Effect of importance sampling on robust segmentation of audio-cough events in noisy environments

Monge-Álvarez, Jesús; Hoyos Barceló, Carlos; Lesso, Paul; Escudero, Javier; Dahal, Keshav; Casaseca, Juan Pablo

Published in:
IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC), 2016

DOI:
10.1109/EMBC.2016.7591541

Published: 18/10/2016

Document Version
Peer reviewed version

Effect of Importance Sampling on Robust Segmentation of Audio-cough Events in Noisy Environments

Jesús Monge-Álvarez[1*], Carlos Hoyos-Barceló[1], Paul Lesso[2], Javier Escudero[3], Keshav Dahal[1], Pablo Casaseca-de-la-Higuera[1]

***Abstract*— This paper proposes a new cough detection system based on audio signals acquired from conventional smartphones. The system relies on local Hu moments to characterize cough events and a *k*-NN classifier to distinguish cough events from non-cough ones (speech, laugh, sneeze, etc.) and noisy sounds. To deal with the unbalance between classes, we employ Distinct-Borderline2 Synthetic Minority Oversampling Technique and a bespoke cost matrix. The system additionally features a post-processing module to avoid isolated false negatives and, this way, increases sensitivity. Evaluation has been carried out using a database comprising a variety of cough and non-cough events and different types of background noise. In this study, we specifically focused on noise likely to appear when the user is carrying the smartphone in daily activities. Different Signal to Noise Ratio values were tested ranging between -15 and 0 dB. Our experiments confirm that local Hu moments are suitable not only for characterizing cough events but also for coping with noisy environments. Results show a sensitivity of 94.17% and a specificity of 92.16% at -15 dB. Thus, our system shows potential as a reliable and place-ubiquitous monitoring device that helps patients self-manage their own respiratory diseases and avoids unreported or fabricated symptoms.**

I. Introduction

Cough is a mechanism of defense of the human body which protects airways from debris, especially mucus, and foreign material [1]. Cough can be considered as "a forced expulsive manoeuvre or manoeuvres against a closed glottis that are associated with a characteristic sound or sounds" [2].

This symptom is connected to many respiratory illnesses such as pneumonia, asthma or laryngitis, and some other generic pathologies (cold, *flu*, allergies, etc.). In addition, cough can be associated to the patients' lifestyle (smokers, sedentary people, etc.). Treatment of conditions related to cough constitutes an important burden for national health systems and economies (e.g. $40 billion per annum in the USA from direct and indirect costs of the common cold [2]). Until recently, the lack of golden-methods to objectively assess cough has limited its study to the use of subjective measurement tools like cough diaries or questionnaires [3]. This approach and other secondary aspects of the underlying diseases (e.g. physical and psychological morbidity) may lead to misinterpretation of cough symptoms [4].

On the basis on the above, international institutions and governments have promoted the potential of telemedicine in the management of respiratory conditions [5]. Current cough monitors rely on pattern recognition engines based on features extracted from cough sounds and other biomedical signals (chest movement, electrocardiogram, etc.). Previous proposals have achieved fair performance figures both in terms of sensitivity and specificity. For instance, the Lifeshirt system offered a sensitivity of 78.2% in laboratory conditions [6] whereas the Leicester Cough Monitor reported a sensitivity of 85.7% in a study with 26 subjects [7]. The Hull Automated Count Counter achieved a sensitivity of 80% in a group of 33 patients [8]. Specificity results in these studies were generally around 90%. Despite these figures, these systems are expensive (ad-hoc design and manufacturing) and uncomfortable (non-wearable during daily activity). To overcome these limitations, tele-health is currently moving towards more generic readily available sensors. The recent advances in smartphone and watch technology additionally allow employing these daily use devices as intelligent cough monitoring systems since they feature a number of embedded sensors able to measure cough sounds and related movement. Moreover, the computational capability of these devices is growing while, at the same time, they feature real-time connectivity to offload complex operations to higher performance computing systems.

In this paper, we propose an automatic cough detection system based on audio signals acquired from conventional smartphones. The system is able to detect cough events with high sensitivity and specificity in a noisy environment, i.e. the signals are recorded while the smartphone is inside the pocket of trousers, shirts or coats. To evaluate our proposal we have synthesized a database comprising a variety of cough and non-cough (speech, laugh, sneeze, etc.) events and different types of background noise. In this study, we specifically focused on noise likely to appear when the user is carrying the smartphone in daily activities. Signal to Noise Ratio (SNR) values ranging between -15 and 0 dB were tested. The final system relies on local Hu moments as a feature feeding a *k*-NN classifier. Hu moments were recently imported from the field of image processing to speech emotion recognition [9]. Our results confirm that local Hu moments allow the segmentation of cough events even in very noisy environments.

