This is a repository copy of *Deep neural network based cough detection using bed-mounted accelerometer measurements*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/221839/

Version: Accepted Version

DEEP NEURAL NETWORK BASED COUGH DETECTION USING BED-MOUNTED ACCELEROMETER MEASUREMENTS

Madhurananda Pahar[1†*], *Igor Miranda*[1⊕], *Andreas Diacon*[2†], *Thomas Niesler*[1†]

[1]Department of Electrical and Electronic Engineering, University of Stellenbosch, South Africa
[2]TASK Applied Science, Cape Town, South Africa
[†]{mpahar, ahd, trn}@sun.ac.za; [⊕]igordantas@ufrb.edu.br

ABSTRACT

We have performed cough detection based on measurements from an accelerometer attached to the patient's bed. This form of monitoring is less intrusive than body-attached accelerometer sensors, and sidesteps privacy concerns encountered when using audio for cough detection. For our experiments, we have compiled a manually-annotated dataset containing the acceleration signals of approximately 6000 cough and 68000 non-cough events from 14 adult male patients in a tuberculosis clinic. As classifiers, we have considered convolutional neural networks (CNN), long-short-term-memory (LSTM) networks, and a residual neural network (Resnet50). We find that all classifiers are able to distinguish between the acceleration signals due to coughing and those due to other activities including sneezing, throat-clearing and movement in the bed with high accuracy. The Resnet50 performs the best, achieving an area under the ROC curve (AUC) exceeding 0.98 in cross-validation experiments. We conclude that high-accuracy cough monitoring based only on measurements from the accelerometer in a consumer smartphone is possible. Since the need to gather audio is avoided and therefore privacy is inherently protected, and since the accelerometer is attached to the bed and not worn, this form of monitoring may represent a more convenient and readily accepted method of long-term patient cough monitoring.

Index Terms— accelerometer, cough detection, Resnet, CNN, LSTM

1. INTRODUCTION

Coughing is the forceful expulsion of air to clear up the airway and a common symptom of respiratory disease [1]. It can be distinctive in nature and is an important indicator used by physicians for clinical diagnosis and heath monitoring in more than 100 respiratory diseases [2], including tuberculosis (TB) [3], asthma [4] and pertussis [5].

Automatic cough detection and classification is possible by applying machine learning algorithms on extracted features from cough sounds [6]. It has also been shown to be possible when using the signals from an accelerometer placed on the patient's body [7]. Since the accelerometer is insensitive to environmental and background noise, it can be used in conjunction with other sensors such as microphones, ECG and thermistors [8].

A cough monitoring system using a contact microphone and an accelerometer attached to the participant's suprasternal (jugular) notch has been considered by [9]. This system allows participants to move around within their homes while the cough audio and vibration is recorded. In related work, an ambulatory cough monitoring system using an accelerometer attached to the skin of the participant's suprasternal notch using a bioclusive transparent dressing was developed in [10]. Here the recorded signal is transmitted to a receiver carried in a pocket or attached to a belt.

Throat-mounted accelerometers have been used successfully to detect coughing in [11] and in [12], and an accelerometer placed at the laryngeal prominence (Adam's apple) in [7]. Two accelerometers, one placed on the abdomen and the second on a belt wrapped at dorsal region, have been used to measure cough rate in the research carried out by [13]. Abdominal placement (between the navel and sternal notch) of the accelerometer was also investigated in [14], and it was applied to patients who were children. Finally, multiple sensors, including ECG, thermistor, chest belt, accelerometer and audio microphones were used for cough detection in [15].

Attaching an accelerometer to the patient's body is however inconvenient and intrusive. Thus, we propose the monitoring of coughing based on the signals obtained from the accelerometer inbuilt in an inexpensive consumer smartphone firmly attached to the patient's bed, as shown in Figure 1, thereby eliminating the need to wear a measuring equipment. We have trained and evaluated deep neural network (DNN) classifiers such as convolutional neural networks (CNN), long-short-term-memory (LSTM) networks, and a residual neural network (Resnet50) [16] architecture using leave-one-out cross-validation on a dataset, prepared for this purpose,

which consists of cough and non-cough events such as sneezing, throat-clearing and getting in and out of the bed. The Resnet50 produces the highest AUC of 0.9888 after 50 epochs with corresponding accuracy 96.71% and sensitivity 99% for 32 sample (320 msec) long frames and grouping them in 10 segments. This shows that it is possible to discriminate between cough events and other non-cough events by using state-of-the-art classifiers such as a Resnet50 architecture; where the accelerometer is no longer attached to the patient's body, rather built inside an inexpensive consumer smartphone attached to the headboard of the patient's bed.

