

# Assessment of support vector machines and convolutional neural networks to detect snoring using Emfit mattress

Jose M. Perez-Macias – *IEEE-EMBS Student Member*, Sharath Adavanne – *IEEE Student Member*,  
Jari Viik, Alpo Värri, Sari-Leena Himanen, and Mirja Tenhunen.

**Abstract**— Snoring (SN) is an essential feature of sleep breathing disorders, such as obstructive sleep apnea (OSA). In this study, we evaluate epoch-based snoring detection methods using an unobtrusive electromechanical film transducer (Emfit) mattress sensor using polysomnography recordings as reference. Two different approaches were studied: a support vector machine (SVM) classifier fed with a subset of spectral features and convolutional neural network (CNN) fed with spectrograms. Representative 10-min normal breathing (NB) and SN periods were selected for analysis in 33 subjects and divided in thirty-second epochs. In the evaluation, average results over 10 fold Monte Carlo cross validation with 80% training and 20% test split were reported. Highest performance was achieved using CNN, with 92% sensitivity, 96% specificity, 94% accuracy, and 0.983 area under the receiver operating characteristics curve (AROC). Results showed a 6% average increase of performance of the CNN over SVM and greater robustness, and similar performance to ambient microphones.

## I. INTRODUCTION

Snoring (SN) is a breathing sound usually defined as a noise produced during the breathing cycle [1] and an early sign of upper-airway dysfunction. Previous research has found that 10–60% of the population suffer from frequent snoring [2]–[4]. Its prevalence increases with age and it is more frequent in men than women [2], [4]. Snoring has been related to with sleepiness, poor work performance, traffic accidents, insomnia, and cardiovascular diseases [2], [5]–[8]. The scoring guidelines of the American Academy of Sleep Medicine (AASM) recommend three different methods for SN detection [9]: a piezoelectric sensor, microphones, and nasal pressure transducer. Different studies have proposed methods for the detection of snoring using these sensors. Detachment of the piezo-sensor, moisture on the nasal prongs, and the availability of different microphone set-ups are among their disadvantages. A recent study comparing

This study was financially supported was supported by Yrjö ja Kalle Väisälän foundation and by Competitive State Research Financing of the Expert Responsibility area of Tampere University Hospital (Grants 9P013, 9R007, 9S007).

Jose Maria Perez-Macias and Jari Viik are with the Faculty of Biomedical Sciences and Engineering, University of Technology, Tampere, 33720 Finland (phone: +358 414913404; e-mail: jose.perez-macias.eng@ieee.org).

Sharath Adavanne and Alpo Värri are with the Department of Signal Processing, Tampere University of Technology, Tampere, 33720 Finland.

Mirja Tenhunen is with the Department of Clinical Neurophysiology, Medical Imaging Centre and Hospital Pharmacy, Pirkanmaa Hospital District, Tampere, Finland.

Sari-Leena Himanen is with the School of Medicine, University of Tampere, Tampere and the Department of Clinical Neurophysiology, Medical Imaging Centre and Hospital Pharmacy, Pirkanmaa Hospital District, Tampere, Finland.

these methods, emphasized the need of a standardized method to measure snoring and audio measuring techniques were suggested as the best candidate [10].

The use of mattress sensors for the diagnosis of sleep-disordered breathing (SDB) is increasing. The most commonly used mattress-like sensors are the Emfit and the polyvinylidene fluoride (PVDF). In Finland, the Emfit mattress is being used in the diagnosis of SDB: it has been found to reliably detect obstructive apneas and hypopneas as well as periods of prolonged partial obstruction [11], [12]. Snoring detection using a PVDF mattress sensor has recently been introduced by Hwang *et al.* They developed a method using two spectral features and a support vector machine (SVM) [13]. Our previous work using Emfit mattress, characterized the spectral differences between normal breathing (NB) and SN thirty-seconds epochs on three bands of interest: 6–16 Hz (BW1), 16–30 Hz (BW2), and 69–100 Hz (BW3) [14].

In the last years, amid rapid hardware developments, deep neural networks (DNN) have gained popularity in the fields of image classification [15] and speech recognition [16]. Convolutional neural networks (CNN) is a type of DNN that performs feature extraction by means of convolving a collection of filters. The first modern CNN architectures appeared on the late 90's [17], [18]. Since its introduction, several new studies have applied different CNN architectures in other fields of study.

