ERJ open research
THE BEST IN OPEN ACCESS BASIC, TRANSLATIONAL & CLINICAL RESPIRATORY RESEARCH

Early View

Original research article

A novel infrasound and audible machine-learning approach for the diagnosis of COVID-19

Guy Dori, Noa Bachner-Hinenzon, Nour Kasim, Haitem Zaidani, Sivan Haia Perl, Shlomo Maayan, Amin Shneifi, Yousef Kian, Tuvia Tiosano, Doron Adler, Yochai Adir

Please cite this article as: Dori G, Bachner-Hinenzon N, Kasim N, *et al*. A novel infrasound and audible machine-learning approach for the diagnosis of COVID-19. *ERJ Open Res* 2022; in press (https://doi.org/10.1183/23120541.00152-2022).

A Novel Infrasound and Audible Machine-Learning Approach for the Diagnosis of COVID-19

Guy Dori[1], *Noa Bachner-Hinenzon[2], Nour Kasim[1], Haitem Zaidani[3], Sivan Haia Perl[4], Shlomo Maayan[5], Amin Shneifi[6], Yousef Kian[5], Tuvia Tiosano[1], Doron Adler[2], Yochai Adir[7]

Affiliations:

1. HaEmek medical center, Afula, Israel

2. Sanolla, Nesher, Israel

3. Rambam medical center, Haifa, Israel

4. Shamir medical center, Zerifin, Israel

5. Barzilai medical center, Ashkelon, Israel

6. Clalit health services, Israel

7. Carmel medical center, Israel

Corresponding author: Noa Bachner-Hinenzon, Hamesila 19, Nesher, Israel, Email: Noa@sanolla.com

Tel: +972-50-8610414

Take home message

AI applied to full spectrum auscultation (infrasound and audible) provides superior morbidity detection compared to standard narrow-band auscultation restricted to the audible spectrum.

Abstract

The COVID-19 outbreak has rapidly spread around the world, causing a global public health and economic crisis. A critical limitation in detecting COVID-19 related pneumonia is that it is often manifested as a “silent pneumonia”, i.e., pulmonary auscultation, using a standard stethoscope, sounds "normal". Chest CT is the gold standard for detecting COVID-19 pneumonia; however, radiation exposure, availability and cost preclude its utilization as a screening tool for COVID-19 pneumonia. In this study we hypothesized that COVID-19 pneumonia, “silent” to the human ear using a standard stethoscope, is detectable using a full spectrum auscultation device that contains a machine-learning analysis.

Lung sounds signals were acquired, using a novel full spectrum (3-2,000Hz) stethoscope, from 164 patients with COVID-19 pneumonia, 61 non-COVID-19 pneumonia and 141 healthy subjects. A machine-learning classifier was constructed, and the data was classified into 3 groups: 1. Normal lung sounds 2. COVID-19 pneumonia 3. Non-COVID-19 pneumonia.

Standard auscultation found that 72% of the non-COVID-19 pneumonia patients had abnormal lung sounds, compared to only 25% for the COVID-19 pneumonia patients. The classifier’s sensitivity and specificity for the detection of COVID-19 pneumonia were 97% and 93%, respectively, when analyzing the sound and infrasound data, and they were reduced to 93% and 80% without the infrasound data ($P<0.01$ difference in ROC with and without infrasound).

This study reveals that useful clinical information exists in the infrasound spectrum of COVID-19 related pneumonia, and machine-learning analysis applied to the full spectrum of lung sounds is useful in its detection.

Introduction

Severe Acute Respiratory Syndrome due to the novel coronavirus was initially reported in China, December 2019, and is now termed COVID-19 disease[1]. COVID-19 may lead to severe pneumonia (COVID-19PN), requiring specialized management in the intensive care unit[2-3]. Diagnosing pneumonia in general, and specifically due to SARS-CoV-2 infection, relies on clinical evaluation, including physical examination and imaging studies. However, COVID-19PN presents diagnostic challenges. Clinically, this respiratory condition has been termed “silent pneumonia” due to the paucity of pathological pulmonary auscultation findings using a standard stethoscope[4]. One possible cause for this “silence” may lie with the expertise required of the physician to diagnose auscultation findings associated with pneumonia[5-6]. Another cause may be related to the frequency of sound waves produced in COVID-19PN, which may actually reside outside the hearing spectrum of the human ear.

Chest computerized tomography (CT) has become the gold standard for diagnosing typical lung pathologies associated with COVID-19PN[7-8]. However, in addition to the radiation exposure, availability and cost of CTs, recent studies have shown that although their sensitivity is high, chest CT specificity is rather low[9]. Available CT data show that COVID-19 patients with both mild and severe clinical presentations demonstrate bilateral patchy infiltrations or ground glass opacities[10-11], for which the differential diagnosis is wide. Surprisingly, similar CT abnormalities were also observed in asymptomatic patients of COVID-19[12]. Thus, predicting the course of COVID-19PN and possible pulmonary deterioration based on CT becomes a difficult task. As a result, some patients who initially present with only mild symptoms and maintain well preserved mechanical characteristics of the lungs with no breathing difficulties, are suddenly diagnosed with severe hypoxemia and breathing difficulties only late in the course of the disease[13-14].

In a recent meta-analysis, overall pooled sensitivity for lung auscultation was 37% with 89% specificity[15], in identifying pneumonia type illnesses. Due to the paucity of suspicious breath sounds, it could be that

sensitivity and specificity of COVID-19PN with a regular stethoscope would be even lower. Pathophysiologically, it is possible that COVID-19PN is "silent" because of its diffuse, peripheral nature of pulmonary inflammation[3], compared with the localized consolidation of a non-COVID-19 pneumonia[16]. CT and lung ultrasound are proposed as first line screening for COVID-19PN[17]. However, during a pandemic, as the pressure to hospitalize patients increases and imaging resources become depleted, resource allocation becomes a critical issue. Patients may benefit from a medical device, based on artificial intelligence, which supports less trained medical teams (triage staff and novices) with reliable diagnostic measures and allows for specialists to attend the most difficult patients.

There have been advances in the use of the electronic stethoscope in the past decade, and new systems and adaptations are already available[18], along with new tools for analyzing lung sounds by using artificial intelligence[19-20]. Simple machine-learning methods succeeded in diagnosing COVID-19 from breath and cough sounds crudely collected on a mobile application with an area under the curve of around 70%[20].

As standard methods for diagnosing COVID-19PN (pulmonary auscultation and imaging) often lack specificity and sensitivity, we investigated the utility of a novel smart digital stethoscope (VoqX, Sanolla), with which pulmonary sounds in the infrasound range (≤ 20 Hz) were recorded and analyzed using machine-learning algorithms. This infrasound range is inaudible to the human ear yet contains vital information for the diagnosis of lung pathologies[21][22][23]. Therefore, the VoqX may support the unmet need for diagnosing COVID-19PN without relying on a physician's auscultation expertise.

