Asthmatic Wheezes Detection - What Contributes the Most to the Role of MFCC in Classifiers Accuracy?

Mario Milicevic, Igor Mazic, and Mirjana Bonkovic

***Abstract*—** Asthma is one of the top five chronic diseases globally and the most common chronic disease among children. It is the most likely cause of recurrent wheezing in children, so computerized respiratory sound analysis is an important diagnostic aid. This research compares the efficiency of the classification algorithms applied both on signals available on the internet and signals recorded on children in real-life clinical settings. The paper proves that the features with logarithmic distribution of energy filter bank along the frequency domain embedded in MFCC, result in better wheezes recognition in an auscultatory breathing signal than spectral features and the similar energy filter bank features which do not have logarithmic distribution along the frequency domain. Furthermore, the paper demonstrates that the SVM classifier performs better than other classifiers applied on signals acquired under ideal and suboptimal conditions.

***Keywords*—** asthma, classification, MFCC, SVM.

I. Introduction

Asthma is a public health problem not just for high-income countries; it occurs in all countries regardless of the level of development. The World Health Organization (WHO) estimates a number of 235 million people currently suffering from asthma [1].

Asthma is considered to be the most common chronic disease among children in nearly all industrialized countries [2]. It is more prevalent in children with a family history of atopy, and symptoms and worsening thereof are frequently provoked by a wide range of triggers, which can include viral infections, indoor and outdoor allergens, exercise, tobacco smoke and poor air quality. A large number of infants and preschool children experience recurrent episodes of bronchial symptoms, especially wheezing and coughing, beginning at a few months of age. Since a clinical diagnosis of asthma can usually be made with certainty by the age of 5, early diagnosis, monitoring and treatment of respiratory symptoms are essential.

M. Milicevic is with the Department of Electrical Engineering and Computing, University of Dubrovnik, C.Carica 4, 20000 Dubrovnik, Croatia (phone: 385-20- 445766; e-mail: mario.milicevic@unidu.hr).

I. Mazic is with the Department of Electrical Engineering and Computing, University of Dubrovnik, C.Carica 4, 20000 Dubrovnik, Croatia (e-mail: igor.mazic@unidu.hr).

M. Bonkovic is with the Faculty of Electrical Engineering Mechanical Engineering and Naval Architecture, University of Split, Rudjera Boskovica 32, 21000 Split, Croatia (e-mail: mirjana.bonkovic@fesb.hr).

Lung auscultation is helpful in providing information concerning the patient’s respiratory function. The presence of wheezing in infants is used as an important parameter in assessing the predisposition to asthma [3]. History of repeated episodes of wheezing is a symptom universally accepted as the starting point for asthma diagnosis in children [4]. The required number of such episodes is generally unspecified, although an arbitrary number of three or more has been proposed. Typical symptom patterns are significant in establishing the diagnosis.

A wheeze can be described as an unintentional and continuous sound [5]. Acoustically, it is characterized by periodic waveforms with a dominant frequency usually over 100 Hz (or 400Hz [6]) and with duration of $\geq$100 ms. Wheezes are usually associated with airways that are obstructed due to various causes. Wheezes with a single peak or with the harmonics of a single basal peak are called monophonic wheezes, while those with variable peaks that differ in harmonics are called polyphonic wheezes [7].

Recently developed computer based respiratory sound analysis methods serve as a powerful tool to diagnose the whole spectrum of disorders and abnormalities in the lungs, including asthma.

However, special attention should be made when this kind of analysis is applied to signals recorded on children. Due to their lack of cooperation, such signals have a number of artifacts and the usual methods of analysis often do not provide good results [8]. In this paper the efficiency of the classification algorithms applied both on signals available on the internet and signals recorded on children in real-life clinical settings are compared.

The paper demonstrates that the SVM classifier performs better than other classifiers applied on signals acquired under ideal and suboptimal conditions. Additionally, it proves that the features with logarithmic distribution of energy filter bank along the frequency domain embedded in MFCC result in better wheeze recognition in an auscultatory breathing signal than spectral features and the similar energy filter bank features which do not have logarithmic distribution along the frequency domain. In the following section the problem is explained in detail and the possible risks of research goals are stated. Section 3 describes applied measuring procedures, signal pre-processing techniques and the classification strategy. Same section presents the experimental results, while Section 4 concludes the paper.

II. Problem Formulation

Phonopneumograms (acoustic breathing records) depend on anatomical and physiological parameters such as sex, age, type and stage of disease. Digital methods of collecting, processing and analyzing phonopneumograms have been in wide use for more than 30 years. Measuring systems consist, among others, of transducers that are put on the chest or trachea, which then collect acoustic signals during breathing.

Several factors affect the results of auscultation signal analysis and make comparing the research results obtained at various institutions more difficult [9] due to the differences in age and corpulence of the patient, air volume changes in the lungs, location of sound capturing, breathing flow, position of the patient and characteristics of the measurement equipment.