The goal of this study is to compare the performance of SN detection methods using an Emfit mattress: (1) a SVM classifier fed with spectral features chosen by a feature selection step and (2) a CNN classifier fed with a time-frequency representation of the Emfit signal.

## II. SUBJECTS AND SIGNALS

### A. Subjects and Signals

The study sample consisted of 33 patients, from which three patients were discarded from the study due to bad signal quality. A cohort of 24 male and 6 female were used in the analysis. Their age ranged 25–60 years and their body mass index (BMI) ranged 22.2–54 kgm<sup>-2</sup>. The polysomnographies (PSGs) were recorded in the sleep laboratory of Tampere University Hospital, Finland. Informed consents were obtained before recording and the study was approved by the ethical committee of the Pirkanmaa Hospital District.

The PSGs were recorded and scored according to the AASM recommendations [9]. The recording comprised of the following signals: EEG (F3-A2, F4-A1, C3-A2, C4-A1, O1-A2, and O2-A1), ECG, two EOG channels, EMGs of the

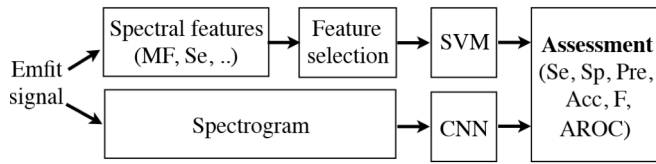


Figure 1. Diagram of the assessment procedure.

submental and tibialis muscles, pulse oximetry ( $SpO_2$ ), and body position. Breathing airflow was obtained using a nasal pressure cannula and a thermistor. Respiratory movements were recorded using two inductive belts placed across the thorax and then abdomen. SN was measured with a ambient microphone fixed on the ceiling of the patient room. The Emfit mattress sensor with dimensions  $32\text{ cm} \times 62\text{ cm} \times 0.4\text{ cm}$  was placed underneath the mattress situated under the thorax. Sampling frequency of 2 Hz was used for  $SpO_2$ , 10 Hz for respiratory movements, 100 Hz for the piezo sensor, 500 Hz for the ECG, and 200 Hz for the Emfit sensor and the remaining signals.

A clinical neurophysiologist selected periods of SN and NB with a maximum length of 10 min for each subject. The selection of these periods was performed by the following protocol: initial periods of SN were selected based on high-frequency components occurring from the nasal pressure transducer. At a second stage, the piezo-sensor placed on the neck was checked for the presence of the snoring signal using the envelope technique and a threshold of  $10\ \mu\text{V}$ . Finally, the resulting SN periods were validated audio signal from recorded video. A total number of 1007 epochs (median 34 and inter-quartile range (IQR) 10) were selected: 463 NB epochs (median 17.5, IQR 8) and 544 SN epochs (median 20, IQR 4). Twelve epochs (10 NB, 2 SN) were discarded due to the presence of artifacts.

The Emfit signal was filtered using a notch filter at 50 Hz to remove the power line interference. A 500th order finite impulse response filter (FIR) designed with a Hamming window and cut-off frequency of 6 Hz. The high order was used to achieve a short transition bandwidth.

### III. METHODS

Two approaches for the SN epoch detection were compared: a SVM fed from a subset of spectral features and a CNN fed with thirty-second epochs spectrograms. A diagram of the procedure is illustrated in Fig. 1.

#### A. Feature extraction and selection

Based on previous work [14], the Emfit recording was parametrized using a set of spectral features on three spectral bands. These were derived from the normalized power spectral density (PSD) estimated using a nonparametric method known as the Welch method [19]. This is a suitable spectral estimation method for non-stationary signals reducing the variance of the power spectrum [20]. A window of 100 samples with 50% overlap along a 512-sample discrete Fourier transform (DFT) was employed.

