ISSN 2723-6471

Performance Improvement of Covid-19 Cough Detection Based on Deep Learning with Segmentation Methods

Suyanto[1], Zanjabila[2,*], Bagus Tris Atmaja[3,*], Wiranto Argo Asmoro[4]

[1,2,3,4]*Department of Engineering Physics, Faculty of Industrial Technology and System Engineering, Sepuluh Nopember Institute of Technology, Surabaya, Indonesia*

[3]*Artifcial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Japan*

(Received: February 1, 2024; Revised: March 5, 2024; Accepted: April 9, 2024; Available online: May 31, 2024)

Abstract

COVID-19 is an emergency problem that is being widely discussed in the world, one of which is the deep learning-based COVID-19 detection method which has been developed based on images of the patient's chest or cough. In this research, we propose a way to improve the performance of deep learning-based COVID-19 cough detection by using a segmentation method to produce several audio files containing one cough signal from one audio file containing several cough sounds signals. In addition, we enabled two automatic cough segmentation methods, namely a Hysteresis Comparator based on the power spectrum and an RMS threshold based on the RMS energy value. The results obtained show that using the segmentation method for cough sounds can improve the model's performance in detecting COVID-19 coughs by 4% to 8%. The segmentation process can also remove noise between cough sound signals and provide a standard input model in the form of one cough signal. In addition, the segmentation results show information related to the characteristics of COVID-19 cough. The evaluation results show that the hysteresis comparator method has better results with an unweighted accuracy (UA) value of 83.19% compared to the RMS threshold method with a UA value of 79.06%.

Keywords: Segmentation, Deep Learning, Signal Processing, COVID-19

1. Introduction

Respiratory disease is the third leading cause of death in the world after heart disease and cancer [1]. At the end of 2019, there was a pandemic of a type of pneumonia called COVID-19. Until February 04, 2024, WHO noted that there were 774 million confirmed cases of COVID-19 in the world and 7 million people died from COVID-19 from 2020 to 2021 [2]. The virus that causes COVID-19, SARS-CoV-2, changes over time. Some of these changes affect the properties of the virus, such as the rate of spread of the virus, the severity of the disease, the level of effectiveness of vaccines, medicines and public health measures. So far, WHO has classified 5 main variations of COVID-19, namely Alpha, Beta, Gamma, Delta, and Omicron, and several other sub-variations in order to prioritize monitoring and research on changes in the virus [3].

There are several symptoms experienced by sufferers of COVID-19, one of which is a dry cough [4]. Cough symptoms that appear in sufferers of COVID-19 are the main symptoms during the acute infection phase which is called an acute cough accompanied by fever and loss of taste and smell which often occurs in sufferers and advanced symptoms during the post-infection phase which is called a chronic cough [5]. There are several initial diagnoses that are often used today, namely RT-PCR and rapid antibody test, and both have their own weaknesses. Diagnosis by RT-PCR is still relatively expensive, laborious with four to six hours of work and low sensitivity after five days of symptom onset, whereas diagnosis by rapid antibody test has low sensitivity especially on the first day of illness and requires rigorous reactivity testing [6]. Therefore, with cough symptoms that are common in people with COVID-19, there is potential to make an initial diagnosis in detecting COVID-19 in someone based on the sound of their cough.

[1]***Corresponding author: Zanjabila, Bagus Tris Atmaja (zanjabilaabil@gmail.com, b-atmaja@aist.go.jp)**
DOI: https://doi.org/10.47738/jads.v5i2.205

 ISSN 2723-6471

There have been many studies studying disease detection using deep learning, such as the detection of dyslexic [7], lung disease detection [8], heart disease prediction [9], classification of tumors and cancer [10], and many more. In a study involving 69 studies regarding the use of deep learning in the medical world, it was found that the results of deep learning had a sensitivity of 9.7% to 100% with an average sensitivity of 79.1% and a specificity of 38.9% to 100% with an average specificity of 88.3%. This shows that the use of deep learning in diagnosing diseases has enormous potential, even the accuracy obtained by deep learning is equivalent to that of professional health nurses [11]. In cases of COVID-19 disease, several studies have developed methods for detecting COVID-19 as in [12], [13], [14], [15], [16] which is based on a CT-scan image of the patient's chest. On the other hand, there are studies that detect COVID-19 based on coughing sounds from sufferers of COVID-19 as in the study [17], [18], [19], [20].

In the deep learning process, especially in the use of sound to carry out classification and detection, it is necessary to pre-process the data. One of the methods in audio pre-processing is the segmentation method. This method has been used in several studies such as sound event detection [21], analysis and detection of heart sounds [22], speech analysis for speech recognition [23], and audio classification [24].

In this study, we propose a way to improve the performance of deep learning-based COVID-19 cough detection using a segmentation method. In addition, we will also evaluate two segmentation methods to improve performance in deep learning, namely the hysteresis comparator method and the RMS (Root Mean Square) threshold. In this study, we used pre-trained PANNS as a deep learning model which was then finetuned using a combination of several COVID-19 cough datasets. Evaluation of the segmentation method begins with several stages, namely determining the combination of datasets, determining the distribution of datasets, and determining the threshold for the data selection process. We use unweighted accuracy as a metric evaluation to avoid bias due to the unbalanced of data.

2. Method

2.1. Datasets

One of the challenges and limitations of using Deep Learning, especially in the health sector, is data availability. To achieve an effective and robust deep learning model, large data is needed to develop the model [25]. However, in the health sector, available data is usually limited and biased. This is because the amount of data for healthy conditions is much greater than the data for sick conditions, the amount of data, especially image data for each disease category, is uneven, or the use of deep learning in specific fields is limited [26].

