Negative	1351	133
Positive	145	907

SVM

Model 8

True Class \ Predicted Class	Negative	Positive
Negative	1360	124
Positive	128	924

3NN

Model 10

True Class \ Predicted Class	Negative	Positive
Negative	1431	53
Positive	54	998

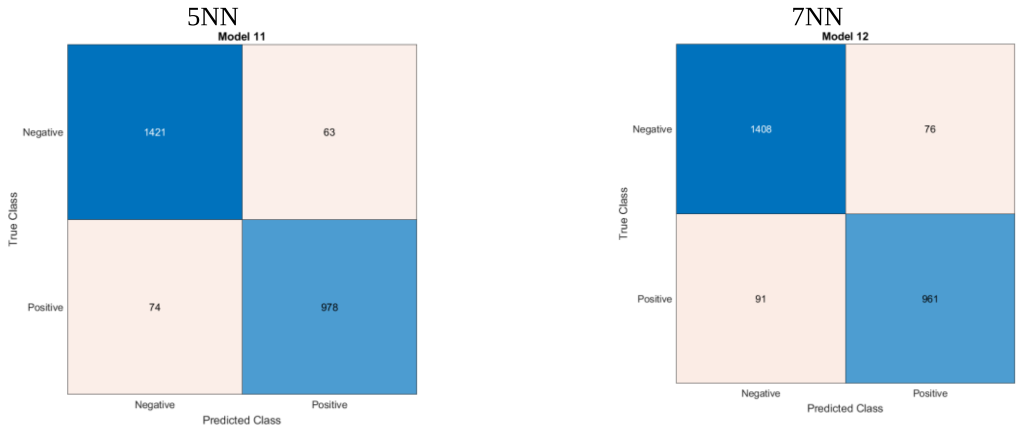

5NN 7NN

Table 7. ROC curves and AUC values using mRMR

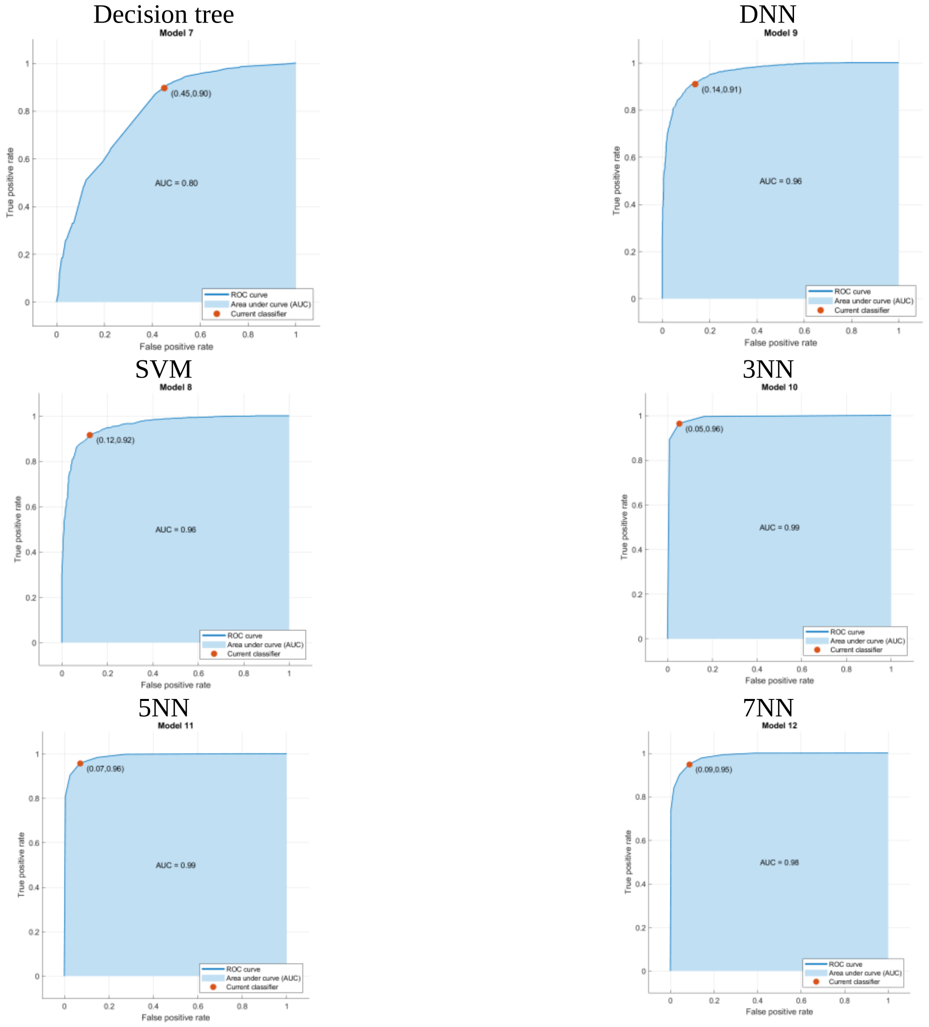

Decision tree DNN SVM 3NN 5NN 7NN

Table 8. Evaluation metrics using mRMR

	Sensitivity %	Specificity %	Precision %	Accuracy %	F1-score %	MCC
DT	55.04	89.7	79.1	75.3	64.91	0.47
DNN	86.22	91.04	87.21	89.04	86.71	0.77
SVM	87.83	91.64	88.17	90.06	88	0.79
3NN	**94.39**	**96.43**	**94.48**	**95.63**	**94.44**	**0.91**
5NN	92.97	95.75	93.95	94.6	93.45	0.88
7NN	91.35	94.88	92.67	93.41	92	0.86

importance score compared to the other ones, and after the fourth, the score is very low.

