



Identification of COVID-19 from Cough Sounds Using Non-Linear Analysis and Machine Learning

Fatma Zehra Solak ^{1*}

^{1*} Konya Technical University, Faculty of Engineering and Natural Sciences, Department of Software Engineering, Konya, Turkey, (ORCID: 0000-0001-5035-7575), fzdogus@ktun.edu.tr

(1st International Conference on Applied Engineering and Natural Sciences ICAENS 2021, November 1-3, 2021)

(DOI: 10.31590/ejosat.1010723)

ATIF/REFERENCE: Solak, F. (2021). Identification of COVID-19 from Cough Sounds Using Non-Linear Analysis and Machine Learning. *European Journal of Science and Technology*, (28), 710-716.

Abstract

Automatic diagnosis of COVID-19 has an active role in reducing the spread of the disease by minimizing interaction with people. Machine learning models using various signals and images form the basis of automatic diagnosis. This study presents the machine learning based models for detecting COVID-19 infection using 'Virufy' dataset containing cough sound signals labeled as COVID-19 and Non-COVID-19. Since the number of COVID positive coughs in the set is less than those of COVID negative, firstly, data balancing was performed with the ADASYN oversampling technique in the study. Then, features were extracted by non-linear analysis of cough sounds using Multifractal Detrended Fluctuation Analysis (MDFA), Lempel-Ziv Complexity (LZC) and entropy measures. Later, the most effective features were selected by ReliefF method. Finally, five machine learning algorithms, namely Support Vector Machine with Radial Basis Function (SVM-RBF), Random Forest (RF), Adaboost, Artificial Neural Network (ANN), k Nearest Neighbor (kNN) were used to identify cough sounds as COVID-19 or Non-COVID19. As a result of the study, the cough sounds of COVID-19 patients and Non-COVID19 subjects were identified with 95.8% classification accuracy thanks to the RBF kernel function of SVM and the selected effective features. With this classifier, 93.1% sensitivity, 98.6% specificity, 98.6% precision, 0.92 kappa statistical values and 93.2% area under the ROC curve were obtained.

Keywords: COVID-19, Cough Sounds, Entropy, Machine Learning, MDFA, LZC, SVM.

Doğrusal Olmayan Analiz ve Makine Öğrenimi Kullanılarak Öksürük Seslerinden COVID-19'un Tanımlanması

Öz

COVID-19'un otomatik teşhisi, insanlarla etkileşimi en aza indirerek hastalığın yayılmasını azaltmada aktif bir role sahiptir. Çeşitli sinyal ve görüntüleri kullanan makine öğrenmesi modelleri, otomatik tanılamamanın temelini oluşturur. Bu çalışma, COVID-19 ve COVID-19 değil olarak etiketlenmiş öksürük ses sinyallerini içeren 'Virufy' veri setini kullanarak COVID-19 enfeksiyonunu tespit etmek için makine öğrenmesi tabanlı modeller sunmaktadır. Veri setindeki COVID pozitif öksürük sayısı, COVID negatif olanlardan daha az olduğu için çalışmada öncelikle ADASYN aşırı örnekleme tekniği ile veri dengeleme yapılmıştır. Ardından, Çokfraktallı Eğimden Arındırılmış Dalgalanma Analizi (Multifraktal Detrended Fluctuation Analysis - MDFA), Lempel-Ziv Karmaşıklık (Lempel-Ziv Complexity-LZC) ve entropi ölçümleri kullanılarak öksürük seslerinin doğrusal olmayan analizi ile öznelikler çıkarılmıştır. Daha sonra ReliefF yöntemi ile en etkili öznelikler seçilmiştir. Son olarak, öksürük seslerini COVID-19 veya değil olarak tanımlamak için, Radyal Tabanlı Çekirdek fonksiyona sahip Destek vektör Makineleri (Support Vector Machine with Radial Basis Function-SVM-RBF), Rastgele Orman (Random Forest-RF), Adaboost, Yapay Sinir Ağları (Artificial Neural Network -ANN), k En Yakın Komşuluk (k Nearest Neighbor -kNN) olmak üzere beş makine öğrenme algoritması kullanılmıştır. Çalışma sonucunda, radyal tabanlı çekirdek fonksiyonuna sahip destek vektör makinesi ve seçilen etkin öznelikler sayesinde COVID-19 hastalarının ve COVID19 olmayan deneklerin öksürük sesleri %95.8 sınıflandırma doğruluğu ile belirlenmiştir. Bu sınıflandırıcı ile %93.1 duyarlılık, %98.6 özgüllük, %98.6 kesinlik, 0.92 kappa istatistik değerleri ve %93.2 ROC eğrisi altında kalan alan değeri elde edilmiştir.