With these problems, we use three datasets, namely the COVID Cough Sound (CSS) dataset from Computational Paralinguistic Challenge (ComParE) 2021 [17] available privately, and the Coswara dataset [27] as well as the COUGHVID dataset [28] which is openly available to the public. The three datasets used contain recordings of coughing sounds that have been labeled as positive coughs for COVID-19 and negative coughs for COVID-19. The three datasets are combined with several combined variations and then divided into three data, namely train data used to train deep learning models, development data used to develop deep learning models, and test data used to test the performance of deep learning models that have been trained. It is hoped that by combining these three datasets, we can handle the large data needs in developing deep learning models.

The ComParE CCS dataset has a total of 725 recorded cough sounds consisting of one to three coughs taken from 343 participants for a total of 1.63 hours [17]. The cough recording has a sampling rate of 16 kHz consisting of positive COVID-19 cough sounds and negative COVID-19 cough sounds. ComParE CCS is divided into three parts, namely data train with a total of 286 audio (215 audio of negative cough and 71 audio of positive cough), data development with a total of 231 audio (183 audio of negative cough and 48 audio of positive cough), and test data with a total of 208 audio (169 negative cough audio and 39 positive cough audio).

The Coswara dataset is an open dataset released on August 7,2020 with a total of 6,507 clean audios. The dataset consists of nine categories, namely breathing (shallow and deep), cough (shallow and heavy), vowel phonation (a, i, u), and one to twenty digits counting (normal and fast paced) [27]. Audio recordings were taken from 941 participants who were healthy and unhealthy, including those identified as positive for COVID-19. Audio samples are taken via a web browser from a laptop or mobile phone with a sampling rate of 48 kHz. In this study we only took the type of cough sound.

ISSN 2723-6471

The COUGHVID dataset is a large cough sound dataset with 25,000 voices recorded and 1,155 tested positive for COVID-19. Data was collected between April 1, 2020 to December 1, 2020 through a web application developed by ecole polytechnique federale de lausanne (EPFL), Switzerland [28]. Cough sounds will be selected using cough detection and 4.6% of cough recordings will be obtained with cough sound detection above 0.8. Of these, there are 25% recordings with COVID status, 35% recordings with symptomatic status, 25% recordings with healthy status, and 15% recordings without status. We only use data with COVID status and healthy status.

2.2. Pre-trained Model

One solution to overcome the problem of lack of data when developing deep learning models is to use transfer learning [26]. Transfer learning is a promising technique in the world of deep learning that can improve model performance in several tasks such as computer vision or image classification for use as a feature extraction from images or videos. The use of transfer learning in the sound domain for several tasks such as classification has begun to be used, one of which is pre-trained audio neural networks (PANNs) which support several tasks such as audio tagging, acoustic scene classification, music classification, speech emotion classification, and sound event detection [29]. There are nearly 1.9 million audios from 527 different classes that are used to train the PANNs model using log-mel spectrograms extracted using fast fourier transform (FFT) and with a window type Hamming window size of 1024, hop length of 1024, and using a 64 mel filter bank. There are several models that are trained and used for transfer learning provided by PANNs, one of which is CNN14 which has the best performance among the other models [29]. In classifying coughs for COVID-19 and non-COVID-19, CNN14 using transfer learning has good performance with an unweighted accuracy of 75.90% and this proves that transfer learning is an effective method in developing deep learning models [30]. The CNN14 model consists of six convolution blocks with each convolution block having two convolution layers and ending with two fully connected layers.

2.3. Data Selection

Selection of cough sounds is the stage of elimination and filtering of cough sound data so that the data used in this study is only cough sound data and discards data that is not a cough sound. This process is one step to overcome problems in developing deep learning, namely related to data quality. Data in the healthcare field is very heterogeneous and noisy, so developing deep learning with healthcare data is challenging [25]. Because data selection is needed to obtain quality data for developing deep learning. At this stage, the data selection process is based on the use of a cough detection algorithm that can predict the value of the degree of certainty indicating that the detected audio contains cough [28]. The selection of this algorithm is a data selection process because the data used to train the algorithm has gone through selection and filtering by expert annotation. The cough detection algorithm used is an eXtreme Gradient Boosting (XGB) type classification model that has been trained with 215 recorded audio files categorized as coughing and non-coughing sounds, and has gone through an extraction process of 68 acoustic features. The cough detection algorithm is proven to have a precision performance of 95.4% in detecting cough sounds [28].

The cough detection process begins by extracting 68 acoustic features in the normalized cough sound file, then the extraction results are entered into the cough detection algorithm. Furthermore, the algorithm will process and predict sounds that are detected to contain coughing or not by showing the value of the degree of certainty for the sound. To get the threshold for the optimum degree of certainty, a computational experiment was carried out related to the threshold for cough detection.

2.4. Data Segmentation

One of the problems faced when developing deep learning in the Healthcare sector is very heterogeneous data [25]. The available data is mixed, resulting in unstructured data and it is difficult to optimize deep learning development. The process of overcoming this is also a challenge for some researchers in developing deep learning in the healthcare sector. Therefore, we carry out segmentation by standardizing the deep learning input in the form of 1 cough signal to overcome the problem of heterogeneous data. The cough segmentation stage aims to get audio containing one cough by cutting audio containing several coughs as shown in figure 1. Audio cutting is done automatically using two different methods, namely the hysteresis comparator and RMS threshold methods with audio input in the form of cough sounds resulting from cough detection. The performance of each of these methods can be seen in detail in previous research [31].

