Aykanat et al. EURASIP Journal on Image and Video Processing (2017) 2017:65
DOI 10.1186/s13640-017-0213-2

EURASIP Journal on Image and Video Processing

RESEARCH

Open Access

# Classification of lung sounds using convolutional neural networks

Murat Aykanat[1*], Özkan Kılıç[1], Bahar Kurt[2] and Sevgi Saryal[3]

**Abstract**

In the field of medicine, with the introduction of computer systems that can collect and analyze massive amounts of data, many non-invasive diagnostic methods are being developed for a variety of conditions. In this study, our aim is to develop a non-invasive method of classifying respiratory sounds that are recorded by an electronic stethoscope and the audio recording software that uses various machine learning algorithms.

In order to store respiratory sounds on a computer, we developed a cost-effective and easy-to-use electronic stethoscope that can be used with any device. Using this device, we recorded 17,930 lung sounds from 1630 subjects. We employed two types of machine learning algorithms; mel frequency cepstral coefficient (MFCC) features in a support vector machine (SVM) and spectrogram images in the convolutional neural network (CNN). Since using MFCC features with a SVM algorithm is a generally accepted classification method for audio, we utilized its results to benchmark the CNN algorithm. We prepared four data sets for each CNN and SVM algorithm to classify respiratory audio: (1) healthy versus pathological classification; (2) rale, rhonchus, and normal sound classification; (3) singular respiratory sound type classification; and (4) audio type classification with all sound types. Accuracy results of the experiments were; (1) CNN 86%, SVM 86%, (2) CNN 76%, SVM 75%, (3) CNN 80%, SVM 80%, and (4) CNN 62%, SVM 62%, respectively.

As a result, we found out that spectrogram image classification with CNN algorithm works as well as the SVM algorithm, and given the large amount of data, CNN and SVM machine learning algorithms can accurately classify and pre-diagnose respiratory audio.

**Keywords:** Machine learning, Convolutional neural networks, Deep learning, Lung sounds

## 1 Introduction

Diagnosis or classification requires recognizing patterns. But most of the time, it is very hard to spot these patterns, especially if the data is very large. Data collected from the environment is usually non-linear, so we cannot use traditional methods to find patterns or create mathematical models. In the past decade, various technologies, such as expert systems, have been used to attempt to solve this problem. However, for critical systems, the error rate for the decision was too high [1].

The latest technology that is attempting to solve this problem is machine learning. Over the years, various successful algorithms were developed and now with the deep learning algorithms, error rate became close to negligible. Especially in computer vision and speech recognition, machine learning is reaching human levels of detection.

Research in this area attempts to make better representations and create models to learn these representations from large-scale unlabeled data [2]. Some of the representations are inspired by advances in neuroscience and are loosely based on interpretation of information processing and communication patterns in a nervous system, such as neural coding which attempts to define a relationship between the stimulus and the neuronal responses and the relationship among the electrical activities of the neurons in the brain [3, 4].

Deep learning is a branch of machine learning based on a set of algorithms that attempt to model high-level abstractions in data by using model architectures, with complex structures, composed of multiple non-linear transformations [3, 5]. An observation (e.g., an image)

* Correspondence: aykanatm@gmail.com
[1]Faculty of Engineering, Department of Computer Engineering, Yıldırım Beyazıt University, Ankara, Turkey
Full list of author information is available at the end of the article

can be represented in many ways including a vector of intensity values per pixel, or in a more abstract way as a set of edges, regions of a particular shape, and various other features. Some representations make it easier to learn tasks (e.g., face recognition or facial expression recognition) from examples [6–8]. One of the promises of deep learning is replacing handcrafted features with efficient algorithms for unsupervised or semi-supervised feature learning and hierarchical feature extraction [9].

Various deep learning architectures such as deep neural networks, convolutional deep neural networks, deep belief networks, and recurrent neural networks have been applied to fields like computer vision, automatic speech recognition, natural language processing, audio recognition, and bioinformatics where they have been shown to produce state-of-the-art results on various tasks [5, 10].

The convolutional network architecture is a remarkably versatile yet conceptually simple paradigm that can be applied to a wide spectrum of perceptual tasks. Convolutional networks are trainable, multistage architectures. The input and output of each stage are sets of arrays called feature maps [11]. Convolutional neural networks (CNNs) are designed to process data that come in the form of multiple arrays. There are four key ideas behind CNN that take advantage of the properties of natural signals: local connections, shared weights, pooling, and the use of many layers. The architecture of a typical CNN is structured as a series of stages. The first few stages are composed of two types of layers: convolutional layers and pooling layers. Units in a convolutional layer are organized in feature maps, within which each unit is connected to local patches in the feature maps of the previous layer through a set of weights called a filter bank. Although the role of the convolutional layer is to detect local conjunctions of features from the previous layer, the role of the pooling layer is to merge semantically similar features into one [12]. The CNN has been found highly effective and has been commonly used in computer vision and image recognition. More recently, with appropriate changes from designing CNN for image analysis to taking into account speech-specific properties, the CNN is also found effective for speech recognition [13].

Auscultation, which is the processes of listening to the internal sounds in the human body through a stethoscope, has been an effective tool for the diagnosis of lung disorders and abnormalities. This process mainly relies on the physician. Using a stethoscope, the physicians may hear normal breathing sounds, decreased or absent breath sounds, and abnormal breath sounds (e.g., rale, rhonchus, squawk, stridor, wheeze, rub) [14, 15]. Auscultation is a simple, patient-friendly and non-invasive method which is widely used but is of low diagnostic value due to the inherent subjectivity in the evaluation of respiratory sounds and to the difficulty involved in relating qualitative assessments to other people [16].

