Effect of audio pre-processing technique for neural network on lung sound classification

Krittaya Chaiyot¹, Supattara Plermkamon², Thana Radpukdee³*

¹,²Department of Mechanical Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen, 40002, Thailand
³Department of Industrial Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen, 40002, Thailand

* Corresponding Author: tthanar@gmail.com

Abstract
In this paper, we intended to study the effect of two audio pre-processing methods to multilayer feedforward network performance on lung sound classification problem. The first one is Mel Frequency Cepstral Coefficients (MFCC) and the second is the MFCC supplement with Linear Discriminant Analysis (LDA). Datasets applied in this study came from Kaggle Respiratory Sound Database, which is the largest resource for machine learning. As a result, MFCC supplement with LDA has both good performance and significant improvement than conventional MFCC.

Keywords: Neural Network, Machine learning, Lung Sound Classification, MFCC, LDA

1. Introduction
Over the past century, physicians have diagnosed lung pathology using respiratory sounds. Those sounds can be related to respiratory health and respiratory disorders that need to have both special skill and reliable stethoscope to precision diagnose. The last two decades have seen a growing trend towards the automatic analysis of respiratory sounds. The reason is that it is not only necessary for ones to have special skills but it has also the potential to detect abnormalities in the early stages of respiratory dysfunction. Consequently, it can increase the effectiveness of decision making [1, 2], and may be of great support to the physicians.

This study set out to investigate the usefulness of automatic analysis of respiratory sounds, particularly sound classification task. This is an audio pre-processing and plays a significant role in preparation of training data on neural network. These processes can be used to identify the components of the audio signal, in the form of numeric data. It is also practical for identifying the audio content and rejecting redundant and unwanted stuff which carries information like background noise [3].

This paper can be divided into four sections. First section is introduction. Sections 2 describe the material and methods used in this study. There are four topics: Mel Frequency Cepstral Coefficient (MFCC), Linear Discriminant Analysis (LDA), Multilayer Perceptron and Data Preparation. Sections 3 include the result and discussion of the current study. The last section is the conclusion of the study.
2. Materials and Methods

Figure 1 illustrates the main framework used in this study. The lung sounds dataset in this study are preprocessed into two types. The first one, recorded lung sounds data just has been preprocessed and extracted features by MFCC so it can be used as training data for neural network, multilayer perceptron (MLP), and to perform classification task. Second, recorded lung sounds data was prepared like the first data type supplement with LDA for reducing dimension. Then, projecting the data into a subspace can be used to improve the class separability, and send as training data into multilayer perceptron. This current study applied Matlab as the main tool to implement the methods and use in the experimental for comparing classification results from different data types.

![Figure 1. The framework used in this study.](image)

2.1 Mel Frequency Cepstral Coefficient (MFCC)

MFCC is the most popular technique that is chosen to extract feature vectors in automatic speech recognition (ASR) [4-7]. The main point of MFCC is an extract of lung sound features without losing important information and rejecting redundant and unwanted stuff. The steps to perform MFCC computation process consist of Pre-Emphasis, Sampling and Windowing, Fast Fourier Transform, Mel Filter Bank, Discrete Cosine Transform, [8, 9] (see Figure 2 below).

![Figure 2. MFCC feature extraction process.](image)
2.2 Linear Discriminant analysis (LDA)
Linear Discriminant Analysis is a dimensionality reduction and projecting the data into a subspace that improves the class separability technique adapted in supervised learning. This technique will project the original data matrix into low dimensional space. There are three steps involved in the LDA performance process. To calculate the between-class matrix according to the equation is used to begin the process (1), then the calculation within-class matrix as represented by equation (2). The last process includes the construction of the lower dimensional space, as set out in Equation (3), and it can maximize between-class variance and minimize within the class variance [10]. In this study, LDA function are adapted from MATLAB Central File Exchange [11].

\[(m_i - m)^2 = W^T S_{Bi} W\]  
(1)

Where \(m_i\) = the projection of the mean of the ith class, \(m\) = the projection of the total mean of all classes, \(W\) = represents the transformation matrix of LDA, \(S_{Bi}\) = between-class variance of the ith class.

\[\sum (W^T x_i - m_i)^2 = \sum W^T S_{Wi} W\]  
(2)

Where \(S_{Wi}\) = within-class variance for each class, \(m_j\) = the projection of the mean of the jth class, \(x_i\) = samples of each class, \(W\) = represents the transformation matrix of LDA.

\[Y = X V_k\]  
(3)

Where \(Y\) = new data matrix, \(X\) = original data matrix, \(V_k\) = lower dimensional space.

2.3 Multilayer Perceptron (MLP)
Multilayer Perceptron are one type of Artificial Neural Network (ANN) that use supervised training to train a network. In this study a classifier architecture (MLP type) is R-10-4 with tansig activation function in hidden layer and softmax activation function in output layer as in Figure 3. This network had been trained 10 times for each data types, and then chooses only one the best training and test result in order to compare network accuracy onward.

![Figure 3. Classification model’s architecture R-10-4 (MLP type)](image)

2.4 Data Preparation
The dataset of the lung sound chosen in this study was from Kaggle Respiratory Sound Database [12]. This was created by two research teams in Portugal and Greece. It includes 920 annotated recordings of lung sounds. These recordings were taken from 126 patients. There is a total of 5.5 hours of recordings containing crackles, wheezes and both crackles and wheezes. Despite the limit of our computation resource, we use only 200 annotated recordings of a length 20 second each, which consist of 50 sounds
recorded for each class. Also, each class has frequency content (see Figure 4). (Normal < 400 Hz, Crackles < 1000 Hz, wheezes < 100 Hz and both wheezes and crackles < 1000 Hz).

