Automatic detection of wheezes by evaluation of multiple acoustic feature extraction methods and C-weighted SVM

Germán D. Sosa, Angel Cruz-Roa, Fabio A. González

MindLab Research Group, Universidad Nacional de Colombia, Bogotá, Colombia

ABSTRACT

This work addresses the problem of lung sound classification, in particular, the problem of distinguishing between wheeze and normal sounds. Wheezing sound detection is an important step to associate lung sounds with an abnormal state of the respiratory system, usually associated with tuberculosis or another chronic obstructive pulmonary diseases (COPD). The paper presents an approach for automatic lung sound classification, which uses different state-of-the-art sound features in combination with a C-weighted support vector machine (SVM) classifier that works better for unbalanced data. Feature extraction methods used here are commonly applied in speech recognition and related problems thanks to the fact that they capture the most informative spectral content from the original signals. The evaluated methods were: Fourier transform (FT), wavelet decomposition using Wavelet Packet Transform bank of filters (WPT) and Mel Frequency Cepstral Coefficients (MFCC). For comparison, we evaluated and contrasted the proposed approach against previous works using different combination of features and/or classifiers. The different methods were evaluated on a set of lung sounds including normal and wheezing sounds. A leave-two-out per-case cross-validation approach was used, which, in each fold, chooses as validation set a couple of cases, one including normal sounds and the other including wheezing sounds. Experimental results were reported in terms of traditional classification performance measures: sensitivity, specificity and balanced accuracy. Our best results using the suggested approach, C-weighted SVM and MFCC, achieve a 82.1% of balanced accuracy obtaining the best result for this problem until now. These results suggest that supervised classifiers based on kernel methods are able to learn better models for this challenging classification problem even using the same feature extraction methods.

Keywords: Wheeze detection, Lung sounds classification, Support Vector Machine, biomedical signal analysis.

1. MOTIVATION AND PURPOSE

Around 4 million of deaths are caused by lung diseases every year. These diseases had affected over 300 million people around the world between 2010 and 2011.[1] Studies have related the growing spread of these diseases with respiratory airways exposition to vicious environments and social factors as obesity and continued smoking. A lot of people suffer some kind of lung disease, their symptoms are underestimated and often lack any medical treatment.

Obstructive lung diseases are mainly characterized by the presence of wheezes, an adventitious sound that is related to various types of narrowing of lung airways. In the clinical practice, wheezes are manually detected by auscultation. However, it means that a level of subjectivity is involved and totally depends on the physician experience, which affects the diagnosis and the reproducibility in clinical routine due to the lack of computerized methods for lung sound analysis.

In the 90's a European Commission funded the CORSA project (Computerized Respiratory Sound Analysis) to standardize all measures, terms and computational techniques related to lung sound characterization.[2] Since CORSA project started, there have been some previous works which had used CORSA standard, e.g. Mayorga et al.,[3] Amjad Hashemi et al.[4] and Bahoura.[5]

Further author information: (Send correspondence to Fabio González)
Fabio González: E-mail: fagonzalezo@unal.edu.co, Telephone: +57 (1)3165000 Ext. 14077

Taking into account that the previous works did not explore the potential capabilities of better machine learning classifiers, we explore in this paper the integration of a more reliable method for automatic classification, a C-weighted Support Vector Machine, evlauating it with different state-of-the-art feature extraction methods. For validation we compared our approach against the best results reported in literature for wheeze detection in respiratory sounds.

2. METHOD: AUTOMATIC DETECTION OF WHEEZES IN LUNG SOUNDS

The proposed method for automatic detection of wheeze lung sounds is summarized in Figure 1. A set of manually labeled lung sounds is used for training. Three different features are extracted from each sample sound: short time Fourier transform (FT), wavelet packet transform (WPT), and mel-frequency cepstral coefficients (MFCC), which are described in Section 2.1. Then, a C-weighted SVM model is trained to classify between wheeze or normal sounds using different type of features as input. During prediction, a new sound sample is processed to extract the features and the learned C-weighted SVM classifier is applied to predict its sound class either normal or wheeze.