Differences due to age are all the more visible with infants. Audible respiratory sounds in early childhood have acoustic characteristics which are recognizably different from those generally heard in adults. Therefore, Mazic et al. [10] propose to use more objective methods to automatically detect wheezing in asthmatics infants, during forced breathing.

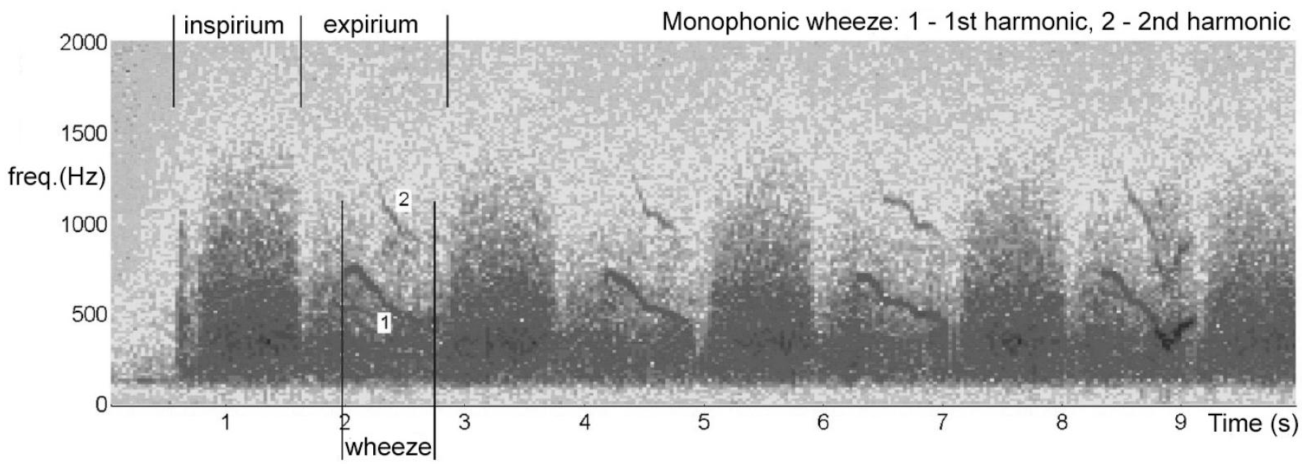

Fig. 1 2D spectrogram of the signal from INT dataset

Many methods were used by previous researchers during the past three decades to process the lung sounds for detecting wheezing [11]. Various types of extracted features have been used, such as the time-frequency spectrum, entropy, Mel Frequency Cepstral Coefficients (MFCC), power spectral density (PSD), standard deviation (SD), Peak Frequency (FP), ske Most authors used phonopneumograms to automatically detect wheezing from different internet databases (INT) or media [12] [13] [14], the purpose of which is first and foremost education, so the data about the measuring system, transducers, position of the measuring point, the age of the subject and breathing technique were mostly not published. A 2D spectrogram of such a signal [12] is shown on Fig. 1.

It is clear from the spectrogram that inspiration lasts almost as long as expiration in which wheezing is present. The signal to noise ratio (S/N), breathing regularity, the expiration/inspiration time ratio, as well as the presence of wheezing in all expirations point to the conclusion that this is a state of controlled forced breathing of an adult in ideal laboratory conditions. It is not possible to achieve these conditions with children under the age of 6.

In this paper, we compare the predictive abilities of models built from publicly available phonopneumograms (internet and CDs) to those built from phonopneumograms recorded in the Dubrovnik General Hospital (DGH) in realistic, suboptimal conditions with children aged from one to six.

Phonopneumograms recorded with children contain not only muscular and cardiovascular sounds, but also many physiological and non-physiological artifacts, such as sounds which are results of forced breathing, or stridor and wheezing which do not originate in the bronchia, but are the consequence of infections in the upper part of the respiratory system, which is a common occurrence with children of that age. All these signals are superimposed in the transducer, which only adds to the difficulty of their recognition and classification. Additionally, to ensure sufficient acoustic power for the respiratory sound, the children were usually encouraged to perform forced breathing, which often resulted in specific physiological artifacts such as inspiratory stridor, which sounds similar to asthmatic wheezes.

It can be expected that recording lung sounds in a noisy hospital versus laboratories under controlled conditions can demand more rigorous preprocessing techniques to combat the noise and other artifacts present in the acoustic signal. Also, there is the possibility that the classification algorithms are less effective.wness, kurtosis, etc. There is no consensus on which features are the best to be extracted, because the final system performance and classification accuracy are a consequence of different signal processing and classification techniques applied.

III. Modelling and Classification

Phonopneumograms from the above mentioned internet databases included 1026 samples (signal segments) of wheezes, and 1374 samples of breathing without wheezes. Each signal segment lasts 100 ms. The signal segments were obtained using 50% overlapping windows along the distinctive signal sequences.