Anahtar Kelimeler: COVID-19, Öksürük Sesleri, Entropi, Makine Öğrenimi, MDFA, LZC, SVM.

* Corresponding Author: fzdogus@ktun.edu.tr

1. Introduction

COVID-19 caused by the corona virus (SARS-CoV-2) emerged in Wuhan, China, in December 2019 and became a serious public health problem by spreading into worldwide in a short time (Narin et al., 2021). There have been 233.503.524 confirmed cases of COVID-19, including 4.777.503 deaths, reported to World Health Organization (WHO) (as of 1th October 2021). Deaths arising from Covid-19 are increasing day by day. For this reason, early diagnosis and measures such as isolation, social distancing, hand washing frequently, not touching the face as much as possible and care for patients are key strategies for a better management of this epidemic (Horry et al., 2020; Manshouri, 2021).

Dry cough, headache, fever, fatigue, shortness of breath, muscle and sore throat are common symptoms of COVID-19 according to WHO (Manshouri, 2021; Mahmoud et al., 2021; Mouawad et al., 2021; Pal and Sankarasubbu, 2021). Dry cough is amongst more common as it is one of the early symptoms of respiratory tract infections and it occurs in 68% to 83% of the people who come for a medical examination (Mouawad et al., 2021; Pal and Sankarasubbu, 2021). Therefore, automatic diagnosis of COVID-19 from cough sound recordings using machine learning and signal processing methods has been one of the popular and important fields of study recently.

Imran et al. (2020) studied for diagnosis of COVID-19 from cough sound samples using classical machine learning and deep learning algorithms by distinguishing COVID-19 coughs and several types of Non-COVID19 coughs. Non-COVID19 cough samples and COVID-19 cough samples were identified with 92.85% accuracy using two deep learning-based classifiers. In another study (Bagad et al., 2020), researchers demonstrated the probability of finding COVID-19 through cough sounds with a 72% area under of ROC curve (AUC) by developing a CNN-based framework. Fakhry et al. (2021) use cough sounds with Mel- spectrogram, mel-frequency cepstral coefficients (MFCC), clinical features, Multi-Branch Deep Learning Network and obtain average AUC of 91%. Coppock et al. (2021) studied with cough and breathing sounds. They used Mel- spectrogram and ResNet. As a result of their study, 84.6% AUC were accessed. In a study (Pahar et al., 2021), authors present a machine learning based COVID-19 cough classifier which can discriminate COVID-19 positive coughs from both COVID-19 negative and healthy coughs recorded on a smartphone. They extracted MFCCs, log frame energies, zero crossing rate (ZCR) and kurtosis features from cough sounds. 98% and 94% AUC values were obtained with Resnet50 and LSTM respectively. Chaudhari et al. (2020) obtained 77.1% AUC by using cough sounds with Mel-spectrogram, MFCC, Clinical features and Ensemble Deep Learning Model. One of the data studied by Mahmoud et al. (2021) regarding the diagnosis of Covid-19 is cough sounds. In addition to coughing, they also examined breathing and speech sounds. MFCC, Spectral Bandwidth, Spectral Centroid, Spectral Roll-off, Zero Crossing Rate, RMS Energy, Skewness, Kurtosis, Variation, SEM features were used with Deep Model and Shallow classifiers. As a result of the study, average 96.4% AUC performance were obtained (Mahmoud et al., 2021). In the study (Manshouri, 2021), Power Spectral Density (PSD) based on Short Time Fourier Transform (STFT) and MFCC were chosen as the efficient feature extraction method for cough sounds. The coughs of subjects with COVID-19 and Non-COVID19 have