ISSN 2723-6471

In the Hysteresis comparator method, cough is segmented using a digital hysteresis comparator algorithm on the strength of the cough audio signal [28]. This method applies a lower segmentation threshold of 0.1 times the RMS cough sound signal and an upper segmentation threshold of 2 times the RMS cough sound signal. The use of this threshold is intended as a determinant of the start and end of a cough, so that when the signal strength of the coughing sound is greater than the upper segmentation threshold, it will be considered the beginning of the cough and if the signal strength after the initial determination of the cough is less than the lower threshold, it will be considered the end of cough.

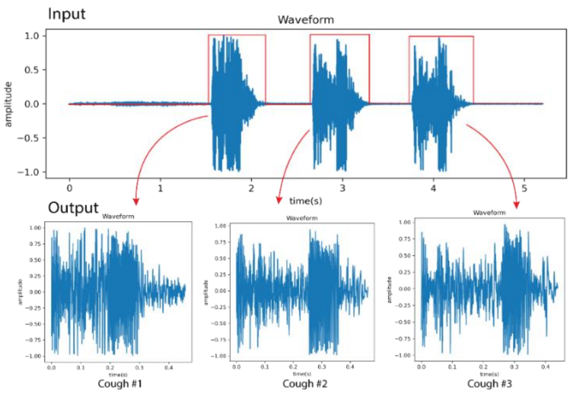


Figure 1. Illustration of the results of the segmentation process in one cough audio produces three cough audios, based on [31]

In the RMS threshold method, the cough sound is segmented based on RMS [32] value of the cough signal that meets the RMS threshold of 0.09 [31]. The RMS value on the cough signal will be preceded by the normalization process so that the RMS has a range of [0,1], then if the RMS signal value is greater than a predetermined threshold it will be considered as the beginning of a cough and if the next RMS value is less than the threshold it will consider the end of the cough [31].

2.5. Evaluation Metric

Unweighted accuracy (UA) or also known as unweighted average recall (UAR) is defined as the unbalanced average of the prediction accuracy of each class [33]. Unweighted accuracy is the amount of accuracy (recall) based on the class then divided by the number of classes [34]. Therefore, this evaluation depends on the number of classes to be classified. If there are two classes, then the probability level for the two classes is 50.0% UAR so class one has a 50% chance of accuracy from UAR and class two also has a 50% chance of UAR. The use of UAR in this research is to prevent accuracy values due to bias from unbalanced data. Therefore, by using UAR, we can show the performance of the model without the influence of data bias. The following is the UAR equation for the two classes.

$$UAR = \frac{1}{2} \times \left(\frac{TP}{TP+FP} + \frac{TN}{TN+FN} \right) \quad (1)$$

TP is true positive, FP is false negative, TN is true negative, and FN is false negative [35].

3. Results and Discussion

We present our results in several sections: choosing the right composition for combining the dataset, choosing the right portion for dividing the dataset, choosing the right cough detection threshold, and the effect of cough segmentation on the performance of deep learning-based classification of COVID-19 cough sounds. The sections were carried out sequentially by running the experiment on the seed “42”. The results reported are the results of experimental variations in each section and the best results are selected from each variation except when comparing cough segmentation. Evaluation of deep learning performance is based on testing the test data taken from ComParE CCS 2021.

2.1. Selection of the Right Combination of Dataset Compositions

As a first step, we select the composition of the dataset using three datasets with the aim of knowing the composition of the combined dataset that has the best effect on deep learning. We evaluated several variations in the composition of the combined datasets by using only positive data in each dataset except the CCS 2021 ComParE dataset or by using both positive and negative data in each dataset. Selection of variations in the composition of the dataset allows bias problems to arise and affects the performance of the model, especially in model generalization. However, this is helped by the evaluation metric that we use, namely UA, which provides accuracy information for each category, thereby minimizing bias due to selection of variations in dataset composition.

Table 1 shows the results of variations in the composition of the combined dataset on deep learning performance in classifying the cough sound of COVID-19. We report the results in the form of UA. Based on the reported results, the composition of the combined dataset consisting of positive and negative CCS 2021 ComParE, positive Coswara, and positive COUGHVID has the best influence on the classification of COVID-19 cough sounds with the highest UA results.

Table 1. Evaluation of variations in the composition of the combined dataset on deep learning performance

Dataset	UA
ComParE, Coswara, COUGHVID	56,31%
ComParE, Coswara, COUGHVID (Positive)	44,67%
ComParE, Coswara (Positive), COUGHVID	62,91%
ComParE, Coswara (Positive), COUGHVID (Positive)	**71,30%**

2.2. Selection of the Right Portion of Dataset

We know that there are no precise rules regarding the portion of data sharing in deep learning. The purpose of this study is to evaluate the portion of the dataset that can improve deep learning performance in the classification of COVID-19 cough sounds. The dataset is divided into training data and development data. In this study, the best dataset composition was used according to the results of table 1. The distribution portions evaluated were 70:30, 75:25, 80:20, 85:15, 90:10, 95:5.

Table 2 shows the results of an evaluation of the portion of the dataset distribution for training data and development data. We only report the results in the form of UA with the portion of the division in percent units. We did not include the amount of data distribution results in this study. Based on the results of the evaluation, the portion of the distribution of the dataset for use as training data and development data is 85:15 with the highest UA acquisition. Table 3 below shows the results of the evaluation of the cough detection threshold

Table 2. The results of the evaluation of the distribution of datasets as training data and development data

Training Data	Development Data	UA
70%	30%	65,18%
75%	25%	67,65%
80%	20%	68,14%
85%	15%	**72,68%**
90%	10%	68,54%
95%	5%	69,14%

Table 3. The results of the evaluation of the distribution of datasets as training data and development data

