TIPTEKNO'24 MEDICAL TECHNOLOGIES CONGRESS
10-12 October 2024
Yasmin Bodrum Resort, Muğla, Türkiye

The Relevance of Auscultation Site in Healthy versus Restrictive Pulmonary Disease Classification

Serhat İsmet Sargın, İpek Şen

Electrical and Electronics Engineering Department, Istanbul Bilgi University, Istanbul, Turkey
ismet.sargin@bilgi.edu.tr, ipek.sen@bilgi.edu.tr

***Abstract*—The aim of this study was to perform a quantitative analysis to assess the relevance of the auscultation site on the chest wall in classifying restrictive lung diseases versus the healthy class. Accordingly, 41 features were calculated from 14-channel pulmonary sounds acquired on the posterior chest wall of 20 healthy subjects and 20 subjects with restrictive lung disease (5 with pneumonia and 15 with interstitial lung disease, ILD). These features were then used to train Bayes, decision tree (DT), and support vector machine (SVM) classifiers in leave-one-subject-out scheme. The two symmetric sites at the bases of the right and left lungs achieved the highest subject-level F1 scores of 0.974 and 0.976, respectively, while also maintaining the smallest deviation in flow cycle-level accuracies across subjects. These results were obtained using the DT classifier with f_{50} and MFCC5.**

***Keywords*—pneumonia; interstitial lung disease; restrictive diseases; classification; microphone analysis; pulmonary sounds.**

I. Introduction

For the past decades, computerized pulmonary sound analysis studies have increasingly been conducted, with an aim to provide a quantifiable, hence objective, substitute for the conventional stethoscope auscultation. Stethoscopes have long been used as one of the first-consulted tools in clinical practice, as they provide valuable information in terms of acoustic changes that occur in the body (lungs, bowels, etc.) due to the potential pathology. The shortcomings of the conventional stethoscopes can be listed as follows: (i) they have a limited and/or non-flat frequency response even when used in the extended mode [1], [2], (ii) their frequency response shows differences across different manufacturers [3], (iii) although Bluetooth transmission and recording are now possible with newly added technological features, there are no analysis methods adapted to clinical practice for the sounds recorded by these stethoscopes, (iv) they can only be used for successive single-channel measurements, later to be synthesized in the physician's mind, and consequently (v) conventional stethoscopes do not offer objective analysis or comparison, and their effectiveness depends on the physician's hearing ability and experience.

The aforementioned shortcomings of conventional stethoscopes are aimed to be mitigated with the development of computerized methods. These methods employ (i) the acquisition hardware for wide and flat frequency response for the minimum loss of acoustic information, and (ii) the software (algorithms) for quantification, detection, and analysis of discriminative features of disorders that are characterized by auscultation findings.

979-8-3315-2981-9/24/$31.00 ©2024 IEEE

Many studies have been conducted in the literature to classify pathological conditions using pulmonary sounds. The studies that concentrate on the classification of restrictive diseases versus the healthy case can be listed as [4], [5], [6], and [7]. The study in [4] focuses on multiple binary classification scenarios, including pneumonia (restrictive) versus healthy classification (overall accuracy of 79.7%). Another study [5] performs healthy versus interstitial lung disease (ILD, restrictive) classification (individual accuracies of 75% and 93% for the healthy and ILD groups, respectively). Two other studies, [6] and [7], both perform three-class classification including the bronchiectasis (obstructive) group besides the healthy and ILD (restrictive) (the former study reports individual accuracies of 95% and 100% and the latter of 90% and 100% for the healthy and ILD groups, respectively). The study in [4] do not report the microphone count, the one in [5] uses 25 microphones, while those in [6] and [7] use 14 microphones. Although all used multi-channel recordings, none examined which microphone placement on the chest wall would lead to more successful disease detection.