A collection of 35 features were extracted from each window in three bands: 6–16 Hz (BW1), 16–30 Hz (BW2), and 60–100 (BW3). In the following notation, we use subscripts 1–3 to represent BW1–3 bands. The absolute

power ( $AP_{1-3}$ ), the peak amplitude ( $PA_{1-3}$ ), the relative power ( $RP_{1-3}$ ), the spectral entropy ( $SE_{1-3}$ ), four statistical moments ( $m1f_{1-3}$ ,  $m2f_{1-3}$ ,  $m3f_{1-3}$ , and  $m4f_{1-3}$ ), and the form factor ( $F_{1-3}$ ). Additionally, power ratios between bands (PR1–5). To summarize the information from the epoch, the first three statistical moments of all features were estimated per epoch, resulting in  $35 \times 3$  features. A  $p$ -dimensional feature vector ( $p=105$ ) was constructed for each epoch. A feature selection algorithm known as max-relevance min-redundancy (MRMR) was selected to reduce the dimensionality of the problem. MRMR selects features by reducing the maximal statistical dependency criterion using mutual information [21]. This method it was fast and yielded reasonable results in line with our previous study [14]. The joint mutual information and the conditional mutual information maximization criterion were also tested and discarded due to lower performance.

#### B. Spectrogram estimation

Each epoch of the Emfit signal was represented using a time-frequency representation called the spectrogram or short-time Fourier transform (STFT). A Hanning window of length 32 samples, 50% overlap, and a 512-sample DFT was used to estimate the spectrogram. The magnitude spectrogram was estimated as the absolute value of the STFT and normalized by the median energy in the 6–10 Hz spectral band. This normalization takes this band as the reference intensity. Finally, the normalized spectrogram was *log*-transformed.

#### C. Classification

##### 1) SVM using selected features

The subset of features were fed to the SVM classifier with a radial basis (RBF) kernel function and soft margin cost equal to one. The SVM classifier finds the optimal hyperplane between two classes by maximizing the distance to the nearest data points called support vectors. The SVM was chosen because it has been shown to perform reliably under different datasets [22].

##### 2) CNN architecture

The proposed CNN architecture (Fig. 2) contains three convolutional layers, each of which is followed by subsampling layers (or max-pooling). The feature parameters from the CNN are flattened before the final fully connected layer with one node. Thus, the architecture comprises of four learned layers: three 2D convolutional layer and one fully-connected layer with one output.

The three 2D convolutional layers extract local shift-invariant features. We used 32 filters with  $[3 \times 3]$  kernel size were employed. To ensure the same input and output size, zero-padding was employed in the 2D convolution. Each convolutional layer was followed by max-pooling layers of size  $[6 \times 5 \times 5]$  in frequency and  $[5 \times 5 \times 5]$  in time dimension. Pooling in between convolutional layers allows more compact representations and increases robustness [23]. All three convolutional layers used batch normalization [24] and a rectified liner unit (ReLU) as an activation function [25], [26]. The use of the ReLU activation allows faster training times than the equivalent sigmoid activation functions [15].

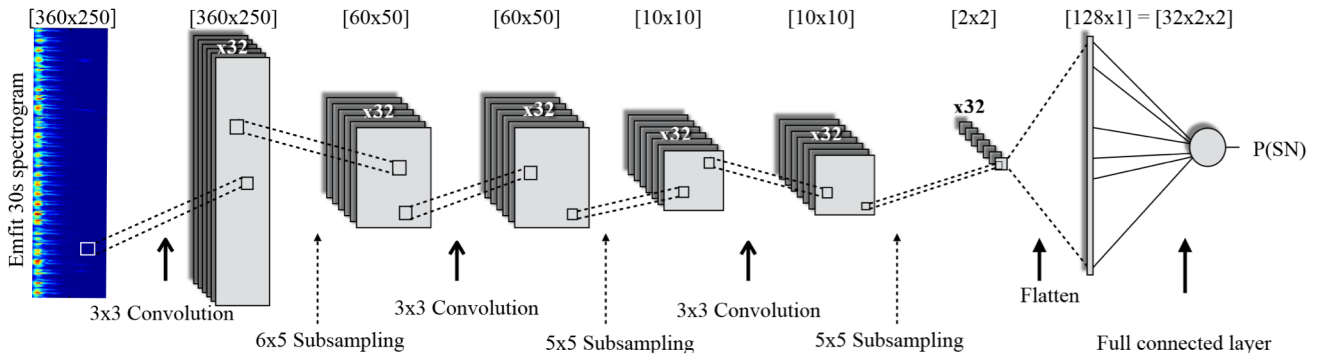


Figure 2. Architecture of the Convolutional Neural Network. The input is the spectrogram of a thirty-seconds Emfit signal. The output is the probability of snoring,  $P(\text{SN})$ .

