Comparison of Cough, Wheeze and Sustained Phonations for Automatic Classification between Healthy subjects and Asthmatic patients

Shivani Yadav, Kausthubha NK, Dipanjan Gope,Uma Maheswari Krishnaswamy, Prasanta Kumar Ghosh

***Abstract*— In this work, we consider the task of automatic classification of asthmatic patients and healthy subjects using voice stimuli. Cough and wheeze have been used as voice stimuli for this classification task in the past. In this work, we focus on sustained phonations, namely /ɑː/, /iː/, /uː/, /eɪ/, /oʊ/ and compare their classification performances with the cough and wheeze. Classification experiments using 35 asthmatic patients and 36 healthy subjects show that sustained vowel /iː/ achieves the highest classification accuracy of 80.79% among five vowels considered. However, it is found to be higher and lower than the classification accuracies of 78.72% and 90.25% obtained using cough and wheeze respectively. This suggests that for speech-based asthma classification, /iː/ would be a better choice compared to other vowels considered in this work. However, when non-speech sounds are included for classification, wheeze is a better choice than sustained /iː/.**

I. Introduction

Asthma is an inflammatory disease of the airways resulting in a number of symptoms including obstruction of the airways, chest discomfort or pain, cough, and wheezes or other peculiar sounds during breathing [1]. 235 million people currently suffer from asthma, with 250k annual deaths according to World Health Organization (WHO) [2]. Spirometry is the most common pulmonary function test which measures the severity of asthma. During spirometry, patients are asked to wear a nose clip, take a deep breath to the best of their capacity, and then exhale into the spirometer as fast as and as long as possible, preferably for at least six seconds. Maneuver primarily depends on patient's effort and cooperation causing the spirometry readings to vary depending on how diligently a patient does the inhalation and exhalation in the suggested manner. It becomes difficult to obtain spirometry readings for children and elderly people [3] as it is strenuous for them to follow the guidance properly given by the technician.

Peak flow meter (PFM) [4] is used as a substitute which measures peak expiratory flow rate (PEFR) through the major airways of patient's lungs, but it fails to measure the same through minor airways, which also swells causing typical asthma symptoms. Therefore, an alternate technique to diagnose and monitor asthma is required to overcome limitations of available methods.

The cough and wheeze based asthma detection and monitoring is convenient for people irrespective of their age and medical conditions, unlike spirometry. A cough is produced by closing the glottis till the pressure builds up below the glottis followed by a sudden release of pressure once the glottis opens. Wheeze, on the other hand, is a continuous musical whistling sound produced in the respiratory tract during breathing. Several works in the literature have proposed techniques for classifying a subject into asthmatic or healthy person based on his/her cough and wheeze sound identified in a recording. For example, Wisniewski et al. used pulmonary wheezes to monitor asthma by using tonal index [5]. A segmentation scheme of respiratory sounds for the detection of wheezes in asthma patient was proposed by Akram et al. [6].

Study carried out by Bentur et al. [7] showed how wheeze monitoring provides quantitative information that correlates well with asthma activity of children. Hiew et al. [8] proposed algorithm for automatic coughs identification and counting for asthma. Automatic wheezing detection based on spectrogram processing and back-propagation neural network were also performed by Lin et al. [9]. There are several works that classify asthma using respiratory sound based on pitch [10], dominant frequency range [11] and duration of the breath [12]. Igor et al. [13] did respiratory sound analysis by using Mel-frequency cepstral coefficients (MFCC) with cascaded Support Vector Machine (SVM) to detect wheezing in asthmatic children. Achuth et al. [14] did asthma severity classification by automatic prediction of spirometry readings from cough and wheeze.