Restrictive conditions that occur in the lungs alter lung structure and functioning by causing thickening of the alveolar walls, decreased lung elasticity, and reduced oxygen exchange [8]. As a result of these changes, restrictive diseases are known to produce fine crackles [9]. In the presence of fine crackles, they are mostly observed in the bilateral lower back area of the lungs [10]. Accordingly, the microphones at the bilateral bases of the lungs are expected to demonstrate a higher discriminative ability as compared to those positioned at the upper regions.

Although the medical literature indicates relevance concerning the auscultation site, computerized pulmonary sounds literature lacks the corresponding quantitative analysis. This study aims to conduct such quantitative analysis as an attempt to fill this gap. Accordingly, various features were calculated from 14-channel pulmonary sounds acquired on the chest wall, and are fed to classifiers. Individual microphone performances are reported from two different perspectives to help assess the relevance of the auscultation site in classifying restrictive diseases versus the healthy case, while also shedding light on the best mathematical features and the classifier.

II. Methodology

A. Data

The pulmonary sounds data used in this study were acquired using the 14-channel data acquisition and processing system described in [12]. Sounds were recorded from 40 subjects: 20 healthy volunteers in Boğaziçi University Lung Acoustic Laboratory - LAL, and (ii) 20 volunteers diagnosed with a restrictive type of disease (5 with pneumonia and the remaining 15 with the broad disease group named interstitial lung disease, ILD) in Yedikule Chest Diseases and Thoracic Surgery Training and Research Hospital, Istanbul, Turkey. The system consists of 14 air-coupled electret microphones in capsules attached on the posterior chest wall, and a pneumatachograph that simultaneously measures the flow cycle for synchronization. Signals are passed through an amplifier-filter and then sampled at a rate of 9600 samples per second. Further details about the data acquisition system can be found in [12] and the microphone positions on the chest wall can be seen in Figure 1. An acquisition session lasted 15 seconds during which the subject sat upright and breathed spontaneously through the mouthpiece of the flow meter while wearing a nose clip. Informed consent was obtained from each volunteer before starting the session. The study was approved by the Istanbul Clinical Research Second Ethics Committee (in compliance with the Declaration of Helsinki).

B. Features

In this study, a total of 41 features (5 main feature types in both time and frequency domains) were used. First, respiratory sounds were divided into the six sub-phases of the flow cycle, namely, early/mid/late inspiration and expiration, as explained in [11]. Then, the features listed below were calculated for each sub-phase. The six sub-phases were then combined such that each flow cycle was represented by a feature vector in a six-dimensional space.

1) Autoregressive Model Coefficients (AR): To satisfy the stationarity assumption of the AR model, each sub-phase is further divided into segments of 250 samples with 50% overlap. Then for each such segment, the sixth-order [13] model coefficients were estimated using the Burg's method. Averages were then taken over the segments to obtain a common set of coefficients for the corresponding sub-phase.

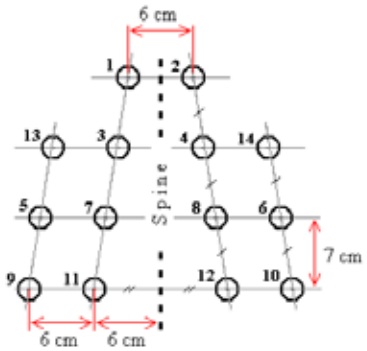


Fig. 1. Microphone Locations on the Posterior Chest Wall [11]

2) Mel Frequency Cepstral Coefficients (MFCC): First, the short-time Fourier transform (STFT) was obtained from the entire sub-phase using 250-point Hanning windows with 50% overlap. Then, the Mel scale filter bank of 32 triangular filters was applied. After taking the logarithm of the Mel scaled spectrum, discrete cosine transform (DCT) was performed and the first 13 MFCC coefficients were taken to continue with, a coefficient count successfully used in most acoustic applications. Finally, averages of the coefficients were taken over the time windows. For more detailed information, please refer to the studies in [14] and [15].