Murphy et al. built a system for automatically providing an accurate diagnosis based upon an analysis of recorded lung sounds. The sound input comes from a number of microphones that are placed around a patient's chest. The system also has a signal processing circuit to convert data from analog to digital. This data is then recorded, organized, and displayed on a computer monitor using an application program. From each microphone, sound data was gathered both in inspiration and in expiration, combined and separately, so that abnormal sounds could be determined easily. The collected data is then manually analyzed, and a diagnosis is reached [17]. This invention proves that respiratory audio data can be collected from patients in a non-invasive way. However, this invention does not use an automated analysis technique to analyze the data.

In this study, we aim to improve on this invention by analyzing audio data with machine learning algorithms and by classifying respiratory sounds. Our data consists of audio recordings of lung sounds that were recorded by chest physicians. We believe, using machine learning, audio data can be analyzed for patterns that will lead to the detection of various pathological lung sounds and help in the diagnosis of respiratory conditions.

## 2 Materials and methods

### 2.1 Building the electronic stethoscope

First of all, since we needed a device to record respiratory audio, we started by researching all commercially available electronic stethoscopes. Two models are currently used in medicine: the Littman 2100 electronic stethoscope [18] and the Thinklabs One electronic stethoscope [19]. These devices simply receive audio signals from the head of the stethoscope by a microphone and a series of electronic circuits and transmit this digital signal into the computer by the 3.5-mm microphone jack commonly found on computers and mobile devices. However, the key difference was Littman 2100 electronic stethoscope required proprietary software, so it was constrained to certain platforms. On the other hand, Thinklabs One electronic stethoscope transmits the audio signal to any device using any software [20]. After analyzing the capabilities of these devices, we decided to build our own custom electronic stethoscope which has a directional microphone strapped inside the head of a stethoscope with a 3.5-mm microphone jack.

Since we do not have a signal enhancing hardware, we needed a good, small and directional microphone to

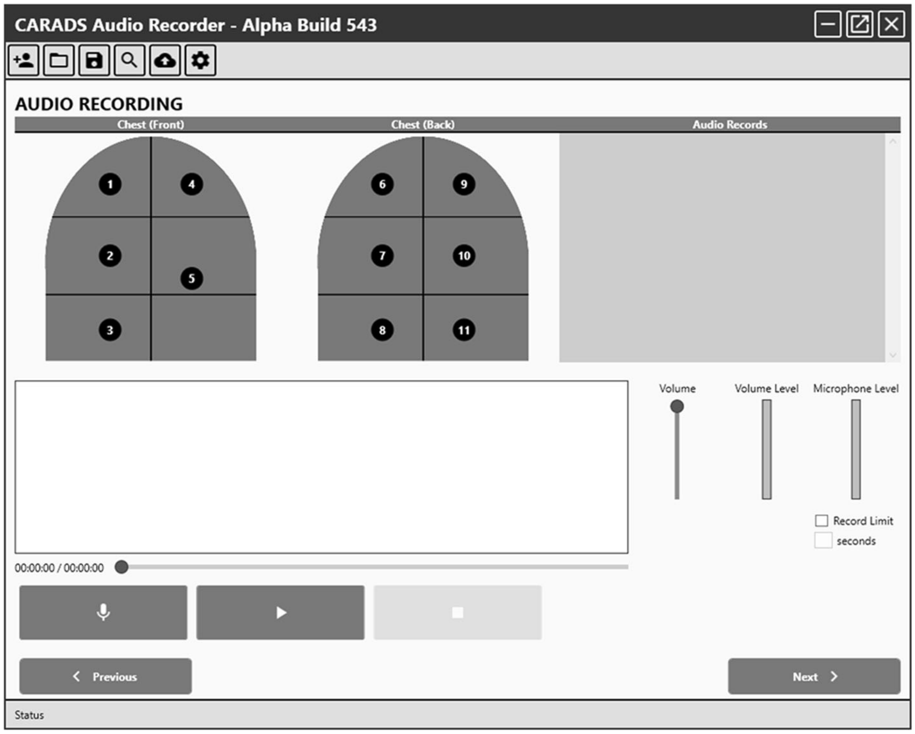


**Fig. 1** Audio recorder interface

obtain the perfect signal. However, the audio was still noisy because of several reasons:

- Hospital environments are naturally very noisy: people talking, phones, noisy devices, ambulance, police sirens, etc.
- There is a scratching noise when the diaphragm of the stethoscope comes in contact with dry skin and body hair.

The first problem is difficult to solve because it is impossible to sound proof the rooms where patients are. But the second problem can be solved simply by lubricating the area of contact. We also discovered that this method increases the reception of low-frequency audio by the microphone.

## 2.2 Software for data acquisition

We needed an application to record audio and save patient data. To this end, we developed a .NET application that creates patient records and uses open source audio library "NAudio" to record, play, and modify audio. It has two main sections:

- Patient information: first name, last name
- Audio recording: audio recordings from 11 areas of the patient's chest (Fig. 1).

The application and the hardware are tested together by recording respiratory audio and showing the results to the chest physicians.

## 2.3 Data acquisition

After receiving positive feedback from all the chest physicians, we decided to move to data acquisition. In the end, three hospitals agreed to participate in our research in their respiratory diseases department: Ankara University, Yıldırım Beyazıt University, and Yıldırım Beyazıt Education and Research Hospital.