The measuring equipment for DGH samples consists of a transducer, 4m long microphone cable, preamplifier, stable 5V source, and a personal computer with an integrated audio card [15]. The resolution of the AD/DA converter is 16 bits, signal-to-noise ratio (SNR) is 90 dB, and the total harmonic distortion at 1 VRMS was 0.01%. Sampling rate is 8 KHz. During a few weeks 369 samples (signal segments) of wheezes were recorded, and 495 samples of breathing without wheezes.

Fig. 2. shows a 3D spectrogram of a three-year-old child breathing. During forced breathing an inspiratory stridor was also recorded. To reduce the impact of cardiovascular and muscular noise, phonopneumograms were first filtered with the Yule–Walker 50th-order high pass filter, with the lower cut-off frequency of 100 Hz. Then, STFT is performed with 50% overlapping Hamming window, using 256 samples, which corresponds to 32 ms frame size.

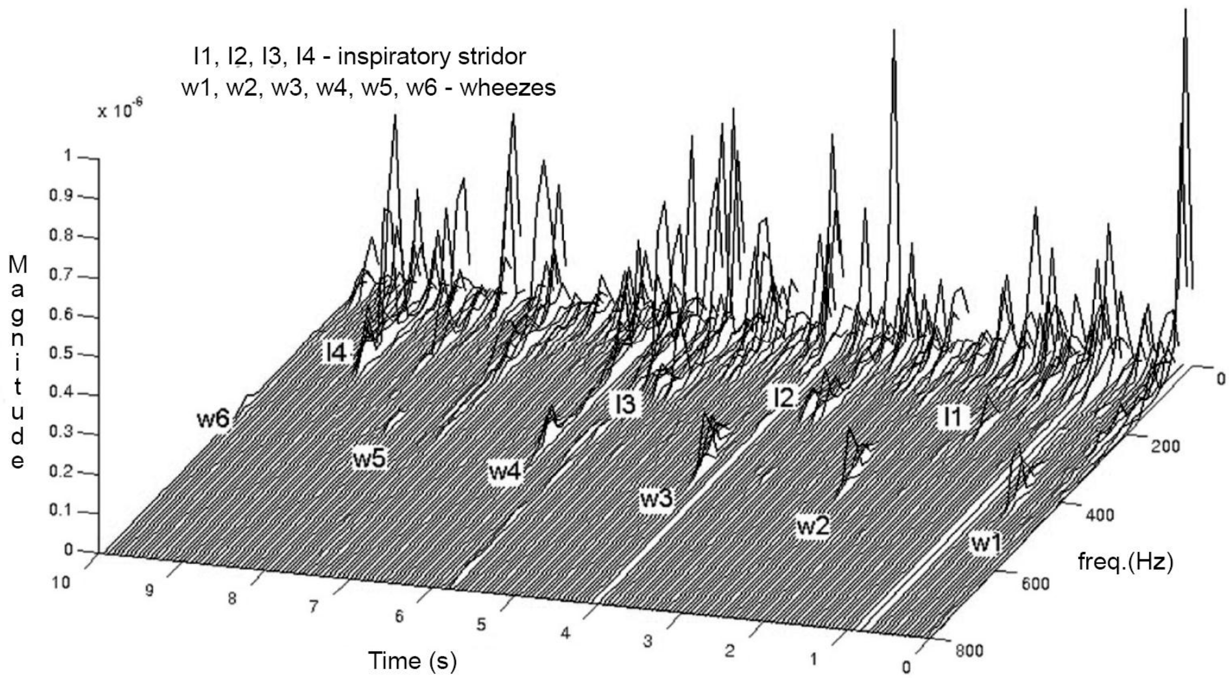

Fig. 2 3D spectrogram of a three-year-old's breathing

Fig. 3. shows a 2D spectrogram of the same phonopneumogram, where the acoustic power of the inspiratory stridor and wheezing within the same order of magnitude can be seen, while the 3D spectrogram shows that the inspiratory stridor appears at lower frequencies than the wheezing, which does not always appear to be the case. Researchers show that by using MFCC as features [16] [17] 18], wheezing detection can achieve an accuracy higher than 95%. There is no standard number of MFCCs for recognizing the lung sounds. For both dataset sources (INT, DGH) we experimented to obtain the optimal number of MFCCs resulting in the maximum classification accuracy. For the INT signals, we also investigated previous works that are directly related to this topic [19]. Finally, we used 15 MFCCs for the INT signals in frequency range from 100 Hz to 1500 Hz, and 12 MFCCs for our DGH signals in frequency range from 100 Hz to 1000 Hz.

For comparison purposes, we also used standard statistical features computed from spectral components calculated using FFT : Renyi entropy, Kurtosis, Spectral Flatness (SF), Skewness, Mean Crossing Irregularity (MCI), Standard Deviation (SD) and f50/f90 ratio [20].

The models are built based on several known machine learning methods: Support Vector Machine (SVM) [21], k-nearest neighbor algorithm (k-NN) [22], Neural Network (NN) [23], Random Forests (RF) [24], Logistic Regression (LR) [25], Naive Bayes (NB) [26].

