Personalized Detection of Explosive Cough Events in Patients With Pulmonary Disease

Bruno M. Rocha	Diogo Pessoa	Alda Marques	Paulo Carvalho	Rui Pedro Paiva
Univ. Coimbra	*Univ. Coimbra*	*Univ. Aveiro*	*Univ. Coimbra*	*Univ. Coimbra*
CISUC	*CISUC*	*Lab3R*	*CISUC*	*CISUC*
DEI	*DEI*	*ESSUA, iBiMED*	*DEI*	*DEI*
Coimbra, Portugal	Coimbra, Portugal	Aveiro, Portugal	Coimbra, Portugal	Coimbra, Portugal
bmrocha@dei.uc.pt	dpessoa@dei.uc.pt	amarques@ua.pt	carvalho@dei.uc.pt	ruipedro@dei.uc.pt

***Abstract*— We present a new method for the discrimination of explosive cough events based on a combination of spectral and pitch-related features. The method was tested on 16 distinct partitions of a database with 9 patients. After a pre-processing stage where non-relevant segments were discarded, we have extracted eight features from each of the other segments and have fed them to the classifiers. Four types of algorithms were implemented to classify the events, with Bayesian classifiers achieving the best performance. Preliminary results showed that performance increased when the analysis was performed on individual subjects and when specific sensor locations were chosen. These results demonstrate that personalizing the analysis is a promising approach and shed some light on where to put sensors when automatic analysis is performed in the future.**

Keywords*— *cough detection; sound analysis; biomedical signal processing

I. Introduction

Cough is the most common symptom for which patients seek medical advice [1]. It occurs naturally as a defense mechanism to protect the respiratory tract and it is one of the most common symptoms of pulmonary disease [2]. It can be characterized by an initial contraction of the expiratory muscles against a closed glottis, followed by a violent expiration as the glottis opens suddenly [3]. The cough sound can be split in three phases: an explosive phase, an intermediate period, and a voiced phase. Technological advances have enabled the development of automated and ambulatory cough monitors, but there are currently neither standardized methods for recording cough nor adequately validated, commercially available, and clinically acceptable cough monitors [4-5].

The main goal of this work was to design a method for the automatic recognition and counting of coughs solely from sound recordings, ideally removing the need for trained listeners. The secondary goal was to find whether subjects' individual characteristics and specific sensor locations influence the performance of the algorithms.

II. Material and Methods

A. Data Collection

Pulmonary signals were recorded at the General Hospital of Imathia - Health Unit of Naousa, Greece, which provided ethics permission to conduct the study. All participants were diagnosed with chronic obstructive pulmonary disease. Auscultation data were acquired with a 3M Littmann 3200 stethoscope sequentially in six different positions: four in the back and two in the front of the chest with each participant in a sitting position. Fig. 1 shows the locations on the chest wall that were used for the recordings. During the acquisition, the volunteers were asked to simulate cough and then to count from one to ten. The dataset is comprised of 54 mono 16-bit recordings (6 per subject) with a sample rate of 4000 Hz and mean duration of approximately 19 s. A detailed description of the dataset can be found in Table I. The physicians who supervised the data acquisition annotated the different events in the timeline and we assigned them to four classes: (1) explosive cough, (2) voiced cough, (3) speech, and (4) other, a class composed of background noises, body rubs, wheezes, crackles, laughter, throat clears, and other artifacts. Segments of cough and speech were the predominant events in this database.

TABLE I. Description of Database

Population/City	Naousa
Sampling Rate (Hz)	4000
Stethoscope	3M Littmann 3200
Number of Patients	9
Average Signal Duration (s)	19

B. Pre-processing

In the pre-processing stage, the audio signal was filtered using an 8th-order infinite impulse response (IIR) high-pass filter with 80 Hz (below the lower bound of the typical adult human voice [6]) as the half-power frequency, followed by normalization. Then, near-silent segments were discarded using the following process: given a threshold for length (40 ms) and another for amplitude (5%), segments whose length and amplitude were both below their respective thresholds were classified as near-silent and discarded, i.e., a segment was considered near-silent if its number of consecutive samples with absolute amplitude below 5% added up to less than 40 ms. These parameter settings were chosen because, following the approach of Drugman et al. [7], we have focussed on the detection of the explosive phase of cough. This phase is characteristic of the beginning of any cough event, while the intermediate phase is very similar to a forced expiration and the voiced phase is not present in all cough events. Fig. 2 shows the audio signal before and after pre-processing.