The neural network was trained by backpropagation through time [27] using a first-order gradient based optimizer (Adam) [28] and the root mean-square error (RMSE) as the objective function. Additionally, due to the heterogeneous nature of our cohort, we used the dropout technique to reduce overfitting [29].

To keep the matrices factorable during the max-pooling operations the thirty-seconds epoch spectrograms with dimensions  $[372 \times 256]$  were trimmed. Thus, the input of the network were spectrograms of size  $[360 \times 250]$  and one output representing the probability of SN.

#### D. Training and testing of the classifiers

We evaluated the results over 10-fold Monte Carlo cross validation (CV) with 80% training and 20% test split (10 random 24 and 6 splits for training and testing, respectively) [30].

In the feature-SVM scenario, the resulting features from the feature selection stage were  $z$ -scored. In the spectrogram-CNN scenario we used a dropout of 0.50, a batch size of 8, and a maximum number of 300 iterations; our tests showed that beyond this this number of iterations the network stopped learning.

#### E. Data and performance assessment

MATLAB (R-2013b, The MathWorks, Inc., Natick, MA, USA) was used for signal processing, feature extraction feature selection, and the SVM classifier. The feature selection step was performed using FEAST ToolBox [31]. The CNN was implemented using the Keras framework [32] and Theano backend [33]. The diagnostic performance was assessed with the following performance indicators: sensitivity (percentage of SN epochs correctly identified), specificity (percentage of NB epochs correctly identified), precision (proportion of SN epochs that that are true positives), accuracy (proportion of SN and NB epochs correctly classified), the F-measure (the weighted harmonic mean between sensitivity and precision), and the area under the receiver operating characteristics curve (AROC).

## IV. RESULTS

Results are displayed in Table I. Performance scores are given using their mean and standard deviation. Overall, both classifiers show high diagnostic performance reaching higher scores than 85% for sensitivity, specificity, and F-measure.

Highest diagnostic performance was achieved using CNN performing above 91% for the sensitivity, specificity and the F-measure; average AROC was  $0.983 \pm 0.002$ . In comparison with the feature-SVM architecture, the CNN performed an average of 5.6% better and 3.4% less variability between Monte Carlo simulations (these percentages were estimated as the mean of the differences of all performance indicators)

The feature selection algorithm (MRMR) was tested on 5–20 number of features. Best classification results were obtained with six features: the first statistical moment of the PR<sub>3</sub>, the second statistical moments of RP<sub>3</sub> and  $m1f_1$ , and finally the third statistical moments of AP<sub>3</sub>, SE<sub>3</sub>, and PR<sub>4</sub>.

## V. DISCUSSION AND CONCLUSIONS

We have evaluated the suitability of modern DNNs over feature-based classification approaches on an Emfit mattress. For this purpose, a subset of spectral features was fed into SVM classifier. Its performance was compared with a DNN-CNN which architecture was specifically designed for this problem. Results were evaluated using a 10-fold Monte Carlo CV with 80-20 splits.

The CNN outperformed the SVM in all performance indicators with significantly less variability, suggesting greater robustness. However, the lack of time and non-linear features on the dataset might have improved slightly the SVM performance. Also, variability between runs was expected, specially due to the heterogeneous nature of the cohort. The CNN uses low-level features obtained from the spectrogram; when compared with known spectral features used by the SVM, it outperforms it with the added advantage

TABLE I. DIAGNOSTIC ASSESSMENT OF THE TWO CLASSIFICATION SCHEMAS BY MEANS OF 10-FOLD MONTE CARLO CV WITH 80-20 SPLIT (MEAN, SD)

	Sensitivity	Specificity	Precision	Accuracy	F-measure	AROC
SVM	$0.908 \pm 0.033$	$0.854 \pm 0.094$	$0.890 \pm 0.063$	$0.885 \pm 0.038$	$0.898 \pm 0.038$	$0.935 \pm 0.020$
CNN	$0.916 \pm 0.025$	$0.965 \pm 0.016$	$0.962 \pm 0.016$	$0.942 \pm 0.006$	$0.938 \pm 0.008$	$0.983 \pm 0.002$

CV, cross-validation, SVM, Support Vector Machine; CNN, convolutional neural network; AROC, area under the receiving operating characteristics curve; SD, standard deviation.

of auto-generating most suitable low-level set of features.