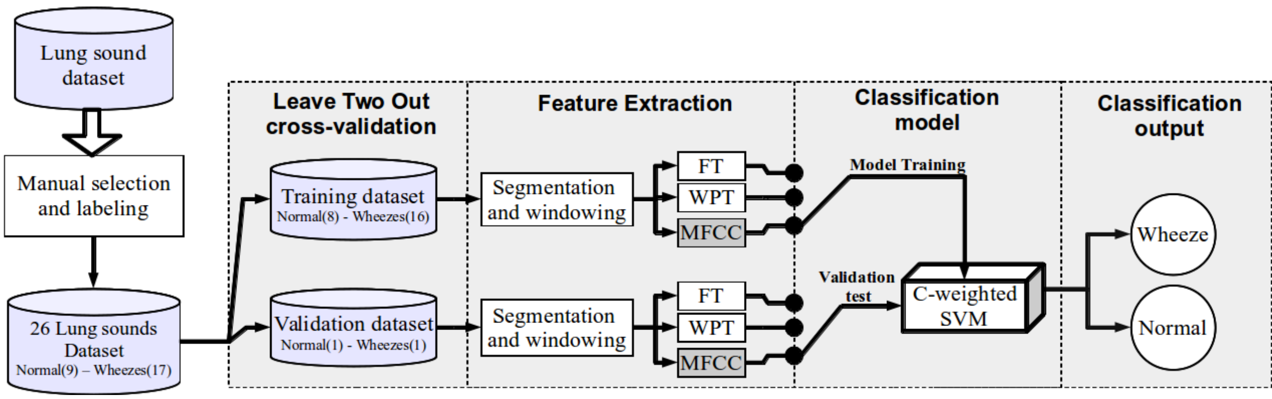

Figure 1. Overview of the proposed method for automatic detection of wheezes in lung sounds by using different acoustic state-of-the-art feature extraction methods and a C-weighted SVM classifier.

2.1 Acoustic feature extraction methods

Given a 1D-signal represented by a d-dimensional vector, a feature extraction method aims to find a new more efficient feature vector representation of m elements where $m < d$. In this work we used three of the state-of-the-art acoustic feature extraction methods which are described below.

Fourier Transform (FT): A variant of the traditional Fourier transform is the Short Time Fourier Transform (STFT), which provides temporal and frequency information by calculating the FT over a set of segments of the original signal, and it is defined for discrete time signals as follows:

$$STFT_i[n] = F[\tau,\omega] = \sum_{n=1}^{f} x[n]w[n-i\tau]e^{-j\omega n} \text{ with } \omega = \frac{2\pi}{N} \tag{1}$$

For the i-th STFT segment analyzed, the signal $x[n]$ is multiplied element by element by a window function $w[n,\tau]$ which is nonzero only in a portion of the signal and help to diminish the spectral leakage effect, this window function is shifted i times for some shifting constant τ. Then the FT is applied over the windowed portion of signal of length f. Being k the number of segments to be analyzed, the full STFT transform will be built concatenating every single iteration of STFT into a $f \times k$ matrix, where the i-th column corresponds to the Fourier content of i-th segment of signal with $i = 1,2,...,k$.

In our case, we set $f = 1024$ (0.1ms with $Fs = 10240$Hz) and $\tau = f/2$ for 50% overlap between segments. The Hamming function was used as window function for this feature extraction process as well as for the other ones. Particularly, for the FT feature set we subdivided spectra for each segment into 26 partitions taking power spectrum density (PSD) of each one, given by:

$$PSD[i,j] = \frac{1}{N}|STFT_{i,j}| \text{for the } j\text{-th partition being } j = 1,2,...,26 \tag{2}$$

Wavelet Package Transform (WPT): Another widely used option for signal analysis is the wavelet transform, where a 1D-signal can be represented as the sum of short-time signals generated by a set of wavelet basis. Wavelet basis are time-amplitude scaled versions of a mother wavelet where all of them satisfy orthogonality among them and can be used as basis to represent variety of signals in time domain. The wavelet transform is defined as follows:

$$WT(x(t)) = \int_{-\infty}^{\infty} x(t)\psi_{a,b}(t)dt \quad \text{with} \quad \psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi(\frac{t-b}{a}) \tag{3}$$

Where a, b are the scale and position coefficient respectively. Particularly, we used the discrete wavelet transform, where the a and b parameters vary in a dyadic sequence through a set of j decomposition stages of filtering and subsampling.[6]

In this work we set $j = 7$ as well as the symlet (sym8) wavelet function as mother wavelet for each segment. Then, a set of 19 features was taken in the same way as it was performed for baseline methods.[4,5] This feature set was built taking the mean of absolute values (μ_j), average power of values (p_j), standar deviation (σ_j) for each subband, and ratio of median of absolute values for adjacent subbands (μ_j/μ_{j+1}) using $j = 3,4,5,6,7$