Threshold	UA
60%	71,40%
70%	71,99%
80%	72,48%
90%	**75,54%**

ISSN 2723-6471

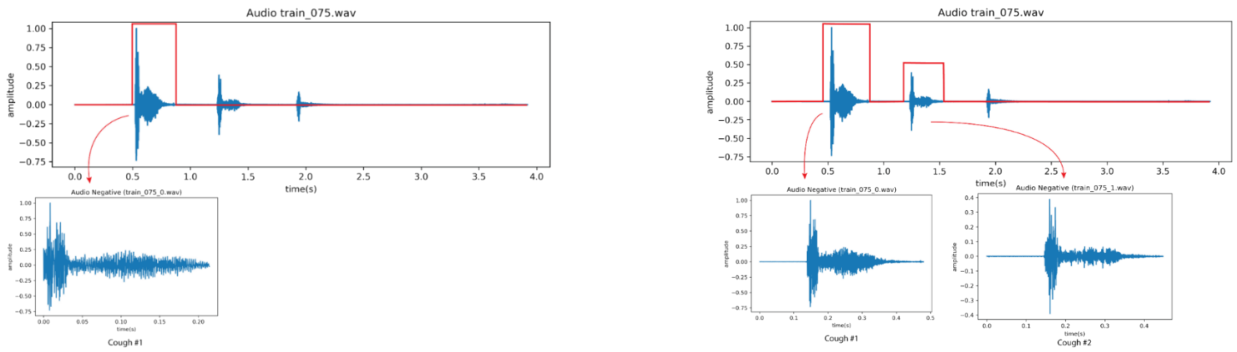
2.3. Selection of Appropriate Cough Detection Thresholds

It should be noted that the three datasets contain data that contains coughing sounds with background noise or data that contains non-coughing sounds. So, in this study we filtered the data by removing non-cough voice data using a cough detection algorithm [28]. The results of the cough detection algorithm are in the form of a degree of certainty which represents that the detected audio contains coughing sounds. However, there are no precise rules regarding the threshold of optimal cough detection to use. The threshold in question is the limit to the value of the degree of certainty of the cough detection results so that if an audio has a value above this threshold, it is considered a cough sound. We evaluate several thresholds namely 60%, 70%, 80%, and 90%.

The results of the evaluation of the cough detection threshold are shown in Table 3. The reported results show an increase when the threshold is larger. The 90% threshold for cough detection gets the best results with the highest UA value. This study proves that the use of cough detection with a threshold with the aim of filtering data can improve deep learning performance in classifying the cough sound of COVID-19.

2.4. Effects of Cough Segmentation on Deep Learning

Segmentation is a pre-processing that affects deep learning performance. The segmentation stage has been used in several cases such as in the classification of lung sounds [36] and on audio classification optimizations [24]. In the case of the classification of cough sounds for COVID-19, we used two segmentation methods for coughing, namely the hysteresis comparator method [28] and the method we developed is RMS threshold [31]. The method developed is based on the RMS value of the normalized cough sound signal with the range [0,1]. Then given an RMS threshold so that if the RMS value crosses the threshold, then the signal can be considered as a coughing sound. The cough segmentation process for each method can be seen in figure 2a and figure 2b. We evaluated the two segmentation methods using data with the best composition at the time of cough detection.



(a) Hysteresis comparator method

(b) RMS threshold method

Figure 2. Segmentation results from Hysteresis Method and RMS Threshold.

The amount of cough segmentation data from the two methods is shown in figure 3. The data segmented by the RMS threshold was more than the data generated by the hysteresis comparator for each data division. Table 4 shows the results of deep learning performance in classifying the cough sound of COVID-19 on the effect of using cough segmentation. Based on the reported results, the performance of deep learning has been enhanced by applying the segmentation process of the two methods. The hysteresis comparator method is the best method with the highest UA when compared to the RMS threshold method. According to [31], his study shows that the performance of the hysteresis comparator method has a higher precision of 73.33% while the RMS threshold method has a precision of 70%.

Table 4. Unweighted accuracy of cough segmentation method evaluation

Methods	UA
Hysteresis comparator [28]	**83,19%**
RMS threshold [31]	79,06%

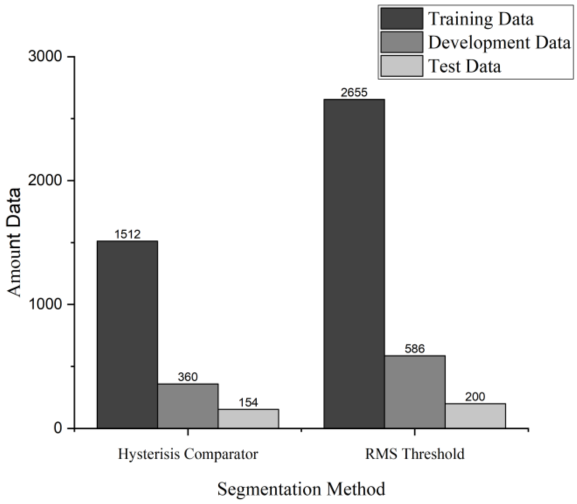
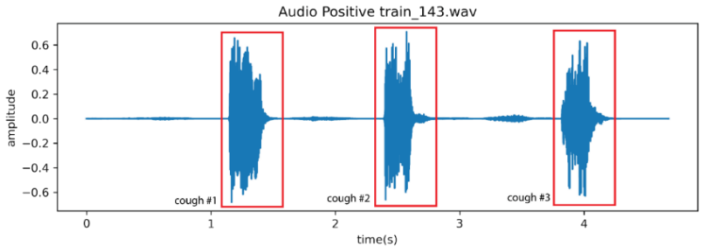
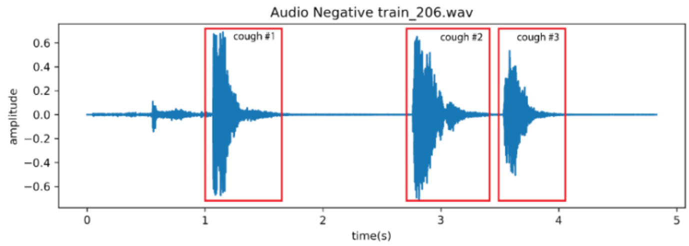


Figure 3. Total data results from the two cough segmentation methods on training data, development data, and test data.