Methods

Subjects and study design

This study was conducted among patients admitted to selected tertiary medical centers in Israel. The study protocol adhered to the principles of the Declaration of Helsinki and was approved by the institutional ethics board of each hospital (IRB Ethics committee of: Shamir medical center No. 0136-20-ASF, Rambam medical center 0631-18-RMB, HaEmek medical center 0136-19-EMC and 0160-19-EMC, Barzilai medical

center 0036-20-BRZ). Written informed consent was obtained from all patients. The following data were collected from each patient: age, gender, height, weight, smoker status and the results of lung auscultation. Diagnosis of non-COVID-19 pneumonia was confirmed by one of the following: anamnesis, physical examination, chest x-ray (or CT), blood test (blood count) and oximetry < 94%. Diagnosis of COVID-19 was performed according to the criteria based on the WHO's recommendation[24]. Patients were excluded if they met at least one of the following criteria: under 18 years of age, pregnant, had any type of chest malformation, under another's guardianship or weighed more than 150 kg. The Healthy group was composed of healthy volunteers that accompanied patients to the hospital and did not suffer from any chronic or acute respiratory disease.

Respiratory sound data

Real time respiratory sound recordings were performed during the physical examination while the subjects were in a sitting position. The VoqX stethoscope was placed on top of the patient's clothing. Acquisition of respiratory sounds was performed over a single layer of clothing for all patients, with no need for direct skin contact. For every patient, 14 sites (A-N) were auscultated, within the anterior, posterior, and lateral chest walls (Figure 1). For every one of these fourteen sites, 16 seconds were recorded. At the end of each examination, data were transferred to a hard drive and stored. In total, 1,974 signals were recorded from healthy subjects (141 *14 locations), 2,296 from patients with COVID-19PN (164*14 locations) and 854 from patients with non-COVID-19 pneumonia (61*14 locations).

The auscultation was non-blinded for the COVID and non-COVID pneumonia patients:

A. In the COVID group, clinicians were aware patients had COVID, and despite the possible bias due to knowing the diagnosis, they have reported abnormal breathing sounds in only 25% of cases.
B. In the non-COVID pneumonia group, clinicians suspected that patients had pneumonia after anamnesis and auscultation. The X-rays defining the presence of pneumonia were performed afterwards.

C. In the Healthy group, the auscultation was part of a different study recording healthy subjects, COPD and asthma patients. The physician auscultated to the respiratory sounds blindly and reported whether abnormal sounds were detected.

VoqX characteristics

Diagnosis by auscultation

Using the stethoscope, the examining physician was able to capture acoustic waves between 3Hz – 2000Hz, as well as amplify the sound using a simple control button.

Diagnosis by visual representation

VoqX created a graphical image of the recorded sound, a "sound signature". The image illustrates the sound of breathing and provided an immediate visual tool for the evaluation of abnormalities.

Diagnosis by machine-learning

The VoqX provides a statistical model of acoustic waves associated with various diseases. After conducting signal analysis, the device provides statistical information as for the probability of the existence (or non-existence) of several diseases. In addition, it estimates the severity of the clinical condition of the examined subject.

Connectivity

The VoqX can connect to a master device through a Bluetooth interface. Thereafter, it is possible to download and exchange data between the device and a computer for backup, analysis and patient tracking (Figure 2).

Signal processing and machine-learning

To optimize the results of the classifier, all recorded acoustic waves were pre-processed by:

1. Removing ambient noise by using a second microphone embedded in the VoqX that detected the ambient noise and dynamically reduced it from the data by an LMS (least mean square) adaptive filter.
2. Truncating each acoustic signal at the beginning and termination of the signal, to eliminate noise generated by placing and removing the stethoscope from body surface. It was performed by a threshold on the first derivative of the signal in the first and last second of the data.
3. Removing clicking noises that were generated from movements of the stethoscope on the patient's clothing during examination. This process was performed by finding spikes in the signal that were greater than 3*median amplitude of the signal.

Thereafter, 164 features were calculated in the time and frequency domains (Appendix A). The features of the time domain were statistical features such as: average, standard deviation, median etc. and features related to the shape of the sound wave (e.g., skewness, kurtosis). The features of the frequency domain included the dominant frequencies at various ranges, including their magnitudes and areas under the curves. Moreover, in the frequency domain, we calculated Mel-Frequency Cepstrum Coefficients (MFCC) features. Heart and breathing rates were extracted from the data and added to the classifier. The algorithm is described in the flowchart in Figure 3. The features were calculated for windows of 4 seconds and the output for every chest location was the median value. Next, the 25% and 75% percentile of the features (14 measurements for 14 locations) were calculated and served as the final features for every patient. Finally, the final-features were ranked according to their p-values for distinguishing between the three groups (Appendix); 144 out of 164 final-features had a significant p-value (P<0.05) and the best 50 were chosen. These 50 features were analyzed by a Gaussian Support Vector Machine (SVM) classifier [25] to classify: Normal, non-COVID-19 pneumonia and COVID-19PN. The Gaussian SVM was chosen among 32 different methods since it provided the best performance. For a new breathing sound, the classifier calculates the distance between its features and the features of all three groups and chooses which group is the closest. The

Gaussian SVM was performed with an automatic kernel scale and a box-constraint of 2. The features were standardized before the analysis. Since the sample size was different for the 3 groups, a standard method of a cost function was applied to the error calculation to equalize the weight of every group. Moreover, in order to generate an automatic objective process of investigating the performance, the data was randomly divided 12 times to 60% of the data set used as training signals and 40% used as a test group.

Statistical analysis

The p-value of the features was calculated by a one-way analysis of variance.

The sensitivity, specificity, accuracy, positive predictive value, and negative predictive value were calculated for each run of the classifier: for the cross-validation group (during the learning process in the creation of the classifier) and for the test group that did not participate during the learning. ROC analysis was performed and the area under the curves with and without the infrasound were compared[26].

Results

Sixty-one patients with non-COVID-19 pneumonia, 164 patients with COVID-19 pneumonia and 141 healthy patients were included in the study (Table 1). In 44 of 61 patients (72%) with non-COVID-19 pneumonia, the physical examination revealed an abnormality on lung auscultation; 59% had only crackles, 7% had only wheezes, 3% had both crackles and wheezes, and in 3%, the intensity of breathing sounds was perceived as reduced. In contrast, in only 41 of 164 (25%) patients with COVID-19PN, abnormalities on lung auscultation were detected. Nineteen percent had crackles, 4% had wheezes, and in 2% breathing sounds were perceived as reduced.

