C3-PO: A Convolutional Neural Network for COVID Onset Prediction from Cough Sounds

Xiangyu Chen[1], Md Ayshik Rahman Khan[2], Md Rakibul Hasan[3], Tom Gedeon[1,3], and Md Zakir Hossain[1,3(✉)]

[1] Australian National University, Canberra, ACT 2601, Australia
[2] La Trobe University, Bundoora, VIC 3083, Australia
[3] Curtin University, Perth, WA 6102, Australia
{Rakibul.Hasan,Tom.Gedeon,Zakir.Hossain1}@curtin.edu.au

Abstract. This study presents a novel approach to diagnosing the highly contagious COVID-19 respiratory disease. Traditional diagnosis methods, such as polymerase chain reaction (PCR) and rapid antigen test (RAT), have been found to be resource-intensive and expensive, prompting the need for alternative diagnostic methods. Existing machine learning-based diagnosis approaches, such as X-rays and CT scans, suffer from suboptimal performance, primarily due to data imbalance and data paucity. To this end, this study proposes **C3-PO**, **C**ough sounds on **C**onvolutional neural network (CNN) for **C**OVID-19 **P**redicti**O**n. The framework utilises data augmentation and segmentation techniques to increase the volume of data to more than three times the original size. It includes an ensemble method to further mitigate the impacts of data paucity and data imbalances. Our CNN model was tested on the crowdsourced Coswara dataset and validated by the Russian dataset. It achieved an accuracy rate of 92.7% and an area under the receiver operating characteristics curve (AUC-ROC) of 98.1% on the Russian dataset, exceeding the existing works by 22% in terms of accuracy. On the Coswara dataset, the method achieved an accuracy rate of 72.3% and an AUC-ROC of 80.0%. Codes and evaluations are publicly available at https://github.com/ZakirANU/C3-PO-CovidCough-CNN.

Keywords: COVID-19 · Diagnosis · Cough sound · Convolutional neural network

1 Introduction

COVID-19 is a novel acute respiratory disease that emerged at the end of 2019 and spread worldwide. As of August 2023, more than 769 million people have been infected [25], with a death toll reaching 6.9 million. However, the statistics may fail to illustrate the severity of COVID-19 as there were many notable cases of under-reporting the infection and mortality data [19].

S. Rudinac et al. (Eds.): MMM 2024, LNCS 14556, pp. 355–368, 2024.
https://doi.org/10.1007/978-3-031-53311-2_26

In addition to its impact on human health, the quarantine policies and travel bans implemented to address COVID-19 have dealt a massive blow to global society and economy. Although the impact of the pandemic on the world is now fading, COVID-19 has sounded the alarm for the world.

COVID-19 is typically diagnosed using polymerase chain reaction (PCR) and rapid antigen test (RAT). PCR tests require specialised equipment and medical personnel, while RAT tests require mass production and are less accurate than PCR. Traditional detection methods, when dealing with such highly contagious diseases, not only consume huge medical resources and delay the timely treatment of other patients but also are inefficient and cannot produce quick results. With the rapid development of machine learning (ML), several studies [5,10,13] have proposed using ML algorithms to classify biological features and achieved some success. Furkan et al. [5] utilised CT images combined with machine vision algorithms to identify COVID-19, reaching an area under the receiver operating characteristics curve (AUC-ROC) of 95%. Kassania et al. [10] used X-ray and CT scan images for diagnosis and achieved up to 99% accuracy. Other studies [1,15] used sound signals for diagnosis. To this end, we aim to diagnose COVID-19 through cough sounds since sounds are easier to collect and far more cost-effective than other alternative signals, such as chest X-rays [10] and CT scans [5].

This paper has three major contributions. **(1)** To address the challenges posed by limited data and imbalanced datasets, we employ data augmentation and segmentation techniques to augment the volume of data. **(2)** We employ an ensemble of models to effectively leverage the samples from the majority class. **(3)** We propose a novel convolutional neural network (CNN)-based framework, **C3-PO**, to diagnose COVID-19 using cough sounds. Our method yields promising results on the Coswara [21] and Russian [6] datasets.