Based on the analysis of the cough sound signal without segmentation process, the cough sound used as input data has noise or sound other than coughing. Figure 4a and figure 4b shows two positive and negative cough sound signals, each containing three coughs, one cough signal can be seen in the area inside the red box while outside the box area is a sound signal that is not a cough sound. Figure 5a and figure 5b shows the segmentation results of the positive and negative cough sound signals, the segmentation results each get three cough signals from one full cough sound signal.

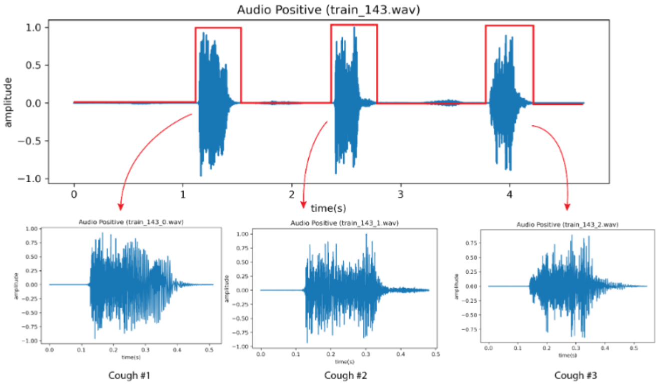


(a) Lots of coughing on positive

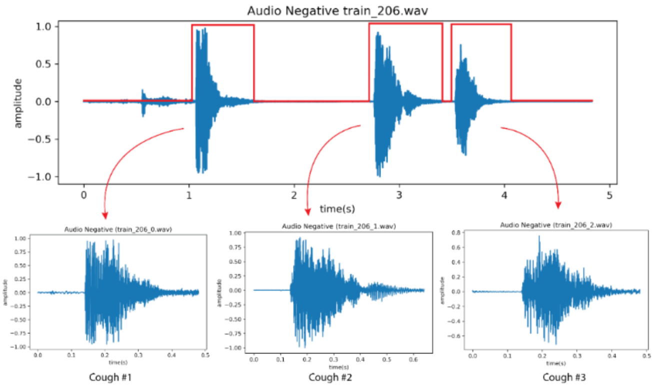


(b) Lots of coughing on negative cough

Figure 4. Lots of coughing on (a) Positive Cough and (b) Negative Cough sound audio without pre-processing is marked with a red box.



(a) Segmentation results on positive cough



(b) Segmentation results on negative cough

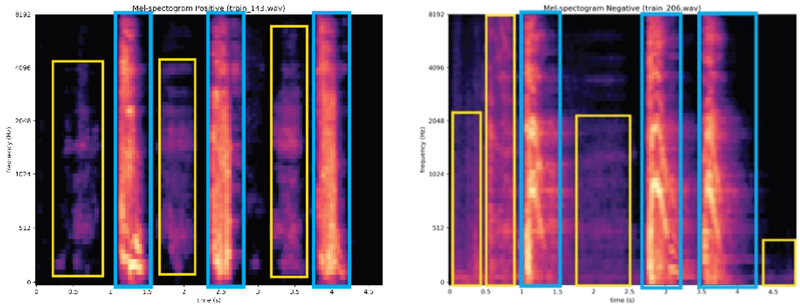
Figure 5. Segmentation results on (a) Positive Cough and (b) Negative Cough which both produce 3 cough audios

We also analyze the effect of segmentation based on the mel spectrogram used as an acoustic feature of the coughing sound. Figure 6a and figure 6b shows the visualization of the mel spectrogram of positive and negative coughing sounds from Figure 4a and figure 4b. There is an area in the image that is given a yellow box as the noise area with the purple mel spectrogram visualization and the blue box as the area of the cough signal with the mel spectrogram visualization orange & yellow. Based on Figure 6a and figure 6b, it is known that both positive and negative cough sound data contain noise or sounds other than coughing. Therefore, segmentation is carried out to get the sound of one cough without any noise. Figure 7a and figure 7b is a visualization of the mel spectrogram of one positive and negative cough segmentation results. Mel spectrogram analysis of positive and negative coughing sounds provides new information regarding the characteristics of a positive and negative cough for COVID-19. Cough characteristics are indicated in the area given a black box at the end of the cough sound. A positive COVID-19 cough is characterized by a high mel spectrum at a frequency of 0 Hz to 4096 Hz while a negative COVID-19 cough has a high mel spectrum at a frequency of 512 Hz to 2048 Hz. These characteristics also appear in other positive and negative cough data as shown in Figure 8a and figure 8b.