Mel-Frequency Cepstral Coefficients (MFCC): The cepstral analysis was mainly motivated by speech recognition problems wehre it is important to separate the complete speech signal $y[n]$ into two signals, glottal speech $s[n]$ and vocal tract response $h[n]$, being the total speech signal the convolution of these ones. Hence, the complete speech signal is modeled in the discrete time domain as follows:

$$y[n] = s[n] * h[n] \tag{4}$$

Taking into account this formulation, the log-transformation of the spectra of the convoluted signal $y[n]$ can be expressed as:

$$log(|Y(\omega)|) = log(|S(\omega)|) + log(|H(\omega)|) \tag{5}$$

where, $|Y(\omega)|$, $|S(\omega)|$, and $|H(\omega)|$ are the absolute values of the corresponding FT of signals $y[n]$, $s[n]$ and $h[n]$. The cepstrum coefficients are obtained by performing the inverse FT of $log(|Y(\omega)|)$ to the time domain. Particularly, Mel-Frequency Cepstral Coefficients are obtained in a similar way, but involving a filtering stage through a sequence of triangular filters in the frequency domain separated by the mel-scale that make the coefficients more consistent to the human hearing.

In this case, we will treat wheezes as superimposed signals over the normal breath sounds, 24 MFCC were used onto a frequency band from 0 to 5120 Hz for each segment.

2.2 C-weighted Support Vector Machine Classifier (C-weighted SVM)

Support Vector Machine (SVM): SVM are the most popular and successful margin-based machine learning method used for regression and classification tasks. The key part of this approach is the "kernel trick", which implicitely transform the original data representation, using a kernel function, into a high dimensional space, a.k.a. feature space, where data is expected to be linearly separable by hyperplanes for the classification task. Independently of the type of kernel used, in training stage, a SVM solves an optimization problem to find the optimal hyperplane to linearly separate the classes, which is defined by:

$$\tilde{L}(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j k(\mathbf{x}_i, \mathbf{x}_j) \quad \text{, s.t. (for any } i = 1, \ldots, n\text{): } 0 \leq \alpha_i \leq C, \text{ and } \sum_{i=1}^{n} \alpha_i y_i = 0 \tag{6}$$

where $k(\mathbf{x}_i, \mathbf{x}_j)$ is the kernel function and C is the complexity parameter.

C-weighted SVM: Imbalanced data, the typical scenario in biomedical applications, may affect the performance of a typical SVM. The complexity parameter C is related with the tolerance of model to misclassification, the larger the complexity of the model the fewer instances misclassified and viceversa. However, the concern for a problem with class imbalance has to do with how to adjust C parameter to let the model be prone to misclassify some instances of majority class while keeping almost all instances of minority class well labeled. As a solution to this problem, the C-weighted SVM uses a weighted C for each class in the data[7] where each C is weighted inversely proportional to the frequency of data for each class. Thus, classes with lots of instances will get a less complex separation margin, while the class with fewer data will be adjusted in a better way with a more complex margin.

3. EXPERIMENTAL EVALUATION

3.1 Lung Sounds Dataset

The database used is the R.A.L.E. Repository provided by the University of Manitoba, in Winnipeg Canada.[8] This dataset consists of more than 50 lung recordings which exhibits a variety of adventitious sounds as wheezes, crackles, rhonchi and other ones. All of these sounds were recorded at 10240 Hz, and then, a post-processing stage was applied: first, a portion of the signal is extracted corresponding to one full breathing cycle; and second, signal amplitude is normalized and filtered using a 4000 Hz high-pass filter for noise suppression. From the original 50 recordings, 26 were chosen for the final dataset, where 17 correspond to wheezes and 9 to normal breath sounds according to the the diagnosis provided by an expert. After that, all recording were partitioned in segments of 0.1ms with 50% overlap where each one was labeled as wheeze or normal segment by the expert. From the 26 recordings dataset, 1188 were extracted, 898 labeled as normal or non-wheeze segments and 290 wheeze segments. The performance of the classifier was evaluated by classifying these segments correctly.

3.2 Experimental Setup

In order to evaluate the classification performance of proposed approach, it was compared against baseline methods described in previous works[3–5] using the same experimental setup and evaluation criteria to obtain comparable results as it is detailed below.