When performing FS by the mRMR, the 3NN has the best classification results:

For the negatives, the model has correctly classified 1431 observations, with 53 false positives,

In the class of the positives, 998 observations have been identified, and 54 are misclassified.

When performing FS by the mRMR, the 3NN has the highest sensitivity (94.39 percent), specificity (96.43 percent), precision (94.48 percent), accuracy (95.63 percent), f1-score (94.44 percent), and MCC (0.91). Followed by the other kNN models; the 5NN and the 7NN have 94.6 percent and 93.41 percent accuracy, 93.45 percent and 92 percent f1-score, and 0.88 and 0.86 MCC, respectively. As for the other classifiers the SVM and the DNN have a good performance for all metrics, and the worst results are achieved by the DT. The ROC curve and AUC values for all classifiers are presented in table 7.

3.4. ReliefF

Figure 4 represents the feature sorted based on the importance scores using the ReliefF FS algorithm. By applying the ReliefF, we have used: GTCC_coeff_1, MFCC_coeff_13, GTCC_coeff_4, MFCC_coeff_9, GTCC_coeff_6, MFCC_coeff_6, MFCC_coeff_5, MFCC_coeff_1, GTCC_coeff_10, GTCC_coeff_13, MFCC_coeff_4, and GTCC_coeff_5.

Reducing the space features by mapping only the 12 first features sorted by the ReliefF, the 3NN has the higher accuracy (97.6 percent). The other kNN models achieved 96.77 percent for k=5, and 95.86 percent for k=7. The SVM has 91.48 percent, the DNN 89.79 percent, and DT obtained 76.38 percent. The 3NN detailed performance for each class is as follows:

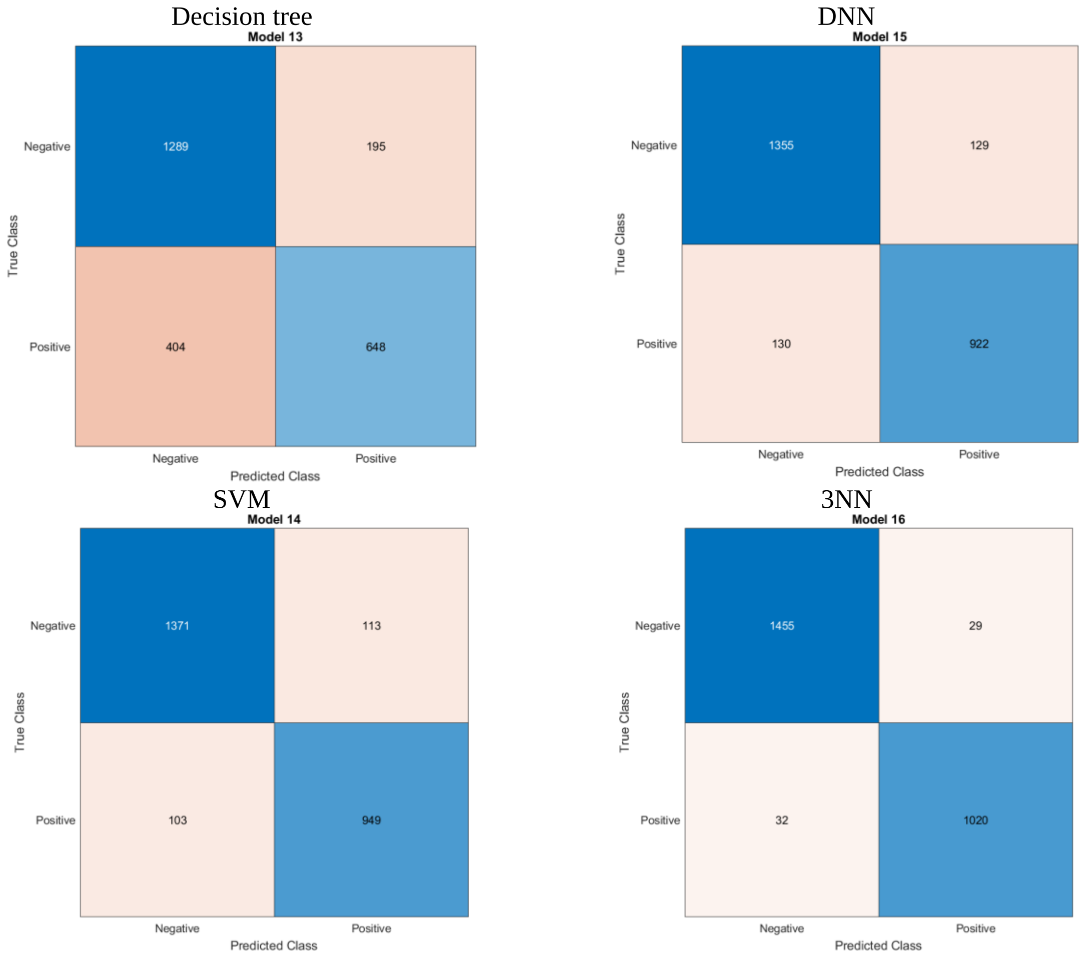
1456 observations negatives have been correctly classified, with 29 false positives.