defined with 95.86% classification accuracy thanks to the RBF kernel function of SVM and the MFCC method.

As a result of the studies in the literature, it has been seen that cough sounds are of great importance in the diagnosis of COVID-19. Based on this inference, the aim of this study was established and the automatic diagnosis of COVID-19 was carried out using appropriate machine learning algorithms and “GitHub” open source cough dataset (GitHub n.d) “Virufy”. In this study, diagnosis is based on classification focusing on cough sounds of COVID-19 patients and Non-COVID19 subjects. Since previous studies generally used deep learning algorithms, spectrograms derived from cough sounds and MFCC features were considered. Whereas, the cough is a non-stationary acoustic event and can exhibit non-linear dynamics. Therefore, non-linear features of cough sounds should also be examined.

Within the scope of the study, non-linear analysis of cough sound signals was performed by Multifractal Detrended Fluctuation Analysis (MDFA) method, Lempel–Ziv Complexity (LZC) and entropy measures. Thus, 29 non-linear features were extracted from coughs. By using these features with different machine learning methods, the cough sounds of COVID-19 patients were separated from those of Non-COVID19. The results of the study show that COVID-19 cough sound samples can be identified with acceptable classification accuracy and AUC.

2. Material and Method

2.1. Dataset

Virufy dataset (Chaudhari et al., 2020) is used in this study. The dataset available on “GitHub” (Virufy COVID-19 Open Cough Dataset, GitHub). In the dataset, 121 cough sound segments exist belonging to sixteen subjects. 48 of these segments are labeled as COVID-19 since they belonged to 7 subjects with a positive PCR test. The other 73 segments belonging to 9 subjects whose tests are negative are labelled as Non-COVID19 coughs. Each segment is approximately 1 second long and sampling frequency of 48000 Hz.

2.2. Pre-processing

COVID-19 positive and negative cough sound segments in the Virufy dataset is unbalanced. Insufficient number of positive segments may result in poor COVID-19 and Non-COVID19 classification success. To effectively identify COVID-19 coughs, the ADASYN (He et al., 2008) oversampling technique is applied. The minor segments (positive) were balanced to the major segments (negative) by generating synthetic positive samples. Thus, a balanced dataset with a positive/negative ratio of approximately 1:1 (71 positive, 73 negative) was obtained. In total, 144 cough sounds were reached, which will be determined as COVID-19 or Non-COVID 19 through classification.

2.3. Feature Extraction and Selection

Cough sounds occurs by passing air flows from the lungs to the mouth and nasal cavities. Since the cough sound is produced by respiratory system, it can exhibit non-linear, irregular and unpredictability characteristics. Also, in general, the nature of many biomedical signals, including cough sounds, show spatial and temporal variations. For these reasons, nonlinear analysis of sounds should be carried out and thus their nonlinear features should be extracted to identify COVID-19 coughs.

This study, firstly, applied the MDFA method which is one of the nonlinear analysis methods to cough sounds. In this way, multifractal spectrum of each sound was created using the following steps of MDFA. Then, the first 23 features appearing in Table 1 were extracted from the spectrums. Fig. 1 shows a multifractal spectrum and how some features are extracted from it (Göğüş et al., 2020).