Leave-Two-Out Cross-Validation: The traditional Leave-One-Out Cross Validation (LOOCV) is a well know strategy to evaluate the prediction ability of a model when the evaluation data set is small. LOOCV is equivalent to a n-fold Cross-Validation where n is equal to the number of instances present in dataset. As it was described in the previous section, the data set is composed of several segmests extracted from 26 different cases. To make a stricter evaluation, segments from the same case are not divided among the training and validation folds, so the LOOCV is done at the case level. This means that the evaluation fold is made of several segments extracted from the same case, which could be normal or wheeze. I order to make this fold more balanced, we perform a Leave-Two-Out Cross-Validation (LTOCV), where each fold takes one sample for each class (wheeze and normal) in the validation data set, whilst the remaining samples are used to train the model. Hence, multiple runs are done by randomly choosing the validation samples. Since we have 17 wheeze and 9 normal cases, a bootstrap sample of 17 elements was made from the normal cases. Thus, we can carry out LTOCV with 17 folds.

Evaluation criteria: The traditional measures reported in different publications related with lung sound classification are sensitivity and specificity; being sensitivity the model capabilty to predict a present condition correctly and specificity its counterpart, i.e. the ability to reject a condition correctly whenever it is no present. The equations below show how these measures are calculated, where TP (True Positives), TN (True Negatives), FN (False Negatives) and FP (False Positives) are the outcome of a confusion matrix.

$$Sensitivity = \frac{TP}{TP + FN} \qquad Specificity = \frac{TN}{TN + FP} \qquad BAC = \frac{Sensitivity + Specificity}{2} \tag{7}$$

Being the presence of an illness condition, wheeze in this case, the positive class, it is usual to choose models with high sensitivity because they are highly accurrate to detect the positive condition. Meaning that it is more costly to reject a patient with presence of wheezes evaluated as normal in comparison to a normal condition evaluated as ill. However, high sensitivity models can involve a very poor specificity performance which is also undesirable. Thus, balanced accurracy (BAC) has a reliant value as evaluation measure for classification models in a medical context since it averages the sensitivity and specificity outcome. Hence, BAC was choosen as evaluation criteria for all methods, even so, sensitivity and specificity measures are also reported. The results in Table 1 are the average over all 17 folds in LTOCV. A 95% confidence interval based on a t-test with 17 degrees of freedom for all measures is also provided.

Baseline and proposed methods: Our proposed approach was evaluated by combining each state-of-the-art feature extraction method described in Subsection 2.1 with the SVM classifier described in Subsection 2.2 resulting in three different strategies: Fourier Transform with C-weighted SVM (FT+SVM), Wavelet Packet Transform with C-weighted SVM (WPT+SVM) and Mel-Frequency Cepstral Coefficients with C-weighted SVM (MFCC+SVM). Linear, polynomial and Gaussian kernels were evaluated but Gaussian yielded the best results. After perform exploration of parameters into a 10 based log-scale grid search going from $1\text{x}10^{-4}$ to $1\text{x}10^{4}$, the best combination for the Gaussian kernel per each feature was: $C = 1$, $\gamma = 0.1$ for MFCC, $C = 1$, $\gamma = 1$ for FT and $C = 0.1$, $\gamma = 0.1$ for WPT. For comparison, the best methods from the state of the art for lung sound classification were used as baselines.[3–5] In,[5] 24 MFCC were combined by a Gaussian mixture model for 8 Gaussians, In,[3] the number of Gaussians was varied from 1 to 20 using 7 MFCC where 9, 10, and 11 Gaussians reported the best results (MFCC+GMM). In[4,5] a Wavelet Packet Transform with 7-level decomposition and taking different sets of Wavelet Transform features (19[5] and 5,10,15[4]) was combined with a Multi-layer Perceptron with 30 hidden units were used as classifier (WPT+MLP). The parameters for each baseline method were set according to the ones reported in those papers.

3.2.1 Results

Table 1 presents the performance results of the baseline methods, WPT+MLP and MFCC+GMM, and the proposed approach C-weighted SVM combined with each type of feature (FT, WPT, MFCC). The best performance is achieved by MFCC+SVM with Gaussian kernel obtaining 82.1% of BAC. This result outperforms both best baseline methods MFCC+GMM and WPT+MLP. Although WPT+MLP yield better specificity than the other evaluated methods, it carries a low sensitivity which is a more valuable measure for the target task in this work. Interestingly, the second best results is produced by our proposed strategy using the WPT instead of MFCC. Comparing the same features, MFCC and WPT, with our suggested C-weighted SVM classifier against the GMM and MLP used in the literature, our approach outperforms these previous results. This suggests that using the same features is possible to obtain better performance by applying a better machine learning classifier. In both cases MFCC features get better results than WPT, whereas FT features is the less appropriate to capture the relevant information in wheeze detection.