And 1020 positive subjects have been detected, and 32 misclassified, for a sensitivity of 96.96 percent, a specificity of 98.05 percent, a precision of 97.24 percent, an f1-score of 97.1 percent, and an MCC of 0.95.

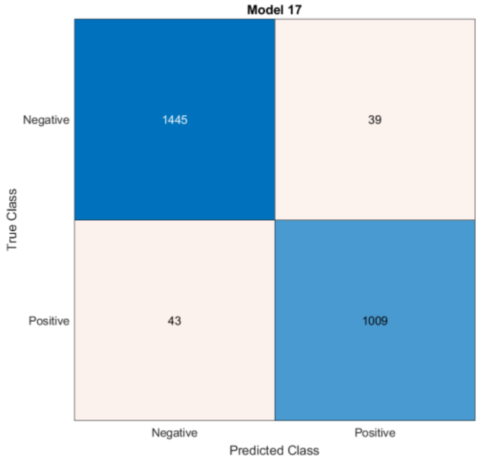
The ROC curve and AUC values for all classifiers are presented in table 10.

The other kNN models have a percentage above 95 percent for all metrics, with 0.93 MCC for 5NN and 0.91 MCC for 7NN. Both SVM and DNN have achieved good results; f1-score of 87.68 percent and 0.79 MCC by using the DNN, and 89.68 percent f1-score and 0.82 MCC by the SVM (Table 11).

Table 9. Confusion matrices using the ReliefF algorithm



5NN



7NN

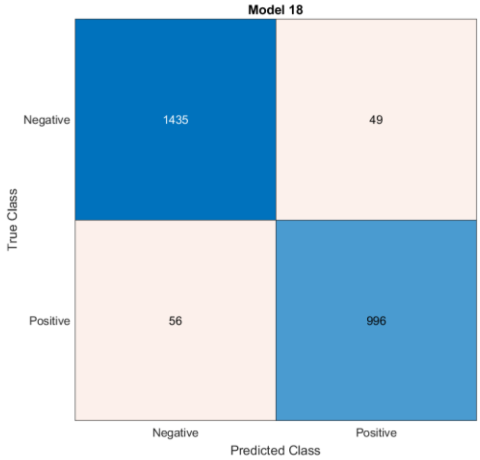
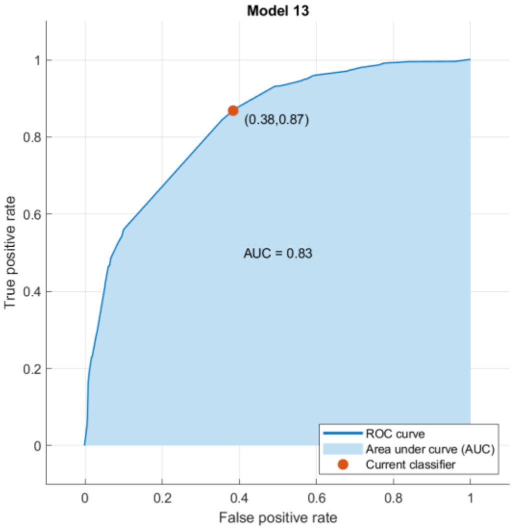
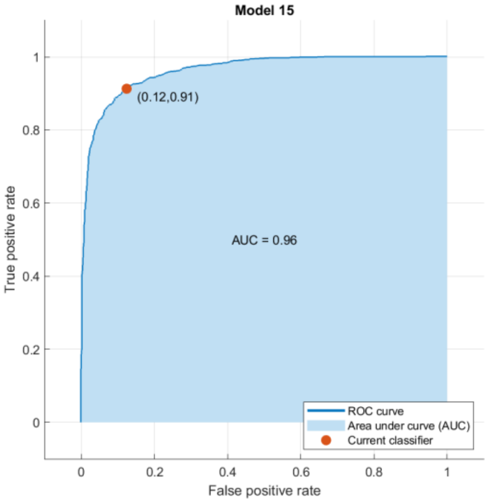