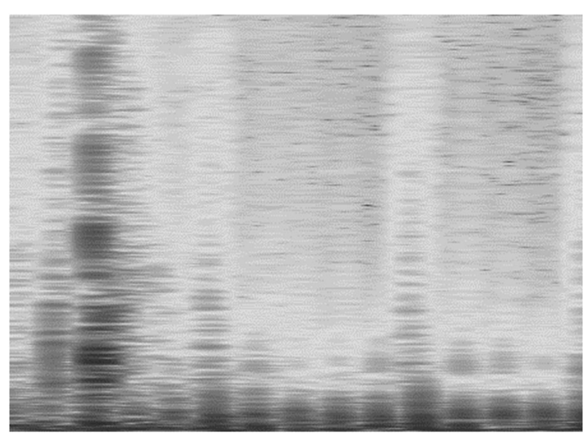

**Fig. 2** Example of respiratory sound spectrogram

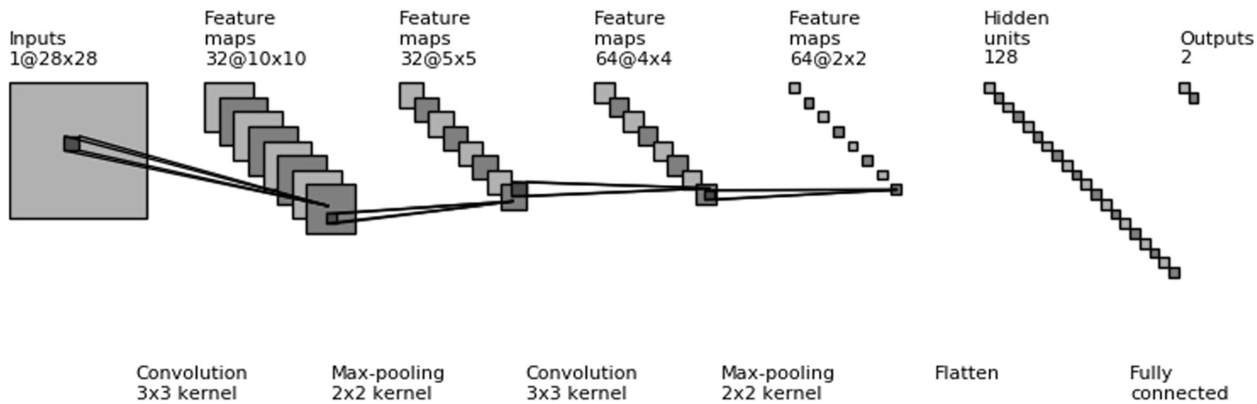


**Fig. 3** CNN structure for classifying pathologic and normal sound types

To start the data acquisition, we needed a laptop with a good audio card. Lenovo ThinkPad E550 Laptop offered the best audio card for our purposes. So we purchased the computer. We also purchased two Seagate Expansion 1 TB external hard drives for backup storage. Once we were set with the equipment, we started the data acquisition. We recorded respiratory audio from 1630 subjects and 11 positions from each patient, totaling to 17,930 audio clips, each 10-s long.

### 2.4 Experiments

In this study, we used two feature extraction methods: mel frequency cepstral coefficient (MFCC) feature extraction and spectrogram generation using short-time Fourier transform (STFT).

In sound processing, the mel frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a non-linear mel scale of frequency. MFCCs are coefficients that collectively make up an MFC. They are derived from a type of cepstral representation of the audio clip [17].

A spectrogram is a visual representation of the spectrum of frequencies in a sound or other signal as they vary with time or some other variable. They are used extensively in the fields of music, sonar, radar, and speech processing and seismology [21].

Since MFCC features are widely used in audio detection systems, the experiments we ran using the MFCC features enabled us to find a base value for accuracy, precision, recall, sensitivity, and specificity. Spectrogram images are also used in audio detection. However, they were never tested in respiratory audio with CNNs. We wanted to see if we can match or exceed the audio detection accuracies with MFCC features.

MFCC datasets were built using SciPy library. We used support vector machines to process these datasets. The spectrogram dataset was built using a combination of open source graph generation library Pylab and various open source image processing libraries. The original spectrograms generated were 800 × 600 RGBA, and they were too large for our computer's memory. We changed the algorithm to generate them 28 × 28 grayscale to fit them into the memory for CNN to process (Fig. 2).