Compared with previous pieces of research [13], our feature-SVM scenario reached slight worse results on 2.1% average, however, the CNN scenario improved available scores by 3.7%. Compared with tracheal and ambient microphones our method performs close to ambient microphones (sensitivity 93.1% and precision 95.9%), whereas with tracheal microphones (sensitivity 98.6%, precision 94.8) [34]. However, Emfit has the added benefit of being unobtrusive, it cannot be accidentally detached (unlike tracheal microphones), and produces less than 2% of data —requiring less computational and storing resources— in contrast to audio measuring techniques.

Having tested both methods, we conclude that both yield good results with the CNN providing a significant boost of performance over the other. Additionally, it poses as a suitable candidate as a sensor to measure snoring.

#### REFERENCES

- [1] D. Pevernagie, R. M. Aarts, and M. De Meyer, "The acoustics of snoring.," *Sleep Med Rev*, vol. 14, no. 2, pp. 131–144, Apr. 2010.
- [2] E. Lindberg, A. Taube, C. JANSON, T. GISLASON, K. Svärdsudd, and G. Boman, "A 10-Year Follow-up of Snoring in Men," *Chest*, vol. 114, no. 4, pp. 1048–1055, Oct. 1998.
- [3] M. Svensson, K. A. Franklin, J. Theorell-Haglöw, and E. Lindberg, "Daytime Sleepiness Relates to Snoring Independent of the Apnea-Hypopnea Index in Women From the General Population," *Chest*, vol. 134, no. 5, pp. 919–924, Nov. 2008.
- [4] S. Spörndly-Nees, P. Åsenlöf, J. Theorell-Haglöw, M. Svensson, H. Igelström, and E. Lindberg, "Leisure-time physical activity predicts complaints of snoring in women: a prospective cohort study over 10 years.," *Sleep Med.*, vol. 15, no. 4, pp. 415–421, Apr. 2014.
- [5] J. Ulfberg, N. Carter, M. Talbäck, and C. Edling, "Excessive daytime sleepiness at work and subjective work performance in the general population and among heavy snorers and patients with obstructive sleep apnea.," *Chest*, vol. 110, no. 3, pp. 659–663, Sep. 1996.
- [6] D. J. Gottlieb, Q. Yao, S. Redline, T. Ali, and M. W. Mahowald, "Does snoring predict sleepiness independently of apnea and hypopnea frequency?," *Am J Respir Crit Care Med*, vol. 162, no. 4, pp. 1512–1517, Oct. 2000.
- [7] M. Koskenvuo, J. Kaprio, K. Heikkilä, S. Sarna, T. Telakivi, and M. Partinen, "Snoring as a risk factor for ischaemic heart disease and stroke in men.," vol. 294, no. 6572, pp. 643–19, Mar. 1987.
- [8] F. J. Nieto, T. B. Young, B. K. Lind, E. Shahar, J. M. Samet, S. Redline, R. B. D'Agostino, A. B. Newman, M. D. Lebowitz, and T. G. Pickering, "Association of sleep-disordered breathing, sleep apnea, and hypertension in a large community-based study. Sleep Heart Health Study.," *JAMA*, vol. 283, no. 14, pp. 1829–1836, Apr. 2000.
- [9] R. B. Berry, R. Budhiraja, D. J. Gottlieb, D. Gozal, C. Iber, V. K. Kapur, C. L. Marcus, R. Mehra, S. Parthasarathy, S. F. Quan, S. Redline, K. P. Strohl, S. L. Davidson Ward, M. M. Tangredi, American Academy of Sleep Medicine, "Rules for scoring respiratory events in sleep: update of the 2007 AASM Manual for the Scoring of Sleep and Associated Events. Deliberations of the Sleep Apnea Definitions Task Force of the American Academy of Sleep Medicine.," *Journal of clinical sleep medicine : JCSM : official publication of the American Academy of Sleep Medicine*, vol. 8, no. 5, pp. 597–619, 15-Oct-2012.
- [10] E. S. Arnardottir, B. Isleifsson, J. S. Agustsson, G. A. Sigurdsson, M. O. Sigurgunnarsdottir, G. T. Sigurdarson, G. Saevarsson, A. T. Sveinbjarnarson, S. Hoskuldsson, and T. GISLASON, "How to measure snoring? A comparison of the microphone, cannula and piezoelectric sensor," *J Sleep Res*, vol. 25, no. 2, pp. 1–11, Nov. 2015.