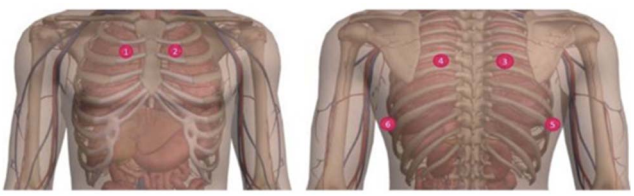

Fig. 1. Chest wall locations for the recording of respiratory sounds

978-1-7281-5200-4/20/$31.00 ©2020 IEEE

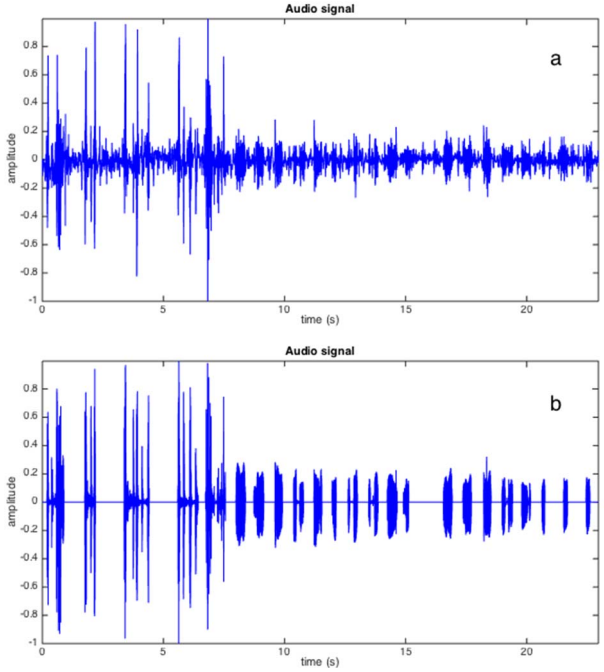

Fig. 2. Audio signal (a) before and (b) after pre-processing

C. Feature Extraction

Two types of features were extracted in this stage: global features (computed for each frame of the signal) and local features (computed for each relevant segment).

First, the magnitude spectrum was computed using the short time Fourier Transform (STFT) in frames of 50 ms with an overlap of 80%. An example of the magnitude spectrum of some events is shown in Fig. 3. Three global features were extracted from this magnitude spectrum: the spectral flux, calculated as the Euclidian distance between the magnitude spectrum of each successive frame; the first Mel-frequency cepstral coefficient (MFCC), which is a measure of the spectral shape of the sound; and the pitch inharmonicity, which estimates the number of partials that were not multiple of the fundamental frequency [8]. Then, for each non-silent segment, the peaks of the waveform envelope were computed and new segments with a duration of 100 ms were defined around the peaks. The MIR Toolbox [8] was used to perform these operations. Five local features were extracted at this stage: Mean and Maximum Spectral Flux, Mean and Maximum MFCC, and Mean Pitch Inharmonicity.

The computation of the other pitch features involved some additional steps. First, an 8th-order IIR low-pass filter with half-power frequency of 300 Hz was applied (the typical adult human voice fundamental frequency range, considering both sexes, is between 85 Hz and 255 Hz [6]). The magnitude spectrum of the low-pass filtered signal was then computed using the STFT and the peaks corresponding to the fundamental frequency at each frame were estimated. Finally, three local features were extracted: Pitch Coverage, the ratio of the number of frames where a fundamental frequency is detected to the total number of frames of each segment; Pitch Mean and Standard Deviation, the average and standard deviation of the estimated fundamental frequencies of each segment.

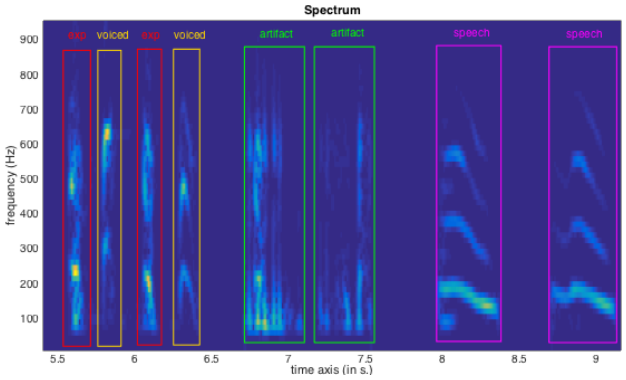

Fig. 3. Magnitude spectrum of eight segments of the following classes: explosive cough (red), voiced cough (yellow), artifact (green), and speech (pink)

D. Classification

To classify the audio signals we compared several classification algorithms from the WEKA data mining software [9]. Four types of classifiers were used in this work: Bayesian, Support Vector Machines (SVM), Propositional Rule Learner, and Bootstrap Aggregation. 10-fold cross validation was performed for each classification algorithm, and the experiment was repeated 10 times, with the average being reported.