All breathing sounds were analyzed using machine-learning software. The input was 164 unique features that were calculated for every signal. Out of 164 features, 144 were able to distinguish between the 3 groups by analysis of variance ($P< 0.05$), however, only fifty features were selected to reduce the calculation time, as adding more than 50 features did not improve classification performance. The 50 features were selected

according to their level of performance in distinguishing between the three study groups. An example of features that improved the performance is given in Figure 4. Figure 4A shows that adding infrasound data to the analyses helps distinguishing between the three groups.

Figures 5A-C demonstrate a visual sound signature created by the device. Figure 5A shows three dense blue columns that represent three respiratory cycles of a healthy subject. Figure 5B shows the visual sound signature of a COVID-19 patient for whom detecting the breathing cycle is difficult, and higher intensities for various frequencies are found. Figure 5C represents a non-COVID-19 pneumonia patient, in which very high intensities are shown along all frequencies, due to crackles.

Sixty percent of the signals were used to train the classifier. The other forty percent served as a test group to investigate the performance of the classifier. This process was randomly performed 12 times. The results are described in Table 2. Table 2 contains three kinds of results: detection in the test group that did not participate in the learning process, the cross-validation that is performed during the learning process to optimize the classifier, and the results of auscultation by using an acoustic stethoscope. The results show that the performance of the classifier is statistically improved when adding the infrasound to the analysis ($P<0.01$) and the sensitivity of detecting COVID-19PN has reached 97% with a specificity of 93% (Table 2).

In addition, we collected data from eight asymptomatic subjects who were COVID-19 positive by nasopharyngeal PCR examination (112 signals) and were living in a quarantine facility. These subjects served as a second test group to validate our ability to diagnose asymptomatic patients. The classifier classified seven of them as COVID-19PN, and one was diagnosed with non-COVID-19 pneumonia. The X-rays of these patients identified three patients with COVID-19PN, while the other five (although diagnosed with the COVID-19 virus) did not show any COVID-19 findings using the x-ray.

Finally, the Receiver Operating Characteristics (ROC) curves with and without infrasound were statistically compared and found to be significantly different for the COVID-19PN group (Figure 6A, P <0.01, AUC=0.98 with infrasound and 0.92 without infrasound), while for the patients with non-COVID-19

pneumonia, the ROC curve difference was insignificant (Figure 6B, AUC=0.92 with infrasound and 0.90 without infrasound).

Discussion

In this study, we demonstrated that the use of the novel digital stethoscope VoqX to acquire breathing signals in the sound and infrasound range of frequencies and analyzing the signals with machine-learning methods, can distinguish between patients with COVID-19PN, non-COVID-19 pneumonia and healthy subjects with an accuracy of 92%, while a standard auscultation by an acoustic stethoscope has reached the accuracy of 52% (Table 2). Infrasound plays an important role in detecting "silent pneumonia" and the VoqX provides this ability (Table 2). This phenomenon is further emphasized when the first MFCC is drawn vs. the Magnitude of breathing frequency (Figure 4). The MFCC approximates the human auditory system's response at different frequency ranges. The Magnitude of breathing frequency is the magnitude at the frequency domain of the peak that was generated from the breathing rate. Figure 4A shows that adding infrasound data to the analyses helps distinguishing between the three groups. Without infrasound analysis, the features of healthy subjects and patients with COVID-19PN are indistinguishable (Figure 4B). Comparing classification results of breathing signals of data containing infrasound with data having limited bandwidth shows improvement of the specificity from 80% for limited bandwidth data without infrasound to 93% for the data with the full spectrum including infrasound (Table 2).

To the best of our knowledge, the VoqX is the first device that utilizes machine-learning on the infrasound and sound of breathing to detect COVID-19PN more accurately, reaching area under the curve of 94% (Figure 6).

In summary, the current study shows that by using the full range of respiratory sound waves (audible and infrasound) it is possible to diagnose COVID-19PN. We have shown that adding the infrasound data of breathing improves the accuracy of COVID-19PN detection (silent pneumonia). Furthermore, since

COVID-19PN is significantly manifested in the infrasound, while non-COVID-19PN has most of its acoustic energy in the sound range, using the novel digital stethoscope VoqX helps distinguishing between COVID-19PN and non-COVID-19 pneumonia.

Limitations and future work

Acquiring data at 14 locations on the chest wall is time consuming (16 seconds were used for the processing for each location). Thus, currently a new classifier is being developed to diagnose lung diseases by using only 4 locations at the back. Acquiring the data will require 4*16= 64 seconds and the analysis will take 30 seconds. Such a time frame for diagnosis enables the integration of the VoqX in the in-clinic practice.

It is noted that there was a significant difference between the age of the healthy subjects and the patients. However, as far as we know, there is no need to consider the patient's age while auscultating to lung sounds [27].

Conflict of interest - Noa Bachner-Hinenzon and Doron Adler are full-time employees of Sanolla. Other authors do not have any conflict of interests.

Table 1 – Selected Patient Characteristics and Physical Findings by Lung Auscultation. STD=standard deviation

							Lung auscultation				
	N	**Average Age ± STD [Years]**	**Average Height ± STD [cm]**	**Average Weight ± STD [kg]**	**Smoker**	**Gender**	**Crackles [%]**	**Wheezes [%]**	**Crackles &Wheezes [%]**	**Reduced Intensity Breathing Sound [%]**	**Vesicular Breathing Sounds [%]**
Healthy	141	42±19	168±11	73±21	52	F-58 M-83	16	0	0	8	76
Non COVID-19PN	61	62±17*	166±9†*	81±20	38†*	F-33 M-28	59†*	7†*	3†*	3	28†*
COVID-19PN	164	58±12*	170±11	78±28	32	F-77 M-87	19*	4*	0	2	75

*P<0.05 vs Healthy, †P<0.05 vs COVID-19PN

Table 2 – Performance Measures [%] from 12 Runs of Random Sets

		Specificity	Sensitivity non-COVID-19 pneumonia	Sensitivity COVID-19 pneumonia	NPV	PPV Non-COVID-19 pneumonia	PPV COVID-19 pneumonia	Accuracy
Test group -not in the learning process – (Average±std)	**With IFS**	93±4**	70±7	97±2**	90±4**	90±4*	94±3**	92±1**
	No IFS	80±7	69±11	93±4	85±5	84±6	85±3	85±2
Cross-validation - optimization of the learning process – (Average±s	**With IFS**	87±2**	71±7*	93±2**	88±3*	78±6**	89±1**	87±2**
	No IFS	75±4	65±8	85±3	79±2	68±5	80±2	78±2

td)								
Auscultation by an acoustic stethoscope (for all data –%, [CI]		76%, [68-83]	72%, [59-82]	25%, [17-32]	43%, [37-50]	71%, [62-79]		52%, [44-66]