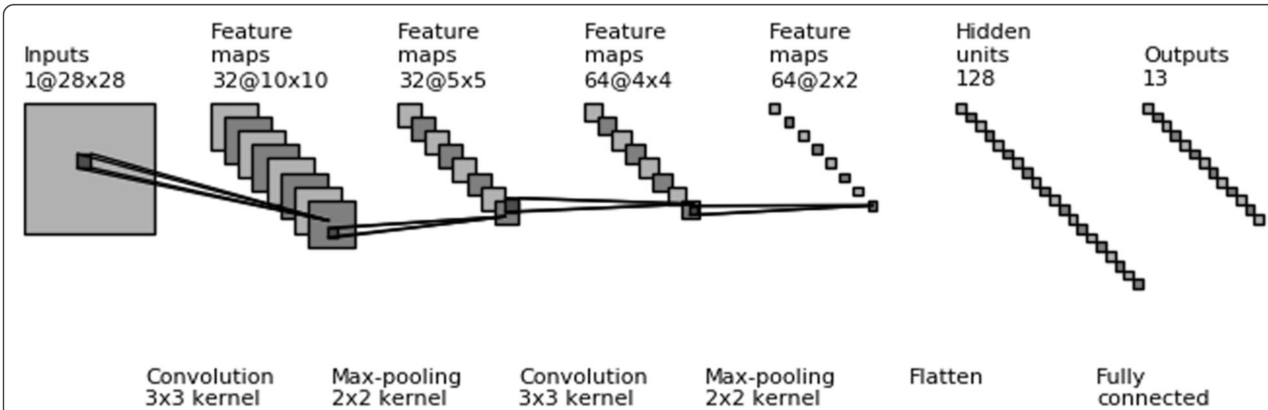


**Fig. 4** CNN structure for classifying all singular sound types

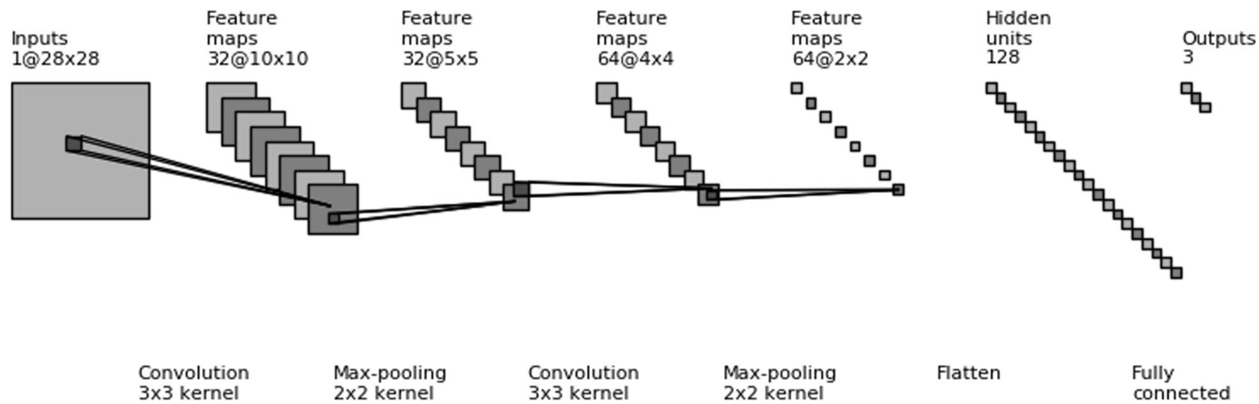


**Fig. 5** CNN structure for classifying rale, rhonchus, and normal sounds

We built eight datasets, four for support vector machines (SVMs) and four for convolutional neural networks (CNNs):

- Two datasets to predict whether respiratory sounds were normal or pathological (17,930 audio clips, two classes)
- Two datasets to classify respiratory sounds into: normal, rhonchus, squeak, stridor, wheeze, rales, bronchovesicular, friction rub, bronchial, absent, decreased, aggravation, or long expirium duration (LED) (14,453 audio clips, 13 classes)
- Two datasets for classification of respiratory sounds labeled with rale, rhonchus, or normal (15,328 audio clips, 3 classes)
- Two datasets for classification of respiratory sounds with all labels including ones with multiple labels (17,930 audio clips, 78 classes)

The CNN structures that we used in our experiments are shown below in Figs. 3, 4, 5, and 6.

## 3 Results and discussion

Our results are in Table 1.

A number of investigations demonstrating the usefulness of computerized lung sound analysis have been reported [22–24]. However, there is a small number of studies available on the clinical utility of auscultation and computerized lung sound analysis for the classification of abnormal lung sounds (Table 2).

As shown in Table 2, the studies in the literature have very limited datasets with a maximum of 2127 audio samples from 34 subjects [25]. Therefore, their accuracy results were either very high when there was a very distinct set of audio data or very low when the audio data was similar [16, 25–37]. This is a major problem as these systems deal with a critical decision in patient's diagnosis. In our study, we collected 11 audio recordings from each of the 1630 healthy and sick subjects totaling to 17,930 audio clips. Because of the larger size of our dataset, we managed to get consistent results in all our experiments.