The rest of the paper is organized as follows: Section II provides a description of the database used for evaluation and the synthesis process to obtain different SNR values. The

This work was supported by the Digital Health & Care Institute as part of the Factory Research Project *SmartCough/MacMasters*.

**Asterisk indicates corresponding author*

[1]J. Monge, C. Hoyos, K. Dahal, and P. Casaseca are with the Center for Artificial Intelligence, Visual Communications and Networking (AVCN) School of Engineering and Computing, University of the West of Scotland, Paisley Campus, Paisley, PA1 2BE, United Kingdom. jesus.monge@uws.ac.uk, jsmonge@outlook.es.

[2]P. Lesso is with Cirrus Logic, Edinburgh, EH11 2QB, United Kingdom.

[3]J. Escudero is with the Institute for Digital Communications, School of Engineering, The University of Edinburgh, Edinburgh, King's Buildings, EH9 3JL, United Kingdom.

proposed method, including the algorithm used to calculate local Hu moments and other employed techniques are also presented. Results are introduced in Section III while Section IV is devoted to the discussion and conclusions.

II. Materials and methods

A. Cough database

The synthesis of the signal database including daily activity noise, cough and non-cough events was performed as follows. First, we recorded different types of background noise in typical daily monitoring contexts. To do so, we introduced several smartphones inside the pockets of jeans, trousers, shirts, t-shirts and coats and left them recording. We also introduced some other common items such as keys, coins, candies, etc. The recordings took place during daily life activities, i.e. walking, sitting down, standing up several times, etc. Later on, cough and non-cough events were recorded in a quiet room in different positions and with the smartphone places in different parts of the body.

A placed power analysis showed that the SNR between cough/non-cough events and background noise was indeed very low, around -6 dB (average power of cough and non-cough events versus average power of recorded noise due to friction with fabric, keys, coins, etc.). Consequently, we focused on lower SNR values and synthesized the signals as follows:

- We equalized all the raw signals to have the same average power.
- Then, we synthesized signals for different SNR values. We used -15, -12, -9, -6, -3, and 0 dB for this experiment. Cough and non-cough events were collated one after the other in a longer signal. Between each event, zero samples with random duration between 0.25 and 1 s were inserted. The reason why we included these gaps is due to the fact that two front events of different nature are very unlikely to occur one immediately after the other. Next, we calculated a gain value, *G*, to be applied to the noisy sounds to achieve the desired SNR (1). Finally, both events and noisy sounds were added.

$$SNR_{dB} = 10 \cdot \log_{10}\left(1/G\right) \Rightarrow G = 10^{-\frac{SNR_{dB}}{10}} \tag{1}$$

The negative SNR values we have used for evaluation align with the expected physics of the acquired noise levels. The farther the source, the lower the perceived intensity. In our case, the noise source (friction with fabric), is closer to the microphone than the cough emitter.

Each SNR version of the database (all signals) lasts 650 s. The sampling frequency was 44.1 KHz and all the sounds were recorded in lossless WAV format with 16 bits per sample. Neither any noise-sound signal nor event (cough and non-cough) were used more than once in the synthesis. Fig. 1 shows 4 SNR versions of one of the signals in our database.

B. Local Hu moments:

Hu moments are widely employed features for object recognition [10] or watermarking [11] in images, among other uses. Recently Sun et al. extrapolated their use to the signal processing field by using the time-frequency dimensions as coordinates for speech emotion recognition [9]. We followed their approach and applied local Hu moments to our signals in an attempt to characterize cough events within noisy environments. We expected that the main properties of Hu moments which showed successful in object recognition – invariance to rotation, scaling and translation – could easily be translated here as a way to cope with noise.

To compute the local Hu moments we performed the following steps:

First, all the signals were downsampled by 5 since cough sounds are typically located between 0 and 2 KHz, and thus, there is no need of a higher sampling frequency (see [12]). Then, each signal is windowed using a Kaiser window with $\beta = 3.5$. As in [12], we used a window-length of 50 ms. The shift of the window is 25 ms.

Secondly, the Power Spectral Density (PSD) of each window was estimated as the Fourier transform of the autocorrelation function, according to the Wiener-Khinchin-Einstein theorem [13]. After normalization, only the one-sided PSD was selected.