Steps of MDFA to produce multifractal spectrums (Ihlen, 2012; Marton et al., 2014);

- ✓ Computing the mean of the signals
- ✓ Creating the integrated series by summing the differences obtained by subtracting the mean value from each value in the series
- ✓ Fragmentation of integrated series and computing the local Root Mean Square (RMS) variation
- ✓ Finding local detrending of series
- ✓ Computing multifractal detrending, q-order RMS (qRMS)
- ✓ Computing q-order Hurst exponent (H_q) and q order mass exponent
- ✓ Computing q-order singularity exponent (h_q) and q-order singularity dimension (D_q)
- ✓ Creating the Multifractal Spectrum (the plot of h_q versus D_q)

The repeatability, predictability and complexity of sounds can be described by the LZC (Lempel and Ziv, 1976; Kaspar and Schuster, 1987) and entropy measures (Approximate Entropy – ApEn, Sample Entropy – SamEn, Permutation Entropy – PerEn, Shannon Entropy – ShEn and Renyi Entropy – RenEn) (Kaspar and Schuster, 1987; Pincus, 1991; Richman and Moorman, 2000). Therefore, LZC ApEn, SamEn, PerEn, ShEn and RenEn nonlinear features seen in Table 1 were also extracted from cough sounds. By extracting these features, in total, 29 nonlinear

features were obtained from cough sounds to detect COVID-19 coughs.

Although cough sounds were tried to be represented by 29 features in the study, it was considered that not all features would have the same effect on the distinction between COVID-19 and Non-COVID19, and some of them might be more effective than others. ReliefF (Kononenko, 1994) feature selection method was used to determine these most effective features. As a result, 9 features were selected as the most effective and classification of cough sounds into COVID-19 and Non-COVID19 were performed by using these effective features with different machine learning algorithm.

2.4. Classification

After the feature extraction and selection processes, this study was continued with the classification process. The diagnosis of COVID-19 was based on the classification of obtained features belonging to cough sounds from COVID-19 patients and Non-COVID19 subjects. Support Vector Machine with Radial Basis Function (SVM-RBF), Random Forest (RF), Adaboost, Artificial Neural Network (ANN), k Nearest Neighbor (kNN) machine learning algorithms were used as classifiers.

Assessment of the classifiers was carried out with k cross validation technique. Classically, the value of k is chosen as 10. Performance evaluation was made with classification accuracy (CA) in Eq. (7), sensitivity in Eq. (8), specificity in Eq. (9), precision in Eq. (10), kappa statistics in Eq. (11) and AUC metrics by benefitting the confusion matrix obtained at the end of each classification.

Table 2 represents the confusion matrix structure in the classifications of this study.