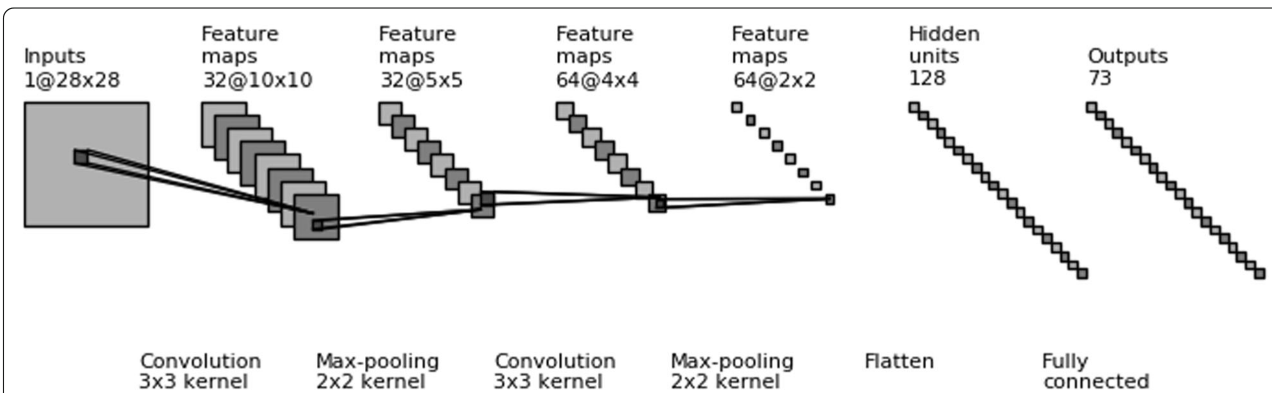


**Fig. 6** CNN structure for classifying all lung sounds

**Table 1** Experiment results

| | Training accuracy | Test accuracy | Training precision | Test precision | Training recall | Test recall | Training sensitivity | Test sensitivity | Training specificity | Test specificity |
|---|---|---|---|---|---|---|---|---|---|---|
| Classification of healthy versus pathological respiratory sounds | | | | | | | | | | |
| CNN (spectrogram) | 87% | 86% | 90% | 86% | 89% | 86% | 89% | 86% | 95% | 86% |
| SVM (MFCC) | 91% | 86% | 94% | 89% | 87% | 87% | 87% | 87% | 87% | 82% |
| Classification of respiratory sounds labeled with a singular type | | | | | | | | | | |
| CNN (spectrogram) | 90% | 76% | 94% | 79% | 86% | 74% | 86% | 74% | NA | NA |
| SVM (MFCC) | 99% | 75% | 99% | 75% | 99% | 99% | 99% | 99% | NA | NA |
| Classification of respiratory sounds labeled with only as type rale, rhonchus, and normal | | | | | | | | | | |
| CNN (spectrogram) | 87% | 80% | 88% | 79% | 85% | 79% | 85% | 79% | NA | NA |
| SVM (MFCC) | 89% | 80% | 89% | 80% | 89% | 89% | 89% | 89% | NA | NA |
| Classification of respiratory sounds with all labels | | | | | | | | | | |
| CNN (spectrogram) | 74% | 62% | 80% | 73% | 66% | 56% | 66% | 56% | NA | NA |
| SVM (MFCC) | 78% | 62% | 78% | 62% | 78% | 78% | 78% | 78% | NA | NA |

*NA* not applicable

In the literature, the audio clip size varies between 8 and 16 s. Similarly, we recorded all our audio clips in 10 s, as suggested by the chest physicians whom we worked with. In other studies, while commercially available devices and software packages were used, we developed our own hardware and software using open source libraries. Previous studies did not mention the audio format used. This can be an issue as some audio formats sacrifice quality for disk space. We used lossless WAV format as we did not want to lose any data.

Rietveld et al. [38] selected clean audio samples, and Baydar et al. [28] recorded their audio clips in a quiet room. However, if one tries to build a system that is trained from these clean data, it would not work in a real environment such as a hospital. Even the quietest hospital rooms have noise that would impact the recording. That is why we developed our electronic stethoscope with as much sound isolation as possible and selected our recording device carefully. In the end, the data we collected had very little external noise but it was collected from a real environment.

In the literature, lung sound classification was made for a maximum of six classes. Kandaswamy et al. [28] implemented a system to classify the lung sounds to one of the six categories: normal, wheeze, crackle, squawk, stridor, or rhonchus. Forkheim et al. [39], investigated to detect only wheezes in isolated lung sound segments. Bahoura et al. [27], Riella et al. [40], and Hashemi et al. [41] classified sounds as whether containing wheezes or normal respiratory sounds. Lu et al. [42] classified fine crackles and coarse crackles. Kahya et al. [15, 30], Flietstra et al. [24], and Serbes et al. [35] classified the presence or absence of a crackle. These studies are very narrow in scope, as they have limited number of classes. Their results are focused on only a few sound types. In our study, we performed 8 different experiments with 2, 3, 13, and 78 classes, diversifying our results greatly.

Previous studies so far used CNNs for classification. In our study, we aimed to use this new classification algorithm on audio and observed that it performs very well and produces consistent results.

Lu et al. [42] acquired their test data set from RALE and ASTRA databases. Riella et al. [40] used lung sounds that were available electronically from different online repositories. The problem with this approach is that the recording hardware and software can be different for each audio clip. This would cause problems in classification because the audio quality is not consistent in all training and test samples. In our study, we used a single recording device and the same recording software on the same device while recording the audio.

While several previous studies [16, 30, 39, 43] compared several algorithms, they did not use a widely accepted audio classification method for benchmarking their neural networks. In our study, we used the classification results of SVMs that use the MFCC features to benchmark our CNN algorithm.

In some studies in the literature in Table 2, the number of audio data or subjects were not mentioned; therefore, it is impossible to compare the results of these studies with our own [39, 40, 42, 44–46].