2 Related Work

2.1 Feasibility

Authors in [24] revealed the difference between medical symptoms of COVID-19 and other respiratory diseases. They observed that even in patients exhibiting fever, the commonly employed pharyngeal swab PCR test may yield negative results. This is attributed to the potential absence of viruses in the upper respiratory tract despite the presence of pneumonia. This observation suggests that COVID-19 may predominantly impact the lower respiratory tract, thereby distinguishing it from other respiratory diseases that primarily affect the upper respiratory tract. Using cough sounds, our study leverages theoretical medical knowledge to diagnose COVID-19.

AI4COVID pipeline [9] uses a nonlinear dimension reduction technique, t-SNE (t-Distributed Stochastic Neighbor Embedding), to perform two-dimensional visualisation of the extracted MFCC (Mel-Frequency Cepstral Coefficients) features from cough sounds. This verifies the feasibility of using cough sounds for COVID-19 diagnosis.

2.2 Cough Classification

In recent years, diagnostic methods have been proposed using cough sounds or respiratory sounds, such as the algorithm proposed in [18]. To detect pertussis, they preprocessed audio using a 3-phase cough model. They leveraged linear regression models to classify audio features such as MFCC [7], Zero-Crossing Rate (ZCR), and crest factor and achieved 92% accuracy. Authors in [11] used a lightweight CNN model on three different modalities (voice, breath and cough) and achieved approximately 92% accuracy in predicting COVID-19. Furthermore, authors in [20] used an ensemble of CNN models to classify COVID-19 and reported an AUC of 80.7%.

Lack of data has always been an issue for diagnosing COVID-19 from cough sounds. To address this, the algorithm proposed in [14] uses data augmentation methods such as pitch shift and time stretch to increase the quantity of data and improve the model's robustness. The algorithm improved the AUC-ROC from 72.23 to 87.07 on the first DiCOVA competition [16] dataset, winning first place in the competition. [2] discovered that each audio clip in the dataset might contain multiple cough sounds. As a result, they significantly increased the number of samples by segmenting the cough audios using root mean square energy (RMSE). The AI4COVID-19 [9] algorithm first trained CNN on cough detection datasets and then fine-tuned them on COVID-19 datasets.

To enhance the model's reliability, several authors have used various techniques. For instance, the AI4COVID-19 [9] algorithm uses three models that require a unanimous agreement for it to provide a definitive result, reducing the misdiagnosis probability. Authors in [12] simulated the effects of muscular degradation on audio samples. They performed the prediction on three different pre-trained models. The models extract audio features and concatenate them for classification, which helped them achieve an accuracy of 97.1% on a verified dataset.

3 Method

3.1 Cough Segmentation

To address issues such as poor generalisation and overfitting caused by the lack of data, audio samples are segmented, extracting single coughs as samples instead of whole audio files. This not only increases the number of samples but also removes useless sound segments, like noise and silence parts. Following the COUGHVID project [17], we also used RMSE to determine valid audio regions:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{1}$$

where x_i denotes each sample value in audio signal, $\bar{x}$ represents the average value of all sample values, and N represents the number of all sample values.

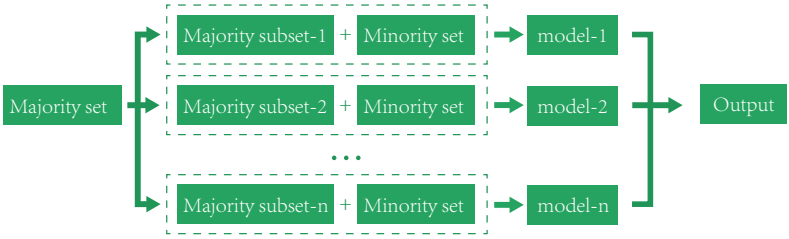


Fig. 1. The process of resampling majority set and ensemble models. The majority subset is divided into multiple subsets to match the minority subset to create multiple balanced datasets.

3.2 Data Augmentation

Data augmentation techniques are used to further enhance the model's generalisation capabilities. The algorithm generates a transformed sample for each original sample, with a variable rate between 0.7–1.40 and 0.9 probability of being stretched in time, 1 probability of pitch shifting (semitones ranging from −2 to 4), horizontal shift between −0.5 and 0.5 with 80% probability, trim and gain probability of 1, and 0.8 probability of polarity inversion. Note that data augmentation is performed on the training set only.

3.3 Split Majority Set and Ensemble Models

Common losses such as cross-entropy loss function do not consider the issue of dataset imbalance, which might result in a lack of generalisability and the model prioritising the majority class samples. To address these issues, a special method is proposed, an approach similar to [4]. The algorithm samples n subsets from the majority data set; each subset has the same number of samples as the minority data set, and these subsets are combined with the minority set to form n balanced data sets. Figure 1 illustrates the process.

