OPEN ACCESS 60

ISSN 2460-9056
http://socj.telkomuniversity.ac.id/ijoict/

Intl. Journal on ICT
Vol. 8, No. 2, June 2022. pp. 60-68
doi:10.21108/ijoict.v8i1.621

Classification Analysis using CNN and LSTM on Wheezing Sounds

Gustav Bagus Samanta [1*], Satria Mandala[2]

[1,2]*School of Computing, Telkom University*
Jl. Telekomunikasi 1 Terusan Buah Batu, Bandung 40257, Indonesia

[*] gustavsmnt@student.telkomuniversity.ac.id

Abstract

Asthma is a public health problem in almost all countries in the world. One of the symptoms that exist in asthmatics is wheezing. Several researchers have conducted research on wheezing classification using machine learning methods, namely the Support Vector Machine (SVM), Mel-frequency cepstral coefficients (MFCC), empirical mode decomposition (EMD), Artificial Neural Network (ANN), ensemble (ENS), K-Nearest Neighbor (KNN) and Short-Time Fourier Transform (STFT) and Convolutional Neural Network (CNN). Of the many studies that have been carried out in detecting wheezing, however, researchers only focus on proposing a new algorithm for wheezing detection. Rarely do researchers focus on comparative analysis to existing algorithms. This study aims to determine the accuracy of the results from wheezing classification of respiratory sounds by comparing the algorithm. In the experiment, 4 algorithm were analyzed namely Classification using Convolutional Neural Network (CNN) with Short-Time Fourier Transform (STFT) extraction feature, Classification using Convolutional Neural Network (CNN) with Mel-Frequency Cepstral Coefficients (MFCC) extraction, Classification using Long-Short Term Memory (LSTM) with Short-Time Fourier Transform (STFT) extraction feature, Classification using Long-Short Term Memory (LSTM) with Mel-Frequency Cepstral Coefficients (MFCC) extraction. Rigorous experiments have been carried out, and it is proven that Classification using Convolutional Neural Network (CNN) algorithm is better than using Long-Short Term Memory (LSTM). Classification with 2 CNN algorithms has 98% accuracy while classification with LSTM and STFT algorithms has 58% accuracy, and LSTM and MFCC algorithms has 100% accuracy.

Keywords: asthma, classification algorithm, wheezing, convolutional neural network, long short-term memory

I. INTRODUCTION

SEVERAL researchers have conducted research on wheezing classification using machine learning methods, namely the Support Vector Machine (SVM) [1], Mel-frequency cepstral coefficients (MFCC) [2], [3], empirical mode decomposition (EMD), Artificial Neural Network (ANN), ensemble (ENS), K-Nearest Neighbor (KNN) and Short-Time Fourier Transform (STFT) [4] and Convolutional Neural Network(CNN) [5]. In research [1] G. D. Sosa et al discusses automatic wheezing detection by evaluating several acoustic feature extraction methods and C-weight SVM. In [6], P. Bokov et al discusses the introduction of wheezing using recorded mouth breathing sounds taken using a smartphone in children. In [4], D. Oletic and V. Rinse discusses wheezing detection based on Compressively Sensed Respiratory Sound Spectra. In research [2] M. Akanat et al discussed about the classification of wheezing with the Convolutional Neural Network method. In the study [7], A. Parkhi and M. Pawar discussed the analysis of lung abnormalities using the Short Time Fourier Transform (STFT)

Received on 20 June 2022. Revised on 7 July 2022. Accepted and published on 2 Aug 2022.

Spectogram analysis of lung sounds. In [8]. Anggoro. et al discussed about speech recognition in the northern Sundanese dialect. In this study, the researcher uses the feature extraction method, namely MFCC for the introduction method using RNN. The researcher has five samples of speech in Sundanese which will be tested for ten times with each part namely epoch testing and mini batch testing. The results obtained in the test that is equal to 74%. In [8], Erwin et al discussed the introduction of Manado dialect using the Recurrent Neural Network. In this study, the researcher used the feature extraction method, namely MFCC. Researchers used ten samples of Manado speech taken from three sources using cellphones. Researchers used three testing methods, namely epoch testing, mini batch, and system testing. Each test method was carried out for ten to eleven repetitions. The accuracy results obtained are 87% using ideal parameters and carried out for ten times of testing. Of the many studies that have been carried out in detecting wheezing, however, researchers only focus on proposing a new algorithm for wheezing detection. Rarely do researchers focus on comparative analysis to existing algorithms. Based on the above problems, this final project is an analytical study of the Convolutional Neural Network and Long-Short Term Memory classification algorithms using the feature extraction method MFCC and STFT.