**Table 2** Machine learning in computerized respiratory sound analysis systems

| Author | Subjects | Classified items | Feature extraction method | Classification method | Accuracy |
|---|---|---|---|---|---|
| Forkheim 1995 [39] | Not specified | Wheeze and normal | Raw signal data, FFT | MLP | The training sets 1 and 2 were 93 and 96% |
| Kahya 1997 [16] | 69n | Normal or abnormal | AR model | k-NN | 69.59% |
| Rietveld 1999 [38] | 60n | Normal and asthma | FT | MLP | 43% |
| Oud 2000 [26] | 10n | Asthmatic patients | Spectral analysis | k-NN | 60 to 90% |
| Waitman 2000 [25] | 17p, 17c | Normal or abnormal | FT | MLP | 73% |
| Bahoura 2003 [27] | 24n | Wheeze | MFCC, FFT, LPC, WPD, SBC | VQ | 75.80 and 77.50% |
| Baydar 2003 [28] | 20n | Normal or abnormal | Periodogram, Welch, Yule-Walker, Burg | Nearest mean classifier | 72% in expiration and 69% in inspiration |
| Kandaswamy 2004 [44] | Not specified | Lung sounds | WT, STFT | MLP | 94.02% |
| Folland 2004 [45] | Not specified | Lung sounds | Spectral computation parametric model, generation linear normalization | MLP, RBFN, CPNN ANN | 97.8% |
| Güler 2005 [47] | 129n | Normal, wheeze, and crackles | Welch | MLP, GANN | ANN 81–91%, GANN 83–93% |
| Martinez-Hernandez 2005 [29] | 19n | Normal or abnormal | Multivariate AR model | MLP | 87.68% |
| Kahya 2006 [5 | 20p, 20c | Rale | WT | k-NN | 46% |
| Lu 2008 [42] | Not specified | Fine and coarse crackles | GMM | GMM VQ | 95.1% |
| Alsmadi 2008 [31] | 42n | Lung sounds | AR model | k-NN and minimum distance classifier | 96% |
| Riella 2009 [40] | Not specified | Wheeze | FFT, STFT | MLP | 92.86% |
| Riella 2010 [46] | Not specified | Lung sounds | DWT | RBFNN | 92.36% |
| Yamamoto 2010 [48] | 114n | Normal or abnormal | Raw data | HMM | 84.2% |
| Charleston-Villalobos 2011 [32] | 27n | Normal or abnormal | AR model | MLP | 75 and 93% |
| Yamashita 2011 [33] | 168n | Normal or emphysema | Segmentation | HMM | 87.4 and 88.7% |
| Feng 2011 [34] | 21n | Normal or abnormal | Temporal–spectral dominance spectrogram | k-NN | 92.4% |
| Serbes 2011 [35] | 26n | Crackles | WT, DWT | SVM | 97.20% |
| Flietstra 2011 [24] | 257n | Pneumonia and CHF | Manual crackle analysis | SVM | Pneumonia 86% and CHF 82% |
| Hashemi 2011 [41] | 140p, 140s | Wheeze | WT | MLP | 89.28% |
| Aras 2015 [36] | 27 pathological, 21 normal s | Rale, rhonchus, and normal | MFCC LFCC | k-NN | The datasets 1 and 2 were 96 and 100% |
| Chen 2015 [37] | 20p | Rale, rhonchus, wheeze, and normal | MFCC | k-NN | 93.2% |

*p* patient, *c* control, *s* sound, *n* subject, *ANN* artificial neural network, *AR* autoregressive, *CHF* congestive heart failure, *COPD* chronic obstructive pulmonary disease, *CPNN* constructive probabilistic neural network, *DWT* discrete wavelet transform, *FFT* fast Fourier transform, *FT* Fourier Transform, *GANN* genetic algorithm-neural network, *GMM* Gaussian Mixture Model, *HMM* Hidden Markov Model, *k-NN* k-nearest neighbor, *LPC* linear predictive coding, *MFCC* Mel frequency cepstral coefficient, *MLP* multi-layer perceptron, *RBFNN* radial basis function neural network, *SBC* subband-based cepstral, *STFT* short-time Fourier transform, *SVM* support vector machine, *VQ* vector quantization, *WPD* wavelet packet decomposition, *WT* wavelet transform

Previous studies' results were not geared toward a practical system. In our study, we developed our device and software to fit into a hospital environment workflow. We are also planning to fit this workflow into a telemedicine system we are developing that allows physicians to remotely listen to and share patient audio data for consultation.

While our results seem numerically lower than the state-of-the-art results, our data set (17,930 audio clips) is the biggest data set when compared with that of the studies done on this field and the audio clips in the data set are not amplified, modified, cleaned, or pre-recorded by a third party which is the case with many of the studies we looked at. We tested our algorithms on eight datasets and obtained consistent results across the board; this was not done in any of the state-of-the-art study so far.

## 4 Conclusions

The goal of this project was to design and construct an electronic stethoscope with an associated software system that can transfer respiratory sounds to a PC for recording and subsequent computer-aided analysis and diagnosis. The hardware-software system was used to collect a dataset of respiratory sounds to train SVM and CNN machine learning algorithms for the automated analysis and diagnosis. The complete system can also be used for all types of body sounds (e.g., lung, heart, intestines) and is expected to be in widespread clinical use.

In this study, we experimented using CNN algorithms in audio classification. Since MFCC features combined with SVM is a generally accepted practice for audio classification, we used it as a benchmark for our CNN algorithm. We found out that spectrogram image classification with CNN algorithm works as well as the SVM system.

CNN and SVM algorithms were run comparatively to classify respiratory audio: (1) healthy versus pathological classification, (2) rale, rhonchus, and normal sound classification, (3) singular respiratory sound type classification, and (4) audio type classification with all sound types. Accuracy results of the experiments were found as (1) CNN 86%, SVM 86%, (2) CNN 76%, SVM 75%, (3) CNN 80%, SVM 80%, and (4) CNN 62%, SVM 62%, respectively.

As a result, we found out that spectrogram image classification with CNN algorithm works as well as the SVM algorithm, and given the large amount of data, CNN and SVM machine learning algorithms can accurately classify and pre-diagnose respiratory audio. This system can be combined with a telemedicine system to store and share information among physicians. We believe our method can improve the results of previous studies and help in medical research.

**Acknowledgements**
I would like to thank the chest physicians Prof. Dr. Turan Acıcan, Prof. Dr. Banu Gülbay, Assoc. Prof. Dr. Bülent Bozkurt, Assoc. Prof. Dr. Gülbahar Yorulmaz Yüce, and Dr. Şilan Işık for their help in the auscultation and for helping us collect the patient and diagnosis data.
I would like to thank lung function test technicians Leyla Ayten, Selçuk Demirtaş, and Hanife Bal and the department nurses for their invaluable help in gathering patient data.

**Funding**
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Availability of data and materials**
The data cannot be shared because patients did not allow the actual data to be released on a repository.