Next, we computed the logarithm of the energies of every window in a series of bands defined by a filterbank in the Mel scale:

$$E_k(m) = \log\left(\sum_{f=f_{\min}}^{f_{mas}} PSD_k[f] \cdot H_m[f]\right) \quad 0 \le m < M \tag{2}$$

where *k* refers to the window whereas *m* denotes each filter within the filterbank. $f_{\min}$ and $f_{\max}$ are the frequencies 0 and 2000 Hz, respectively. The filterbank in the Mel scale is defined as:

$$H_m(f) = \begin{cases} 0, & f < C(m-1) \\ \dfrac{2\cdot(f - C(m-1))}{(C(m+1) - C(m-1))(C(m) - C(m-1))}, & C(m-1) \le f < C(m) \\ \dfrac{2\cdot(C(m+1) - f)}{(C(m+1) - C(m-1))(C(m+1) - C(m))}, & C(m) \le f < C(m+1) \\ 0, & f \ge C(m+1) \end{cases} \tag{3}$$

C(*m*) with $0 \le m \le M$ are the centers of each filter in the filterbank, uniformly spaced between $f_{\min}$ and $f_{\max}$ in the Mel scale. The equations to convert Hz scale to Mel scale and the opposite are (4) and (5) respectively:

$$f[Mel] = 2595 \cdot \log_{10}\left(1 + f[Hz]/700\right) \tag{4}$$

$$f[Hz] = 700 \cdot \left(10^{f[Mel]/2595} - 1\right) \tag{5}$$

A value of $M = 75$ was employed. Consequently, after performing this step for all the windows, we had a $(K \times (M-1))$ matrix, *E*, with *K* the number of windows in the signal.

To compute the local Hu moments of the energy matrix, we split E into $(K \times ((M/w)-1))$ blocks B_{ij}, where w is the size block. We used $w = 5$, as in [9]:

$$B_{ij} = \begin{pmatrix} E_i(w \cdot j) & \dots & E_i(w \cdot j + w - 1) \\ \vdots & \ddots & \vdots \\ E_{i+w-1}(w \cdot j) & \cdots & E_{i+w-1}(w \cdot j + w - 1) \end{pmatrix} \quad (6)$$

$$i = 1 \dots K \quad j = 1 \dots ((M/w)-1)$$

The latest $(w-1)$ blocks, which correspond to $i = K - w + 2, \dots, K$, are padded with zeros up to the size $(w \times w)$ when no more data from the energy matrix E is available.

The first invariant moment θ of each B_{ij} is obtained as:

$$\theta = \eta(p = 2, q = 0) + \eta(p = 0, q = 2) \quad (7)$$

$$\eta(p,q) = \frac{\mu(p,q)}{(\mu(0,0))^{\rho}} \quad \rho = (p + q + 2)/2 \quad (8)$$

$$\mu(p,q) = \sum_{u=1}^{w} \sum_{v=1}^{w} (u - \bar{u})^p \cdot (v - \bar{v})^q \cdot g(u,v) \quad (9)$$

$$g(u,v) = B_{ij}(u,v) \quad p,q = 0,1,2,\dots$$

In (9), $\bar{u}$ and $\bar{v}$ are $\bar{u} = \varphi(1,0)/\varphi(0,0)$ and $\bar{v} = \varphi(0,1)/\varphi(0,0)$, with:

$$\varphi(p,q) = \sum_{u=1}^{w} \sum_{v=1}^{w} u^p \cdot v^p \cdot g(u,v) \quad p,q = 0,1,2,\dots \quad (10)$$

All θ are used to construct a real $(K \times ((M/w)-1))$ matrix, Q. To conclude, we computed the discrete cosine transform of each row of Q and only kept the 2nd to 14th coefficients. The result is a matrix TQ with size $(K \times 13)$, being the rows of this matrix the Hu moments of each window of the signal.

C. Importance sampling, classification and post-processing

In our system, windows are our basic classification unit – i.e. each row of TQ will be considered as an observation by the k-NN classifier. Besides, there are only two classes: windows which belong to cough events (positive class) and windows which belong to non-cough events or noisy sounds which may or may not include a non-cough event (negative class).

The k-NN classifier was configured with $k = 1$, standardized Euclidean distance as a distance metric, the inverse of the distance as weighting function and the distances were exhaustively computed, in other words, from each observation to the rest. The classification was based in a 5-fold cross validation partition of the feature space.