II. LITERATURE REVIEW

A. Wheezing

Wheezing is a marker of lung symptoms experienced by every individual, be it parents, children, or health care providers. In the International Study of Asthma and Allergies in Childhood (ISAAC) wheezing is used as a substitute for asthma [9]. Wheezing occurs spontaneously repeatedly, with a high pitch and often heard on expiration [10].

Based on [11]., study found that 48% of children within the first 6 years have at least one wheezing disease, 34% before age 3 years usually is said to be early wheezing and half of the children continue at the age of 6 years

Wheezing sounds can be distinguished from waveforms that transformed into a spectrogram, Fig.1 is an example of a sound waveform transformation of wheezing, crackle, and rhonchi.

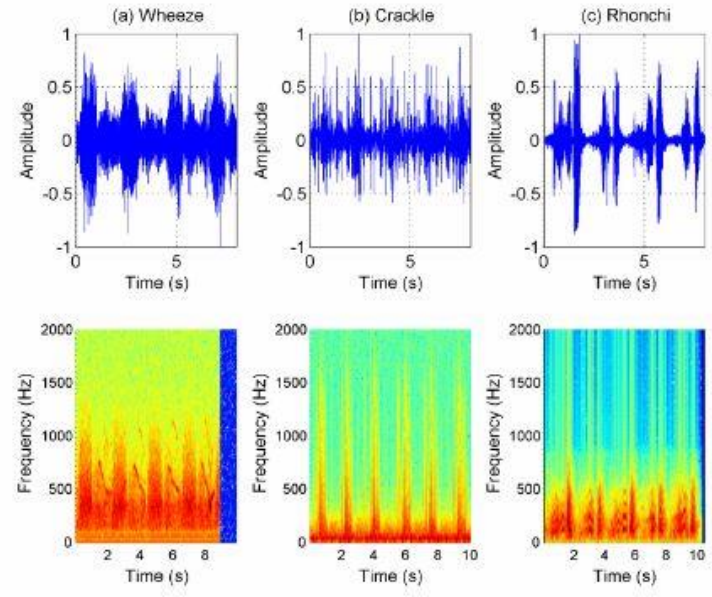

Fig.1 Visualization adventitious lung sounds (ALS) waveforms (top) and spectrogram (bottom) of (a) wheeze, (b) crackle, and (c) rhonchi.

B. *Short-Time Fourier Transform (STFT)*

Short-time Fourier Transform is a digital signal processing algorithm which is the development of the fast Fourier transform algorithm. Algorithm STFT works by picking up a signal in t seconds then decoding into the frequency domain so that the position as well as the time and domain can be known signals. Short-time Fourier Transform has calculation calculations based on window function [12]:

$$STFT(t,f) = \int x(\tau) * (\tau - t)e^{-2f\pi f\tau}d\tau$$

The window function (w) will be placed on the first signal with t = 0 width (T) min $\frac{T}{2}$ maximum frequency.

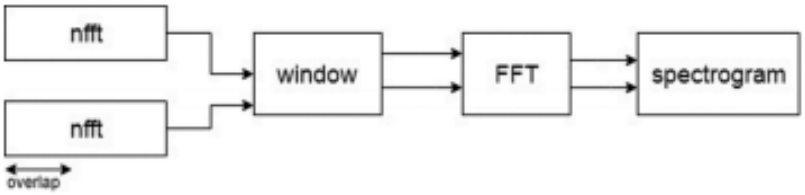

Fig. 2 Block diagram of STFT ability to overlap, while constructing a spectrogram.

As can be seen in Fig.2, STFT can also be calculated using the Fast Fourier Transform with the signal variables x(n) and the window w(n) must be discrete, where m is the high-resolution time of n. So it can be expressed in spectogram with problem

$$\text{spectogram}\{x(n,\omega)\} = |X(n,\omega)|^2$$

The most important part in STFT lies in the width of the function window because it will affect the frequency resolution and time resolution [13].