Table 1. Classification performance for automatic detection of wheeze lung sounds in terms of balanced accuracy (BAC), sensitivity and specificity for each of one of the strategies.

	BAC	Sensitivity	Specificity
MFCC+SVM (Ours)	**0.821 ±0.07**	0.815 ±0.10	0.826 ±0.07
WPT+SVM (Ours)	**0.818 ±0.05**	0.944 ±0.06	0.692 ±0.09
FT+SVM (Ours)	0.791 ±0.07	0.760 ±0.11	0.823 ±0.07
MFCC+GMM[3,5]	0.807 ±0.06	0.803 ±0.10	0.811 ±0.06
WPT+MLP[4,5]	0.752 ±0.08	0.650 ±0.16	0.853 ±0.05

It is worth to say that C-weighted SVM offers another advantage over other methods by being a powerful and highly efficient method. Since the margin of decision for classifictaion is based on support vectors which are built on a subset of instances of training set, the training process for a SVM is faster than a neural networks. This yields a faster training stage for SVM taking less than a minute to train and evaluate all 17 LTOCV folds. All three feature extraction methods were performed using MATLAB R2010b. On the other hand, the proposed C-Weighted SVM as well as baseline methods were implemented in Python 2.7.3 using different Machine Learning modules like Pybrain[9] for MultiLayer Perceptron and scikit-learn[7] for GMM and SVM running on a dual-core laptop with 2.4GHz CPU.

4. NOVEL CONTRIBUTIONS

The main contribution of this work is a more accurate automatic wheeze detection method based on a C-weighted SVM classifier applied for the first time in this domain. Our experimental evaluation shows that using better machine learning algorithm for pattern recognition it is possible to achieve better performance with the same feature extraction methods. Whereas SVM classifiers have been used to detect cracks and squawks lung sounds, these were not used for wheeze detection, thus accentuating the contribution of this paper in the CORSA framework.

5. CONCLUSIONS

We applied for the very first time a C-weighted SVM algorithm as classifier for automatic detection of wheeze sounds over a variety of the best feature extraction methods in the CORSA framework. Our experimental results show that using a better classifier for this problem yields in a performance improvement independently of the feature extraction method selected. The best results were achieved by combining MFCC and C-weighted SVM with Gaussian kernel obtaining 82.1% of BAC. Additionally, we proved the high efficiency of SVM algorithms to train the whole setup of 17 LTOCV folds, which include 1188 segments, in less than a minute. In similar way, prediction took less than a second to evaluate each one of the folds. Future work includes feature combination strategies to train kernel-based models for automatic classification of more types of lung sounds like ronchi, fine and coarse crackles.

ACKNOWLEDGMENTS

Sosa and Cruz-Roa thank for young researcher and doctoral grants from Administrative Department of Science, Technology and Innovation of Colombia (Colciencias) numbers 566/2012 and 528/2011 respectively.

REFERENCES

1. Organization, W. H. et al., "The top 10 causes of death: The 10 leading causes of death in the world, 2000 and 2011," (2013).
2. Sovijärvi, A., Vanderschoot, J., and Earis, J., [*Computerized Respiratory Sound Analysis (CORSA): Recommended Standards for Terms and Techniques: ERS Task Force Report*], Munksgaard (2000).
3. Mayorga, P., Druzgalski, C., Morelos, R., Gonzalez, O., and Vidales, J., "Acoustics based assessment of respiratory diseases using gmm classification," in [*Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*], 6312–6316, IEEE (2010).
4. Hashemi, A., Arabalibiek, H., and Agin, K., "Classification of wheeze sounds using wavelets and neural networks," in [*International Conference on Biomedical Engineering and Technology. Singapore: IACSIT Press*], 127–131 (2011).
5. Bahoura, M., "Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes," *Computers in biology and medicine* **39**(9), 824–843 (2009).
6. Mallat, S., [*A wavelet tour of signal processing*], Academic press (1999).
7. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E., "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research* **12**, 2825–2830 (2011).
8. Ward, J. J., "Rale lung sounds 3.1 professional edition," *Respiratory Care* **50**(10), 1385–1388 (2005).
9. Schaul, T., Bayer, J., Wierstra, D., Sun, Y., Felder, M., Sehnke, F., Rückstieß, T., and Schmidhuber, J., "Py-Brain," *Journal of Machine Learning Research* **11**, 743–746 (2010).