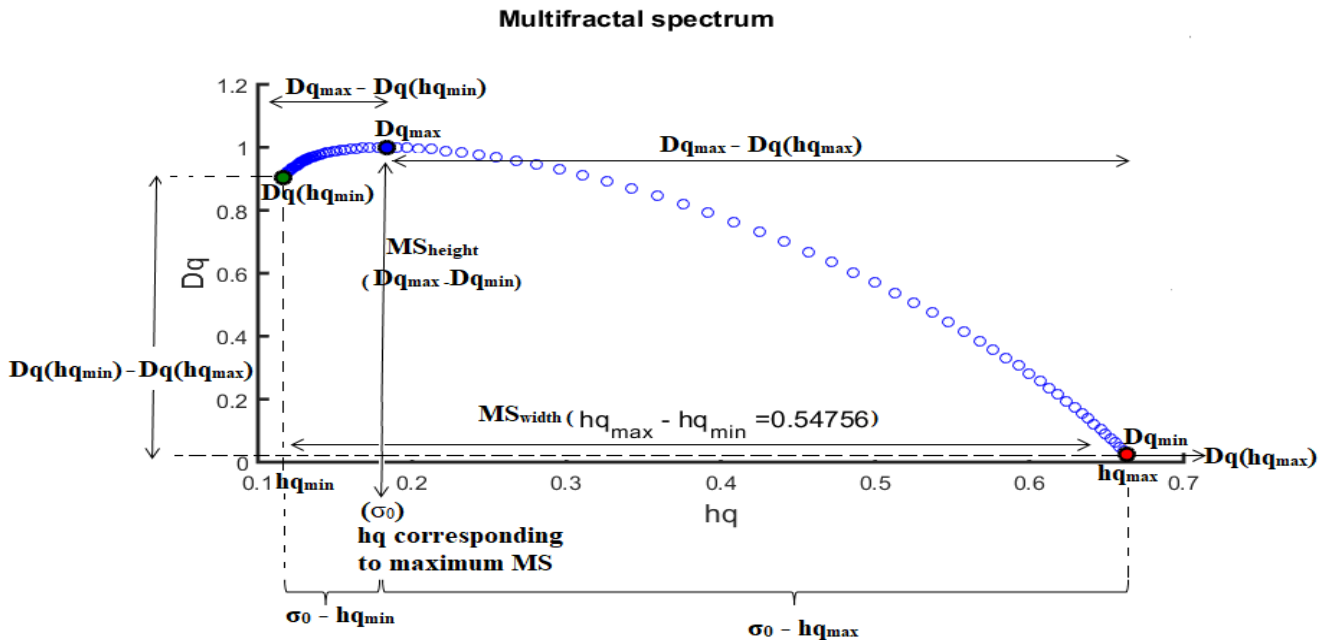


Fig. 1. Multifractal spectrum and extraction of some features

Table 1. Extracted nonlinear features of cough sounds

Feature Num.	Feature Name
F1	The mean of Hurst exponents.
F2	Maximum Hurst exponent value (Hq_{max}).
F3	Minimum Hurst exponent value (Hq_{min}).
F4	Generalized Hurst exponent.
F5	Maximum singularity exponent value (hq_{max}).
F6	Minimum singularity exponent value (hq_{min}).
F7	Mean value of singularity exponent values.
F8	Singularity exponent corresponding to maximum multifractal spectrum (σ_0) (hq value corresponding to max Dq value).
F9	Asymmetric index (AI) calculated using Eq. (1). $AI = \frac{\Delta\sigma_L - \Delta\sigma_R}{\Delta\sigma_L + \Delta\sigma_R}$ $\Delta\sigma_L = \sigma_0 - hq_{min}$ $\Delta\sigma_R = hq_{max} - \sigma_0$ Eq.(1)
F10	Horizontal distance between singularity exponent value corresponding to the maximum of multifractal spectrum and minimum singularity exponent ($\Delta\sigma_L$ in Eq. (1)).
F11	Horizontal distance between singularity exponent value corresponding to the maximum of multifractal spectrum and maximum singularity exponent ($\sigma_0 - hq_{max}$).
F12	Maximum value of multifractal spectrum (Dq_{max}).
F13	Mean value of multifractal spectrum values (Mean Dq).
F14	Skewness of Multifractal spectrum.
F15	Kurtosis of Multifractal spectrum.
F16	Variance of Multifractal Spectrum.
F17	Multifractal spectrum corresponding to maximum singularity exponent ($Dq(hq_{max})$).
F18	Multifractal spectrum corresponding to minimum singularity exponent ($Dq(hq_{min})$).
F19	Multifractal spectrum width calculated with Eq. (2). $MS_{width} = hq_{max} - hq_{min}$ Eq.(2)
F20	Multifractal spectrum height calculated with Eq. (3). $MS_{height} = Dq_{max} - Dq_{min}$ Eq.(3)
F21	Multifractal spectrum with left truncation as in Eq. (4). (Difference between maximum spectrum value and spectrum value corresponding min singularity exponent). $MS_{left} = Dq_{max} - Dq(hq_{min})$ Eq.(4)
F22	Multifractal spectrum with right truncation as in Eq. (5) (Difference between maximum spectrum and spectrum corresponding max singularity exponent). $MS_{right} = Dq_{max} - Dq(hq_{max})$ Eq.(5)
F23	Vertical distance in multifractal spectrum calculated as seen in Eq. (6). $MS_{vertical_distance} = Dq(hq_{min}) - Dq(hq_{max})$ Eq.(6)
F24	Lempel–Ziv Complexity (LZC)
F25	Approximate Entropy (ApEn)
F26	Sample Entropy (SamEn)
F27	Permutation Entropy (PerEn)
F28	Shannon Entropy (ShEn)
F29	Renyi Entropy (RenEn)