C. *Convolutional Neural Network (CNN)*

Convolutional Neural Network is a branch of deep learning used to perform visual recognition. Not only visuals CNN recognition can also be used in audio processing, especially in classification sound. However, they have architectural differences that are practical and significant consequences.

1) *Layer Architecture*

Convolutional Neural Network has several layers that arranged in its architecture which consists of:

a) Convolutional Layer

Layers arranged over neurons to form a filter that has a pixel size. Convolutional Layer used as an extraction feature by sliding the filter there is by using a dot operation between the input and the filter value, so it gets an output called activation map. As can be seen in 0 the convolution layer symbolized as CONV, is where the feature learning happens, filters are just small patches that represent visual feature of a CNN and convolve it over the input volume to get a single activation map.

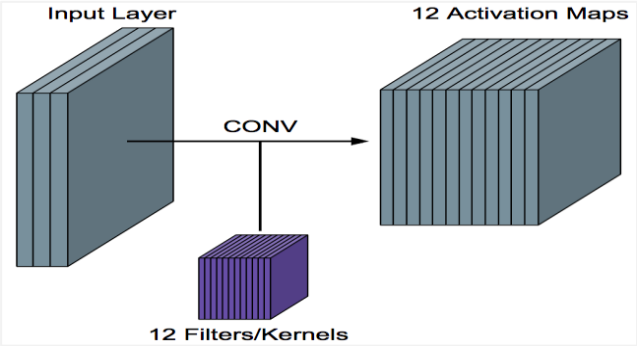

Fig. 3 The convolutional layer.

b) Pooling Layer

Layer that serves to reduce the size of the feature map that has been obtained on the convolutional layer and increase position invariance of the features. The method used in determine pooling is Max pooling whereby searching the largest value from the feature map, and Average Pooling by finding the average value on the feature map. Fig.3 shows an example of max pooling operation and average pooling with a 2x2 pixel filter size from 4x4 pixel input for max pooling, inside of the window choose the maximum value in that window. For average pooling, take the average of the values in the window.

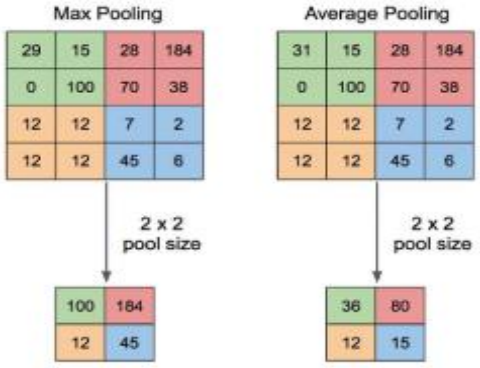

Fig.4 Pooling layer operation.

c) Fully Connected Layer

The function of the Fully Connected Layer is to perform transformation of data dimensions so that classification can be carried out linear [14]. Inside the fully connected layer there are several layers it includes: hidden layer, activation function, output layer and loss function.

2) Rectified Linier Units

Rectified Linear Units is one of the activation functions that thresholding the input pixel with a value of zero with the function $f(x) = \max(0, x)$ [15], [16], 0.5 is a visualization of the function.

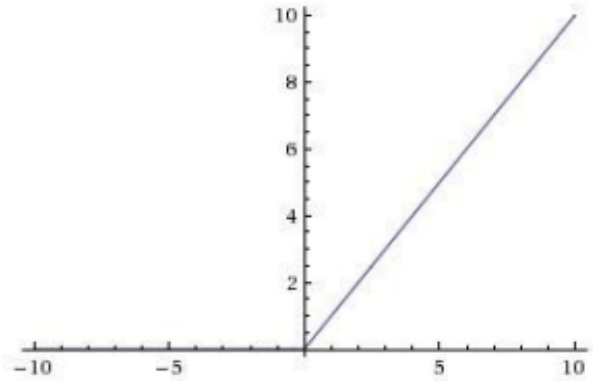

Fig.5 Visualization of function $f(x) = \max(0, x)$

D. Mel-Frequency Cepstral Coefficients (MFCC)

MFCC feature extraction aims to obtain features in the form of parameters. The MFCC has seven stages. The first stage is pre-emphasis, the second is frame blocking, third windowing, fourth Fast Fourier Transform (FFT), fifth Mel Frequency Wrapping (MFW), the sixth Discrete Cosine Transform (DCT), and the seventh cepstral lifting [17].