2. DATASET PREPARATION

Data collection was performed at a small 24h TB clinic near Cape Town, South Africa, which can accommodate approximately 10 staff and 30 patients. Each ward has four beds and can thus accommodate up to four patients at one time. The overall motivation of our work is to develop a practical method of automatic cough monitoring for the patients in this clinic, in order to assist with the monitoring of recovery.

The recording setup is shown in Figure 1. An enclosure housing an inexpensive consumer smartphone is firmly attached to the back of the headboard of each bed in a ward. A data gathering Android smartphone application, developed for this study, continuously monitors the accelerometer and audio signals from an external microphone (also shown in Figure 1). The 3-axis accelerometer has a sampling frequency of 100 Hz and only magnitudes were recorded. This energy-threshold-based detection for both audio and acceleration signals results in a large volume of data being captured. In addition, continuous video recordings were made using ceiling-mounted cameras in order to assist with data annotation.

This work considers automatic classification of the acceleration signals. The audio signals and video recordings were used only during the manual annotation process in order to unambiguously confirm the presence or absence of a cough. We define an 'event' to be any detected accelerometer or audio activity. An example of the accelerometer magnitudes captured for a cough and for a non-cough event are shown in Figure 2. Annotation was performed using the multimedia software tool ELAN [17], which allowed easy consolidation of the audio and video for accurate manual labelling.

Our final dataset, summarised in Table 1, contains approximately 6000 coughs and 68000 non-coughs from 14 adult male patients totalling 3.16 and 32.20 hours of data respectively. No other information regarding patients are recorded due to ethical constraints. This dataset was used to train and evaluate the classifiers within a cross-validation framework.

Table 1 shows that coughs are underrepresented in our dataset. To compensate for this imbalance, which can detrimentally affect machine learning [18, 19], we have applied SMOTE data balancing during training [20, 21]. This technique oversamples the minor class by generating synthetic
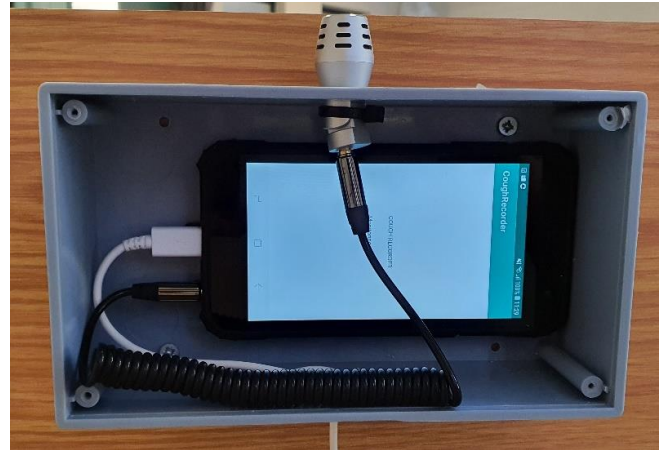
A Samsung Galaxy J4 smartphone with an inbuilt accelerometer connected to an external BOYA BY-MM1 cardioid microphone by a 3.5 mm audio jack
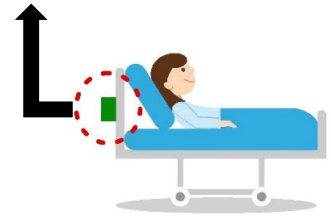
Fig. 1. **Recording Process:** A plastic enclosure housing an inexpensive smartphone running data gathering software is attached behind the headboard of each bed. The audio signal from the external microphone is also monitored, but only used for the purpose of annotation of the acceleration signal.
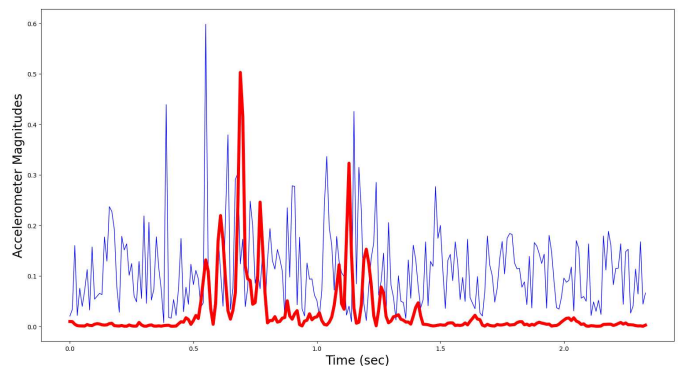
Fig. 2. **Example accelerometer magnitudes** for a cough event (red) and a non-cough event (blue). In this case, the non-cough event was the patient moving in the bed.

samples (instead of for example random oversampling). SMOTE has previously been successfully applied to cough detection and classification based on audio recordings [22].