For the aforementioned machine learning methods and both datasets, an optimization of important parameters which influence the model build was made. For some algorithms, this is a very important step, such as the SVM with RBF kernel, for which it is important to set the C and gamma parameters. This is why a grid search using cross-validation was used to find optimal values in a parameter space [27].

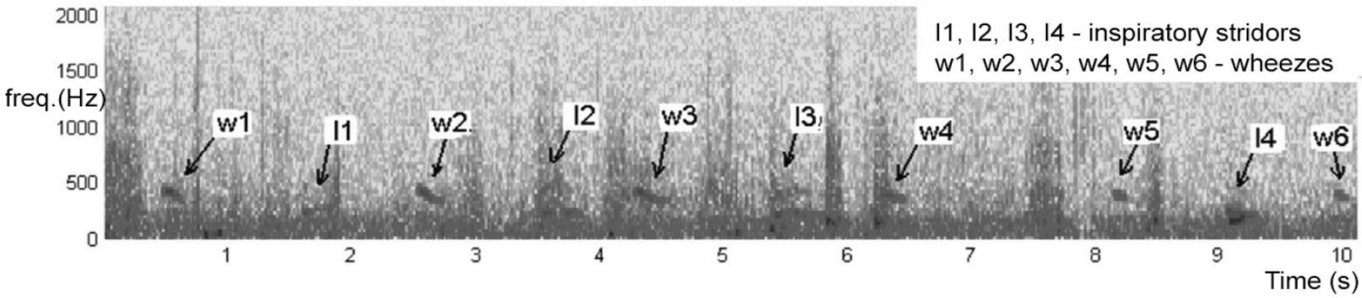

Fig. 3 2D spectrogram of a three-year-old's breathing

A. Results

Fig. 4. shows the results of the grid search procedure used to determine the optimal gamma and C parameters for the SVM algorithm (RBF kernel). One can see that it is very important to pay attention to the choice of values for the mentioned parameters for the SVM algorithm. Despite this, many papers comparing the SVM algorithm to other algorithms can be found which have no clear methodology of the parameter optimization.

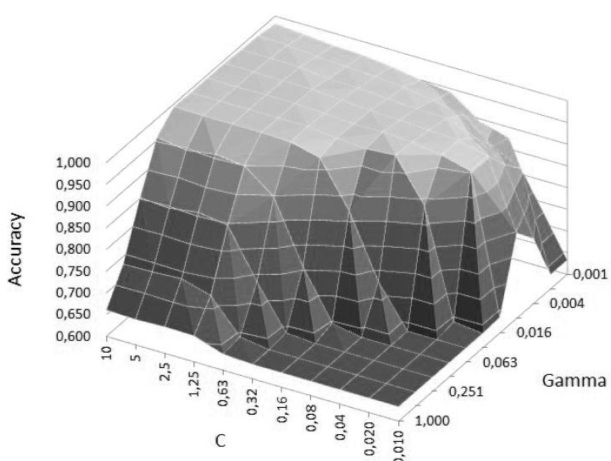

Fig. 4 Results of the grid search procedure

Fig. 5. shows the results of the classification for both datasets based on 7 statistical features. The best results (accuracy 93.62% for INT signals, or 91.77% for DGH signals) were achieved with the Neural network algorithm with 2 hidden layers. The SVM and k-NN algorithms, pointed out in numerous articles as a good choice for pulmonary acoustic signals classification, scored somewhat lesser results here (accuracy between 80% and 85%). Looking at all 6 algorithms, one can conclude that the classification is only marginally more successful with internet data.

On the other hand, Fig. 6. shows the classification results for both datasets based on MFCC features. The results are, as expected, much better, and the most successful classifiers are SVM and k-NN, which had the accuracy of 99%. At the same time, all classifiers have achieved somewhat better results for DGH signals.

Additionally, Fig. 7. shows the performance of binary classifiers for DGH signals (MFCC features) in the form of Receiver Operating Characteristic (ROC) curves [28]. By analyzing the shown curves, as well as the Area Under the Curve (AUC), it can be seen that the best results are achieved with the SVM algorithm.