*IFS-Infrasound, NPV-Negative predictive value, PPV–Positive predictive value, * P<0.05 vs no infrasound, ** P<0.01 vs. no infrasound*

Data availability

All the data and codes are available: GitHub - Sanolla/Lung_Classifier: Sanolla lungs classifier - Matlab code

Figure legends

Figure 1 – A Map of locations for acquiring the acoustic data

Figure 2 - Optional cloud connectivity

Figure 3– Flowchart of the preprocessing and machine-learning algorithm

Figure 4 – Visual sound signature of VoqX recorded from: a healthy subject (A), a COVID-19PN patient (B) and a patient with non-COVID-19 pneumonia (C). The x axis is Time and contains 8.5 seconds. The Y axis is Frequency from 0 to 1000 Hz. The colors are the intensity of breathing sounds in dBFS (dB-full-scale)

Figure 5 – Two features (out of 164) that strongly depend on the infrasound: MFCC1 as function of Magnitude of breathing frequency for COVID-19 (blue), Normal (red) and non-COVID-19 (yellow). A. With infrasound B. Without infrasound.

Figure 6 – ROC curves after classification with and without infrasound. A. Detection of "silent" COVID-19PN. B. Detection of non-COVID-19 pneumonia.

References

1. Bonilla-Aldana DK, Dhama K, Rodriguez-Morales AJ. Revisiting the one health approach in the context of COVID-19: A look into the ecology of this emerging disease. *Adv. Anim. Vet. Sci.* 2020; .
2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* 2020; .
3. The Lancet. Emerging understandings of 2019-nCoV. Lancet 2020.
4. Tan B, Zhang Y, Gui Y, Wu S, Li Y. The possible impairment of respiratory-related neural loops may be associated with the silent pneumonia induced by SARS-CoV-2. *J. Med. Virol.* 2020; .
5. Loudon R, Murphy RLH. Lung sounds. Am. Rev. Respir. Dis. 1984.
6. Pasterkamp H, Kraman SS, Wodicka GR. Respiratory sounds: Advances beyond the stethoscope.

Am. J. Respir. Crit. Care Med. 1997.

7. Ye Z, Zhang Y, Wang Y, Huang Z, Song B. Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review. *Eur. Radiol.* 2020; .
8. Garin N, Marti C, Scheffler M, Stirnemann J, Prendki V. Computed tomography scan contribution to the diagnosis of community-acquired pneumonia. Curr. Opin. Pulm. Med. 2019.
9. Xu B, Xing Y, Peng J, Zheng Z, Tang W, Sun Y, Xu C, Peng F. Chest CT for detecting COVID-19: a systematic review and meta-analysis of diagnostic accuracy. Eur. Radiol. 2020.
10. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z, Xiong Y, Zhao Y, Li Y, Wang X, Peng Z. Clinical Characteristics of 138 Hospitalized Patients with 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA - J. Am. Med. Assoc.* 2020; .
11. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; .
12. Zhang R, Ouyang H, Fu L, Wang S, Han J, Huang K, Jia M, Song Q, Fu Z. CT features of SARS-CoV-2 pneumonia according to clinical presentation: a retrospective analysis of 120 consecutive patients from Wuhan city. *Eur. Radiol.* 2020; .
13. Rodriguez-Morales AJ, Cardona-Ospina JA, Gutiérrez-Ocampo E, Villamizar-Peña R, Holguin-Rivera Y, Escalera-Antezana JP, Alvarado-Arnez LE, Bonilla-Aldana DK, Franco-Paredes C, Henao-Martinez AF, Paniz-Mondolfi A, Lagos-Grisales GJ, Ramírez-Vallejo E, Suárez JA, Zambrano LI, Villamil-Gómez WE, Balbin-Ramon GJ, Rabaan AA, Harapan H, Dhama K, Nishiura H, Kataoka H, Ahmad T, Sah R. Clinical, laboratory and imaging features of COVID-19: A systematic review and meta-analysis. Travel Med. Infect. Dis. 2020.
14. Xie J, Tong Z, Guan X, Du B, Qiu H, Slutsky AS. Critical care crisis and some recommendations during the COVID-19 epidemic in China. *Intensive Care Med.* 2020; .
15. Arts L, Lim EHT, van de Ven PM, Heunks L, Tuinman PR. The diagnostic accuracy of lung auscultation in adult patients with acute pulmonary pathologies: a meta-analysis. *Sci. Rep.* 2020; .
16. Franquet T. Imaging of pneumonia: Trends and algorithms. *Eur. Respir. J.* 2001; .
17. Prachand VN, Milner R, Angelos P, Posner MC, Fung JJ, Agrawal N, Jeevanandam V, Matthews JB. Medically Necessary, Time-Sensitive Procedures: Scoring System to Ethically and Efficiently Manage Resource Scarcity and Provider Risk During the COVID-19 Pandemic. *J. Am. Coll. Surg.* 2020; .
18. Vasudevan RS, Horiuchi Y, Torriani FJ, Cotter B, Maisel SM, Dadwal SS, Gaynes R, Maisel AS. Persistent Value of the Stethoscope in the Age of COVID-19. Am. J. Med. 2020.

19. Mlodzinski E, Stone DJ, Celi LA. Machine Learning for Pulmonary and Critical Care Medicine: A Narrative Review. *Pulm. Ther.* 2020; .
20. Brown C, Chauhan J, Grammenos A, Han J, Hasthanasombat A, Spathis D, Xia T, Cicuta P, Mascolo C. Exploring Automatic Diagnosis of COVID-19 from Crowdsourced Respiratory Sound Data. *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.* 2020.
21. Gavriely N, Cugell DW. Breath Sounds Methodology. Breath Sounds Methodol. 2019.
22. Gavriely N, Nissan M, Rubin AHE, Cugell DW. Spectral characteristics of chest wall breath sounds in normal subjects. *Thorax* 1995; 50: 1292–1300.
23. Kasim N, Bachner-hinenzon N, Brikman S, Cheshin O, Adler D, Dori G. Biomedical Signal Processing and Control A comparison of the power of breathing sounds signals acquired with a smart stethoscope from a cohort of COVID-19 patients at peak disease , and pre-discharge from the hospital. *Biomed. Signal Process. Control* [Internet] Elsevier Ltd; 2022; 78: 103920Available from: https://doi.org/10.1016/j.bspc.2022.103920.
24. WHO. Clinical management of severe acute respiratory infection when novel coronavirus (2019-nCoV) infection is suspected. interim guidance. *WHO* 2020; .
25. VLADIMIR VAPNIK. Support-Vector Networks. *Mach. Learn.* 1995; 20: 273–297.
26. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982; .
27. Gross V, Dittmar A, Penzel T, Schüttler F, Von Wichert P. The relationship between normal lung sounds, age, and gender. *Am. J. Respir. Crit. Care Med.* 2000; 162: 905–909.