3) Fine Crackles: Since restrictive diseases (ILD and pneumonia) are associated with the production of fine crackles [9], it was deemed appropriate to use the amount of fine crackles as another feature. Fine crackles were first detected using the algorithm proposed in [12] and then counted per sub-phase.

4) Power Spectral Density (PSD) Based Features: The PSD was estimated for each sub-phase using Welch's method, with 250-point and 50% overlapping Hamming windows, and 1024-point FFT. Then, the features listed below were calculated from the PSD estimate:

- Percentile Frequencies (f_x): Percentile frequency f_x is the frequency at which the PSD reaches x ($x = 25, 50, 75, 90, 95$) percent of the total area under it.
- Maximum Power in PSD (P_{max})
- Frequency Corresponding to P_{max} (F_{max})
- Energy in Frequency Bands (BIN): The PSD was divided into frequency intervals and the power per interval (band) was calculated. Since the frequency range from zero to 4800 Hz (half of the sampling rate by Nyquist theorem) is expressed by 512 points (half of the number of FFT points, 1024), the frequency difference between two points is 9.375 Hz. It was chosen for each band to contain 10 points, resulting in bands of 93.75 Hz each (approximately 100 Hz). The first 10 bands out of a total of 51 were used in this study, as higher frequencies are less relevant in terms of information content.

5) Statistical Features: Renyi entropy [16], variance, kurtosis and skewness are calculated.

These features were chosen for their expected abilities to model the lungs as filters (AR), to emphasize the spectral patterns in the acoustic information content (MFCC and the PSD based features), and to describe the amplitude distribution (statistical features) of pulmonary sounds in the presence of a restrictive disease.

C. Classifiers

Given a microphone recording, the features calculated from sub-phases were combined to form the six-dimensional feature vectors per flow cycle. As multiple flow cycles fit in one recording session of 15 seconds, multiple such vectors represent every subject in the feature space. The vectors of all subjects were brought together and used to train the classifier, and this was done separately for each feature (each of 41)

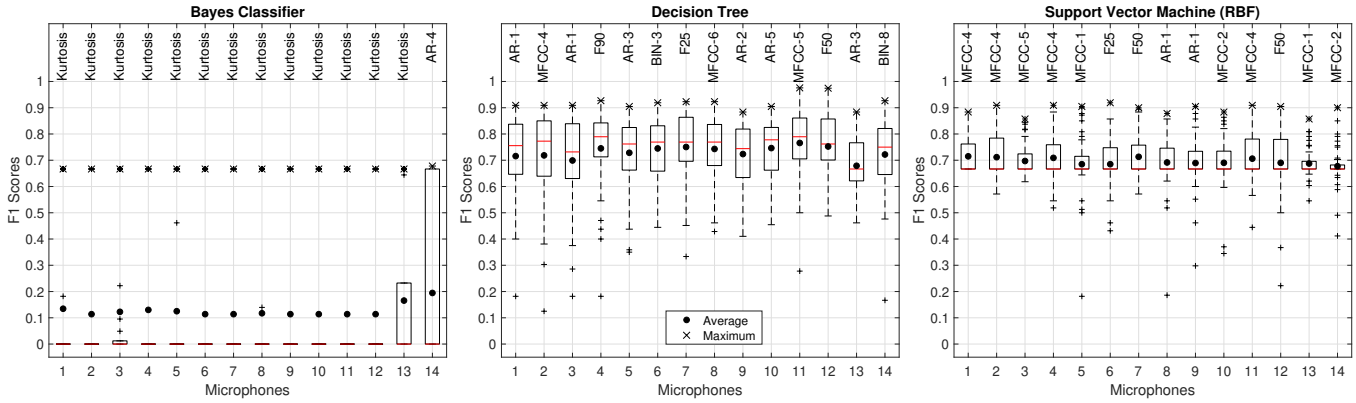


Fig. 2. F1 scores of the three classifiers. Each box contains the scores of 41 features. The dot and cross show the average and maximum of the 41 scores, respectively. For each microphone, the name of the feature that gives the maximum score is written at the top.

and each microphone (each of 14). Three different types of classifiers listed below were employed. For more information about the classifiers, please refer to [17].