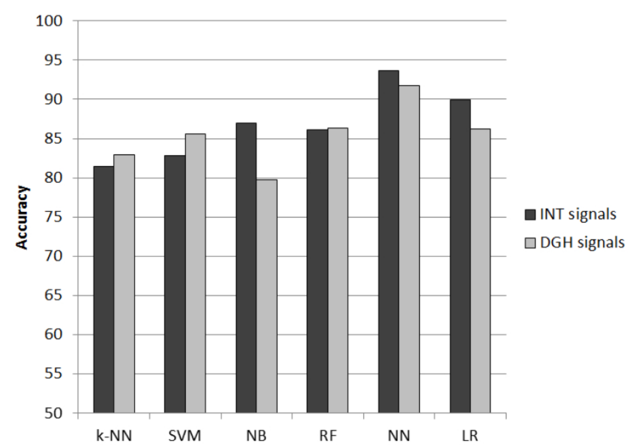

Fig. 5 Classification accuracy for INT/DGH signals represented with 7 statistical features

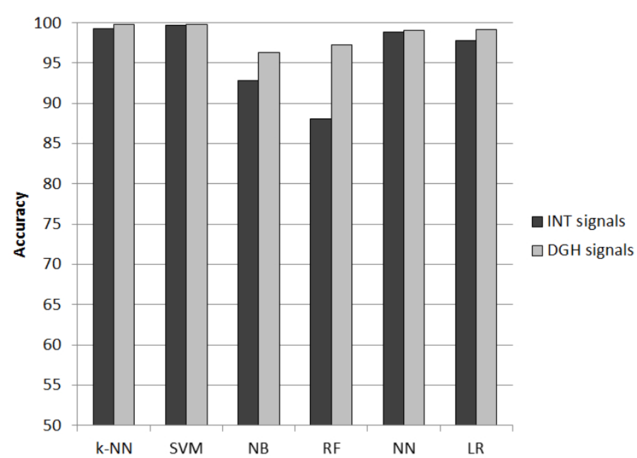

Fig. 6 Classification accuracy for INT/DGH signals represented with MFCC features

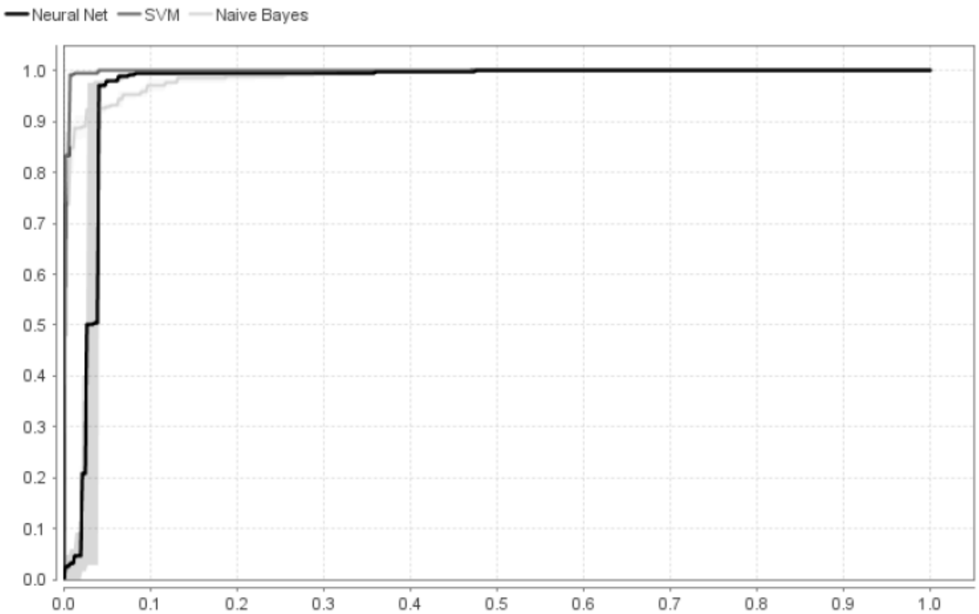

Fig. 7 ROC curves for binary classifiers (DGH signals)

B. Comparison of classification performance for different energy filter bank features

High value evaluation measures for wheeze classification using MFCC justified our expectations, with feature-based models providing good wheeze recognition. However, in addition, we were interested in which of these feature components has the greatest influence on the results of classification. Therefore, a different experiment was performed in which the classification efficacy was compared using three different features, whose values were defined through the energies of the respective frequency bands.

The classification obtained with these features was compared to the classification results obtained with the features described for MFCC. The three new features include (Fig. 8): filter based energy coefficients - $FBEC_{MEL}$, energies in evenly distributed overlapping "db4" filters using wavelet packet transform - $FBEC_{WPT}$, and energies in evenly distributed non-overlapping rectangular filters - $FBEC_{R}$. Below are the values of the four aforementioned features, and at the end of this subsection the conclusion will emphasize the advantages of specific features based on MFCC.

$$ACC = \frac{1}{k}\sum_{i=1}^{k}\frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}. \qquad (1)$$

The accuracies of wheezing recognition (1) achieved with the SVM classifier and multiple sets of features were compared using standard k-fold (k=10) cross-validation [29].

TP denotes the number of true positives, TN the number of true negatives, FP the number of false positives, and FN the number of false negatives.