The results of the analysis prove that the segmentation process is useful for retrieving the necessary data and removing noise from the data or known as data cleaning. According to a study from [37], data cleaning can significantly improve model performance. In addition, the cough characteristics that appear in the visualization of the segmented mel spectrogram also help deep learning in recognizing the characteristics and differences in the positive and negative cough sound data of COVID-19. This is in accordance with the principles of deep learning in classifying data, namely by recognizing patterns from the input data [38], [39]. Therefore, the segmentation process can improve the performance of the model by 4% to 8% in detecting positive and negative COVID-19 coughs. Apart from that, when compared with several previous studies, by carrying out the segmentation process there is clearly an improvement. As in research conducted with the ComParE dataset and the Fusion of Best Configuration method, UA was obtained at 72.90% [17]. Then previous research using the ComParE dataset and the same using pre-trained PANNS to carry out

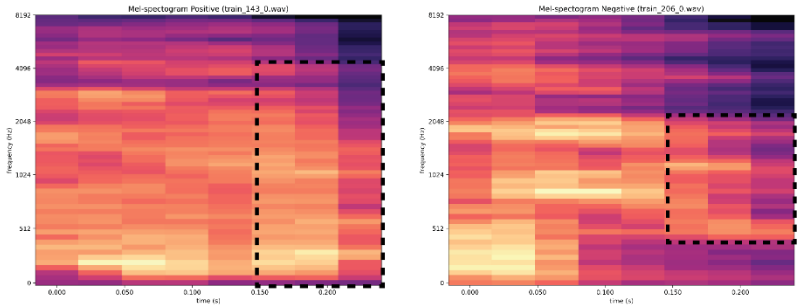
ISSN 2723-6471

classification obtained results of 75.90% [30]. This shows that the addition of the segmentation process provides improvements to the model.



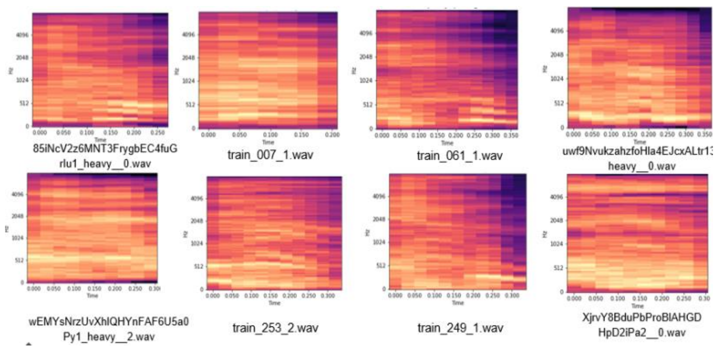
(a) Positive cough (b) Negative cough

Figure 6. Mel spectrogram on (a) Positive Cough and (b) Negative Cough

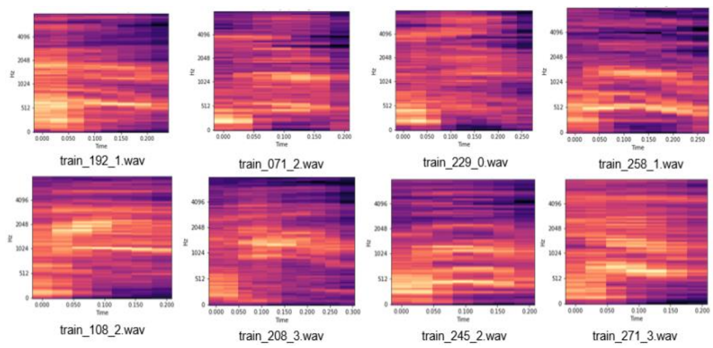


(a) One of the positive coughs (b) One of the negative coughs

Figure 7. Mel spectrogram on one of (a) the positive cough and (b) the negative cough segmentation results



(a) Several positive cough segmentation results



(b) Several negative cough segmentation results

Figure 8. Mel spectrogram results on several (a) positive cough and (b) negative cough segmentation results

 ISSN 2723-6471

4. Conclusion

In this paper, we propose a way to improve deep learning performance in detecting COVID-19 using the segmentation method. We evaluated two segmentation methods, namely RMS threshold method and the hysteresis comparator method and compared them with the COVID-19 cough detection process without going through the cough sound data segmentation process. The results of this study indicate that the use of segmentation in deep learning-based cough detection can improve model performance and this is in line with the increase in UA value of deep learning models by 4% to 8%. This increase is due to the segmentation process only focusing on taking one cough sound and removing noise from the cough sound so that it can be one of the techniques for data cleaning. Apart from that, the segmentation results on cough sounds provide information regarding the pattern or characteristics of positive and negative coughs so that this also helps the model in detecting COVID-19 coughs. The results of the evaluation of the segmentation method that we use show that the hysteresis comparator method has better performance in increasing the UA compared to the RMS threshold method. In the future, with the results we obtained, especially improving performance through the segmentation process, it can be implemented in cases of other respiratory diseases characterized by coughing, such as Chronic obstructive pulmonary disease (COPD) or Tuberculosis TB. This makes it easier for other researchers to develop deep learning models based on cough sounds by standardizing the deep learning input in the form of 1 cough signal and improving the performance of the deep learning model.

Apart from that, with this cough segmentation study, it is hoped that there will be several further research directions. One is to find differences between cough signals caused by various respiratory diseases so that it can become an important feature for further development. Another direction is to study the potential limitations and sources of error in the segmentation process, such as variability in cough sound characteristics among individuals to improve the performance of cough segmentation. The performance improvement could be carried out by study the potential strategies for addressing false positives or misclassifications in the segmentation results.

5. Declarations

5.1. Author Contributions

Conceptualization: S., Z., B.T.A., W.A.A.; Methodology: B.T.A.; Software: S.; Validation: S., B.T.A.; Formal Analysis: S., B.T.A.; Investigation: S.; Resources: W.A.A.; Data Curation: W.A.A.; Writing Original Draft Preparation: S. and B.T.A.; Writing Review and Editing: B.T.A. and S.; Visualization: S. All authors have read and agreed to the published version of the manuscript.

5.2. Data Availability Statement

The data evaluated in this study are obtained from third parties. Coswara is publicly available at http://github.com/iiscleap/Coswara-Data. The COUGHVID dataset requeires user access request, which can be requested at https://zenodo.org/records/5126606. The ComParE CCS is requested through ComParE 2021 Challenge [17].