1) Bayesian classifier: Naïve Bayes

This classifier is a kind of Bayesian network that is termed naive because it relies on two simplifying assumptions: predictive attributes are conditionally independent given the class; no hidden or latent attributes influence the prediction process [10]. Two implementations of this classifier were tested, which only differ in the method used for density estimation of continuous attributes: one uses a single Gaussian to estimate the density of each variable while the other uses a kernel method.

2) SVM: Sequential Minimal Optimization

Sequential Minimal Optimization (SMO) is an algorithm for training SVM that quickly solves the SVM Quadratic Programming (QP) problem by decomposing it into QP sub-problems [11]. Two kernel functions were used in this project: linear and radial basis function. For each kernel, the complexity parameter C assumes the values of 0.1, 1, and 10. This parameter allows one to trade off training error vs. model complexity. The higher the value for C is, the smaller the number of training errors become, leading to a behavior approaching that of a hard-margin SVM [12]. Data were standardized before all computations.

3) Propositional Rule Learner: RIPPER

Repeated Incremental Pruning to Produce Error Reduction (RIPPER) [13] is an optimization of the Incremental Reduced Error Pruning (IREP) rule-learning algorithm [14]. IREP integrates reduced error pruning with a separate-and-conquer algorithm. For each class, RIPPER grows rules in a greedy fashion and selects the conditions with highest information gain. Each rule is then pruned by deleting the conditions that maximize the function

$$v(Rule, PrunePos, PruneNeg) \equiv \frac{p + (N - n)}{P + N} \quad (1)$$

where P (respectively N) is the total number of examples in PrunePos (PruneNeg) and p (n) is the number of examples in

PrunePos (PruneNeg) covered by Rule. This process is repeated until no deletion improves the value of v. After generating the initial rule set, the rules are revised and optimized using a Maximum Description Length heuristic. Three versions of this classifier were tested, with the following number of optimization runs: 2, 5, and 10.

4) Bootstrap Aggregation: Bagging

Bagging predictors is a method for generating multiple versions of a predictor and using these to get an aggregated predictor [15]. This method is used to reduce the classifier variance. Bootstrapping the training set forms the multiple versions and the replicates are used as new learning sets. Bagging uses ten replicates of three base classifiers in this project: OneR [16], which uses the minimum-error attribute for prediction, discretizing numeric attributes, Naive Bayes with Gaussian fitting [10], and J48, an implementation of the C4.5 decision tree [17].

III. Evaluation

A. Datasets

The original database gave origin to 16 datasets: one for each sensor position (P1, P2, ... P6), one for each subject (S1, S2, ... S9), and one containing all recordings (SP). The number of events per class in each dataset is presented in Table II.

B. Evaluation metrics

Classification accuracy was not a good metric to use in this project as the datasets were not balanced. Additionally, there was only one relevant class; therefore, metrics that are useful in binary decision problems are presented for the positive class: the area under the Receiver Operator Characteristic (ROC) curve and the area under the Precision-Recall Curve (PRC). The relationship between these two curves is discussed thoroughly elsewhere [18].

C. Results

The results for the SP data set are shown in Table III.

Considering Naive Bayes with Gaussian fitting (NB Gaussian) as the baseline (it is the fastest algorithm of this group and its interpretability is high), only two other classifiers perform significantly better on this data set: NB with Kernel fitting (NB Kernel) and Bagging with J48 trees (Bagging J48).

TABLE II. Number of events per class in each dataset

Datasets	Cough	Voiced	Speech	Artifact
All	470	333	479	113
P1	75	59	68	11
P2	66	49	100	20
P3	94	71	86	11
P4	75	62	76	14
P5	78	43	83	21
P6	82	49	66	36
S1	73	32	75	15
S2	29	21	63	17
S3	47	42	63	7
S4	80	67	48	18
S5	37	39	62	15
S6	67	42	51	23
S7	36	34	52	4
S8	62	18	43	10
S9	39	38	22	4

TABLE III. Average ROC and PRC for the SP dataset; statistically significantly (SS) better results than baseline (*) in bold; SS worse results than baseline in italic

Algorithm	ROC	PRC
NB Gaussian*	0.88	0.77
NB Kernel	**0.91**	**0.82**
SMO Linear C=0.1	0.88	*0.71*
SMO Linear C=1	0.88	*0.71*
SMO Linear C=10	0.89	*0.71*
SMO RBF C=0.1	*0.84*	*0.63*
SMO RBF C=1	*0.87*	*0.68*
SMO RBF C=10	0.90	0.74
RIPPER opt=2	0.87	0.75
RIPPER opt=5	0.88	0.75
RIPPER opt=10	0.88	0.75
Bagging OneR	*0.85*	*0.68*
Bagging Naive Bayes	0.88	0.77
Bagging J48	**0.93**	**0.86**