Table 10. ROC curves and AUC values using ReliefF

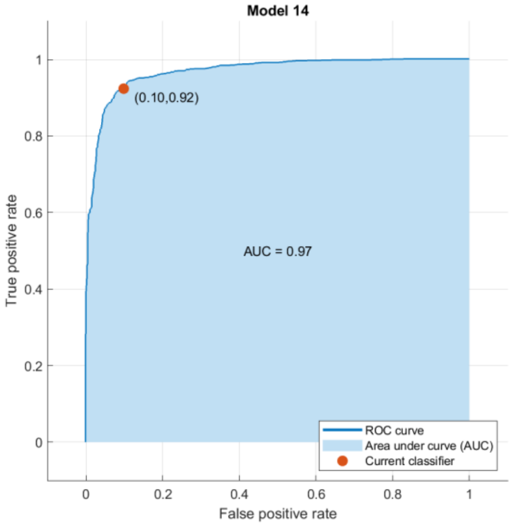
Decision tree



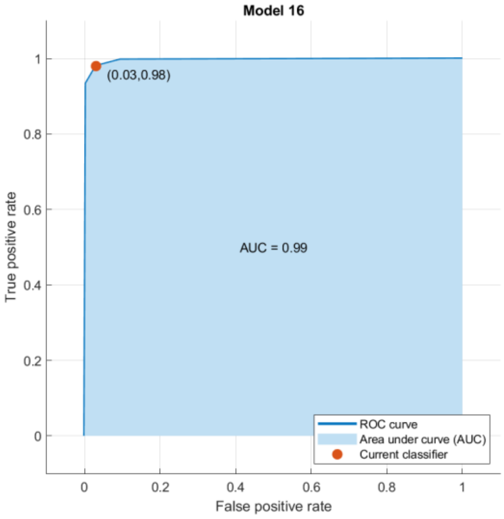
DNN



SVM



3NN



5NN

7NN

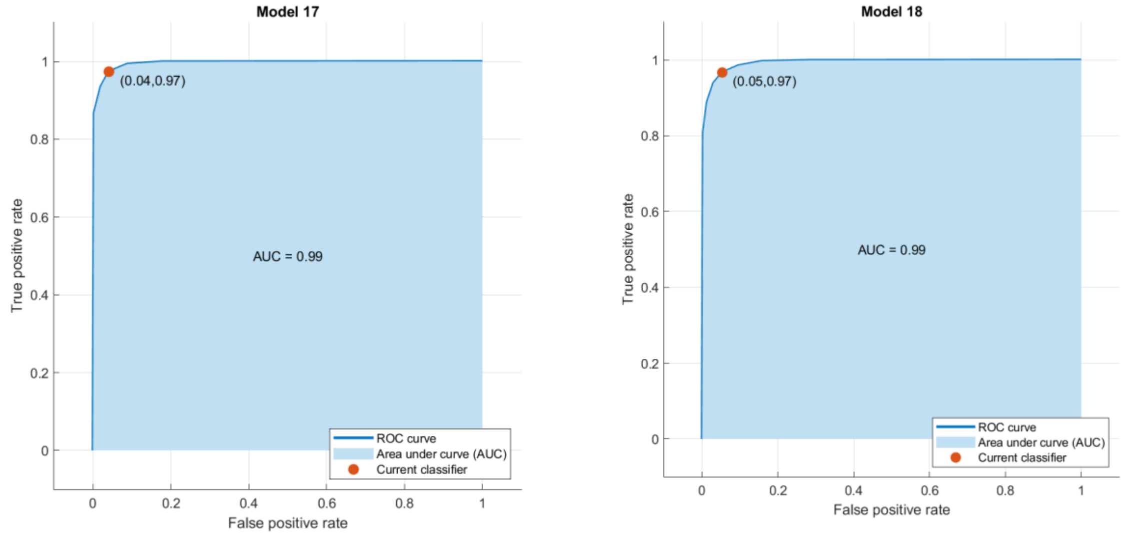


Table 11. Evaluation metrics using ReliefF

	Sensitivity %	Specificity %	Precision %	Accuracy %	F1-score %	MCC
DT	61.6	86.86	76.87	76.38	68.39	0.51
DNN	87.64	91.3	87.73	89.79	87.68	0.79
SVM	90.21	92.39	89.36	91.48	89.78	0.82
3NN	**96.96**	**98.05**	**97.24**	**97.6**	**97.1**	**0.95**
5NN	95.91	97.37	96.28	96.77	96.1	0.93
7NN	94.68	96.7	95.31	95.86	95	0.91

3.5. ANOVA

By applying the ANOVA FS method, the features are listed based on their importance scores in Figure 5. The first 12 selected features are: MFCC_coeff_13, MFCC_coeff_9, MFCC_coeff_12, GTCC_coeff_5, GTCC_coeff_6, GTCC_coeff_11, GTCC_coeff_4, MFCC_coeff_6, GTCC_coeff_12, GTCC_coeff_13, GTCC_coeff_7, and GTCC_coeff_2.