E. Long-Short Term Memory (LSTM)

Long-Short Term Memory (LSTM) is a special type of neural network, where including part of the Recurrent Neural Network (RNN). Unlike feed forward neural network conventionally, RNN uses and feedback from the output layers back to the input layers, where each feedback connection can be used as a time-delay gate. RNN architecture able to represent explicitly the effect of past output values on current output calculations, making it ideal for modeling the autocorrelation structure of time series or time series data [18].

III. RESEARCH METHOD

The following are some explanations of the methods used in this study.

A. Data

In this experiment, the data used has a total of 73 sound recordings for data test consisting of 32 normal breath sounds and 33 abnormal breath sounds, and for data test consisting of 4 normal breath sounds and 4 abnormal breath sounds. The data was obtained from kaggle with research named "Respiratory Sound Database for the Development of Automated Classification". The data used is a voice recording in the form of .wav format.

The first thing to do in the experiment is to label the sound recording data with a normal or abnormal label as can be seen in TABLE I. The next step is to insert the sound recording data that has been labeled to the algorithm.

TABLE I. EXAMPLE OF DATASET EXPERIMENTS

No.	Data	Label
1	102_1b1_Ar_sc_Meditron.wav	normal
2	121_1p1_Tc_sc_Meditron.wav	normal
3	122_2b1_Al_mc_LittC2SE.wav	normal
4	121_1b1_Tc_sc_Meditron.wav	normal
5	125_1b1_Tc_sc_Meditron.wav	normal
6	101_1b1_Al_sc_Meditron.wav	abnormal
7	103_2b2_Ar_mc_LittC2SE.wav	abnormal
8	104_1b1_Al_sc_Litt3200.wav	abnormal
9	104_1b1_Ar_sc_Litt3200.wav	abnormal
10	104_1b1_Ll_sc_Litt3200.wav	abnormal

B. Confusion Matrix

Classification model valuation is based on testing to estimate the correct object and false [19], the order of the tests is tabulated in the confusion matrix where the classes are predicted is displayed at the top of the matrix and class observed on the left. Each cell contains a number which shows how many cases are of the observed class for predictable as can be seen in TABLE II. True Positive is predicted positive and it's true, True Negative is predicted negative and it's true, False Positive is predicted positive and it's false, False Negative is predicted negative and it's false.

TABLE II. COMBINATIONS OF PREDICTED AND ACTUAL VALUES

Classification	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positive (FPs)
Predicted Negative (0)	False Negatives (FNs)	False Positives (Fps)

C. Scenario

1) Splitting Data

It is recommended that we divide the observations into training and testing data. Empirical studies show that the best results are obtained if we use 20% of the data for testing, and the remaining 80% of the data for training [20]. In this study provides 20% of the data for testing, and the remaining 80% of the data for training.

2) Analyzing

In the experiment, 4 algorithms were analyzed:

a) Classification using Convolutional Neural Network (CNN) with Short-Time Fourier Transform (STFT) extraction feature.
b) Classification using Convolutional Neural Network (CNN) with Mel-Frequency Cepstral Coefficients (MFCC) extraction.
c) Classification using Long-Short Term Memory (LSTM) with Short-Time Fourier Transform (STFT) extraction feature.
d) Classification using Long-Short Term Memory (LSTM) with Mel-Frequency Cepstral Coefficients (MFCC) extraction.

IV. RESULTS

The evaluation of the classification model is based on testing to estimate normal or abnormal, then each algorithm will be compared. The method used to compare algorithms are using different learning rates and different epoch. Learning rate affects the testing and training accuracies of algorithm and therefore researchers have to explore different learning rates before settling on one, this experiment uses the default learning rate of 0.0001 [21]

A. Learning Rate 0.0001

There is a limit to the times one can decrease the learning rate. To avoid wasting time at such points, one should avoid repeating the same steps while taking the same path that results in the same minimum.