Table IV presents the best results achieved for each sensor location with each type of classifier: NB Kernel, SMO RBF with C=10, RIPPER with 10 optimizations, and Bagging J48. The best performance was achieved with Bagging J48 for almost all positions, although the difference to NB Kernel was not statistically significant in any case. As expected, the best performance was achieved with the first sensor location, in the right side of the chest. If we consider explosive cough as the desired signal, this location is optimal because the signal-to-noise ratio is the highest possible and the possible interference from heart sound is smaller than in the left side.

Table V presents the best results achieved for each subject with each type of classifier. Areas under the ROC and PRC curves are both higher than 0.9 in seven patients when NB Kernel is used, and almost perfect performance was reached with subject 5. However, performance drops substantially with subjects 2 and 8. This drop in performance might be related to the inadequacy of the features for discriminating the sounds of this patient.

TABLE IV. Average ROC and PRC of each type of classifier on the sensor position datasets

Datasets	NB	SMO	RIPPER	Bagging
P1	0.94/0.91	0.91/0.78	0.86/0.71	0.95/0.92
P2	0.95/0.86	0.94/0.80	0.87/0.69	0.94/0.88
P3	0.91/0.86	0.89/0.76	0.84/0.68	0.92/0.87
P4	0.93/0.89	0.92/0.77	0.89/0.77	0.94/0.89
P5	0.92/0.86	0.90/0.76	0.86/0.77	0.93/0.88
P6	0.86/0.78	0.86/0.69	0.80/0.60	0.89/0.83

TABLE V. Average ROC and PRC of each type of classifier on the subject datasets

Datasets	NB	SMO	RIPPER	Bagging
S1	0.97/0.96	0.94/0.86	0.92/0.87	0.97/0.95
S2	0.92/0.79	0.89/0.68	0.88/0.66	0.91/0.79
S3	0.98/0.97	0.98/0.95	0.91/0.85	0.98/0.95
S4	0.96/0.93	0.93/0.85	0.90/0.81	0.95/0.91
S5	1.00/0.99	0.98/0.93	0.94/0.84	0.99/0.96
S6	0.95/0.91	0.94/0.85	0.87/0.74	0.93/0.88
S7	0.96/0.93	0.96/0.86	0.89/0.74	0.96/0.90
S8	0.83/0.80	0.79/0.68	0.82/0.73	0.81/0.76
S9	0.99/0.99	0.97/0.92	0.94/0.89	0.99/0.98

Fig. 4 plots Pitch Inharmonicity against Flux Mean and allows us to compare the discriminative power of these features in the S5 and S8 data sets. When feature selection was

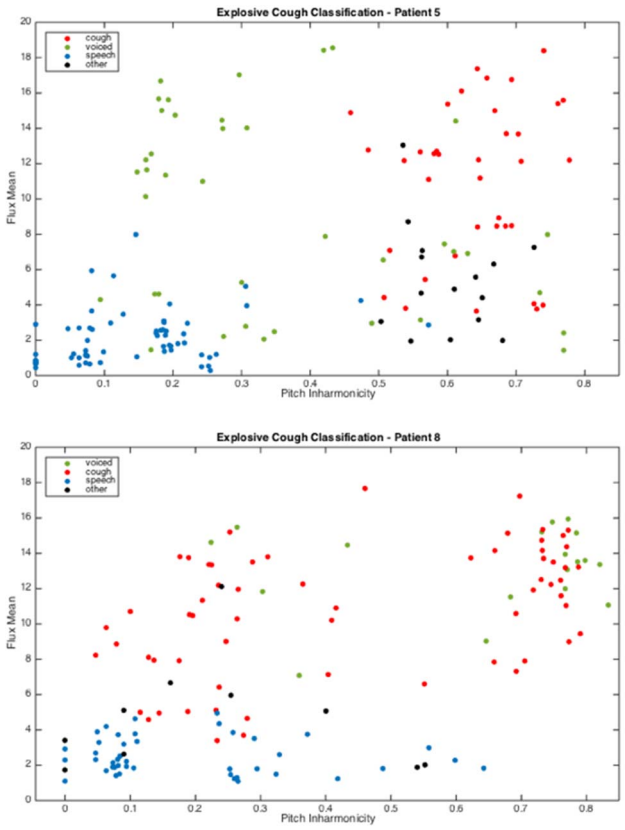

Fig. 4. Pitch Inharmonicity vs. Flux Mean in S5 and S8

performed using the ReliefFAttributeEval algorithm [19] in Weka, these two features were ranked in the top 5 features for every data set except S9. Fig. 4 demonstrates how difficult was to discriminate explosive cough in S8 with these features, as instances of this class are located everywhere, while in S5 they congregate around the right half of the plot.