While the role of non-speech sounds like cough and wheeze for asthma detection has been extensively investigated, there is no work that investigates how effective several speech sounds could be for the same task. An understanding of how asthma signature could be encoded in different speech sounds could help in developing techniques that could detect asthma from natural voice of a subject rather than asking subjects to produce cough and wheeze on demand that may not be natural. Towards this, we, in this preliminary work, explore the role of five sustained vowels, namely, /ɑː/ (as in 'father'), /iː/ (as in 'See'), /uː/ (as in 'Blue'), /eɪ/ (as in 'Say'), /oʊ/ (as in 'Go') instead of directly using natural speech from a subject. Sustained phonations is used in order to quantify the potential of vowel sounds for asthma classification without any co-articulation effect which occurs due to the temporal variation of vocal tract, that often happens in running speech. Since, the signature of asthma is typically present in the lung volume, [15] which, in turn, affects the glottal voice source [16]. In this work we restrict to vowel stimuli because, for non-vowel sounds, the speech is also modulated by the vocal tract constriction

Shivani Yadav, Kausthubha NK, Dipanjan Gope and Prasanta Kumar Ghosh are with Indian Institute of Science, Bangalore, India.(email id: shivaniyadav@iisc.ac.in, kausthubhan@iisc.ac.in, dipanjan@iisc.ac.in, prasantg@iisc.ac.in).
Uma Maheshwari K. is with Pulmonary Medicine, St.Johns National Academy of Health Sciences, Bangalore, India (email id: umamohan99@gmail.com)

unlike that in vowel [17]. We perform an automatic classification of asthmatic patients and healthy subjects using acoustic features from these sustained vowels and compare the classification accuracies with those obtained using cough and wheeze. Classification experiments using 35 asthmatic patients and 36 healthy subjects shows that sustained vowel /iː/ achieves the highest classification accuracy of 80.79% among five vowels considered. However, it is found to be higher and lower than the classification accuracies of 78.72% and 90.25% obtained using cough and wheeze, respectively. This suggests that although sustained /iː/ could be used for asthma detection, wheeze is superior to the sustained /iː/ for the classification task.

II. Dataset

Dataset used in this work consists of total of 71 subjects comprising 35 patients (17 female, 18 male) and 36 healthy subjects (18 female, 18 male) recruited from St. John's National Academy of Health Sciences, Bangalore. The healthy subjects are middle aged with an age range of 19-42 years and average age of 24 years. The age range of the patients are 19-78 years with an average age of 43 years.

Recordings have been taken under doctor's guidance. Prior approval for the recording was taken from the St. John's hospital ethics committee. The consent for recording was taken from each subject. To verify whether a subject is suffering from asthma or any lung disease, spirometry and other lung functional test have been performed and cross-checked by the doctor for confirmation. The patients in the database suffer uniformly from various levels of asthma - mild, moderate, and severe. Forced expiratory volume in 1 sec (FEV1) for 35 patients lies between 0.36 ls^{-1} to 3.72 ls^{-1} with an average of 1.56 and standard deviation (SD) of 0.82. Similarly, FEV1_FVC ratio for patients lies between 50%-97% of their reference values where, FVC denotes forced vital capacity. Following spirometry test, subjects were asked to first cough, next wheeze followed by sustained phonations in the following order: /ɑː/, /iː/, /uː/, /eɪ/, /oʊ/. We refer cough, wheeze and sustained phonations as stimuli from now onward. Each stimuli is recorded for five times in a row. Thus, we obtain 355 recordings (175 for patients and 180 for healthy subjects) for each of the seven stimuli. Recording was done by using ZOOM H6 handy recorder at a sampling rate of 48kHz and 16 bits/sample. All frequencies (till 24kHz) present in the recorded signal is used in this experiment to consider all spectral variations.Recordings was performed in the spirometry lab of the hospital itself. The recording room was moderately noisy because of fan and conversation between technician and patients. While recording, microphone was placed near the mouth to suppress the noise in the room as much as possible and capture the signal of interest with high SNR.

Sufficient breaks are given between recordings of different stimuli to ensure that the patient is not tired of recording. During wheeze and sustained phonations recording, a nose clip is used to block the air flow through the nose so that subjects can exhale to their full capacity. During cough recording, nose clip was not used to ensure free flow of sudden expulsion of cough air. Average time for the entire recording was $\sim$ 8 minutes per subject. Wheeze, cough and sustained phonations boundary were manually marked after listening and examining the waveform using Audacity [18].

III. Automatic classification between asthmatic patients and healthy subjects

The schematic diagram of proposed approach for healthy subject vs asthmatic patient classification is shown in Fig. 1. In the training stage of the classification, representative features from the recording are used along with their respective class labels to train a classifier. Statistics of the MFCC sequence over the entire recording of a stimuli are used as the features for classification. SVM is used as the classifier. Due to limited data samples, we have not used more complex classifier such as deep neural network (DNN). In the testing phase, statistics of the MFCC sequence is computed and is provided to the trained SVM classifier to obtain a decision for the test stimuli.