1) Decision Tree (DT): Multivariate DT was adopted, and the depth was optimized over a range from 1 to 30. Gini index was used as the impurity measure.

2) Support Vector Machine (SVM): Linear and radial basis function kernels (RBF) were used. For the linear kernel, the penalty parameter was optimized over a range between 10^{-5} and 10^{5}. For the RBF kernel, the range was the same for the penalty parameter and it was between 10^{-3} and 10^{3} for the gamma value.

3) Bayes Classifier (BC): Gaussian distribution was assumed for the class likelihoods. The mean vectors and covariance matrices associated with the classes, and the class prior probabilities, were determined by maximum likelihood estimation. Decisions were made to maximize the posterior probabilities calculated using Bayes' rule.

Leave-one-subject-out was adopted, i.e., at each fold, all flow cycles of one participant were used for validation, and the flow cycles of the remaining 39 participants were used for training. In order to form subject-level decisions, flow cycle decisions were combined using majority voting. For performance assessment, F1 score (restrictive class being the positive one), and accuracy were used. F1 score require both positive and negative diagnosis and, therefore can not be calculated at each fold, where the validation set includes data from only one class. However, accuracy can be calculated at each fold as a percentage of correctly classified flow cycles, regardless of the class. Therefore, it is possible to calculate the average $\pm$ standard deviation over the folds (subjects) as a measure of the consistency of the classifier. Accordingly, F1 scores will be at subject-level and accuracies will be at flow cycle-level while reporting the results.

III. Experimental Results & Discussion

Although the linear kernel SVM was also selected for use in this study, it was omitted after being tested on 20 features, based on the observations that (i) the processing times were very long, and (ii) the F1 scores were not above those obtained with the other two classifiers. Throughout the rest of the paper, SVM will stand for the SVM classifier with RBF kernel for simplicity.

F1 scores versus microphone for the three classifiers are given as boxplots in Figure 2. Each box in the figure contains the F1 scores of the 41 features. The average and maximum F1 scores are also indicated in the figure with dots and crosses, respectively, and the names of the features that achieved the maximum F1 scores are written just above the crosses for each microphone.

The plots reveal that the F1 scores obtained with the Bayesian classifier are quite low compared to the other classifiers. Although the maximum F1 score is above 0.600 for all microphones, the average F1 score is low (0.195) even for the 14$^{\text{th}}$ microphone which achieved the highest F1 score (0.678). This indicates that the performance of the Bayes classifier is

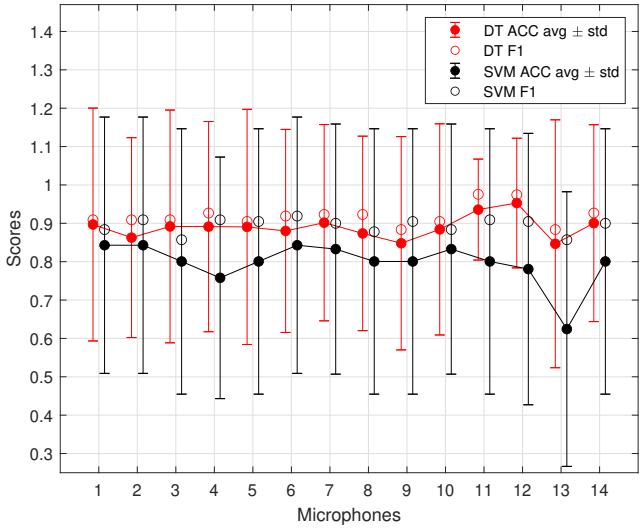


Fig. 3. F1 Scores and accuracies for the best two classifiers. Each classifier uses only its best feature per microphone (see Figure 2). Accuracies are the averages of the flow cycle-level accuracies over 40 subjects. F1 scores are subject-level and are the values shown with crosses in Figure 2.