Appendix – Features for machine learning and their p values for separating the 3 groups

Features	P value
quantile 25 of amplitude of FFT of breathing	<0.05

quantile 25 of amplitude of FFT of heart sounds	<0.05
quantile 25 of heart rate	<0.05
quantile 25 of respiration rate	<0.05
quantile 25 of envelope of breathing sounds	<0.05
quantile 25 of mean Value	<0.05
quantile 25 of median Value	<0.05
quantile 25 of standard Deviation	<0.05
quantile 25 of standard Deviation of Envelope	<0.05
quantile 25 of mean envelope	<0.05
quantile 25 of mean absolute deviation	<0.05
quantile 25 of quantile 25	<0.05
quantile 25 of quantile75	<0.05
quantile 25 of signalIQR	<0.05
quantile 25 of sampleSkewness	<0.05
quantile 25 of sampleKurtosis	<0.05
quantile 25 of signalEntropy	<0.05
quantile 25 of spectral entropy	<0.05
quantile 25 of dominant frequency value (range 0-400 Hz)	<0.05
quantile 25 of dominant frequency magnitude (range 0-400 Hz)	<0.05
quantile 25 of dominant frequency ratio (range 0-400 Hz)	<0.05
quantile 25 of sum FFT (range 0-400 Hz)	<0.05
quantile 25 of dominant frequency value (range 400-800 Hz)	<0.05
quantile 25 of dominant frequency magnitude(range 400-800 Hz)	<0.05
quantile 25 of dominant frequency ratio(range 400-800 Hz)	<0.05
quantile 25 of sum FFT (range 400-800 Hz)	<0.05
quantile 25 of dominant frequency value (range 800-1200 Hz)	<0.05
quantile 25 of dominant frequency magnitude (range 800-1200 Hz)	<0.05
quantile 25 of dominant frequency ratio (range 800-1200 Hz)	<0.05
quantile 25 of sum FFT (range 800-1200 Hz)	<0.05
quantile 25 of dominant frequency value (range 1200-1600 Hz)	<0.05
quantile 25 of dominant frequency magnitude (range 1200-1600 Hz)	<0.05
quantile 25 of dominant frequency ratio (range 1200-1600 Hz)	<0.05
quantile 25 of sum FFT (range 1200-1600 Hz)	<0.05
quantile 25 of dominant frequency magnitude (range 1600-2000 Hz)	<0.05
quantile 25 of dominant frequency ratio (range 1600-2000 Hz)	<0.05
quantile 25 of dominant frequency ratio (range 1600-2000 Hz)	<0.05
quantile 25 of sum FFT (range 1600-2000 Hz)	<0.05
quantile 25 of MFCC1	<0.05
quantile 25 of MFCC2	>0.05
quantile 25 of MFCC3	<0.05
quantile 25 of MFCC4	<0.05
quantile 25 of MFCC5	<0.05
quantile 25 of MFCC6	<0.05
quantile 25 of MFCC7	<0.05
quantile 25 of MFCC8	<0.05
quantile 25 of MFCC9	<0.05

quantile 25 of MFCC10	<0.05
quantile 25 of MFCC11	<0.05
quantile 25 of MFCC12	<0.05
quantile 25 of MFCC13	>0.05
quantile 25 of MFCC14	<0.05
quantile 25 ofMFCC15	<0.05
quantile 25 of MFCC16	<0.05
quantile 25 of MFCC17	<0.05
quantile 25 of MFCC18	<0.05
quantile 25 of MFCC19	<0.05
quantile 25 of MFCC20	<0.05
quantile 25 of sum of FFT 0-2000 Hz	<0.05
quantile 25 of MFCC1 of infrasound	<0.05
quantile 25 of MFCC2 infrasound	<0.05
quantile 25 of MFCC3 infrasound	<0.05
quantile 25 of MFCC4 infrasound	<0.05
quantile 25 of MFCC5 infrasound	>0.05
quantile 25 of MFCC6 infrasound	<0.05
quantile 25 of MFCC7 infrasound	<0.05
quantile 25 of MFCC8 infrasound	>0.05
quantile 25 of MFCC9 infrasound	>0.05
quantile 25 of MFCC10 infrasound	<0.05
quantile 25 of MFCC11 infrasound	>0.05
quantile 25 of MFCC12 infrasound	>0.05
quantile 25 of MFCC13 infrasound	>0.05
quantile 25 of MFCC14 infrasound	<0.05
quantile 25 of MFCC15 infrasound	<0.05
quantile 25 of MFCC16 infrasound	>0.05
quantile 25 of MFCC17 infrasound	<0.05
quantile 25 of MFCC18 infrasound	<0.05
quantile 25 of MFCC19 infrasound	<0.05
quantile 25 of MFCC20 infrasound	<0.05
quantile 25 of sum of FFT 0-20 Hz	<0.05
quantile 75 of amplitude of FFT of breathing	<0.05
quantile 75 of amplitude of FFT of heart sounds	<0.05
quantile 75 of heart rate	<0.05
quantile 75 of respiration rate	<0.05
quantile 75 of envelope of breathing sounds	<0.05
quantile 75 of mean Value	<0.05
quantile 75 of median Value	<0.05
quantile 75 of standard Deviation	<0.05
quantile 75 of standard Deviation of Envelope	<0.05
quantile 75 of mean envelope	<0.05
quantile 75 of mean absolute deviation	<0.05
quantile 75 of quantile 75	<0.05
quantile 75 of quantile75	<0.05