IV. Conclusion

We presented a cough discrimination method based on a combination of spectral and pitch-related features. Four types of algorithms were used to classify segments of explosive cough against segments of three other classes: voiced cough, speech, and artifacts. The original database of 9 patients was partitioned in 16 different ways and the results demonstrated that it might be worth to follow the individualized approach for this problem. Performance improved substantially when each patient was analyzed separately. As high performance was not reached for two of the patients, in future work we intend to develop new features that easily adapt to each patient.

Acknowledgments

The authors would like to thank Luís Mendes, Evangelos Kaimakamis, Eleni Perantoni, and Ioannis Vogiatzis for the data acquisition. This work was partially supported by Fundação para a Ciência e Tecnologia (FCT) Ph.D. scholarship SFRH/BD/135686/2018, by iBiMED - UIDB/04501/2020, and by H2020-ICT-2018-2 programme WELMO under grant agreement 825572.

References

[1] Irwin, R., Boulet, L., Cloutier, M., Fuller, R., Gold, P., Hoffstein, V., & Prakash, U. (1998). Managing cough as a defense mechanism and as a symptom: a consensus panel report of the American College of Chest Physicians. CHEST Journal, 114(2_Supplement), 133S-181S.

[2] Korpáš, J., & Tomori, Z. (1979). Cough and other respiratory reflexes (pp. 81-104). S. Karger.

[3] Evans, J. N., & Jaeger, M. J. (1975). Mechanical aspects of coughing. Lung, 152(4), 253-257.

[4] Korpáš, J., Vrabec, M., Sadlonova, J., Salat, D., & Debreczeni, L. (2003). Analysis of the cough sound frequency in adults and children with bronchial asthma. Acta Physiologica Hungarica, 90(1), 27-34.

[5] Abaza, A., Day, J., Reynolds, J., Mahmoud, A., Goldsmith, W., McKinney, W., & Frazer, D. (2009). Classification of voluntary cough sound and airflow patterns for detecting abnormal pulmonary function. Cough, 5(1), 1.

[6] Baken, R. J., & Orlikoff, R. F. (2000). Clinical measurement of speech and voice. Cengage Learning.

[7] Drugman, T., Urbain, J., Bauwens, N., Chessini, R., Valderrama, C., Lebecque, P., & Dutoit, T. (2013). Objective study of sensor relevance for automatic cough detection. Biomedical and Health Informatics, IEEE Journal of, 17(3), 699-707.

[8] Lartillot, O., & Toiviainen, P. (2007, September). A Matlab toolbox for musical feature extraction from audio. In International Conference on Digital Audio Effects (pp. 237-244).

[9] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. ACM SIGKDD explorations newsletter, 11(1), 10-18.

[10] John, G. H., & Langley, P. (1995). Estimating continuous distributions in Bayesian classifiers. In Proceedings of the Eleventh conference on Uncertainty in artificial intelligence (pp. 338-345). Morgan Kaufmann Publishers Inc.

[11] Platt, J. C. (1999). Fast training of support vector machines using sequential minimal optimization. Advances in kernel methods, 185-208.

[12] Joachims, T. (2002). Learning to classify text using support vector machines: Methods, theory and algorithms. Kluwer Academic Publishers.

[13] Cohen, W. W. (1995). Fast effective rule induction. In Proceedings of the twelfth international conference on machine learning (pp. 115-123).

[14] Fürnkranz, J., & Widmer, G. (1994). Incremental reduced error pruning. InProceedings of the 11th International Conference on Machine Learning (ML-94) (pp. 70-77).

[15] Breiman, L. (1996). Bagging predictors. Machine learning, 24(2), 123-140.

[16] Holte, R. C. (1993). Very simple classification rules perform well on most commonly used datasets. Machine learning, 11(1), 63-90.

[17] Quinlan, J. R. (2014). C4.5: programs for machine learning. Elsevier.

[18] Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. In Proceedings of the 23rd international conference on Machine learning (pp. 233-240). ACM.

[19] Kononenko, I. (1994). Estimating attributes: analysis and extensions of RELIEF. In Machine Learning: ECML-94 (pp. 171-182). Springer Berlin Heidelberg.