**Authors' contributions**
MA is responsible for the data collection, experiment design, algorithm design, and documentation. ÖK did the study design and coordination, performed thesis consultation, and revised the paper. BK and SP provided medical expertise in the data analysis and revision of the paper. All authors read and approved the final manuscript.

**Ethics approval and consent to participate**
This study was approved by the local Human Experiments Ethical Committee of Turgut Özal University (29.12.2015–0123456/0023).
The voluntary declaration form was read to the patient and signed with approval for participation in the study.

**Consent for publication**
The voluntary declaration form was read to the patient and signed with approval for the publication of the study.

**Competing interests**
We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**
[1]Faculty of Engineering, Department of Computer Engineering, Yıldırım Beyazıt University, Ankara, Turkey. [2]Clinic of Chest Diseases, Ankara Dışkapı Yıldırım Beyazıt Education and Research Hospital, Ankara, Turkey. [3]Faculty of Medicine, Department of Chest Diseases, Ankara University, Ankara, Turkey.

Received: 19 June 2017 Accepted: 30 August 2017
Published online: 11 September 2017

**References**

1. BDCN Prasadl, PESNK Prasad, Y Sagar, An approach to develop expert systems in medical diagnosis using machine learning algorithms (asthma) and a performance study. IJSC **2**(1), 26–33 (2011)
2. Y Bengio, A Courville, P Vincent, Representation learning: a review and new perspectives. IEEE Trans. Pattern Anal. Mach. Intell. **35**, 1–31 (2013)
3. Y Bengio, Learning Deep Architectures for AI. (2009), http://www.iro.umontreal.ca/~bengioy/papers/ftml.pdf. Accessed 26 Jan 2016
4. J Schmidhuber, Deep learning in neural networks: an overview. Neural Netw. **61**, 85–117 (2015)
5. AB Olshausen, Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature **381**(6583), 607–609 (1996)
6. K Nasrollahi, T Telve, S Escalera, J Gonzalez, TB Moeslund, P Rasti and G Anbarjafari, Spatio-temporal pain recognition in CNN-based super-resolved facial images. In video analytics. Face and facial expression recognition and audience measurement. Third International Workshop, VAAM 2016, and Second International Workshop, FFER 2016, Cancun, Mexico, Revised Selected Papers, Springer, Vol. 10165, pp. 151, December 4, 2016