quantile 75 of signalIQR	<0.05
quantile 75 of sampleSkewness	<0.05
quantile 75 of sampleKurtosis	<0.05
quantile 75 of signalEntropy	<0.05
quantile 75 of spectral entropy	<0.05
quantile 75 of dominant frequency value (range 0-400 Hz)	<0.05
quantile 75 of dominant frequency magnitude (range 0-400 Hz)	<0.05
quantile 75 of dominant frequency ratio (range 0-400 Hz)	<0.05
quantile 75 of sum FFT (range 0-400 Hz)	<0.05
quantile 75 of dominant frequency value (range 400-800 Hz)	<0.05
quantile 75 of dominant frequency magnitude(range 400-800 Hz)	<0.05
quantile 75 of dominant frequency ratio(range 400-800 Hz)	<0.05
quantile 75 of sum FFT (range 400-800 Hz)	<0.05
quantile 75 of dominant frequency value (range 800-1200 Hz)	<0.05
quantile 75 of dominant frequency magnitude (range 800-1200 Hz)	<0.05
quantile 75 of dominant frequency ratio (range 800-1200 Hz)	<0.05
quantile 75 of sum FFT (range 800-1200 Hz)	<0.05
quantile 75 of dominant frequency value (range 1200-1600 Hz)	<0.05
quantile 75 of dominant frequency magnitude (range 1200-1600 Hz)	<0.05
quantile 75 of dominant frequency ratio (range 1200-1600 Hz)	<0.05
quantile 75 of sum FFT (range 1200-1600 Hz)	<0.05
quantile 75 of dominant frequency magnitude (range 1600-2000 Hz)	<0.05
quantile 75 of dominant frequency ratio (range 1600-2000 Hz)	>0.05
quantile 75 of dominant frequency ratio (range 1600-2000 Hz)	<0.05
quantile 75 of sum FFT (range 1600-2000 Hz)	<0.05
quantile 75 of MFCC1	<0.05
quantile 75 of MFCC2	<0.05
quantile 75 of MFCC3	<0.05
quantile 75 of MFCC4	<0.05
quantile 75 of MFCC5	<0.05
quantile 75 of MFCC6	<0.05
quantile 75 of MFCC7	<0.05
quantile 75 of MFCC8	<0.05
quantile 75 of MFCC9	<0.05
quantile 75 of MFCC10	<0.05
quantile 75 of MFCC11	<0.05
quantile 75 of MFCC12	<0.05
quantile 75 of MFCC13	<0.05
quantile 75 of MFCC14	<0.05
quantile 75 ofMFCC15	<0.05
quantile 75 of MFCC16	<0.05
quantile 75 of MFCC17	<0.05
quantile 75 of MFCC18	<0.05
quantile 75 of MFCC19	<0.05
quantile 75 of MFCC20	<0.05
quantile 75 of sum of FFT 0-2000 Hz	<0.05

quantile 75 of MFCC1 of infrasound	<0.05
quantile 75 of MFCC2 infrasound	<0.05
quantile 75 of MFCC3 infrasound	<0.05
quantile 75 of MFCC4 infrasound	<0.05
quantile 75 of MFCC5 infrasound	>0.05
quantile 75 of MFCC6 infrasound	<0.05
quantile 75 of MFCC7 infrasound	>0.05
quantile 75 of MFCC8 infrasound	<0.05
quantile 75 of MFCC9 infrasound	>0.05
quantile 75 of MFCC10 infrasound	>0.05
quantile 75 of MFCC11 infrasound	<0.05
quantile 75 of MFCC12 infrasound	<0.05
quantile 75 of MFCC13 infrasound	>0.05
quantile 75 of MFCC14 infrasound	>0.05
quantile 75 of MFCC15 infrasound	<0.05
quantile 75 of MFCC16 infrasound	<0.05
quantile 75 of MFCC17 infrasound	<0.05
quantile 75 of MFCC18 infrasound	<0.05
quantile 75 of MFCC19 infrasound	<0.05
quantile 75 of MFCC20 infrasound	<0.05
quantile 75 of sum of FFT 0-20 Hz	<0.05

Figure 1

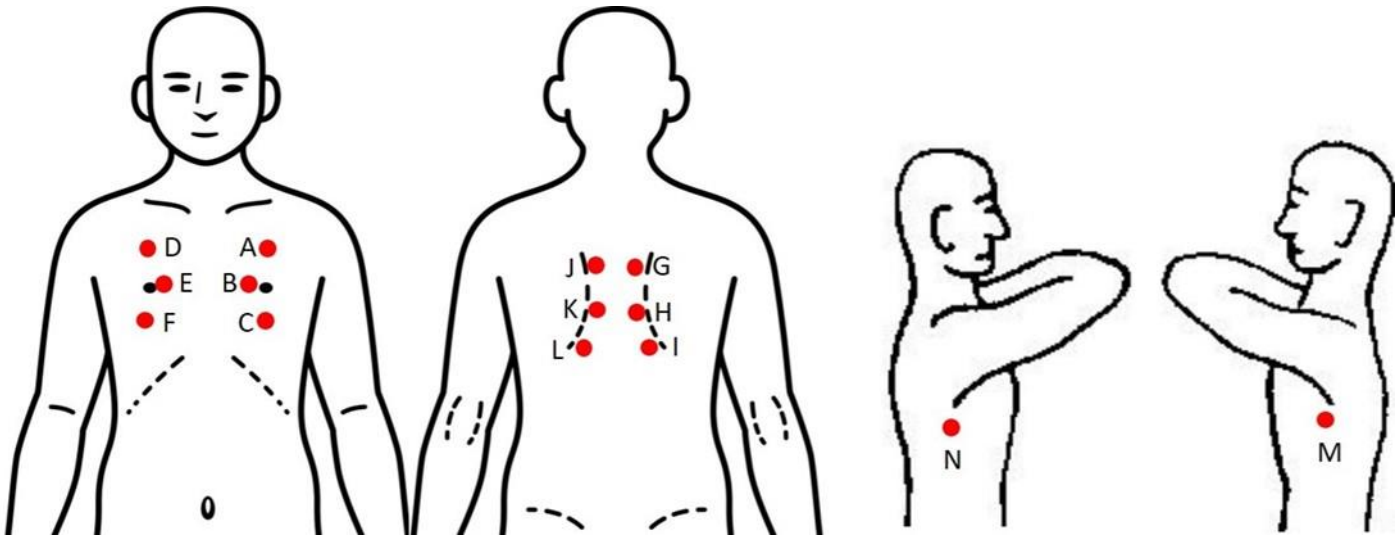


Figure 2

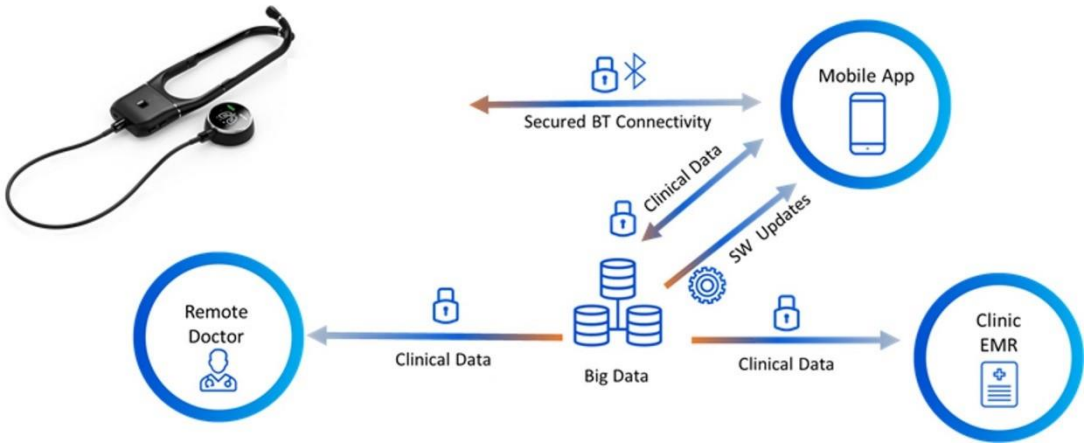

Mobile App
Secured BT Connectivity
Clinical Data
SW Updates
Remote Doctor
Clinical Data
Big Data
Clinical Data
Clinic EMR