3. FEATURE EXTRACTION

Power spectra [23], root mean square value, kurtosis, moving average and crest factor are extracted from the acceleration

Table 1. **Ground Truth Dataset Summary:** 'PATIENTS': list of the patients; 'COUGHS': number of confirmed cough events; 'NON COUGHS': number of confirmed events that are not coughs; 'COUGH TIME': total amount of time (in sec) for cough events; 'NON-COUGH TIME': total amount of time (in sec) for non-cough events.

PATIENTS	COUGHS	NON COUGHS	COUGH TIME	NON-COUGH TIME
Patient 1	88	973	169.16	1660.67
Patient 2	63	1111	117.67	1891.92
Patient 3	469	11025	893.91	18797.32
Patient 4	109	9151	204.06	15596.71
Patient 5	97	7826	188.26	13344.98
Patient 6	192	12437	360.72	21197.35
Patient 7	436	14053	825.23	23953.15
Patient 8	368	2977	702.05	5077.89
Patient 9	2816	3856	5345.27	6569.32
Patient 10	649	2579	1236.84	4400.42
Patient 11	205	527	391.42	901.38
Patient 12	213	323	402.61	547.62
Patient 13	213	712	401.61	1211.75
Patient 14	82	455	158.77	777.64
TOTAL	**6000**	**68005**	**11397.6**	**115928.12**

signals as features for classification. The frame length (Ψ) and number of segments (C) have been used as the feature extraction hyperparameters, shown in Table 2 and 4. The input feature matrix, shown in Figure 3 and 4, has the dimension of (C, $\frac{\Psi}{2} + 5$) and power spectra have ($\frac{\Psi}{2} + 1$) coefficients.

Table 2. **Feature extraction hyperparameters.** Frame lengths (16, 32, 64 samples i.e. 160, 320 and 640 msec long) overlap in such a way that the number of segments (5 and 10) are the same for all events in the dataset.

Hyperparameter	Description	Range
Frame length (Ψ)	Size of frames in samples in which cough is segmented	2^k where $k = 4, 5, 6$
No. of Segments (C)	Number of segments in which frames were grouped	5, 10

Frame length (Ψ) used to extract features from acceleration signal is shorter than those generally used to extract features from audio [24], because the accelerometer in the smartphone (shown in Figure 1) has a lower sampling rate of 100 Hz and longer frames lead to deteriorated performance as the signal properties can no longer be assumed to be stationary.

4. CLASSIFIER TRAINING

Our dataset contains 14 patients and a leave-one-out cross-validation scheme [25] has been used to train and evaluate
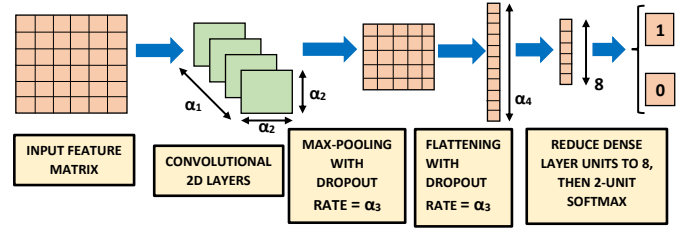

Fig. 3. **CNN Classifier**, trained and evaluated using leave-one-out cross-validation [25], produces results shown in Table 4 for feature extraction hyperparameters shown in Table 2.

Fig. 4. **LSTM classifier**, trained and evaluated using leave-one-out cross-validation [25], produces results shown in Table 4 for feature extraction hyperparameters in Table 2.

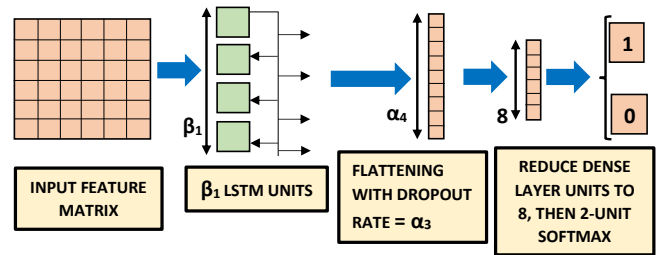our three DNN classifiers: CNN, LSTM and Resnet50.