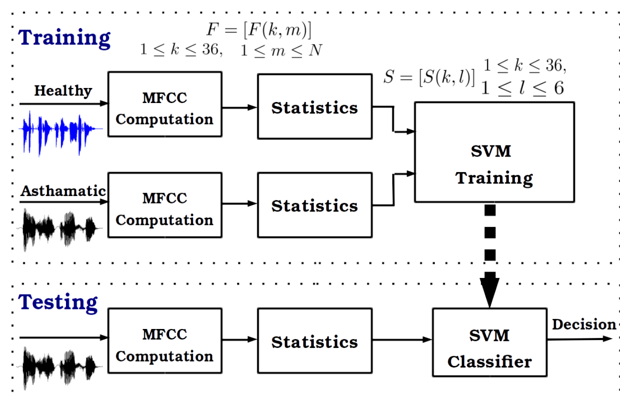

Fig. 1. Schematic diagram for the classification between healthy subjects and asthmatic patients.

The goal of designing features for classification is to capture the lung volume, an indicator for asthma, which could be encoded in the sound generated using different stimuli. The recordings corresponding to the chosen stimuli in this work are vocal sounds. Production of the vocal sounds, particularly speech could be described by the glottal source signal generated due to pressure from lungs. This signal gets modulated by the shape of the vocal tract and nasal cavity. We assume that the influence of asthma on glottal signal characteristics would reflect on the spectral characteristics of the vocal sounds. For this purpose, we compute 12-dimensional MFCC (excluding the energy coefficient) of the recordings with it first and second derivatives. This results in a 36-dimensional MFCC vector in an analysis window. MFCCs are calculated for each of the five repetitions of every stimuli. Let a stimuli instance has N analysis windows with a window length of N_w samples and window shift of N_{sh} sample. Following MFCC computation we obtain a MFCC matrix $F = [F(k,m)], 1 \le k \le 36, 1 \le m \le N$. We used voicebox to calculate MFCC [19].

As cough and wheeze are non-stationary sounds, we hypothesize that the variation in the MFCC vector sequence would provide cues for healthy subjects vs asthmatic patient classification. In the case of sustained vowels, the voice characteristics changes as the lung pressure reduces due to lack of breathing during phonation. We hypothesize that the variation in the spectral characteristics of a sustained vowels would indicate whether a subject is healthy or suffering from asthma. In order to capture the variation in the MFCC vector sequence, we compute six different statistics along each dimension k of MFCC, resulting in six dimensional feature sequence $S(k,l), 1 \leq k \leq 36, 1 \leq l \leq 6$. Computing such statistics from feature vector sequence has been done in the past [20]. Instead of going for a large number of statistics, we stick to six statistics only in this work. These are mean, mode, median, root mean square (RMS), variance and SD as defined below:

$$S(k,1) = \frac{1}{N}\sum_{m=1}^{N} F(k,m), \quad S(k,2) = mode(\{F(k,i)\}_{1\leq i\leq N})$$

$$S(k,3) = median(\{F(k,i)\}_{1\leq i\leq N}), S(k,4) = \sqrt{\frac{1}{N}\sum_{m=1}^{N}(F(k,m))^2}$$

$$S(k,5) = \frac{1}{N-1}\sum_{m=1}^{N}(F(k,m) - S(k,1))^2, \quad S(k,6) = \sqrt{S(k,5)}$$

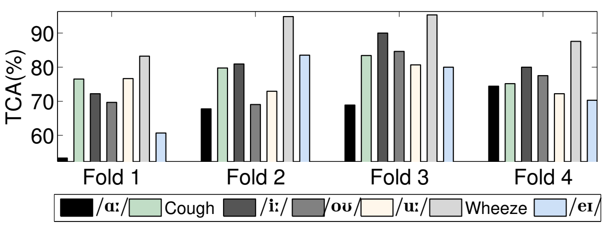

Fig. 2. Bar graph shows fold wise TCA with S_{216} for all stimuli.