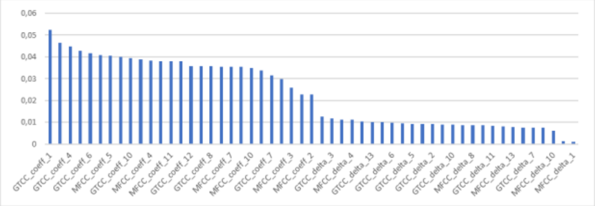


Fig. 4. Feature importance scores sorted using the ReliefF algorithm

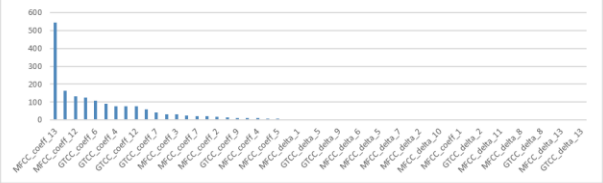


Fig. 5. Feature importance scores sorted using the ANOVA algorithm

Reducing the space features by selecting the 12 features ranked first by ANOVA, the 3NN again has the best results; 97.2 percent accuracy, 96.58 percent sensitivity, 97.64 percent specificity, 96.67 percent precision, 96.62 f1-score, 0.94 MCC. By classifying the dataset with this model, we have:

1449 negative observations correctly classified, 35 false positives, for a recall of 99 percent,

And 1016 positive subjects have been detected, 36 false negatives,

To test the effectiveness of our model to detect COVID-19 patients by using the cough recording, we classified the subjects using the 3NN and the ReliefF FS algorithm, we selected this model since we lose 1 percent of accuracy with only 22 percent of features, (12 features instead of 53). The classification was done based on the most frequent value predicted by frame for each subject, and the results demonstrate how accurate our approach is in predicting the subjects with COVID-19 disease (Figure 6).

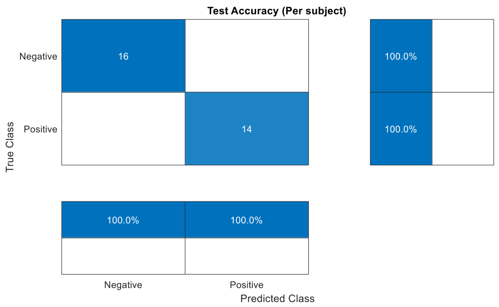
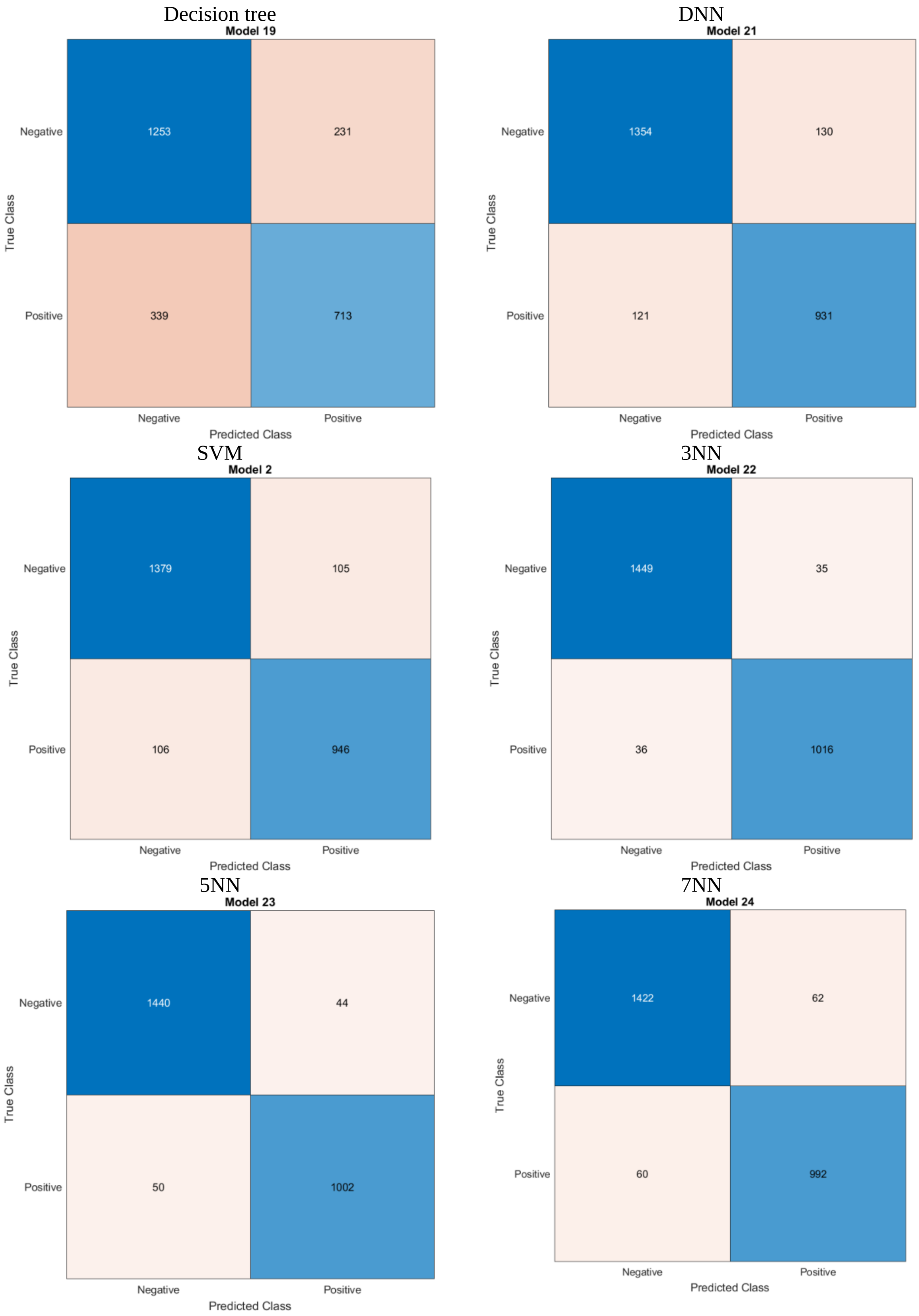