7. R Collobert, Deep Learning for Efficient Discriminative Parsing. (2011), http://www.video.lectures.net. Accessed 26 Jan 2016
8. P Glauner, Comparison of Training Methods for Deep Neural Networks. (2015), https://arxiv.org/abs/1508.06535. Accessed 26 Jan 2016
9. L Deng, D Yu, Deep Learning: Methods and Applications. (2014), http://research.microsoft.com/pubs/209355/DeepLearning-NowPublishing-Vol7-SIG-039.pdf. Accessed 26 Jan 2016
10. L Gome, Machine-Learning Maestro Michael Jordan on the Delusions of Big Data and Other Huge Engineering Efforts. (2014), http://spectrum.ieee.org/robotics/artificialintelligence/machinelearning-maestro-michael-jordan-on-the-delusions-of-big-data-and-other-huge-engineering-efforts. Accessed 26 Jan 2016
11. Y LeCun, K Kavukcuoglu, C Farabet, Convolutional networks and applications in vision. Circuits and Systems (ISCAS), Proceedings of 2010 IEEE International Symposium on, IEEE, pp. 253–256, 2010
12. Y LeCun, Y Bengio, G Hinton, Deep learning. Nature **521**(7553), 436–444 (2015)
13. L Deng, Three Classes of Deep Learning Architectures and their Applications: A Tutorial Survey. APSIPA Transactions on Signal and Information Processing, 2012
14. RG Loudon, The lung exam. Clin. Chest Med. **8**(2), 265–272 (1987)
15. S Reichert, R Gass, C Brandt, E Andres, Analysis of respiratory sounds state of the art. Clin. Med. **2**, 45–58 (2008)
16. YP Kahya, EC Guler, S Sahin, Respiratory disease diagnosis using lung sounds. Engineering in Medicine and Biology Society, Proceedings of the 19th Annual International Conference of the IEEE, pp. 2051–2053, 1997
17. RLH Murphy, U.S. Patent 6, 139, 505, 31 Oct 2000
18. Littmann, Digital stethoscope. http://www.littmann.com/wps/portal/3M/en_US/3M-Littmann/stethoscope/stethoscope-catalog/catalog/~/3M-Littmann-Electronic-Stethoscope-Model-3200-Black-Tube-27-inch-3200BK27?N=5932256+4294958300&rt=d. Accessed 26 May 2016
19. Thinklabs, Digital stethoscope. http://www.thinklabs.com. Accessed 26 May 2016
20. SH Ah, *S Lee, Hierarchical Representation Using NMF Neural Information Processing* (Springer Heidelberg, Berlin, 2013)
21. Acoustics of Speech and Hearing. Spectrograms. UCL/PLS/SPSC2003/WEEK http://www.phon.ucl.ac.uk/courses/spsci/acoustics/week1-10.pdf. Accessed 26 May 2016
22. H Pasterkamp, SS Kraman, GR Wodicka, Respiratory sounds, advances beyond the stethoscope. Am. J. Respir. Crit. Care Med. **156**, 974–987 (1997)
23. JE Earis, BMG Cheetham, Current methods used for computerized respiratory sound analysis. Eur. Respir. Rev. **10**(77), 586–590 (2000)
24. B Flietstra, N Markuzon, A Vyshedskiy, R Murphy, Automated analysis of crackles in patients with interstitial pulmonary fibrosis. Pulm. Med. **2010**, 1–7 (2011)
25. LR Waitman, KP Clarkson, JA Barwise, PH King, Representation and classification of breath sounds recorded in an intensive care setting using neural networks. J. Clin. Monit. Comput. **16**(2), 95–105 (2000)
26. M Oud, EH Dooijes, JS van der Zee, Asthmatic airways obstruction assessment based on detailed analysis of respiratory sound spectra. IEEE Trans. Biomed. Eng. **47**, 1450–1455 (2000)
27. M Bahoura, C Pelletier, New parameters for respiratory sound classification. Electrical and computer engineering, IEEE CCECE, Canadian Conference. IEEE **3**, 1457–1460 (2003)
28. K.S. Baydar, A. Ertuzun, Y.P. Kahya, Analysis and classification of respiratory sounds by signal coherence method. Engineering in Medicine and Biology Society, Proceedings of the 25th Annual International Conference of the IEEE. IEEE, 2950–2953 (2003)
29. HG Martinez-Hernandez, CT Aljama-Corrales, R Gonzalez-Camarena, VS Charleston-Villalobos, G Chi-Lem, Computerized classification of normal and abnormal lung sounds by multivariate linear autoregressive model. Engineering in Medicine and Biology Society, IEEE-EMBS, 27th Annual International Conference of the IEEE, pp. 5999–6002, 2006
30. YP Kahya, M Yeginer, B Bilgic, Classifying respiratory sounds with different feature sets. Conf. Proc. IEEE. Eng. Med. Biol. Soc. **1**, 2856–2859 (2006)
31. S. Alsmadi, Y.P. Kahya, Design of a DSP-based instrument for real-time classification of pulmonary sounds. Comput. Biol. Med. **38**, 53–61 (2008)
32. S. Charleston-Villalobos, G. Martinez-Hernandez, R. Gonzalez-Camarena, G. Chi-Lem, J.G. Carrillo, T. Aljama-Corrales, Assessment of multichannel lung sounds parameterization for two-class classification in interstitial lung disease patients. Comput. Biol. Med. **41**, 473–482 (2011)
33. M Yamashita, S Matsunaga, S Miyahara, Discrimination between healthy subjects and patients with pulmonary emphysema by detection of abnormal respiration. International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 693–696, 2011
34. F. Jin, S. Krishnan, F. Sattar, Adventitious sounds identification and extraction using temporal–spectral dominance-based features. IEEE Trans. Biomed. Eng. **58**, 3078–3087 (2011)
35. G Serbes, CO Sakar, YP Kahya, N Aydin, Feature extraction using time–frequency/scale analysis and ensemble of feature sets for crackle detection. 33rd Annual International Conference of the IEEE EMBS Boston, Massachusetts USA, pp. 3314–3317, 2011
36. S Aras, A Gangal, Y Bülbül, Lung sounds classification of healthy and pathologic lung sounds recorded with electronic auscultation. Signal Processing and Communications Applications Conference (SIU), 2015 23th, IEEE, p. 252–255, 2015
37. C.H. Chen, W.T. Huang, T.H. Tan, C.C. Chang, Y.J. Chang, Using K-nearest neighbor classification to diagnose abnormal lung sounds. Sensors **15**, 13132–13158 (2015)
38. S Rietveld, M Oud, EH Dooijes, Classification of asthmatic breath sounds: preliminary results of the classifying capacity of human examiners versus artificial neural networks. Comput. Biomed. Res. **32**(5), 440–448 (1999)
39. K.E. Forkheim, D. Scuse, H. Pasterkamp, A comparison of neural network models for wheeze detection. WESCANEX 95. Communications, Power, and Computing. Conference Proceedings. IEEE **1**, 214–219 (1995)
40. RJ Riella, P Nohama, JM Maia, method for automatic detection of wheezing in lung sounds. Braz. J. Med. Biol. Res. 42, 674-684 (2009)
41. A Hashemi, H Arabalibiek, K Agin, Classification of wheeze sounds using wavelets and neural networks. 2011 International Conference on Biomedical Engineering and Technology, IPCBEE, vol.11, IACSIT Press, Singapore, 2011
42. X Lu, M Bahoura, An integrated automated system for crackles extraction and classification. Biomed. Signal. Process. Contr. **3**, 244–254 (2008)
43. Z Dokur, Respiratory sound classification by using an incremental supervised neural network. Pattern. Anal. Appl. **12**, 309–319 (2009)
44. A Kandaswamy, CS Kumar, RP Ramanathan, S Jayaraman, N Malmurugan, Neural classification of lung sounds using wavelet coefficients. Comput. Biol. Med. **34**, 523–537 (2004)
45. R Folland, E Hines, R Dutta, P Boilot, D Morgan, Comparison of neural network predictors in the classification of tracheal-bronchial breath sounds by respiratory auscultation. Artif. Intell. Med. **31**, 211–220 (2004)
46. RJ Riella, P Nohama, JM Maia, *Methodology for Automatic Classification of Adventitious Lung Sounds* (Springer, Berlin, Heidelberg/Munich, 2010), pp. 1392–1395
47. İ Güler, H Polat, U Ergün, Combining neural network and genetic algorithm for prediction of lung sounds. J. Med. Syst. **29**, 217–231 (2005)
48. H Yamamoto, S Matsunaga, K Yamauchi, M Yamashita, S Miyahara, *Classification between Normal and Abnormal Respiratory Sounds Based on Maximum Likelihood Approach. Proceedings of 20th International Congress on Acoustics* (ICA, Sydney, 2010), pp. 517–520