TABLE I
NUMERICAL DETAILS OF FIGURE 3. IMPORTANT FIGURES ARE IN BOLD.

Mics	DT ACC	SVM ACC	DT F1	SVM F1
1	0.897 ± 0.303	**0.843** ± 0.334	0.909	0.884
2	0.863 ± 0.260	**0.843** ± 0.334	0.909	0.909
3	0.892 ± 0.303	0.801 ± 0.346	0.909	0.857
4	0.891 ± 0.274	0.758 ± **0.315**	0.927	0.909
5	0.891 ± 0.306	0.801 ± 0.346	0.905	0.905
6	0.880 ± 0.265	**0.843** ± 0.334	0.919	**0.919**
7	0.902 ± 0.256	0.833 ± 0.326	0.923	0.900
8	0.874 ± 0.253	0.801 ± 0.346	0.923	0.878
9	0.848 ± 0.278	0.801 ± 0.346	0.884	0.905
10	0.884 ± 0.275	0.833 ± 0.326	0.905	0.884
11	**0.936 ± 0.132**	0.801 ± 0.346	**0.976**	0.909
12	**0.953 ± 0.169**	0.781 ± 0.354	**0.974**	0.905
13	0.847 ± 0.323	0.624 ± 0.358	0.884	0.857
14	0.900 ± 0.257	0.801 ± 0.346	0.927	0.900

not successful enough for the data.

DT achieved higher scores than SVM. Taking this into account and examining the maximum F1 scores (see the crosses in the plot), the 11th and 12th microphones were observed to be the two most successful ones with F1 scores of 0.976 (achieved with MFCC5, sensitivity: 1.000, specificity: 0.950, DT depth: 1) and 0.974 (achieved with f_{50}, sensitivity: 0.9500, specificity: 1.000, DT depth: 2) respectively. These two microphones are positioned at the bases of the right and left lungs, which is in alignment with the expectations supported by the medical literature. (Please note that each box contains 41 different features, therefore one should not look for statistical significance to be able to say that some microphones are more successful than others based on this plot. Boxplot is only used for better visualisation.)

To analyse the consistency of the classifier through the folds, flow cycle-level accuracies were taken into consideration. Then, the average accuracy and the associated standard deviation were calculated over 40 subjects for each microphone. This is done for both the DT and SVM classifiers, using their best features (see Figure 2) and the corresponding hyperparameters (i.e., depth for the DT and penalty parameter and gamma for the SVM) per microphone. The results are shown in Figure 3. Observation reveals that (i) DT outperforms SVM for all the microphones in terms of both F1 score and accuracy, and (ii) the 11th and 12th microphones not only achieve the highest performance scores but also exhibit the lowest standard deviation around the average accuracy, indicating greater consistency of the classifier across folds.

These should be considered as preliminary results obtained with a limited data size (e.g., due to difficulties in clinical data collection), however in consistency with the medical literature. More complex classifiers do not seem needed, however new features and their combinations can be investigated.

IV. CONCLUSION

In this study, an analysis was performed to assess the relevance of the auscultation sites in classifying restrictive lung diseases versus the healthy class. Features calculated from 14-channel pulmonary sounds acquired on the chest wall were fed to various classifiers to evaluate individual microphone performances. Consequently, the two symmetric sites at the bases of the right and left lungs, specifically, the 11th and 12th microphones, achieved the highest F1 scores, 0.974 and 0.976, respectively. These microphones were also the two most consistent across the folds. These results were obtained using the DT classifier with f_{50} and MFCC5. For a more reliable conclusion, data should be augmented, and the study should be repeated with an improved methodology.

REFERENCES

[1] G.R. Wodicka H. Pasterkamp, S.S. Kraman. "Respiratory sounds: advances beyond the stethoscope. *American journal of respiratory and critical care medicine*, 156(3):974–987, 1997.