Table 2. Confusion matrix structure

	Predicted Class		
		Non-COVID19	COVID-19
True Class	Non-COVID19	TP	FN
	COVID-19	FP	TN

$$CA = \frac{TP+TN}{TP+TN+FP+FN} \tag{Eq.7}$$

$$Sensitivity(Sens.) = \frac{TP}{TP+FN} \tag{Eq.8}$$

$$Specificity(Spec.) = \frac{TN}{TN+FP} \tag{Eq.9}$$

$$Precision (P)= \frac{TP}{TP+FP} \tag{Eq.10}$$

$$Kappa\ statistics(K) = \frac{P_o - P_e}{1 - P_e} \tag{Eq.11}$$

Pe in Eq. (11) is calculated as in Eq. (12).

$$P_e = \frac{[(TP + FP) \times (TP + FN)] + [(FN + TN) \times (FP + TN)]}{(TP + TN + FP + FN)^2} \tag{Eq.12}$$

3. Results and Discussion

In order to achieve the aim of the study, primarily 29 nonlinear features were extracted from cough sounds, than 9 effective features to be used in the classification of coughs into COVID-19 and Non-COVID19 were selected by the ReliefF method. The selected features are shown in Table 3.

Table 3. Selected features by the ReliefF method

Feature Num.	Feature Name
F10	$\Delta\sigma L$
F18	Dq(hq _{min})
F21	MS _{left}
F23	MS _{vertical_distant}
F24	LZC
F26	SamEn
F27	PerEn
F28	ShEn
F29	RenEn

By using these selected 9 effective features with SVM-RBF, RF, Adaboost, ANN, kNN classification algorithms, the performances seen in Table 4 were obtained.

Table 4. Classification performances

Classifier	CA (%)	Sens (%)	Spec (%)	P (%)	K	AUC (%)
SVM-RBF	95.8	93.1	98.6	98.6	0.92	93.2
RF	90.7	83.5	98.6	98.3	0.82	87.5
Adaboost	89.6	83.5	95.8	95.3	0.79	84.7
ANN	91.6	85.0	98.6	98.4	0.83	88.7
kNN	84.0	78.1	90.1	89.1	0.68	77.4

As presented in Table 4, the highest performances for all measures were obtained using SVM classifier with RBF kernel. With 9 selected features and SVM-RBF, a classification accuracy of 95.8%, a kappa value of 0.92 and an AUC value of 93.2% were achieved. While the cough sounds of COVID-19 patients were accurately defined with 98.6% performance, the

coughs of those who did not have this disease could be predicted correctly with 93.1%. The confusion matrix obtained as a result of the SVM-RBF classifier algorithm is shown in Table 5. As seen in this table, only one of the cough sounds that indicate COVID-19 has been misclassified.

Table 5. Confusion matrix obtained by 9 selected features and SVM-RBF

		Predicted Class	
		Non-COVID19	COVID-19
True Class	Non-COVID19	68	5
	COVID-19	1	70

*Parameters= Kernel: RBF, Cost(c):9.90, Gamma(g): 0.49

Following SVM-RBF, higher performances were obtained with RF and ANN classifiers. With both classifiers, classification accuracies over 90% were achieved. These classifiers were able to identify COVID-19 cough sounds with very high accuracy (98.6%), just like SVM-RBF. However, these two classifiers were less successful than SVM-RBF in identifying cough sounds that do not indicate COVID-19. According to Table 4, it could be said that the kNN classifier provided the lowest performance in terms of evaluation criteria.