5.3. Funding

This study is part of a research project with the number 1014/PKS/ITS/2022 funded by the Directorate of Research and Community Service, Institut Teknologi Sepuluh Nopember, Surabaya in 2022. The work of B.T.A is supported by the New Energy and Industrial Technology Development Organization (NEDO), Japan, under Project No. JPNP20006.

5.4. Institutional Review Board Statement

Not applicable.

5.5. Informed Consent Statement

Not applicable.

5.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] H. Wang *et al.*, “Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015,” *Lancet*, vol. 388, no. 10053, pp. 1459–1544, 2016, doi: 10.1016/S0140-6736(16)31012-1.

[2] World Health Organization, “WHO Coronavirus (COVID-19) Dashboard,” 2024. [Online]. Available: https://covid19.who.int/. [Accessed: Feb. 17, 2024].

[3] World Health Organization, “Tracking SARS-CoV-2 variants,” 2024. [Online]. Available: https://www.who.int/activities/tracking-SARS-CoV-2-variants. [Accessed: Feb. 17, 2024].

[4] Y. C. Liu, R. L. Kuo, and S. R. Shih, “COVID-19: The first documented coronavirus pandemic in history,” *Biomed. J.*, vol. 43, no. 4, pp. 328–333, 2020, doi: 10.1016/j.bj.2020.04.007.

[5] W. J. Song *et al.*, “Confronting COVID-19-associated cough and the post-COVID syndrome: role of viral neurotropism, neuroinflammation, and neuroimmune responses,” *Lancet Respir. Med.*, vol. 9, no. 5, pp. 533–544, 2021, doi: 10.1016/S2213-2600(21)00125-9.

[6] J. M. Hernández-Pérez, E. Martín-González, and M. Pino-Yanes, “Strengths and weaknesses of the diagnostic tests for SARS-CoV-2 infection,” *Med. Clínica (English Ed.)*, vol.155, no. 10, pp. 464-465, 2020, doi: 10.1016/j.medcle.2020.05.025.

[7] O. L. Usman, R. C. Muniyandi, K. Omar, and M. Mohamad, “Advance Machine Learning Methods for Dyslexia Biomarker Detection: A Review of Implementation Details and Challenges,” *IEEE Access*, vol. 9, no. 1, pp. 36879–36897, 2021, doi: 10.1109/ACCESS.2021.3062709.

[8] S. Bharati, P. Podder, and M. R. H. Mondal, “Hybrid deep learning for detecting lung diseases from X-ray images,” *Informatics Med. Unlocked*, vol. 20, no. 1, p. 100391, 2020, doi: 10.1016/j.imu.2020.100391.

[9] P. Ramprakash, R. Sarumathi, R. Mowriya, and S. Nithyavishnupriya, “Heart Disease Prediction Using Deep Neural Network,” *Proc. 5th Int. Conf. Inven. Comput. Technol. ICICT 2020*, vol. 5, no. 1, pp. 666–670, 2020, doi: 10.1109/ICICT48043.2020.9112443.

[10] A. M. Hassan, M. B. El-Mashade, and A. Aboshosha, “Deep learning for cancer tumor classification using transfer learning and feature concatenation,” *Int. J. Electr. Comput. Eng.*, vol. 12, no. 6, pp. 6736–6743, 2022.

[11] X. Liu *et al.*, “A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis,” *Lancet Digit. Heal.*, vol. 1, no. 6, pp. e271–e297, 2019, doi: 10.1016/S2589-7500(19)30123-2.

[12] M. Al Mehedi Hasan, J. Shin, and F. Parvin, “Deep Transfer Learning Based Detection of COVID-19 from Chest X-ray Images,” *ACM Int. Conf. Proceeding Ser.*, vol. 1, no. 1, pp. 64–70, 2021, doi: 10.1145/3460238.3460249.

[13] Z. Iklima, T. M. Kadarina, and E. Ihsanto, “Realistic image synthesis of COVID-19 chest X-rays using depthwise boundary equilibrium generative adversarial networks,” *Int. J. Electr. Comput. Eng.*, vol. 12, no. 5, pp. 5444–5454, 2022.

[14] D. Arias-Garzón *et al.*, “COVID-19 detection in X-ray images using convolutional neural networks,” *Mach. Learn. with Appl.*, vol. 6, no. April, p. 100138, 2021, doi: 10.1016/j.mlwa.2021.100138.

[15] A. Musha, A. Al Mamun, A. Tahabilder, J. Hossen, B. Jahan, and S. Ranjbari, “A deep learning approach for COVID-19 and pneumonia detection from chest X-ray images,” *Int. J. Electr. Comput. Eng.*, vol. 12, no. 4, pp. 3655–3664, 2022, doi: 10.11591/ijece.v12i4.pp3655-3664.

[16] P. Mahesh, Y. G. Prathyusha, B. Sahithi, and S. Nagendram, “Covid-19 Detection from Chest X-Ray using Convolution Neural Networks,” *J. Phys. Conf. Ser.*, vol. 1804, no. 1, 2021, doi: 10.1088/1742-6596/1804/1/012197.

[17] B. W. Schuller *et al.*, “The INTERSPEECH 2021 computational paralinguistics challenge: COVID-19 cough, COVID-19 speech, escalation & primates,” *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 6, pp. 4291–4295, 2021, doi: 10.21437/Interspeech.2021-19.

[18] E. Casanova *et al.*, “Deep Learning against COVID-19: Respiratory Insufficiency Detection in Brazilian Portuguese Speech,” *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, vol. 1, no. 1, pp. 625–633, 2021, doi: 10.18653/v1/2021.findings-acl.55.