[2] V. Oliynik. "On potential effectiveness of integration of 3m littmann 3200 electronic stethoscopes into the third-party diagnostic systems with auscultation signal processing. In *2015 IEEE 35th International Conference on Electronics and Nanotechnology (ELNANO)*, pages 417–421, 2015.

[3] R.K. Brown A.M. Stern P.Y. Ertel, M.Lawrence. Stethoscope acoustics. *Circulation*, 34(5):899–909, 1966.

[4] U. Cini I. Sen G. Kilic Y.P. Kahya M. Yeginer, K. Çiftçi. "Using lung sounds in classification of pulmonary diseases according to respiratory subphases. In *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, volume 1, pages 482–485. IEEE, 2004.

[5] R.Gonzalez-Camarena G. Chi-Lem J.G. Carillo T. Aljama-Corrales S.Charleston-Villalobos, G. Martinez-Hernandez. "Assessment of multichannel lung sounds parameterization for two-class classification in interstitial lung disease patients. *Computers in biology and medicine*, 41(7):473–482, 2011.

[6] Y.P. Kahya I.Sen, M.Saraclar. "Computerized diagnosis of respiratory disorders: Svm based classification of var model parameters of respiratory sounds. *Methods of Information in Medicine*, 53(4):291–295, 2014.

[7] Y.P. Kahya I. Sen, M. Saraclar. "A comparison of svm and gmm-based classifier configurations for diagnostic classification of pulmonary sounds. *IEEE Transactions on Biomedical Engineering*, 62(7):1768–1776, 2015.

[8] T.E. Hartman-E.S. Yi J.H. Ryu, C.E. Daniels. "Diagnosis of interstitial lung diseases. In *Mayo Clinic Proceedings*, volume 82, pages 976–986. Elsevier, 2007.

[9] R.G. London-E.E. Lower R.P. Baughman, R.T. Shipley. "Crackles in interstitial lung disease: comparison of sarcoidosis and fibrosing alveolitis. *Chest*, 100(1):96–101, 1991.

[10] Y. Ishimatsu-S. Nakashima H. Ishimoto N. Sakamoto K. Nishitsuji-S. Shiwa T. Sakai S. Miyahara K. Ashizawa H. Mukae R. Kozu T. Fukumitsu, Y. Obase. "The acoustic characteristics of fine crackles predict honeycombing on high-resolution computed tomography. *BMC pulmonary medicine*, 19:1–8, 2019.

[11] Y.P. Kahya I. Sen, M. Saraclar. "Differential diagnosis of asthma and copd based on multivariate pulmonary sounds analysis. *IEEE Transactions on Biomedical Engineering*, 68(5):1601–1610, 2021.

[12] Y.P. Kahya I. Sen. "A multi-channel device for respiratory sound data acquisition and transient detection". In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, pages 6658–6661. IEEE, 2005.

[13] B. Sankur-E.C. Güler T.Engin B. Sankur, Y.P. Kahya. "Comparison of ar-based algorithms for respiratory sounds classification.

[14] E. Rodriguez-Villegas R.X.A. Pramono, S.A. Imtiaz. "Evaluation of mel-frequency cepstrum for wheeze analysis. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4686–4689. IEEE, 2019.

[15] C. Pelletier M. Bahoura. "New parameters for respiratory sound classification. In *CCECE 2003-Canadian Conference on Electrical and Computer Engineering. Toward a Caring and Humane Technology (Cat. No. 03CH37436)*, volume 3, pages 1457–1460. IEEE, 2003.

[16] Y.P. Kahya-M.K. Mihcak S. Aydore, I. Sen. "Classification of respiratory signals by linear analysis. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2617–2620. IEEE, 2009.

[17] E. Alpaydin. *Introduction to Machine Learning, Fourth ed.* Adaptive Computation and Machine Learning Series. MIT Press, 2020.