TABLE III. COMPARISON OF RESULTS WITH 0.0001 LEARNING RATE AND DIFFERENT EPOCHS

Model	Epoch 50		Epoch 100		Epoch 150		Epoch 200	
	Accuracy	Loss	Accuracy	Loss	Accuracy	Loss	Accuracy	Loss
CNN with STFT	83%	0.412	87%	0.3276	90%	0.3144	90%	0.2895
CNN with MFCC	75%	0.575	87%	0.2709	90%	0.2688	98%	0.1325
LSTM with STFT	56%	0.6924	56%	0.6898	56%	0.6922	56%	0.6875
LSTM with MFCC	92%	0.2871	87%	0.316	98%	0.246	98%	0.1979

Based on the Fig.6, the learning rate 0.0001 and different epochs, both of CNN model has increased the accuracy, meanwhile the accuracy of LSTM with STFT model does not change, and the LSTM with MFCC accuracy changes up and down. The graph of the results of each algorithm with learning rate 0.001 can be seen in 0.

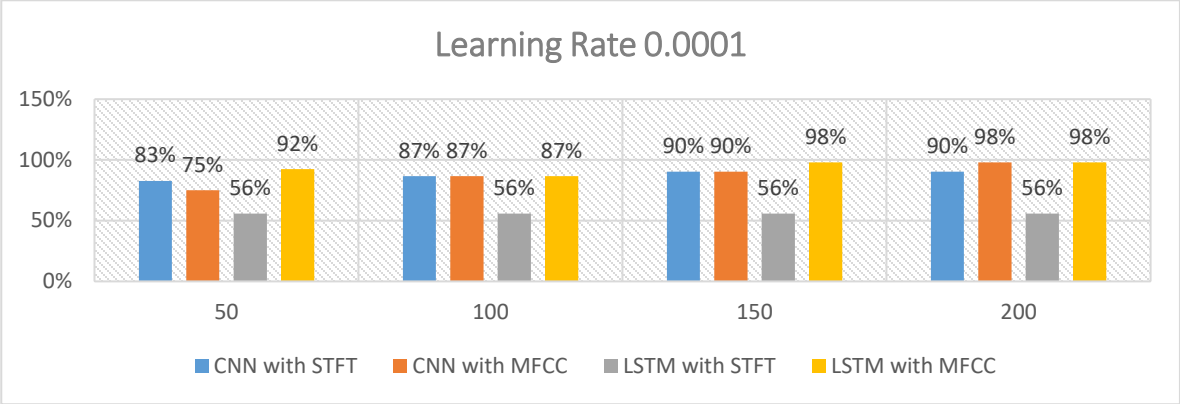

Fig.6 Visualization of learning rate 0.0001.

B. *Learning Rate 0.0001*

Researchers raised the learning rate to high with the learning rate of 0.001 [21].

TABLE IV. COMPARISON OF RESULTS WITH 0.001 LEARNING RATE AND DIFFERENT EPOCHS

Model	Epoch 50		Epoch 100		Epoch 150		Epoch 200	
	Accuracy	Loss	Accuracy	Loss	Accuracy	Loss	Accuracy	Loss
CNN with STFT	92%	0.2709	94%	0.1678	96%	0.1391	98%	0.0993
CNN with MFCC	56%	0	56%	0.6911	98%	0.0906	56%	0
LSTM with STFT	56%	0.6882	56%	0.686	56%	0.6882	56%	0.6864
LSTM with MFCC	100%	0.4138	100%	0.0581	98%	0.179	100%	0.0338

Based on the 0TABLE IV, the learning rate 0.001 and different epochs, the CNN with STFT model has increased the accuracy, CNN with MFCC model accuracy changes up and down, LSTM with STFT model does not change, and LSTM with MFCC model has reached 100% accuracy. The graph of the results of each algorithm with learning rate 0.001 can be seen in Fig.6.