As in other real world problems, our feature space was unbalanced between classes: around 16% of the windows were cough events whereas the rest were negative class windows. Thus, our k-NN classifier would be biased towards the negative class. In order to avoid this problem we used two techniques: oversampling of the positive class and undersampling of the negative class using the recently proposed Distinct-Borderline2 Synthetic Minority Oversampling Technique (DB2SMOTE) [14] as well as a cost matrix to train the k-NN classifier.

Among data-level techniques to deal with unbalanced distributions of classes, DB2SMOTE is an improved version of the basic SMOTE. DB2SMOTE is a hybrid technique (include both oversampling and undersampling) specifically designed for high density data. It considers that boundary observations have more influence on the classification. With this in mind, DB2SMOTE generates synthetic observations in both majority and minority class boundaries. Likewise, to make the border lines more clear, the algorithm removes specific majority class observations [14].

Regarding the cost matrix, we relied on it as a cost-sensitive technique to complement DB2SMOTE. We defined the matrix values following a data-driven approach:

$$\begin{bmatrix} 0 & 1 \\ 1/0.16 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 6.25 & 0 \end{bmatrix} \quad (11)$$

The matrix in (11) means that a false negative is the worst misclassification for us (6.25), since we would be losing a window with cough events. By means of DB2SMOTE we compensated the ratio between classes from 16% to 40%. The remaining 10% is assumed to be treated with the cost matrix.

Finally, our system includes a post-processing module to additionally improve performance. With the aim of improving sensitivity, the post-processing task avoids isolated false negatives by setting every non-cough window surrounded by cough ones to actual coughs. The pipeline of the proposed system is depicted in Fig. 2.

III. Results

The smartphone-based cough detector described in previous sections was trained and tested using a 5-fold cross validation scheme and a database of signals. Sensitivity (SEN), specificity (SPE), accuracy (ACC) and area under the ROC curve (AUC) were used as performance metrics.

Evaluation was performed in three different scenarios: 1) no importance sampling with the costs matrix nor post-processing, 2) only importance sampling and the costs matrix, and 3) using both of them.

Table I (a) shows the overall results for the first case. Even for the worst SNR level (-15 dB), the sensitivity is relatively high, with levels in line with previous proposals (78.65%). Besides, it achieves an improvement of 4.4% when increasing SNR. All specificity values are above 95%.

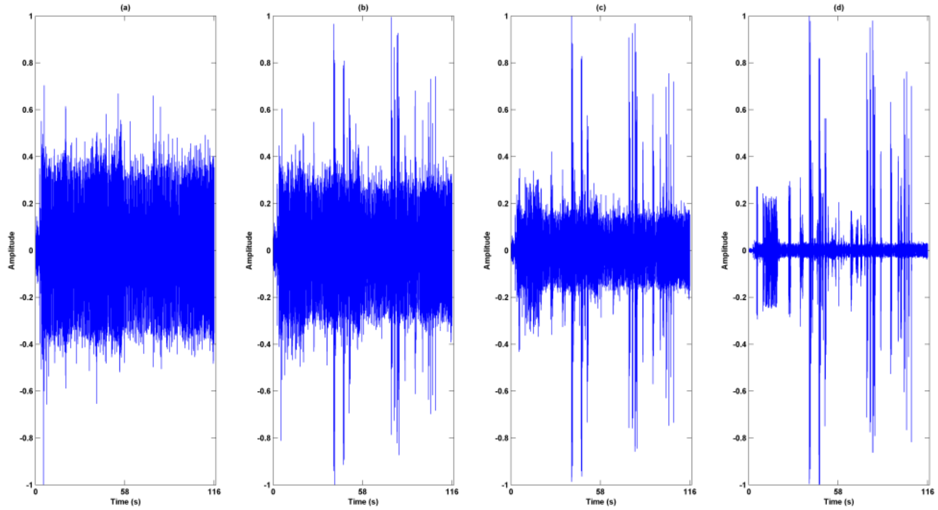

Figure 1. Representation of four SNR version of a synthetised signal: (a) -15 dB; (b) -6 dB; (c) 0 dB; (d) 15 dB. The version in (d) was synthetised to easily appreciate the position of events within the background sounds.