As in this study, there are several studies in literature containing analysis and classification of cough sounds for the diagnosis of COVID-19. However, most of the studies used different datasets. For this reason, it was not possible to compare this study with previous studies in detail. The comparison has been made with very few studies using the verify dataset as one of the datasets they studied as seen in Table 6.

Table 6 shows that this study is competitive with performance to other studies that use the same dataset as one of their sets in terms of different performance measures. However, it should be noted that most of these studies in the literature considered not only the Virufy data set, but also different data sets, as well as sounds other than cough sounds, and gave the performance criteria seen in Table 6 as average values. Studies in the literature have generally used similar feature extraction methods, especially MFCC, for the analysis of cough sound signals. This study, unlike other studies, examined the nonlinear structure of cough sounds and extracted their nonlinear features. With the use of these features, the study achieved its purpose and was able to identify the positive and negative labeled cough sounds of COVID-19 with high success.

4. Conclusions and Recommendations

In this study, it was aimed to identify cough sounds with high accuracy as COVID-19 and Non-COVID19. In accordance with this purpose, 29 nonlinear features were extracted and then the most effective 9 features were selected by ReliefF method. When these effective features are examined, it is seen that, 4 of them are features of multifractal spectrums, other 4 features are entropies and the remaining 1 feature is the LZC. So, besides LZC, both the features extracted from the multifractal spectrums created by MDFA method and the entropy features that analyze the complexities of the sounds were effective in identifying cough sounds.

Table 6. Comparison to previous studies for the cough-based COVID-19 diagnosis.

Research	Dataset	Sound Type	Features	Models/Classifiers	Results
Chaudhari et al., (2020)	Coswara Coughvid Virufy	Cough	Mel- spectrogram MFCC	Ensemble Deep Learning Model	77.1% AUC
Mahmoud et al. (2021)	Coswara Virufy	Cough Breathing Speech	MFCC – RMS – ZCR Spectral Rolloff / Centroid/etc.	Deep Model Shallow classifiers	96.4% AUC 96% CA
Manshoury (2021)	Virufy	Cough	STFT – MFCC	SVM classifier with RBF kernel	95.8% CA 98.6% sens. 91.7% spec.
This study	Virufy	Cough	Nonlinear features (MDFA based and entropy features)	SVM-RBF RF Adaboost ANN kNN	95.8% CA 93.1% sens. 98.6% spec. 98.6% Precision 0.92% Kappa 93.2% AUC

Considering the classification performances of different machine learning algorithms with effective features, it is seen that the most successful classifier algorithm is SVM-RBF. Based on the results, we can say that the SVM-RBF classifier better deal with the nonlinear behavior of COVID-19/Non-COVID19. The study has been promising for automatic diagnosis of COVID-19 using cough sounds. However, the number of data is low. This has been the limitation of the study. Future work will address this constraint involving larger datasets and, as the amount of data increases, focus on the automatic diagnosis of COVID-19 deep learning methods.