C. Analysis using Mel Filter Bank Energy Coefficients – $FBEC_{MEL}$

When, instead of using the MFCC, we used the coefficients obtained at the end of step 4 (see Fig. 8), FBE (various filter bank energies), the results in Table I were gained. Comparing the first three rows of Table I (MFCC features) with rows 4, 5 and 6 ($FBEC_{MEL}$ features), it is evident that the results obtained with MFCC are better than the results achieved with $FBEC_{MEL}$. The results are also shown in Fig. 9. The difference in the calculation of MFCC and $FBEC_{MEL}$ is that logarithms and the DCT onto power signal are not applied in the mel filters in the $FBEC_{MEL}$ calculation. The output of DCT represents a cepstrum coefficient.

Table I Classification accuracy for various filter bank energy filters

Row no.	Features	Accuracy
1	32 MFCC	98.50
2	16 MFCC	98.95
3	8 MFCC	97.82
4	32 $FBEC_{MEL}$	98.00
5	16 $FBEC_{MEL}$	97.84
6	8 $FBEC_{MEL}$	97.11
7	32 $FBEC_{WPT}$	97.04
8	16 $FBEC_{WPT}$	97.19
9	8 $FBEC_{WPT}$	94.49
10	32 $FBEC_{R}$	93.33
11	16 $FBEC_{R}$	93.70
12	8 $FBEC_{R}$	93.81
13	8 $FBEC_{DWT}$	86.58

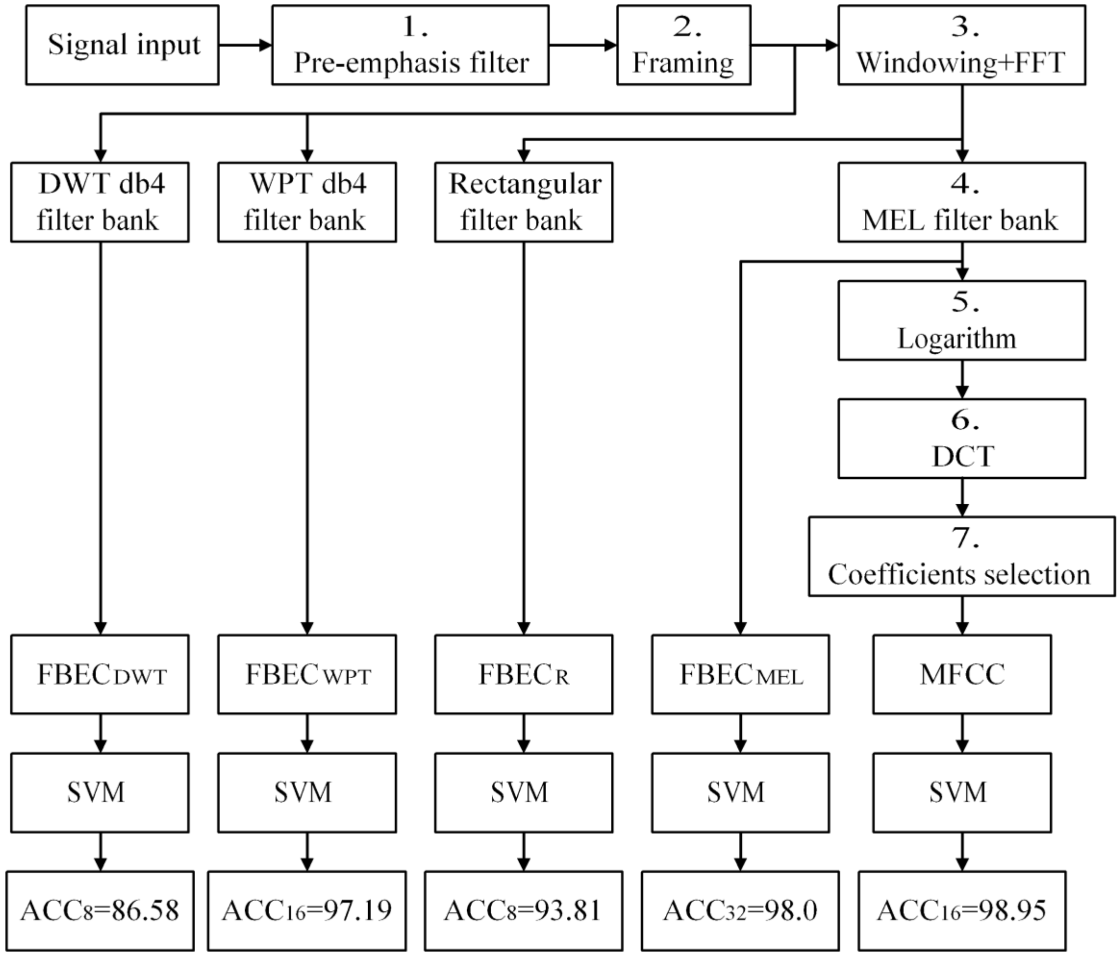

Fig. 8 The algorithm scheme obtaining the energy filter bank features used for classification efficacy comparisons

D. Analysis using features derived from the energies of rectangular and WPT filters

$FBEC_{WPT}$ were obtained using the third, fourth and fifth levels of decomposition with a "db4" wavelet. Generally, each level results in 2^{level} coefficients. Therefore, using an appropriate scale function and the "db4" wavelet, 8, 16 and 32 coefficients were obtained, from which coefficients of lowpass L and highpass H filters were obtained. Summing the squares of sequences at the output of the appropriate decomposition level, the energy contained in the WPT filters was obtained, i.e., $FBEC_{WPT}$.