- [11] M. Tenhunen, E. Rauhala, J. Virkkala, O. Polo, A. Saastamoinen, and S.-L. Himanen, "Increased respiratory effort during sleep is non-invasively detected with movement sensor.," *Sleep Breath*, vol. 15, no. 4, pp. 737–746, Nov. 2011.
- [12] M. Tenhunen, E. Elomaa, H. Sistonen, E. Rauhala, and S.-L. Himanen, "Emfit movement sensor in evaluating nocturnal breathing.," *Respiratory Physiology & Neurobiology*, vol. 187, no. 2, pp. 183–189, Jun. 2013.
- [13] S. H. Hwang, C. M. Han, H. N. Yoon, D. W. Jung, Y. J. Lee, D.-U. Jeong, and K. S. Park, "Polyvinylidene fluoride sensor-based method for unconstrained snoring detection.," *Physiol. Meas.*, vol. 36, no. 7, pp. 1399–1414, Jul. 2015.
- [14] J. M. Perez-Macias, J. Viik, A. Väri, S.-L. Himanen, and M. Tenhunen, "Spectral analysis of snoring events from an Emfit mattress.," *Physiol. Meas.*, vol. 37, no. 12, pp. 2130–2143, Nov. 2016.
- [15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks.," *NIPS*, pp. 1097–1105, 2012.
- [16] G. Hinton, L. Deng, D. Yu, G. E. Dahl, A.-R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups.," *Signal Processing Magazine, IEEE*, vol. 29, no. 6, pp. 82–97, 2012.
- [17] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 5, pp. 530–535, May 1997.
- [18] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition.," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [19] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms.," *IEEE Transactions on Audio and Electroacoustics*, vol. 15, no. 2, pp. 70–73, Jun. 1967.
- [20] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. John Wiley & Sons, 2009.
- [21] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy.," *PAMI*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
- [22] D. Meyer, F. Leisch, and K. Hornik, "The support vector machine under test.," *Neurocomputing*, vol. 55, no. 1, pp. 169–186, Sep. 2003.
- [23] Y.-L. Boureau, J. Ponce, and Y. LeCun, "A Theoretical Analysis of Feature Pooling in Visual Recognition.," *arXiv*, 2010.
- [24] S. Ioffe and C. Szegedy, "Batch Normalization - Accelerating Deep Network Training by Reducing Internal Covariate Shift.," *arXiv*, vol. cs.LG, 2015.
- [25] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines.," presented at the Proceedings of the 27th international ..., 2010.
- [26] X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks.," *AISTATS*, 2011.
- [27] P. J. Werbos, "Backpropagation through time: what it does and how to do it.," *Proc. IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990.
- [28] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization.," *arXiv.org*, vol. cs.LG. 22-Dec-2014.
- [29] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout - a simple way to prevent neural networks from overfitting.," *JMLR*, 2014.
- [30] W. Dubitzky, M. Granzow, and D. Berrar, Eds., *Fundamentals of Data Mining in Genomics and Proteomics*. Boston, MA: Springer US, 2007, pp. 1–281.
- [31] G. Brown, A. Pocock, M.-J. Zhao, and M. Luján, "Conditional Likelihood Maximisation: A Unifying Framework for Information Theoretic Feature Selection.," *JMLR*, vol. 13, no. Jan, pp. 27–66, 2012.
- [32] F. Chollet, "Keras.," *GitHub repository https://github.com/fchollet/keras*, 2015. [Online]. Available: <https://github.com/fchollet/keras>. [Accessed: 13-Feb-2017].
- [33] J. Bergstra, O. Breuleux, and F. Bastien, "Theano: A CPU and GPU math compiler in Python.," *Proc 9th Python ...*, 2010.
- [34] A. Azarbarzin and Z. M. K. Moussavi, "Automatic and Unsupervised Snore Sound Extraction From Respiratory Sound Signals.," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 5, pp. 1156–1162, 2011.