IV. Experiments and Results

A. Experimental Setup

The classification experiments are performed in a four fold cross-validation setup. For this purpose, we make four groups (G1_H, G2_H, G3_H, G4_H) of healthy people each having nine subjects. Similarly, four groups (G1_P, G2_P, G3_P, G4_P) of patients are formed where all groups have nine patients except G4_P, which has eight patients. In i-th fold of the cross validation, Gi_H and Gi_P are used as the test set and the remaining healthy and patients groups are used as the training set. We use an analysis window of duration (N_w) 960 samples and shift of (N_{sh}) 480 samples to compute 12-dimensional MFCC, which is obtained by using 32 sub-bands placed uniformly on Mel-scale in the range of 0Hz-24kHz. We further calculate velocity and acceleration coefficients to obtain $36\times N$ MFCC matrix F for a stimuli with N frames. Using F, the 36×6 dimensional statistical feature matrix S has been calculated. We vectorized S to a 216-dimensional (=36×6) feature vector (S_{216}) for classification. We also experimented with statistical features computed from static MFCCs (72-dimensional S_{72}) and static MFCCs along with its first derivatives (144-dimensional S_{144}) and second derivatives (216-dimensional S_{216}). LIBSVM toolkit is used to implement SVM classifier [26]. The SVM hyper

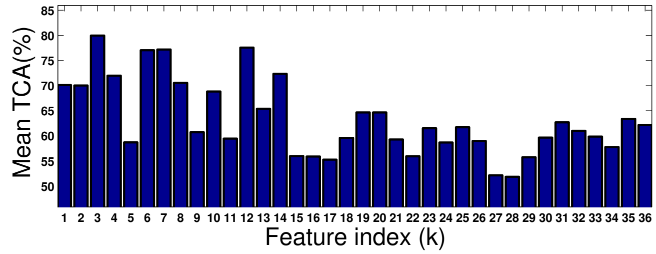

Fig. 3. TCA averaged across four folds using $S(k,l)$, $1 \leq l \leq 6$ from wheeze for different feature index k

parameters γ and C are optimized using grid search using a five fold cross-validation within the training set by using radial basis function. Grid search was performed for $\log_2(\gamma)$ and $\log_2(C)$ in the range of -1 to 20 and -1 to 10, respectively with the step size of 1. The performance of the algorithm was evaluated by using total classification accuracy (TCA), and F_1 score [21].

TABLE I

Mean (SD) of TCA(%) and F_1(%) using S_{72}, S_{144} and S_{216}.

Stimuli	TCA			F_1		
	S_{72}	S_{144}	S_{216}	S_{72}	S_{144}	S_{216}
/ɑː/	54.7(8.8)	61.7(6.9)	66.1(9.0)	52.9(9.8)	62.9(4.3)	67.7(7.7)
cough	77.3(6.1)	81.0(2.5)	78.7(3.7)	75.8(9.2)	80.3(5.1)	77.2(6.8)
/iː/	74.3(3.2)	76.8(9.0)	80.8(7.3)	76.4(3.7)	78.5(8.4)	81.6(7.3)
/oʊ/	72.3(10.7)	75.8(5.3)	75.2(7.4)	72.5(7.7)	76.5(2.5)	75.8(6.7)
/uː/	74.5(4.3)	73.9(2.2)	75.6(3.9)	75.3(3.8)	75.0(3.2)	76.6(3.7)
wheeze	**89.8(8.2)**	**90.5(6.9)**	**90.2(5.9)**	**89.4(8.8)**	**90.3(7.3)**	**90.2(6.2)**
/eɪ/	69.4(12.8)	74.0(8.6)	73.6(10.3)	70.4(15.2)	74.6(8.9)	74.2(9.3)

B. Results and discussion

Fig. 2 shows a bar plot of the TCA using S_{216} with seven different stimuli in each fold. It is clear from the figure that wheeze results in the highest mean TCA consistently in every fold. Among sustained phonations, /iː/ achieves the highest TCA in all folds except Fold1. In order to obtain an average performance across all folds, we present average (SD) values of both TCA as well as F_1 score in Table I using S_{72}, S_{144}, and S_{216}. The highest performance for each feature set is marked bold in each column. From the table it is seen that among all stimuli wheeze results in the highest average TCA and F_1 of 90.2% and 90.2% using S_{216}, respectively. The second highest classification performance is achieved by /iː/ followed by cough. However, when S_{144} is used as the feature, wheeze remains the best performing stimuli with cough being the second best followed by /iː/. Interestingly, /iː/ results in the highest classification performance among all sustained vowels considered using both S_{144} and S_{216}. On the other hand, /ɑː/ results in the lowest TCA and F_1 score using S_{72}, S_{144} as well as S_{216} features. These results suggest that among sustained vowels /iː/ is the best candidate for automatic classification between asthmatic patients and healthy subjects. However, the best F_1 score obtained using /iː/ is 81.6% (absolute) less than that using wheeze.This suggests that for speech-based asthma classification, /iː/ would be a better choice compared to other vowels considered in this work. However, when non-speech sounds are included for classification, wheeze is a better choice than sustained /iː/.