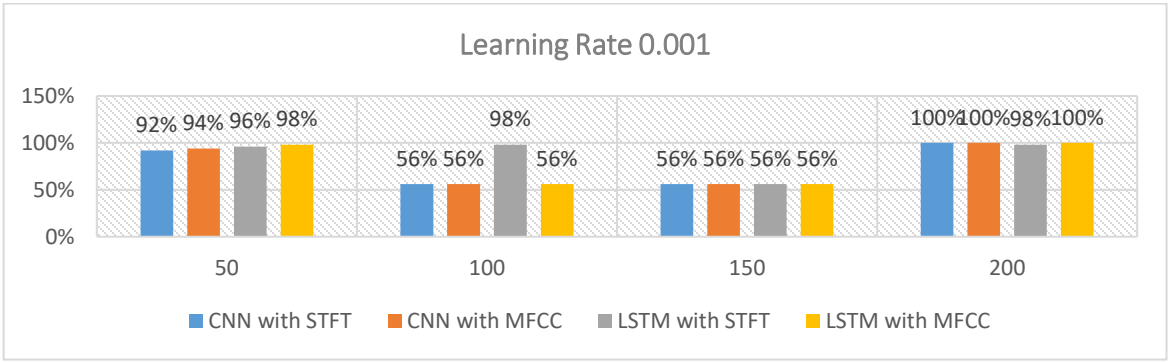

Fig. 7 Visualization of learning rate 0.001.

V. CONCLUSION AND RECOMMENDATIONS

Based on the results of the four algorithms, the most optimal algorithm for classifying wheezing is with CNN because of the accuracy value obtained. And there is a possibility that the algorithm that has 100% accuracy is overfitting, this can happen if the model is trained in an extremely complex way so that its estimation has high variance but low bias [22]. It can be concluded that the CNN algorithm is the most suitable classification algorithm for data like this, one of the reasons that can be seen here is how the algorithm comes with high accuracy for different learning rates and epochs. The challenges in conducting this research are collecting data directly from the hospital and finding methods that can be used for this kind of data. For further study, researchers may be able to use more datasets, the more data used the more accurate the algorithm's performance results will be, adding denoising feature into the algorithm, develop to avoid overfitting.

DATA AND COMPUTER PROGRAM AVAILABILITY

Data and program used in this paper can be accessed in the following site:

1) *https://drive.google.com/drive/folders/1RdQ2d00wk9NasZIFOvVx0ZTRqLOTx741?usp=sharing*
2) *https://drive.google.com/drive/folders/1qBt-5tthbsMKJ8ZoS8PpNrOE96yjYzrS?usp=sharing*

ACKNOWLEDGMENT

The authors would like to thank the reviewers for all useful and helpful comments on our manuscript. The author, however, bears full responsibility for the Paper.

REFERENCES

[1] A. C.-R. a. F. A. G. G. D. Sosa, "Automatic detection of wheezes by evaluation of multiple acoustic feature extraction methods and C-weighted SVM," *10th Int. Symp. Med. Inf. Process. Anal.*, vol. 9287, no. 1, p. 928709, 2015, doi: 10.1117/12.2073614., 2015.

[2] Ö. K. B. K. a. S. S. M. Aykanat, "Classification of lung sounds using convolutional neural networks," *Eurasip J. Image Video Process*, vol. 2017, no. no. 1, 2017, doi: 10.1186/s13640-017-0213-2., 2017.

[3] O. B. a. M. Bahoura, "Efficient FPGA-based architecture of an," *J. Syst. Archit*, vol. 88, no. pp. 54–64, 2018, doi:10.1016/j.sysarc.2018.05.010, 2018.

[4] D. O. a. V. Bilas, "Asthmatic Wheeze Detection from Compressively," *IEEE J. Biomed. Heal. Informatics,*, vol. 22, no. 5, pp. 1406–1414, 2018, doi: 10.1109/JBHI.2017.2781135, 2018.

[5] E. P. S. A. I. A. F. Kirill Kochetov, "Wheeze Detection Using Convolutional Neural Networks," *EPIA Conf. Artif. Intelegent*, vol. 1, no. . d, pp. 87–94, 2017, doi:, 2017.

[6] B. M. P. F. a. C. D. P. Bokov, "Wheezing recognition algorithm using recordings of respiratory sounds at the mouth in a pediatric," *Comput. Biol. Med*, vol. 70, no. pp. 40–50, 2016, doi:10.1016/j.compbiomed.2016.01.002., 2016.

[7] A. P. a. M. Pawar, "Analysis of deformities in lung using short time Fourier transform spectrogram analysis on lung sound," *Proc. - 2011 Int.Conf. Comput. Intell. Commun. Syst. CICN*, no. pp. 177–181, 2011, doi: 10.1109/CICN.2011.35., 2011.