However, each of 200 lung sound used in the training ANN still has a large amount of sample, 882,000 per one recorded lung sound, and it takes very high resource to computation. In order to overcome this each recorded lung sound has scaled down the sample rate from 44.1 kHz to 16 kHz [13], reference to the frequency content in Figure 4, 16 kHz can prevent an aliasing problem. As a result, sample per one recorded lung sound is 320,000 and duration is still 20 second. Then perform MFCC, which will get column vector of 25,974 elements for each recorded lung sound, so for totally, 200 recorded lung sounds we will get a matrix of 25,974 by 200. However, this matrix just still big for sending to ANN, so again all of 200 recorded lung sound have squeezed duration time from 20 seconds to 7.3 seconds by change sample rate to 44.1 kHz (Sample point just still 320,000 per each recorded lung sound). This process will increase the playback speed of recorded lung sound but still maintain the previous sample point without losing any information and also help to reduce data after performing MFCC to 9,399 by 200 matrixes, which use for train ANN so-called data type I. Figure 5 shows an audio waveform for one recorded lung sound.

![Figure 4. Frequency content for a single recorded lung sound.](image-url)
a. Recorded lung sound with 882,000 samples, Fs = 44.1 kHz, time length = 20 seconds

b. Recorded lung sound with 320,000 samples, Fs = 16 kHz, time length = 20 seconds

c. Recorded lung sound with 320,000 samples, Fs = 41.1 kHz, time length = 7.3 seconds

Figure 5. Compare a recorded lung sound in Audio waveform.

It should be noted that the preparation processes of data type II are the same as data type I. While the main difference in data type II is supplement with LDA after performing MFCC.

3. Result from the experiment

After preparing the training data as provided in topic 2.4, we have to train and investigate the network performance. Generally, the confusion matrix was used to measure the performance of a model in the classification task. We also use this tool to investigate the performance of the network.

To begin this training ANN process, each type of data was used to train R-10-4 network for 10 times. After training, the one that shows the best performance was indicated by a confusion matrix. The best result from data type I showed accuracy on the test set as 60% and 100% on the training set (see Figure 6). We can see that data type II indicated the best accuracy for a test set by 96.7% and for a training set is 100%, as set out in Figure 7.

The results from both types of data have good accuracy on training data, it was found that the network can learn training data quite well. Whilst the test data cannot be seen by the network never seen before. Only data type II showed a significant improvement on testing accuracy, resulting in the predominant of MFCC supplement with LDA for use in classification problem.
4. Conclusion
This study has identified two input preprocessing methods that affect the classification performance. As can be seen from the result on topic 3, the MFCC supplement with LDA has the best performance for the classifier. Although this experiment has to deal with multivariable data like audio data a classifier, it can have an ability to identify each class with good performance. The findings of this study suggest that this pre-processing technique can also be applied in many similar classification tasks like a heartbeat sound etc.
References
[1] Saraiva A Santos D Francisco A Sousa J Ferreira N Soares S and Valente A 2020 Proc. the 13th International Joint Conference on Biomedical Engineering Systems and Technologies 3 138-144
[2] Pasterkamp H Kraman S and Wodicka G R 1997 American Journal of Respiratory and Critical Care Medicine 156 974-987
[3] Sharma D and Ali I 2015 Proc. Int. Conf. on Advances in Computing, Communications and Informatics (Kochi) 1052-1057
[4] Winursito A Hidayat R and Bejo A 2018 Proc. Int. Conf. on Information and Communications Technology (Yogyakarta) 379-383
[5] Joshi S C and Cheeran A N 2014 MATLAB Based Feature Extraction Using Mel Frequency Cepstrum Coefficients for Automatic Speech Recognition
[6] Davis S and Mermelstein P 1980 IEEE Transactions on Acoustics, Speech, and Signal Processing 28(4) 357-366
[7] Huang X Acero A and Hon H 2001 Spoken Language Processing: A guide to theory, algorithm, and system development (Prentice Hall)
[8] Magre S Deshmukh R and Shrishrimal P 2013 Proc. Int. Conf. on Recent Advances in Statistics and Their Applications (Aurangabad, Maharashtra: Ambedkar Marathwada University)
[9] Chakraborty K Talele A and Upadhyya S 2014 International Journal of Innovative Research in Advanced Engineering 1 158-161
[10] Tharwat A Gaber T Ibrahim A and Hassani E A 2017 Ai Communications 169-190
[11] Yarpiz 2020 Linear Discriminant Analysis (LDA) aka. Fisher Discriminant Analysis (FDA), https://www.mathworks.com/matlabcentral/fileexchange/53151-linear-discriminant-analysis-lda-aka-fisher-discriminant-analysis-fda), accessed on 17 September 2020
[12] Marsh 2019 Respiratory Sound Database, Version 2, https://www.kaggle.com/vbookshelf/respiratory-sound-database, accessed on 17 September 2020
[13] Messner E Hagmüller M Swatek P and Pernkopf F 2016 Proc. 9th International Conference on Biomedical Electronics and Devices (Rome, Italy) 34-39