From Table I (rows 7, 8 and 9) and Fig. 8, it is evident that the highest overall accuracy of 97.19% occurs with 16 WPT filters. The worst results were obtained using evenly distributed energies in non-overlapping rectangular filters, shown in rows 10, 11 and 12 in Table I, and appropriate bars in Fig. 9.

From Fig. 9 it is clear that there are three key steps to obtaining high classification accuracy. The first step is calculating energies in logarithmically distributed overlapping triangular filters. The second step is using cepstrum coefficients as the output of DCT, and the third step assumes improvements achieved by optimizing the number of coefficients used (filter banks).

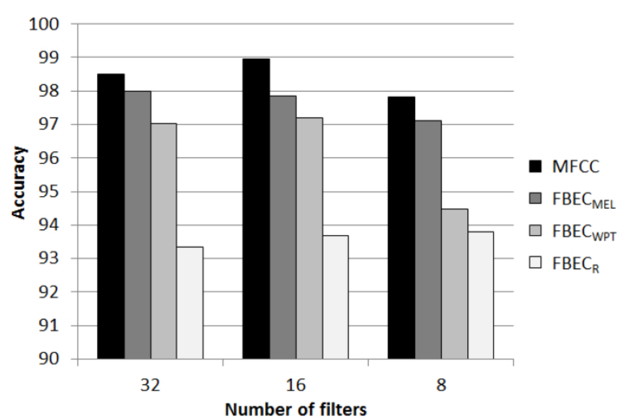

Fig. 9 Comparison of classification accuracy for different energy filter bank features calculated with appropriate numbers of filters

To confirm the importance of logarithmic distribution along the frequency axis of filter banks, the energy distributed along the frequency bands defined by discrete wavelet transformations was calculated. Filter banks were determined by iteratively dividing the low-frequency bands into two equal parts. The last row of Table I indicates that the classification accuracy is the worst using 8 $FBEC_{DWT}$ coefficients.

IV. Conclusion

The results of wheeze detection based on signals available on the internet were compared to the signals recorded in the Dubrovnik General Hospital. The signals were recorded using equipment comprised of standard components, with children in realistic conditions, including the effects of ambient sounds, cardiovascular and muscular noise, and other physiological and non-physiological artifacts. Different pattern recognition methods were used to classify both datasets of respiratory sounds into normal and wheeze classes. The experiments show that, by properly filtering and preprocessing the entry data, and using MFCC features, signals recorded in suboptimal conditions can achieve good results, very similar to those collected from the internet. That is an important prerequisite for the construction of a low cost automated system for monitoring asthma, based on a mobile device and the appropriate application, in which raw data from a transducer is processed and analyzed.

We also compared classification accuracy of the three new features ($FBEC_{MEL}$, $FBEC_{WPT}$ and $FBEC_{R}$) with the results obtained with the features described for MFCC. Our results show that the features with logarithmic distribution of energy filter bank along the frequency domain embedded in MFCC result in better asthmatic wheezes recognition than spectral features and the similar energy filter bank features, which do not have logarithmic distribution along the frequency domain. Having in mind that the human auditory system has logarithmic characteristics, the performed experiment clearly shows that MFCCs represent features adapted to the auditory system. Therefore, the artificial system recognized precisely the pattern that people hear, like most similar patterns with which the system was trained. Although the superiority of the MFCC coefficients is well known from the literature, our experiment presented in this paper shows what contributes the most to MFCC importance.

References

[1] H. J. Zar, and T. W. Ferkol, "The global burden of respiratory disease impact on child health", *Pediatric pulmonology* 49(5), 2014, pp. 430–434.

[2] L. Bacharier, A. Boner, K.-H. Carlsen, P. Eigenmann, T. Frischer, M. Gotz, P. Helms, J. Hunt, A. Liu, N. Papadopoulos, et al., "Diagnosis and treatment of asthma in childhood: a practall consensus repor"t, *Allergy* 63(1), 2008, pp 5–34. DOI: 10.1111/j.1398-9995.2007.01586.x

[3] F. D. Martinez, A. L. Wright, L. M. Taussig, C. J. Holberg, M. Halonen, W. J. Morgan, "Asthma and wheezing in the first six years of life", *New England Journal of Medicine*, 332(3), 1995, pp. 133-138. DOI: 10.1056/NEJM199501193320301

[4] N. Papadopoulos, H. Arakawa, K.-H. Carlsen, A. Custovic, J. Gern, R. Lemanske, P. Souef, et al., "International consensus on (ICON) pediatric asthma", *Allergy* 67(8), 2012, pp. 976–997., DOI:10.1111/j.1398-9995.2012.02865.x

[5] S. Reichert, R. Gass, C. Brandt, and E. Andres, "Analysis of respiratory sounds: state of the art", *Clinical Medicine: Circulatory, Respiratory, and Pulmonary Medicine*, 2008, 2, pp. 45-58.