References

- A. Fakhry, X. Jiang, J. Xiao, G. Chaudhari, A. Han, and A.Khanzada, "Virufy: A Multi-Branch Deep Learning Network for Automated Detection of Covid-19," preprint from arXiv:2103.01806, 2021.
- A. Imran, I. Posokhova, H. N. Qureshi, U. Masoos, M. S. Riaz, K. Ali, C. N. John, M. I. Hussain, and M. Nabeel, "AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app.," *Inform Med Unlocked*, vol. 20, pp. 100378, 2020.
- A. Lempel, and J. Ziv, "On the Complexity of Finite Sequences," *IEEE Transactions on Information Theory*, vol. 22, no.1, pp. 75-81,1976.
- A. Mahmoud, K. H. Rahouma, and S.M. Ramzy, "Pay attention to the speech: COVID-19 diagnosis using machine learning and crowdsourced respiratory and speech recordings," *Alexandria Engineering Journal*, 2021. In press.
- A. Narin, C. Kaya, and Z. Pamuk, "Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks," *Pattern Anal Appl*, pp. 1-14, 2021.
- A. Pal, and M. Sankarasubbu, "Pay Attention to the cough: Early Diagnosis of COVID-19 using Interpretable Symptoms Embeddings with Cough Sound Signal Processing," in *36th ACM/SIGAPP Symposium on Applied Computing (SAC '21)*. March 22–26, 2021.
- E. A. Ihlen, "Introduction to multifractal detrended fluctuation analysis in matlab," *Front Physiol*, vol. 3, pp. 141, 2012.
- F. Kaspar, and H. Schuster, "Easily calculable measure for the complexity of spatiotemporal patterns," *Phys Rev A Gen Phys*, vol. 36, no. 2, pp. 842-848, 1987.
- F. Z. Göğüş, G. Tezel, S. Özşen, S. Küçüktürk, H. Vatansev, and Y. Koca, "Identification of Apnea-Hypopnea Index Subgroups Based on Multifractal Detrended Fluctuation Analysis and Nasal Cannula Airflow Signals," *Traitement du Signal*, vol. 37, no. 2, pp. 145-156, 2020.
- G. Chaudhari, X. Jiang, A. Fakhry, A. Han, j. Xiao, S. Shen, and A. Khanzada, "Virufy: Global Applicability of Crowdsourced and Clinical Datasets for AI Detection of COVID-19 from Cough," *ArXiv*, 2020. 2011.13320.
- H. Coppock, A. Gaskell, P. Tzikaris, A. Baird, L. Jones, and B. W. Schuller, "End-2-End COVID-19 Detection from Breath & Cough Audio," Preprint from arXiv:2102.08359v1, 2021.
- H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 1322-1328, 2008.
- I. Kononenko, "Estimating Attributes: Analysis and Extensions of Relief," in *European Conference on Machine Learning*, pp. 171-182, 1994.
- J. S. Richman, and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *Am J Physiol Heart Circ Physiol*, vol. 278, no. 6, pp. H2039-49, 2000.
- L. F. Márton, S. t. Brassai, L. Bako, and L. Losonczy, "Detrended Fluctuation Analysis of EEG Signals," *Procedia Technology*, vol. 12, pp. 125-132, 2014.
- M. J. Horry, S. Chakraborty, M. Paul, A. Ulhaq, B. Pradhan, M. Saha, and N. Shukla, "COVID-19 Detection Through Transfer Learning Using Multimodal Imaging Data," *IEEE Access*, vol. 8, pp. 149808-149824, 2020.
- M. N. Manshoury, "Identifying COVID-19 by using spectral analysis of cough recordings: a distinctive classification study," *Cogn Neurodyn*, pp. 1-15, 2021.
- M. Pahar, M. Klopper, R. Warren, and T. Niesler, "COVID-19 cough classification using machine learning and global smartphone recordings," *Comput Biol Med*, vol. 135, pp. 104572, 2021.
- P. Bagad, A. Dalmia, J. Doshi, A. Nagrani, P. Bhamare, A. Mahale, S. Rane, N. Agarwal, and R. Panicker, "Cough Against: COVID Evidence of COVID-19 Signature in Cough Sounds," Preprint from arXiv, 2020.
- P. Mouawad, T. Dubnov, and S. Dubnov, "Robust Detection of COVID-19 in Cough Sounds: Using Recurrence Dynamics

- and Variable Markov Model," SN Comput Sci, vol. 2, no. 1, pp. 34, 2021.
- S. M. Pincus, "Approximate entropy as a measure of system complexity," Proc. Nati. Acad. Sci., vol. 88, pp. 2297-2301, 1991.
- Virufy COVID-19 Open Cough Dataset, GitHub - virufy/virufy_data, (n.d.). https://github.com/virufy/virufy_data (accessed October 08, 2021). Virufy, Editor. 2020.