The CNN classifier, shown in Figure 3, has been set up with α_1 2D convolutional layers with kernel size α_2 and rectified linear units as activation functions. A dropout rate α_3 has been applied along with max-pooling, followed by α_4 dense layers with rectified linear units as activation functions, followed by another 8 dense layers, also with rectified linear units as activation functions.

The LSTM classifier, shown in Figure 4, has been set up with β_1 LSTM units with rectified linear units as activation functions and a dropout rate α_3. Then α_4 dense layers have been applied with rectified linear units as activation functions, followed by another 8 dense layers also with rectified linear units as activation functions.

For both LSTM and CNN classifiers, a final softmax function produces one output for a cough event (i.e. 1) and the other for a non-cough event (i.e. 0), as shown in Figure 3 and 4. Features are fed into these two classifiers in batch size of ξ_1 for ξ_2 number of epochs.

The residual network (Resnet) architecture we trained and evaluated has 50 layers and has been found to deliver the state-of-the-art performance in image recognition. We have replicated the 50-layer architecture used in Table 1 of [16] in our experiments. Table 3 lists the classifier hyperparameters that were optimised during leave-one-out cross-validation.

Table 3. **CNN & LSTM classifier hyperparameters**, optimised using leave-one-out cross-validation and shown in Figure 3 and 4.

Hyperparameters	Classifier	Range
Batch Size (ξ_1)	CNN & LSTM	2^k where $k = 6, 7, 8$
No. of epochs (ξ_2)	CNN & LSTM	10 to 200 in steps of 20
No. of Conv filters (α_1)	CNN	3×2^k where $k = 3, 4, 5$
kernel size (α_2)	CNN	2 and 3
Dropout rate (α_3)	CNN & LSTM	0.1 to 0.5 in steps of 0.2
Dense layer size (α_4)	CNN & LSTM	2^k where $k = 4, 5$
LSTM units (β_1)	LSTM	2^k where $k = 6, 7, 8$
Learning rate (β_2)	LSTM	10^k where $k = -2, -3, -4$

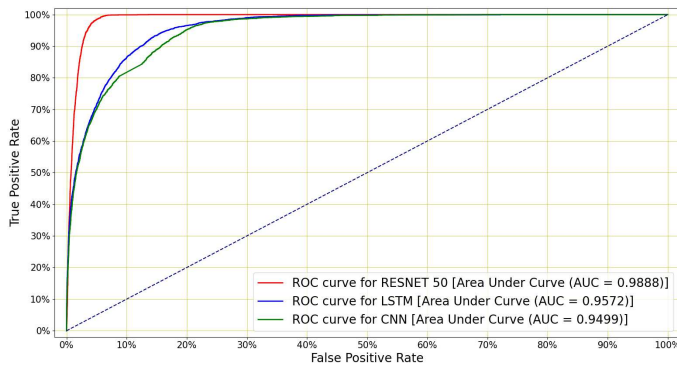**Fig. 5**. **Mean ROC curves for cough detection**, for the best performing DNN classifiers in Table 4. AUC values are averaged over 14 leave-one-patient-out cross-validation folds during hyperparameter optimisation. Resnet50 outperforms the LSTM and CNN over a wide range of operating points and has achieved the highest accuracy of 96.71%.

5. RESULTS

Table 4 lists the performance achieved by our three DNN classifiers for the hyperparameters mentioned in Table 2. These results are averages over the 14 leave-one-patient-out testing partitions during hyperparameter optimisation.

Table 4 shows that the best-performing CNN uses 64 samples (640 msec) long frames and 10 number of segments to achieve an AUC of 0.9499. The optimal LSTM classifier achieves the slightly higher AUC of 0.9572 when using a frame length of 32 samples (320 msec) and 10 number of segments. However, the best performance is achieved by the Resnet50 architecture, with an AUC of 0.9888 after 50 epochs from 32 samples (320 msec) long frames and 10 number of segments. Figure 5 shows the mean ROC curves for the optimal CNN, LSTM and Resnet50 configurations shown in Table 4, where the means were calculated over the 14 cross-validation folds. The Resnet50 classifier is superior to the other two classifiers over a wide range of operating points.