As the highest classification performance is obtained using wheeze, we further investigate the relative role by the sta-

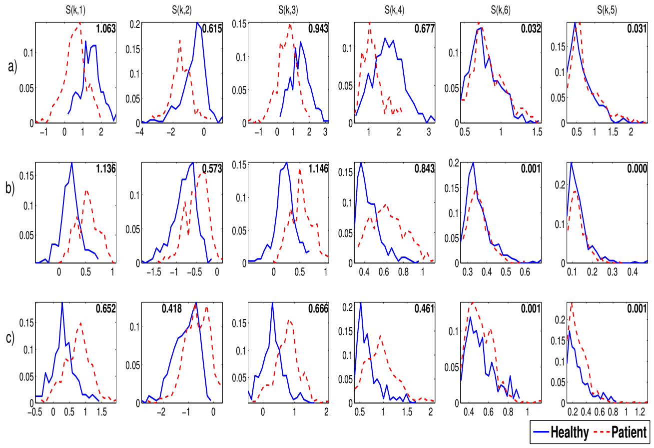

Fig. 4. Histogram plot of mean, mode, median, RMS, SD and variance statistics of wheeze between asthmatic patients and healthy subjects with a) $S(3,l),\ 1 \leq l \leq 6$, b) $S(12,l),\ 1 \leq l \leq 6$, c) $S(7,l),\ 1 \leq l \leq 6$. Fisher discriminant ratio (FDR) of each statistics for above coefficients in each case is shown at the top right corner of each plot.

tistical features from different coefficients in MFCC feature. For this purpose, we conduct the classification experiments using wheeze with 6-dimensional $S(k,l),\ 1 \leq l \leq 6$ feature separately for every $1 \leq k \leq 36$. The TCA obtained for different values of k is shown using a bar plot in Fig. 3. It is clear from the figure that $S(k,l),\ 1 \leq l \leq 6$ for static coefficients of MFCC ($1 \leq k \leq 12$) performs better than the first derivatives ($13 \leq k \leq 24$) and second derivatives ($25 \leq k \leq 36$). This indicates that the nature in which the energy distributions across frequencies vary over time in a wheeze signal encodes the asthma specific signature compared to those for velocity and acceleration coefficients of MFCCs. From Fig. 3 it can also be seen that three highest TCAs are obtained by $S(k,l),\ 1 \leq l \leq 6$ for k= 3, 12, 7.

We further examine the relative role of each of the statistical features in $S(k,l),\ 1 \leq l \leq 6$ for k=3, 12, 7. The histogram of $S(k,l)$ for every $k \in \{3,12,7\}$ and $l(1 \leq l \leq 6)$ are shown in Fig. 4. In order to quantify the discriminative capacity of each statistical feature, we compute Fisher discriminant ratio (FDR) [22] between the two classes (patients vs healthy) for every feature. FDR values (as seen at the top corner in every histogram plot) suggest that the mean, mode and median statistics are relatively more discriminative compared to the RMS, variance and SD statistics. This, in turn, indicates that the an average shape of the energy distribution across frequencies provides a reasonable cues for classification between asthmatic patients and healthy controls.

V. Conclusion and Future Work

In this paper we compare wheeze, cough and sustained vowels for automatic classification between asthmatic patients and healthy subjects with statistics of MFCC as the features and SVM as the classifier. The experimental results demonstrate that wheeze is the best stimuli for classification with a classification accuracy of 90.5%. However, sustained /ɪː/ performs the best among all sustained vowels with an accuracy of 80.8%. As the best performing stimuli is wheeze where there is no voicing, our future plan includes investigation of fricatives as stimuli for asthma classification task. Future work also includes selection of the best feature among all used in the present study and perform classification.