Figure 3

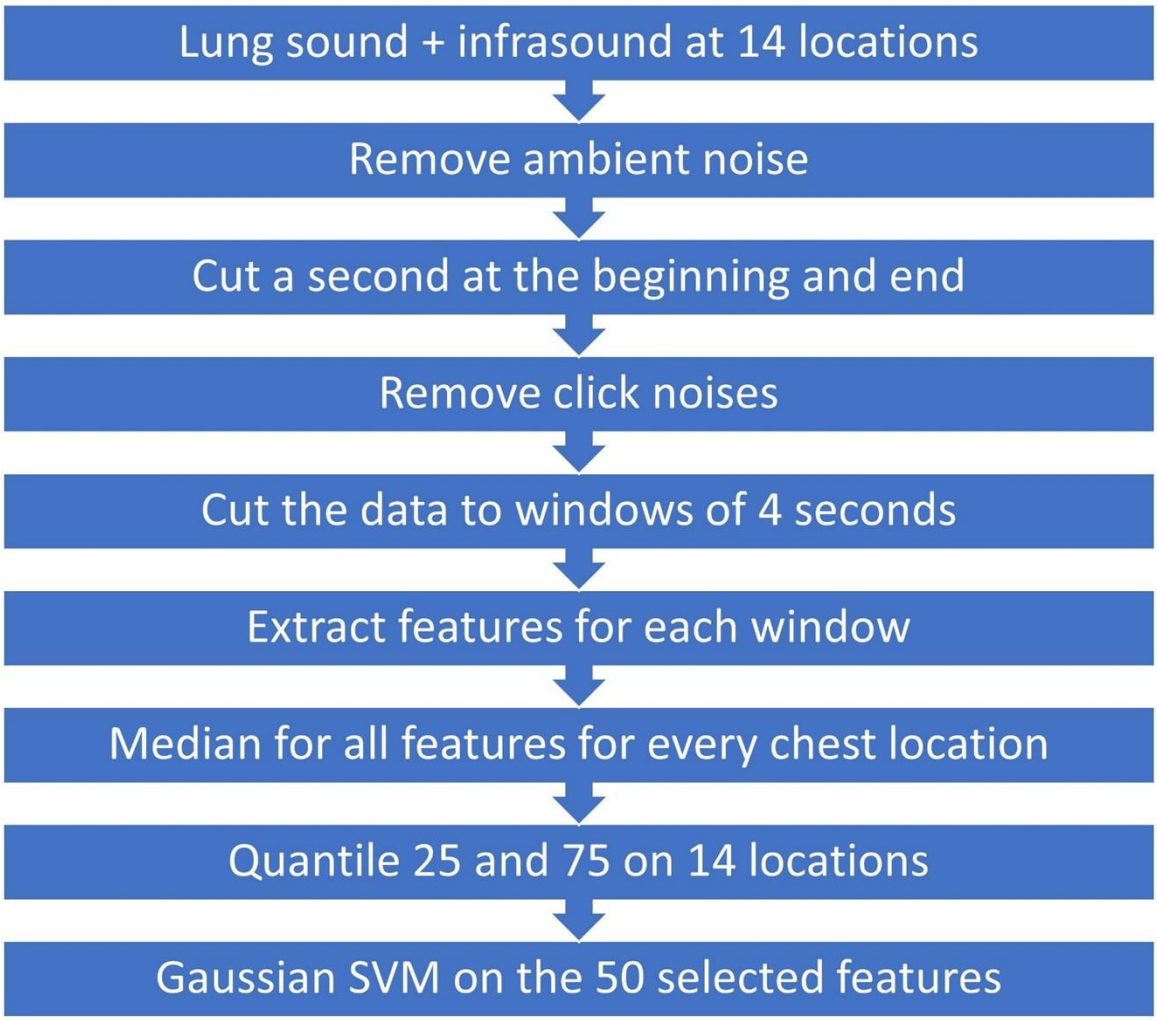

Lung sound + infrasound at 14 locations
Remove ambient noise
Cut a second at the beginning and end
Remove click noises
Cut the data to windows of 4 seconds
Extract features for each window
Median for all features for every chest location
Quantile 25 and 75 on 14 locations
Gaussian SVM on the 50 selected features

Figure 4

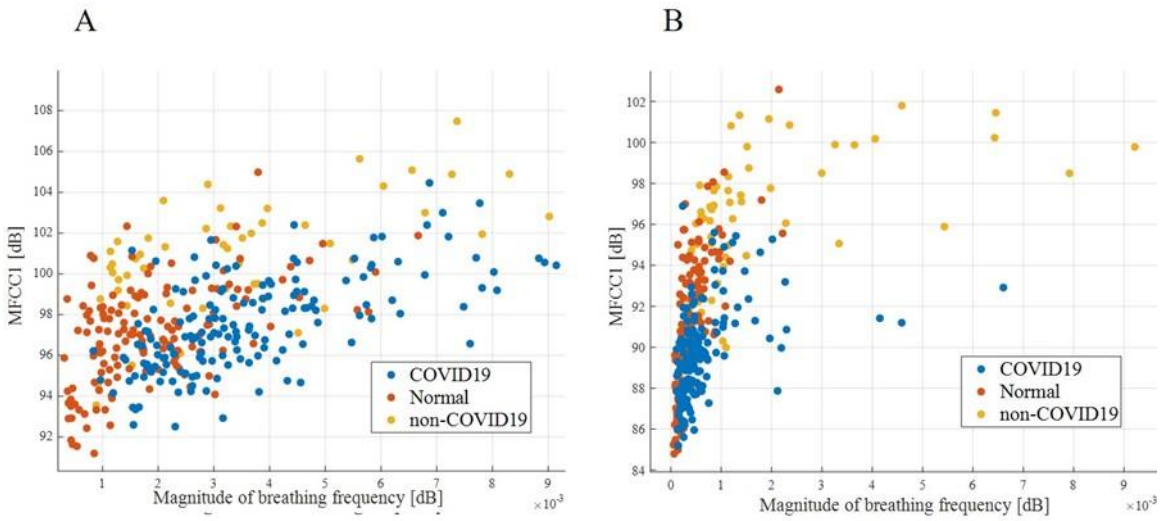

A
B
MFCC1 [dB]
Magnitude of breathing frequency [dB]
×10^-3
COVID19
Normal
non-COVID19