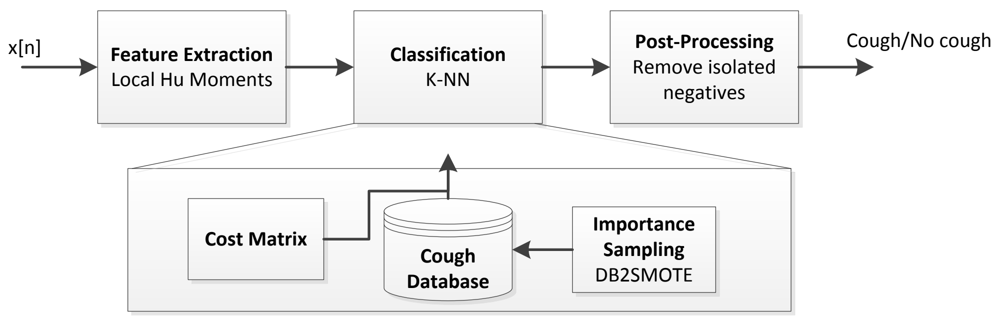

Figure 2. Pipeline of the proposed cough detection system.

TABLE I. CLASSIFICATION RESULTS (%): (A) RAW CLASSIFICATION; (B) WITH IMPORTANCE SAMPLING AND THE COSTS MATRIX; (C) WITH IMPORTANCE SAMPLING, THE COSTS MATRIX AND POST-PROCESSING

	(a)				**(b)**				**(c)**			
SNR (dB)	**SEN**	**SPE**	**ACC**	**AUC**	**SEN**	**SPE**	**ACC**	**AUC**	**SEN**	**SPE**	**ACC**	**AUC**
-15	78.65	96.93	93.83	87.79	86.20	90.58	89.84	88.39	92.97	89.63	90.20	98.75
-12	79.67	97.23	94.26	88.45	86.99	90.94	90.27	88.96	93.76	89.99	90.63	98.75
-9	80.91	97.44	94.63	89.17	87.62	91.60	90.93	89.61	93.97	90.79	91.33	98.88
-6	81.86	97.46	94.81	89.66	88.12	92.16	91.47	90.14	94.15	91.39	91.86	98.94
-3	82.17	97.88	95.22	90.03	88.59	92.41	91.77	90.50	94.17	91.78	92.18	99.06
0	83.03	97.89	95.37	90.46	88.66	92.76	92.06	90.71	94.17	92.16	92.50	99.10

Table I (b) shows the same results when importance sampling and the costs matrix were applied in the training stage. Now, the worst sensitivity is again at -15 dB of SNR but a higher value was achieved (86.20%). Similarly, it improves with SNR increasing up to a value of 88.66%. However, this approach brings along a loss of specificity. Now, these values are between 90.58% and 92.76%.

The overall classification results of our final system are shown in Table I (c). After the inclusion of the post-processing module, the sensitivity is further improved. In addition, this improvement takes place for all SNR values. On the contrary, the specificity undergoes a slight reduction.

IV. DISCUSSION AND CONCLUSION

We have proposed a new cough detection system based on smartphones. The system is composed of feature extraction and classification modules, based on local Hu moments and a *k*-NN classifier respectively, further improved by a costs matrix and an importance sampling of the cough database (to compensate the unbalance between classes) and a post-processing module (to enhance the sensitivity of our system).

The first point to discuss is the robustness of Hu moments against noise. The explanation of this behavior lies in the main properties of Hu moments as a feature set. Hu moments are widely used in image processing because of their invariance against scaling, rotation and translation [15]. Then, if we understand the variation introduced by the noise as the equivalent in signals to scaling, rotation and translation in images, our results would confirm that these properties of Hu moments are translated as noise robustness in our signal processing problem.

As for importance sampling, we decided to use a hybrid technique to keep under control the main drawbacks: overtraining due to oversampling in the minority class and a loss of the induction capacity because of undersampling of the majority class. Furthermore, we used the data-level technique only to compensate the class unbalance up to 40%. Some studies have proved that, for some classifiers such as *k*-NN, a total compensation of class unbalance via a single technique is not always the best choice [16]. Thus, to mitigate the effect of the remaining unbalance, we introduced a cost matrix in our *k*-NN classifier. The benefit of DB2SMOTE plus the cost matrix is the achieved improvement in sensitivity. On the other hand, the drawback is the specificity reduction, probably derived from the undersampling step of DB2SMOTE.

As for the post-processing module, results in Table I (c) confirm its suitability. We introduced this module to take advantage of the within-class distributions. That is to say, when a person suffers a cough episode, several windows with cough events will be recorded, in other words, cough windows appear in bursts. Consequently, if the output labels of our classifier show a negative-class label surrounded by positive-class labels, it is highly probable that the negative-class label is a false negative. We directly convert this false negative into positive-class labels and, by doing so, we enhance the cough detection capabilities of our system without almost no loss in specificity.