[8] A. B. O. a. R. E. S. E. Lapian, "RECURRENT NEURAL NETWORK FOR SPEAKING RECOGNITION IN DITALCES MANADO," *Medicus*, vol. 5, no. 3, pp. 3–4, 2018.

[9] E. M. E. Gershwin and T. E. Albertson, "Bronchial Asthma," *Humana Press*, vol. 53, no. 9, 2013.

[10] A. B. a. B. P. S. F. Syafria, "Lung Voice Recognition with MFCC as Feature Extraction and Backpropagation as Classifier," *Ilmu Komput. dan Agri-Informatika*, vol. 3, no. 1, p. 27, doi: 10.29244/jika.3.1.27-36., 2017.

[11] R. Y. a. D. P. A. Krisdanti, "Imunopatogenesis Asthma," *J.Respirasi*, vol. 3, no. 1, p. 26, 2019, doi: 10.20473/jr.v3-i.1.2017.26-33, 2019.

[12] H. Hasanah, "Comparative Evaluation of Short Time Fourier Transform (STFT) and Wigner Distribution (WD) on Electrocardiogram Classification(ECG)," no. pp. 1–7.

[13] A. R. a. D. K. B. E. T. Handono, "Tune determination Javanese gamelan using the short time fourier algorithm transform," no. pp. 1–12.

[14] A. Y. W. a. R. S. I. W. S. E. Putra, "Image Classification Using Convolutional Neural Network (Cnn) On Caltech 101 Image Classfication Using Convolution Neural Network (Cnn) on Caltech 101," *Inst. Teknol. Sepuluh Novemb*, 2016.

[15] S. I. a. A. Nilogiri, "Implementation of Deep Learning in Identification Types of Plants Based on Leaf Image Using Convolutional Neural Network," *JUSTINDO (Jurnal Sist. dan Teknol. Inf. Indones*, vol. 3, no. 2, pp. 49–56, 2018, doi: 10.32528/JUSTINDO.V3I2.2254., 2018.

[16] K. J. Piczak, "ENVIRONMENTAL SOUND CLASSIFICATION WITH CONVOLUTIONAL NEURAL NETWORKS," *IEEE 25th Int. Work. Mach. Learn. Signal Process*, no. pp. 1–6, 2015, doi: 10.1109/MLSP.2015.7324337., 2015.

[17] S. H. A. E. P. Heriyanto, "CEPSTRAL FREQUENCY MEL FREQUENCY EXTRACTION COEFFICIENT (MFCC) AND AVERAGE COEFFICIENT FOR QUR'AN READING CHECK," *TELEMATIKA*, vol. 15, no. 02, OKTOBER, 2018, Pp. 99 – 108, 2018.

[18] A. S. a. M. Bahoura, "Long Short Term Memory Based Recurrent Neural Network for Wheezing Detection in Pulmonary Sounds," *IEEE International Midwest Symposium on Circuits and Systems (MWSCAS)*, 2021.

[19] P. A. R. R. M. Ardiyansyah, "Comparative Analysis of Data Mining Classification Algorithms For Blogger Dataset With Rapid Miner," *JURNAL KHATULISTIWA INFORMATIKA*, vol. 6, no. 1 p-ISSN: 2339-1928 & e-ISSN: 2579-633X, 2018.

[20] V. K. O. K. Afshin Gholamy, "Why 70/30 or 80/20 Relation Between Training and Testing Sets: A Pedagogical Explanation," *Departement Technical Reports(CS)*, 2018.

[21] D. M. M. B. K. K. E. C. T. Jennifer Jepkoech, "The Effect of Adaptive Learning Rate on the Accuracy of Neural Networks," *(IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 12, no. 8, 2021, 2021.

[22] M. C. Benyamin Ghojogh, "The Theory Behind Overfitting, Cross Validation,Regularization, Bagging, and Boosting: Tutorial," no. arXiv:1905.12787v1, 2019.

[23] A. B. O. a. R. E. S. A. Anggoro, "Recurrent Neural Network For Speech Recognition in Recurrent Neural Network for Speech," vol. 5, no. 3, pp. 6431–6435, 2018.