References

[1] FD Martinez. Genes, environments, development and asthma: a reappraisal. *European Respiratory Journal*, 29(1):179–184, 2007.

[2] World Health Organization et al. Asthma fact sheet n 307,(2013), 2015.

[3] National Asthma Education, Prevention Program (National Heart, Lung, and Blood Institute). Second Expert Panel on the Management of Asthma. *Expert panel report 2: guidelines for the diagnosis and management of asthma*. DIANE Publishing, 1997.

[4] Jeff Baker. Peak flow meter, February 10 2009. US Patent D586,248.

[5] Marcin Wisniewski and Tomasz P Zielinski. Application of tonal index to pulmonary wheezes detection in asthma monitoring. In *Signal Processing Conference, 2011 19th European*, pages 1544–1548. IEEE, 2011.

[6] Akram Belghith and Christophe Collet. An evidence segmentation scheme for asthma detection using a priori wavelet respiratory sound information. In *IS&T/SPIE Electronic Imaging*, pages 753509–753509. International Society for Optics and Photonics, 2010.

[7] L Bentur, R Beck, M Shinawi, T Naveh, and N Gavriely. Wheeze monitoring in children for assessment of nocturnal asthma and response to therapy. *European Respiratory Journal*, 21(4):621–626, 2003.

[8] YH Hiew, JA Smith, JE Earis, Barry MG Cheetham, and AA Woodcock. DSP algorithm for cough identification and counting. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, volume 4, pages IV–3888. IEEE, 2002.

[9] Bor-Shing Lin, Huey-Dong Wu, and Sao-Jie Chen. Automatic wheezing detection based on signal processing of spectrogram and back-propagation neural network. *Journal of healthcare engineering*, 6(4):649–672, 2015.

[10] Gérard Charbonneau, J Racineux, M Sudraud, and E Tuchais. Digital processing techniques of breath sounds for objective assistance of asthma diagnosis. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'82.*, volume 7, pages 736–738. IEEE, 1982.

[11] T Richard Fenton, Hans Pasterkamp, A Tal, and Victor Chernick. Automated spectral characterization of wheezing in asthmatic children. *IEEE transactions on biomedical engineering*, (1):50–55, 1985.

[12] Sandra Reichert, Raymond Gass, Christian Brandt, and Emmanuel Andrès. Analysis of respiratory sounds: state of the art. *Clinical Medicine Insights. Circulatory, Respiratory and Pulmonary Medicine*, 2:45, 2008.

[13] Igor Mazić, Mirjana Bonković, and Barbara Džaja. Two-level coarse-to-fine classification algorithm for asthma wheezing recognition in children's respiratory sounds. *Biomedical Signal Processing and Control*, 21:105–118, 2015.

[14] Achuth Rao MV, Kausthubha NK, Shivani Yadav, Dipanjan Gope, Uma Maheswari Krishnaswamy, and Prasanta Kumar Ghosh. Automatic prediction of spirometry readings from cough and wheeze for monitoring of asthma severity. In *European Signal Processing Conference (EUSIPCO)*, pages 41–45, 2017.

[15] Dan B Teculescu. Lung volume changes in asthma measured concurrently by two methods. *American Review of Respiratory Disease*, 108(2):391–391, 1973.

[16] Jenny Iwarsson, Monica Thomasson, and Johan Sundberg. Effects of lung volume on the glottal voice source. *Journal of voice*, 12(4):424–433, 1998.

[17] Catherine P Browman and Louis Goldstein. Articulatory phonology: An overview. *Phonetica*, 49(3-4):155–180, 1992.

[18] D Mazzoni and R Dannenberg. Audacity [software]. pittsburg, 2000.

[19] Mike Brookes et al. Voicebox: Speech processing toolbox for matlab. *Software, available [Mar. 2011] from www. ee. ic. ac. uk/hp/staff/dmb/voicebox/voicebox. html*, 47, 1997.

[20] Florian Eyben, Martin Wöllmer, and Björn Schuller. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1459–1462. ACM, 2010.

[21] David Martin Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. 2011.

[22] Andrew R Webb. *Statistical pattern recognition*. John Wiley & Sons, 2003.