To sum up, our cough detection system based on smartphones outperforms previous cough monitors in terms of sensitivity as well as noise robustness. Its high noise robustness (up to -15 dB) makes the system place-ubiquitous, which could be helpful to track the daily evolution of patients with respiratory disease, avoiding unreported or fabricated symptoms and reporting the true impact of cough symptoms in their quality of life. Moreover, a smartphone implementation enlarges the number of possible users while reducing the costs of manufacturing. On this basis, our system would be a good candidate for a reliable monitoring device that can help patients self-manage their own respiratory diseases [17].

ACKNOWLEDGMENT

The authors would like to thank the Smartcough clinical team at the University of Edinburgh (Prof. Brian McKinstry, Dr. Hilary Pinnock, Dr. Roberto Rabinovich, and Dr. Lucy McCloughan) for their help and support in clinical matters. Thanks are also given to Chest Heart and Stroke Scotland (Lorna Stevenson, Dave Bertin, and Jill Adams) for her support in setting up the patient panel for the Smartcough projet.

REFERENCES

[1] K. K. Lee and S. S. Birring, “Cough,” *Medicine*, vol. 40, no. 4, pp. 173-176, Apr. 2012.

[2] A. H. Morice, “ERS guidelines on the assessment of cough,” *Eur Respir J*, vol. 29, no. 6, pp. 1256-1276, Jun. 2007.

[3] C. T. French, R. S. Irwin, K. E. Fletcher, and T. M. Adams, “Evaluation of a cough-specific quality-of-life questionnaire,” *Chest*, vol. 121, no. 4, pp. 1123-1131, Apr. 2002.

[4] L. PA. McGarvey et al., “Prevalence of psychomorbidity among patients with chronic cough,” *Cough*, vol. 2, no. 4, Jun. 2006.

[5] Communication from the Commission of the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on telemedicine for the benefit of patients, healthcare systems and society, 2008.

[6] M. A. Coyle, et al., “Evaluation of an ambulatory system for the quantification of cough frequency in patients with chronic obstructive pulmonary disease,” *Cough*, vol. 1, no. 3, Aug. 2005.

[7] S. S. Birring, et al.,“The Leicester cough monitor: preliminary validation of an automated cough detection system in chronic cough,” *Eur Respir J*, vol. 31, no. 5, pp. 1013-1018, May. 2008.

[8] S. J. Barry, et al., “The automatic recognition and counting of cough,” *Cough*, vol. 2, no. 8, Sep. 2006.

[9] Y. Sun, G. Wen and J. Wang, “Weighted spectral features based on local Hu moments for speech emotion recognition,” *Biomed Signal Process Control*, vol. 18, pp. 80-90, Apr. 2015.

[10] L. Zhang, J. Pu and Z. Zhang, “Application of improved HU moments in object recognition,” in *Proc. Int. Conf. on Automotion and Logistics*, Zhengzhou, 2012, pp. 554-558.

[11] G. Duan, X. Zhao, A. Chen and Y. Liu, “An improved Hu moments invariants based classification method for watermarking algorithm,” in *Int. Conf. on Information and Network Security*, Beijing, 2014, pp. 205-209.

[12] P. Casaseca-de-la-Higuera, “Effect of downsampling and compressive sensing on audio-based continuous cough monitoring,” *in Annual Int. Conf. of Engineering in Medicine and Biology Society*, Milan, 2015, pp. 6231-6235.

[13] S. Haykin, *Communication systems*, John Wiley & Sons, New York, 2001.

[14] J. Pengfei, Z. Chunkai and H. Zhenyu, “A new sampling approach for classification of imbalanced data sets with high density,” in *Int. Conf. on Big Data and Smart Computing*, Bangkok, 2014, pp. 217-222.

[15] M. E. Celebi and Y. A. Aslandogan, “A comparative study of three moment-based shape descriptors,” *Int. Conf. on Information Technology: Coding and Computing*, Las Vegas, USA, 2005, vol. 1, pp. 788-793.

[16] D. J. Dittman, T. M. Khoshgoftaar and A. Napolitano, “Selecting the appropriate data sampling approach for imbalanced and high-dimensional bioinformatics datasets,” *IEEE Int. Conf. on Bioinformatics and Bioengineering*, Boca Raton, FL, 2014, pp. 304-310.

[17] P. G. Gibson et al., “Self-management education and regular practitioner review for adults with asthma,” *Cochrane Database Syst Rev*, vol. 1, CD001117, Jul. 2002.