[6] M. Bahoura, and X. Lu, "Separation of crackles from vesicular sounds using wavelet packet transform", *Acoustics, Speech and Signal Processing*, 2006., DOI: 10.1109/ ICASSP.2006.1660533

[7] N. Nagasaka, "Lung Sounds in Bronchial Asthma", *Allergology International*, 2012, 61, pp. 353-363, DOI: 10.2332/allergolint.12-RAI-0449

[8] M. Rossi, A. R. A. Sovijarvi, P. Piirila, L. Vannuccini, F. Dalmasso, and J. Vanderschoot. "Environmental and subject conditions and breathing manoeuvres for respiratory sound recordings", *European Respiratory Review 10*, 77, 2000, pp. 611-615.

[9] J. E. Earis, B. M. Cheetham, "Current methods for computerized respiratory sound analysis", *Eur. Respir. Rev.*, 2000., 10, pp. 586–90.

[10] I. Mazic, S. Sovilj and R. Magjarevic, "Analysis of respiratory sounds in asthmatic infants", *Measurement Science Review*, 2003., 3, pp. 11–21.

[11] S.M. Shaharum, K. Sundaraj, and R. Palaniappan, "A survey on automated wheeze detection systems for asthmatic patients", *Bosnian Journal of Basic Medical Sciences*, 2012. ,12(4), pp. 249-55.

[12] FAEMSE: Mild Expiratory Wheeze Lung Sound, http://www.faemse.org/downloads.html (05.06.2015.)

[13] W. Diane, *Heart & Lung Sounds Reference Library*, http://www.pesihealthcare.com (03.06.2015.)

[14] *Evaluating heart & breath sounds. Nursing know-how*, Wolters Kluwer/Lippincott Williams & Wilkins, 2009

[15] I. Mazic, M. Bonkovic, and B. Dzaja, "Two-Level Coarse-to-Fine Classification Algorithm for Asthma Wheezing Recognition in Children's Respiratory Sounds", *Biomedical Signal Processing and Control*, 21, 2015, pp. 105-118.

[16] R. Palaniappan, K. Sundaraj, and S. Sundaraj, "A comparative study of the svm and k-nn machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals", *BMC Bioinformatics*, 15, 2014, DOI: 10.1186/1471-2105-15-223

[17] M. Bahoura, and C. Pelletier, "Respiratory sounds classification using cepstral analysis and Gaussian mixture models", *26th Annual Conference of the IEEE EMBS*, San Francisco, 2004, DOI: 10.1109/IEMBS.2004.1403077

[18] M. Bahoura, and C. Pelletier, "Respiratory sounds classification using Gaussian mixture models", *Canadian Conference on Electrical and Computer Engineering*, 2004, Niagara Falls, DOI: 10.1109/CCECE.2004.1349639

[19] M. Bahoura, "Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes", *Computers in biology and medicine* 39, 9, 2009, pp. 824-843.

[20] R. Palaniappan, K. Sundaraj, N. U. Ahamed, A. Arjunan, S. Sundaraj, "Computer-based respiratory sound analysis: a systematic review", *IETE Technical Review*, 30 (3), pp. 248-256, DOI: 10.4103/0256-4602.113524

[21] B. Schölkopf, and A. Smola, *Learning with kernels: support vector machines, regularization, optimization, and beyond*, MIT Press, Cambridge, MA, 2002.

[22] D. Aha, D. Kibler, and M. Albert, "Instance-based learning algorithms", *Machine Learning 6*, 1991, pp. 37-66. DOI:10.1007/BF00153759

[23] J. Han, and M. Kamber, *Data Mining*, Morgan Kaufmann, San Francisco, CA, 2001.

[24] L. Breiman, "Random forests", *Machine learning*, 2001, 45.1, pp. 5-32.

[25] D. Hosmer, and S. Lemeshow, *Applied logistic regression*, (3rd ed.), Wiley, New York, 2013, DOI: 10.1002/9781118548387

[26] J.H. George, and P. Langley, "Estimating Continuous Distributions in Bayesian Classifiers", *Eleventh Conference on Uncertainty in Artificial Intelligence*, San Mateo, 1995, pp. 338-345.

[27] C.W. Hsu, C. C. Chang, and C. J. Lin, "A practical guide to support vector classification", 2003, pp.1-16.

[28] A. P. Bradley, "The use of the area under the ROC curve in the evaluation of machine learning algorithms." *Pattern recognition 30*, no. 7, 1997, pp. 1145-1159.

[29] J. D. Rodriguez, A. Perez, and J. A. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation", Pattern Analysis and Machine Intelligence, *IEEE Transactions on*, 32, no. 3, 2010, pp. 569-575.