Fig. 6. confusion matrix by subject using the SVM and the ReliefF FS algorithm

Table 12. Confusion matrices using ANOVA


Decision tree

DNN

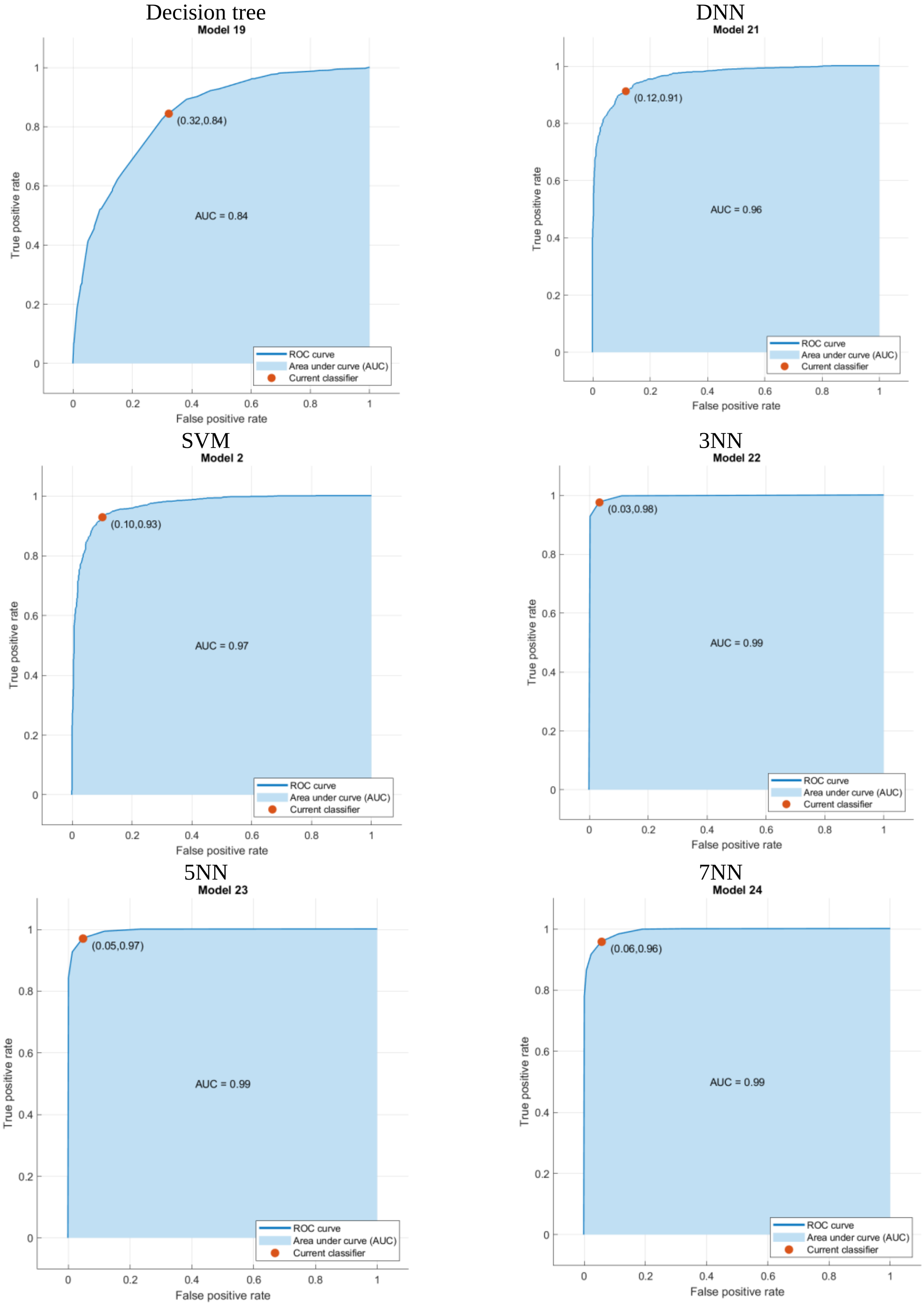
SVM

3NN

5NN

7NN

Table 13. ROC curves and AUC values using ANOVA



Decision tree

DNN

SVM

3NN

5NN

7NN

Table 14. Evaluation metrics using ANOVA

	Sensitivity %	Specificity %	Precision %	Accuracy %	F1-score %	MCC
DT	67.78	84.43	75.53	77.52	71.44	0.53
DNN	88.5	91.24	87.75	90.1	88.12	0.78
SVM	89.92	92.92	90	91.68	89.97	0.83
3NN	**96.58**	**97.64**	**96.67**	**97.2**	**96.62**	**0.94**
5NN	95.25	97.04	95.8	96.3	95.52	0.92
7NN	94.3	95.82	94.12	95.19	94.2	0.9