Figure 5

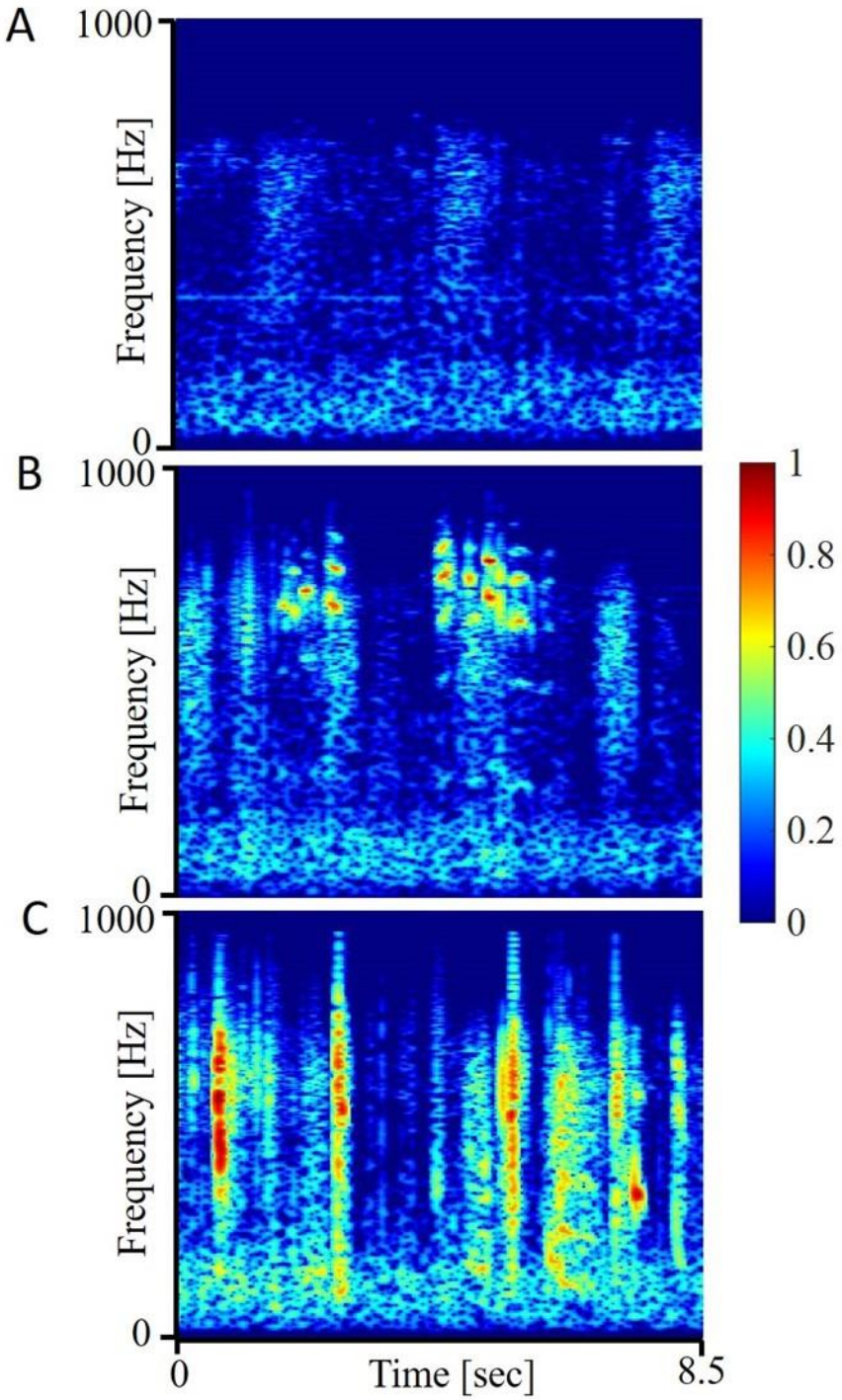
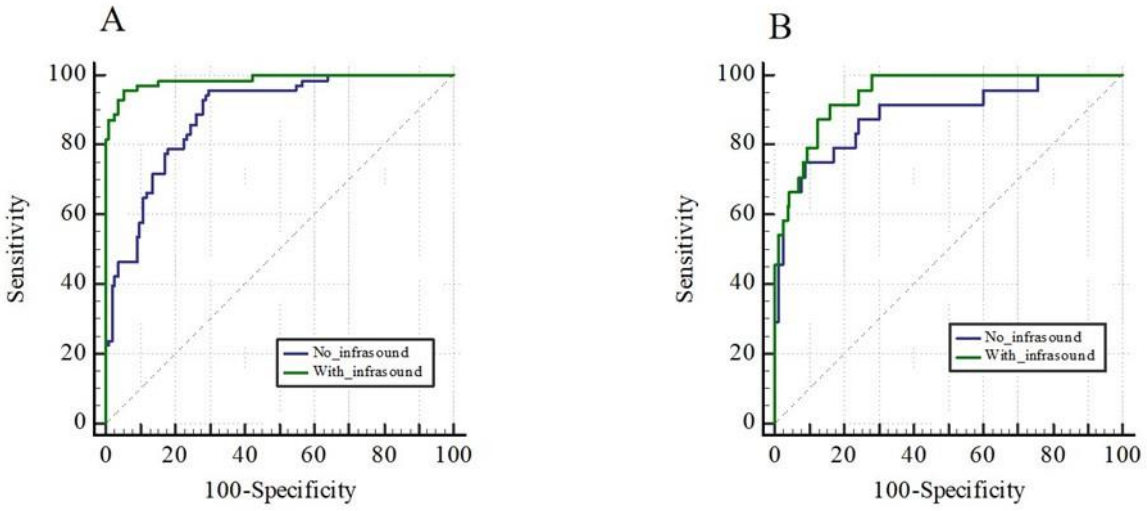

A
1000
Frequency [Hz]
0
B
1000
Frequency [Hz]
0
C
1000
Frequency [Hz]
0
0
Time [sec]
8.5
1
0.8
0.6
0.4
0.2
0

Figure 6


A
B
Sensitivity
100-Specificity
No_infrasound
With_infrasound
0
20
40
60
80
100