Table 4. **Leave-one-out cross-validation results for DNN classifiers.** The values are averaged over 14 cross-validation folds.

Frame (Ψ)	Seg (C)	Clas--sifier	Mean Spec	Mean Sens	Mean Accuracy	Mean AUC
16	5	CNN	83%	86%	84.55%	0.9243
16	10	CNN	87%	84%	85.66%	0.9358
32	5	CNN	76%	93%	84.47%	0.9272
32	10	CNN	84%	86%	85.25%	0.9324
64	5	CNN	85%	87%	86.31%	0.9339
64	*10*	*CNN*	*91%*	*80%*	*85.82%*	*0.9499*
16	5	LSTM	84%	91%	87.58%	0.9444
16	10	LSTM	85%	92%	88.32%	0.9504
32	5	LSTM	79%	95%	87.1%	0.9457
32	*10*	*LSTM*	*86%*	*93%*	*89.21%*	*0.9572*
64	5	LSTM	84%	93%	88.68%	0.954
64	10	LSTM	86%	89%	87.66%	0.9489
16	5	Resnet50	93%	98%	95.43%	0.9802
16	10	Resnet50	94%	99%	96.35%	0.9812
32	5	Resnet50	94%	99%	96.54%	0.9810
32	***10***	***Resnet50***	***94%***	***99%***	***96.71%***	***0.9888***
64	5	Resnet50	94%	99%	96.35%	0.9854
64	10	Resnet50	95%	98%	96.46%	0.9884

6. CONCLUSION AND FUTURE WORK

A deep neural network based cough detector is able to accurately discriminate between the accelerometer measurements due to coughing and due to other movements as captured by a consumer smartphone attached to a patient's bed. The best system, using the Resnet50 architecture, achieves an AUC of 0.9888. These experimental results are based on a specially-compiled corpus of manually-annotated acceleration measurements, including approximately 6000 cough and 68000 non-cough events, gathered from 14 patients in a small TB clinic. Although accelerometer-based detection of coughing has been considered before, it has always made use of sensors worn by the patient, which is in some respects intrusive and can be inconvenient. We have shown that excellent discrimination is also possible when the sensor is attached to the patient's bed. This presents a less intrusive method of cough monitoring, which can be of practical use in monitoring the recovery process of patients, for example in the clinic where the data was collected. Acceleration-based monitoring also has important privacy advantages over the detection of cough sounds by audio- based monitoring, which often raises privacy concerns, and we have found patients to be uncomfortable in the presence of audio-based monitoring equipment. Accelerometer-based monitoring, using a bed-mounted inexpensive consumer smartphone, represents a more discreet and also cost-effective alternative.

In ongoing work, we are attempting to optimise some of the Resnet50 metaparameters and also to incorporate audio [26] along with accelerometer measurements with a view to further improvement on cough detection results.

7. REFERENCES

[1] J Korpáš, J Sadloňová, and M Vrabec, "Analysis of the cough sound: an overview," *Pulmonary Pharmacology*, vol. 9, no. 5-6, pp. 261–268, 1996.

[2] J Knocikova, J Korpas, M Vrabec, and M Javorka, "Wavelet analysis of voluntary cough sound in patients with respiratory diseases," *Journal of Physiology and Pharmacology*, vol. 59, no. Suppl 6, pp. 331–40, 2008.

[3] GHR Botha, G Theron, RM Warren, M Klopper, K Dheda, PD Van Helden, and TR Niesler, "Detection of tuberculosis by automatic cough sound analysis," *Physiological Measurement*, vol. 39, no. 4, pp. 045005, 2018.

[4] Mahmood Al-khassaweneh and Ra'ed Bani Abdelrahman, "A signal processing approach for the diagnosis of asthma from cough sounds," *Journal of Medical Engineering & Technology*, vol. 37, no. 3, pp. 165–171, 2013.

[5] Renard Xaviero Adhi Pramono, Syed Anas Imtiaz, and Esther Rodriguez-Villegas, "A cough-based algorithm for automatic diagnosis of pertussis," *PloS one*, vol. 11, no. 9, pp. e0162128, 2016.

[6] Igor DS Miranda, Andreas H Diacon, and Thomas R Niesler, "A comparative study of features for acoustic cough detection using deep architectures," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2019, pp. 2601–2605.