4. CONCLUSION

In this study, we examine and evaluate the capability of cough analysis to reliably identify COVID-19. After extracting the gammatone and Mel-frequency cepstral coefficients from COVID-19 positive patients and healthy controls, we used a variety of machine learning methods to select the features and classify the observations. The best classification results were obtained when all characteristics and the 3NN classifier were combined. The model has an accuracy and f1-score of about 98 percent. The 3NN has the advantage when applying FS, with accuracy and an f1-score above 97 percent when it was combined with the ReliefF. All model shows promising results except for the DT, which indicate the effectiveness of the approach presented in this paper. The AI-based screening method is important for restricting the transmission of the virus, by developing models with accurate sensitivity and specificity; in our case, we were able to reach 98 percent for both metrics, with such high accuracy we believe the proposed method could help in the development of straightforward, inexpensive, quick, and accurate diagnosis system to reduce the likelihood and risk of COVID-19 spreading, and for better monitoring and management of the disease.

These findings are promising, but they must be validated by a controlled clinical trial conducted by medical specialists. Furthermore, because of the pandemic's rapid and ongoing expansion, there is still a lack of understanding regarding the disease's aetiology and progression, as well as the association between demographic and clinical data of COVID-19 patients. In this investigation, we concentrated primarily on the effects of COVID-19 infection on voice quality. In our upcoming works, we intend to implement these models in embedded systems and assess their real-world effectiveness.

Author contributions: *research concept and design, E,B., J.E-M., A.J.; Collection and/or assembly of data, E,B.; Data analysis and interpretation, E,B., Writing the article, E,B., A.J.; Critical revision of the article, E,B., J.E0M., A.J., A.J.; Final approval of the article, E,B., J.E-M., A,J., A.J.*

Declaration of competing interest: *The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.*

REFERENCES

1. Wu Z, McGoogan JM (2020). Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 2019; 323(13): 1239-1242. https://doi.org/10.1001/jama.2020.2648.
2. Guan WJ, Ni Z, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC, Du B, Li LJ. Clinical characteristics of coronavirus disease 2019 in China. New England Journal of Medicine 2020; 382(18): 1708-1720. https://doi.org/10.1056/NEJMoa2002032.
3. Wang W, Xu Y, Gao R, Lu R, Han K, Wu G, Tan W. Detection of SARS-CoV-2 in different types of clinical specimens. Jama 2020; 323(18): 1843-1844. https://doi.org/10.1001/jama.2020.3786.
4. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, Xia L. (2020). Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology 2020; 296(2): 32-40. https://doi.org/10.1148/radiol.2020200642.
5. Mei X, Lee HC, Diao KY, Huang M, Lin B, Liu C, Yang Y. Artificial intelligence–enabled rapid diagnosis of patients with COVID-19. Nature medicine 2020; 26(8): 1224-1228. https://doi.org/10.1038/s41591-020-0931-3.
6. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, van Smeden M. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 2020; 369. https://doi.org/10.1136/bmj.m1328.
7. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Cao B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet 2020; 395(10223): 497-506. https://doi.org/10.1016/S0140-6736(20)30183-5.
8. Fan DP, Zhou T, Ji GP, Zhou Y, Chen G, Fu H, Shao L. Inf-net: Automatic covid-19 lung infection segmentation from CT images. IEEE Transactions on Medical Imaging 2020; 39(8): 2626-2637. https://doi.org/10.1109/TMI.2020.2996645.
9. Pereira RM, Bertolini D, Teixeira LO, Silla Jr CN, Costa YM. COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios. Computer methods and programs in biomedicine 2020; 194: 105532. https://doi.org/10.1016/j.cmpb.2020.105532.
10. Benmalek E, Elmhamdi J, Jilbab A. Comparing CT scan and chest X-ray imaging for COVID-19 diagnosis. Biomedical Engineering Advances 2021; 1: 100003. https://doi.org/10.1016/j.bea.2021.100003.
11. El Asnaoui K, Chawki Y. Using X-ray images and deep learning for automated detection of coronavirus disease. Journal of Biomolecular Structure and Dynamics 2021; 39(10): 3615-3626. https://doi.org/10.1080/07391102.2020.1767212.
12. Brown C, Chauhan J, Grammenos A., Han J, Hasthanasombat A, Spathis D, Mascolo C. Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data. arXiv preprint arXiv 2020; 05919. https://doi.org/10.48550/arXiv.2006.05919
13. Huang Y, Meng S, Zhang Y, Wu S, Zhang Y, Zhang Y, Cai J. The respiratory sound features of COVID-19 patients fill gaps between clinical data and screening methods. MedRxiv 2020. https://doi.org/10.1101/2020.04.07.20051060.
14. Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, Wang FS. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. The Lancet respiratory medicine 2020, 8(4), 420-422. https://doi.org/10.1016/s2213-2600(20)30076-x.
15. Pan F, Ye T, Sun P, Gui S, Liang B, Li L, Zheng C. (2020). Time course of lung changes on chest CT during recovery from 2019 novel coronavirus

(COVID-19) pneumonia. Radiology 2020. https://doi.org/10.1148%2Fradiol.2020200370.