[19] S. Gupta and A. Gupta, “Dealing with noise problem in machine learning data-sets: A systematic review,” *Procedia Comput.*

Sci., vol. 161, pp. 466–474, 2019, doi: 10.1016/j.procs.2019.11.146.

[20] N. Melek Manshouri, "Identifying COVID-19 by using spectral analysis of cough recordings: a distinctive classification study," *Cogn. Neurodyn.*, vol. 16, no. 1, pp. 239–253, 2022, doi: 10.1007/s11571-021-09695-w.

[21] Q. Kong, Y. Xu, I. Sobieraj, W. Wang, and M. D. Plumbley, "Sound Event Detection and Time-Frequency Segmentation from Weakly Labelled Data," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 27, no. 4, pp. 777–787, 2019, doi: 10.1109/TASLP.2019.2895254.

[22] Q. Liu, X. Wu, and X. Ma, "An automatic segmentation method for heart sounds," *Biomed. Eng. Online*, vol. 17, no. 1, pp. 1–22, 2018, doi: 10.1186/s12938-018-0538-9.

[23] Y. Amirgaliyev, M. Hahn, and T. Mussabayev, "The speech signal segmentation algorithm using pitch synchronous analysis," *Open Comput. Sci.*, vol. 7, no. 1, pp. 1–8, 2017, doi: 10.1515/comp-2017-0001.

[24] S. Zahid, F. Hussain, M. Rashid, M. H. Yousaf, and H. A. Habib, "Optimized Audio Classification and Segmentation Algorithm by Using Ensemble Methods," *Math. Probl. Eng.*, vol. 2015, no. 1, pp. 1-11, 2015, doi: 10.1155/2015/209814.

[25] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," *Brief. Bioinform.*, vol. 19, no. 6, pp. 1236–1246, 2017, doi: 10.1093/bib/bbx044.

[26] S. Yang, F. Zhu, X. Ling, Q. Liu, and P. Zhao, "Intelligent Health Care: Applications of Deep Learning in Computational Medicine," *Front. Genet.*, vol. 12, no. April, pp. 1–21, 2021, doi: 10.3389/fgene.2021.607471.

[27] N. Sharma *et al.*, "Coswara - A database of breathing, cough, and voice sounds for COVID-19 diagnosis," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 2020, no. Octob, pp. 4811–4815, 2020.

[28] L. Orlandic, T. Teijeiro, and D. Atienza, "The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms," *Sci. Data*, vol. 8, no. 1, pp. 2–11, 2021, doi: 10.1038/s41597-021-00937-4.

[29] Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. D. Plumbley, "PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 28, no. 1, pp. 2880–2894, 2020, doi: 10.1109/TASLP.2020.3030497.

[30] E. Casanova *et al.*, "Transfer learning and data augmentation techniques to the COVID-19 identification tasks in ComParE 2021," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 6, no. ii, pp. 4301–4305, 2021.

[31] B. T. Atmaja, Zanjabila and Suyanto, "Comparing hysteresis comparator and RMS threshold methods for automatic single cough segmentations," *Int. J. Inf. Technol.*, vol. 16, no. 1, pp. 5-12, 2023, doi: 10.1007/s41870-023-01626-8.

[32] B. T. Atmaja and M. Akagi, "The effect of silence feature in dimensional speech emotion recognition," *Proc. Int. Conf. Speech Prosody*, vol. 2020, no. May, pp. 26–30, 2020, doi: 10.21437/SpeechProsody.2020-6.

[33] F. Metze, A. Batliner, F. Eyben, T. Polzehl, B. Schuller, and S. Steidl, "Emotion recognition using imperfect speech recognition," *Proc. 11th Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH 2010*, vol. 11, no. 1, pp. 478–481, 2010.

[34] M. A. Nessiem, M. M. Mohamed, H. Coppock, A. Gaskell, and B. W. Schuller, "Detecting COVID-19 from breathing and coughing sounds using deep neural networks," *Proc. - IEEE Symp. Comput. Med. Syst.*, vol. 2021, no. June, pp. 183–188, 2021, doi: 10.1109/CBMS52027.2021.00069

[35] J. C. Kim, H. Rao, and M. A. Clements, "Investigating the use of formant based features for detection of affective dimensions in speech," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6975 LNCS, no. PART 2, pp. 369–377, 2011, doi: 10.1007/978-3-642-24571-8_48.

[36] S. B. Shuvo, S. N. Ali, S. I. Swapnil, T. Hasan, and M. I. H. Bhuiyan, "A Lightweight CNN Model for Detecting Respiratory Diseases from Lung Auscultation Sounds Using EMD-CWT-Based Hybrid Scalogram," *IEEE J. Biomed. Heal. Informatics*, vol. 25, no. 7, pp. 2595–2603, 2021, doi: 10.1109/JBHI.2020.3048006.

[37] P. Li, X. Rao, J. Blase, Y. Zhang, X. Chu, and C. Zhang, "CleanML: A study for evaluating the impact of data cleaning on ml classification tasks," *Proc. - Int. Conf. Data Eng.*, vol. 2021–April, no. 1, pp. 13–24, 2021.

[38] et al. Jan Wira Gotama Putra, "Pengenalan konsep pembelajaran mesin dan deep learning," *Comput. Linguist. Nat. Lang. Process. Lab.*, vol. 4, pp. 1–235, 2020, [Online]. Available: https://www.researchgate.net/publication/323700644

[39] L. Deng and D. Yu, "Deep learning: Methods and applications," *Found. Trends Signal Process.*, vol. 7, no. 3–4, pp. 197–387, 2013, doi: 10.1561/2000000039.