[7] Helia Mohammadi, Ali-Akbar Samadani, Catriona Steele, and Tom Chau, "Automatic discrimination between cough and non-cough accelerometry signal artefacts," *Biomedical Signal Processing and Control*, vol. 52, pp. 394–402, 2019.

[8] Paul Munyard, Carolyn Busst, Ron Logan-Sinclair, and Andrew Bush, "A new device for ambulatory cough recording," *Pediatric Pulmonology*, vol. 18, no. 3, pp. 178–186, 1994.

[9] Lucy Pavesi, Subbu Subburaj, and Kerri Porter-Shaw, "Application and validation of a computerized cough acquisition system for objective monitoring of acute cough: a meta-analysis," *Chest*, vol. 120, no. 4, pp. 1121–1128, 2001.

[10] Ian M Paul, Kitman Wai, Steven J Jewell, Michele L Shaffer, and Vasundara V Varadan, "Evaluation of a new self-contained, ambulatory, objective cough monitor," *Cough*, vol. 2, no. 1, pp. 7, 2006.

[11] Michael Coyle, P Alexander Derchak, and Lance Jonathan Myers, "Systems and methods for monitoring cough," June 1 2010, US Patent 7,727,161.

[12] Jingqi Fan, German Comina, Robert Gilman, Jose Lopez, and Brian H Tracey, "Cough monitoring for pulmonary tuberculosis using combined microphone/accelerometer measurements," *The Journal of the Acoustical Society of America*, vol. 135, no. 4, pp. 2268–2268, 2014.

[13] Johnny Yat Ming Chan, Stephen Anthony Tunnell, and Joshua Aaron Lawrence Jacobs, "Systems, methods and kits for measuring cough and respiratory rate using an accelerometer," Sept. 4 2014, US Patent App. 13/783,257.

[14] Kota Hirai, Hideyuki Tabata, Mariko Hirayama, Tohru Kobayashi, Yasumasa Oh, and Hiroyuki Mochizuki, "A new method for objectively evaluating childhood nocturnal cough," *Pediatric Pulmonology*, vol. 50, no. 5, pp. 460–468, 2015.

[15] Thomas Drugman, Jerome Urbain, Nathalie Bauwens, Ricardo Chessini, Carlos Valderrama, Patrick Lebecque, and Thierry Dutoit, "Objective study of sensor relevance for automatic cough detection," *IEEE journal of Biomedical and Health informatics*, vol. 17, no. 3, pp. 699–707, 2013.

[16] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[17] Peter Wittenburg, Hennie Brugman, Albert Russel, Alex Klassmann, and Han Sloetjes, "ELAN: a professional framework for multimodality research," in *5th International Conference on Language Resources and Evaluation (LREC 2006)*, 2006.

[18] Jason Van Hulse, Taghi M Khoshgoftaar, and Amri Napolitano, "Experimental perspectives on learning from imbalanced data," in *Proceedings of the 24th International Conference on Machine learning*, 2007, pp. 935–942.

[19] Bartosz Krawczyk, "Learning from imbalanced data: open challenges and future directions," *Progress in Artificial Intelligence*, vol. 5, no. 4, pp. 221–232, 2016.

[20] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[21] Guillaume Lemaître, Fernando Nogueira, and Christos K Aridas, "Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 559–563, 2017.

[22] Anthony Windmon, Mona Minakshi, Pratool Bharti, Sriram Chellappan, Marcia Johansson, Bradlee A Jenkins, and Ponrathi R Athilingam, "Tussiswatch: A smart-phone system to identify cough episodes as early symptoms of chronic obstructive pulmonary disease and congestive heart failure," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 4, pp. 1566–1573, 2018.

[23] Bo Liang, SD Iwnicki, and Yunshi Zhao, "Application of power spectrum, cepstrum, higher order spectrum and neural network analyses for induction motor fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 39, no. 1-2, pp. 342–360, 2013.

[24] Gen Takahashi, Takeshi Yamada, Shoji Makino, and Nobutaka Ono, "Acoustic scene classification using deep neural network and frame-concatenated acoustic feature," *Detection and Classification of Acoustic Scenes and Events*, 2016.

[25] Claude Sammut and Geoffrey I Webb, "Leave-one-out cross-validation," *Encyclopedia of Machine Learning*, pp. 600–601, 2010.

[26] Madhurananda Pahar and Leslie S Smith, "Coding and Decoding Speech using a Biologically Inspired Coding System," in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, 2020, pp. 3025–3032.