16. Han J, Brown C, Chauhan J, Grammenos A, Hasthanasombat A, Spathis D, Mascolo C. (2021, June). Exploring automatic COVID-19 diagnosis via voice and symptoms from crowdsourced data. ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2021; 8328-8332. https://doi.org/10.1109/ICASSP39728.2021.9414576.
17. Kumar LK, Alphonse PJA. Automatic diagnosis of COVID-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: cough, voice, and breath. Alexandria Engineering Journal 2022; 61(2): 1319-1334. https://doi.org/10.1016/j.aej.2021.06.024.
18. Sharma N, Krishnan P, Kumar R, Ramoji S, Chetupalli SR, Ghosh PK, Ganapathy S. Coswara--a database of breathing, cough, and voice sounds for COVID-19 diagnosis. arXiv preprint arXiv:2005.10548 2022. https://doi.org/10.48550/arXiv.2005.10548.
19. Valero X, Alias F. Gammatone cepstral coefficients: Biologically inspired features for non-speech audio classification. IEEE Transactions on Multimedia 2012; 14(6):1684-1689. https://doi.org/10.1109/TMM.2012.2199972.
20. Liu JM, You M, Li GZ, Wang Z, Xu X, Qiu Z, Chen S. Cough signal recognition with gammatone cepstral coefficients. In 2013 IEEE China Summit and International Conference on Signal and Information Processing 2013; 160-164. https://doi.org/10.1109/ChinaSIP.2013.6625319.
21. Haton JP, Cerisara C, Fohr D, Laprie Y, Smaïli K. Reconnaissance automatique de la parole: Du Signal à son Interprétation. Dunod 2006; 392.
22. Slaney M. An efficient implementation of the Patterson-Holdsworth auditory filter bank. Apple Computer, Perception Group, Technical Report 1997; 35(8).
23. Moore BC, Glasberg BR. A revision of Zwicker's loudness model. Acta Acustica united with Acustica 1996; 82(2): 335-345.
24. Indyk P, Motwani R. Approximate nearest neighbors: towards removing the curse of dimensionality. Proceedings of the thirtieth annual ACM symposium on Theory of computing 1998; 604-613. https://doi.org/10.1145/276698.276876.
25. Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. Proceedings of the fifth annual workshop on Computational learning theory 1992; 144-152. https://doi.org/10.1145/130385.130401.
26. Quinlan JR. Induction of decision trees. Machine learning 1986; 1(1): 81-106. https://doi.org/10.1007/BF00116251.
27. Dietz WE, Kiech EL, Ali M. (1989, January). Classification of data patterns using an autoassociative neural network topology. International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems 1989; 2: 1028-1036. https://doi.org/10.1145/67312.67378.
28. Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on pattern analysis and machine intelligence 2005; 27(8): 1226-1238. https://doi.org/10.1109/TPAMI.2005.159.
29. Kira K, Rendell LA. A practical approach to feature selection. Machine Learning Proceedings 1992; 249-256. https://doi.org/10.1016/B978-1-55860-247-2.50037-1

Received 2022-11-24
Accepted 2023-05-19
Available online 2023-05-22

Elmehdi BENMALEK is currently a researcher at ENSAM de Rabat, Morocco; he acquired a PhD in Electrical engineering from Mohamed V University, Rabat, Morocco in 2019. He is a member of Electronic Systems, Sensors and Nanotechnologies laboratory. His domains of interest include artificial intelligence, signal processing and embedded systems.

Jamal EL Mhamdi is currently a professor at ENSAM de Rabat, Morocco. he acquired a PhD in Electronics and Telecommunication from Rennes University, France in 1988. He is a member of Electronic Systems, Sensors and Nanotechnologies laboratory. His domains of interest include artificial intelligence, embedded systems and signal processing.

Abdelilah Jilbab is currently a professor at ENSAM de Rabat, Morocco; he acquired a Computer and Telecommunication PhD from Mohamed V University, Rabat, Morocco in 2009. He is a member of Electronic Systems, Sensors and Nanotechnologies laboratory. His domains of interest include artificial intelligence, embedded systems and signal processing.

Atman JBARI is currently a Professor at the electrical engineering department of ENSAM de Rabat, Mohamed V University in Rabat, Morocco. In 2009, he received his PhD in computer and telecommunications from Mohammed 5 University. His current research interests include artificial intelligence, signal processing and embedded electronic systems. He is a member of Electronic Systems, Sensors and Nanotechnologies laboratory.