The n datasets are used to train n independent models, and the results y_i of the n models are combined to generate the final output y:

$$y = \frac{1}{n} \sum_{i=0}^{n} y_i \tag{2}$$

It is worth noting that the segments in the test set may not come from the same audio source as those in the training set when splitting the training and test sets.

3.4 Model Architecture

The model architecture (Fig. 2) includes two convolutional layers and three fully connected layers, with cross-entropy as the loss function. In the network, dropout layers and batch normalisation layers are added to reduce overfitting. We implement the model using the Pytorch framework.

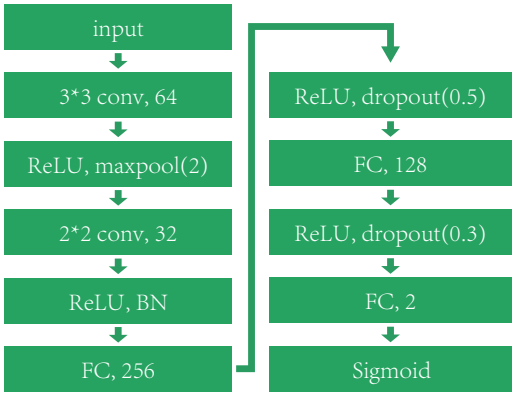


Fig. 2. An illustration of the network architecture. ReLU – Rectified Linear Unit, BN – Batch Normalisation, FC – Fully Connected.

3.5 Feature Selection

To ensure the effectiveness of the features, a forward selection method was used to screen the features. The specific steps are listed below:

1. Combine the k features to be adopted with the already adopted features.
2. Perform cross-validation to obtain the accuracy of the k combined features.
3. Compare the accuracy of the k combined features; if all are lower than the accuracy of the already adopted feature combination, stop selecting new features. Otherwise, select the to-be-adopted feature corresponding to the combination feature with the highest accuracy.
4. Return to the first step and iterate.

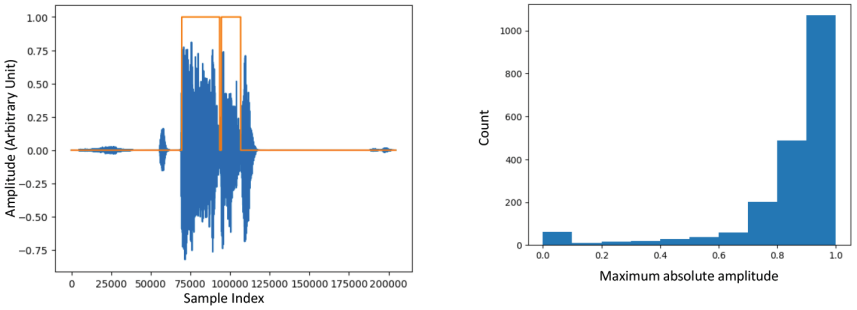
4 Experiments and Results

4.1 Dataset

This study primarily employs the Coswara dataset [21] and Russian dataset [6] to train and test the models. The Coswara dataset represents crowdsourced datasets and has been used in the second DiCOVA Challenge [22]. Each participant provided nine different kinds of sound samples in this dataset [3]. The Russian dataset, on the other hand, is a clinical dataset, implying that a portion of its data has been medically verified. However, the sample size is much smaller due to the higher cost of sample collection. The Russian dataset is a partially validated dataset, so we only take the validated data from the dataset, ensuring the labels are highly credible. The details of the datasets are provided in Table 1.

Table 1. Number of samples in the Coswara and the Russian datasets.

Dataset	Label	Raw	Segmentation	Augmentation
Coswara	Positive	589	1091	2182
	Negative	1405	3341	6682
	Total	1994	4432	8864
Russian	Positive	381	359	718
	Negative	438	914	1828
	Total	819	1273	2546



(a) (b)

Fig. 3. (a) An example of segmenting an audio signal, where the segments marked by orange line are cropped as useful segments. (b) Histogram of the maximum absolute value of sounds in the Coswara dataset. (Color figure online)

4.2 Data Preprocessing

In the Russian dataset, some audio files are recorded in stereo channels. We retain the data from one of the channels based on the size of the RMSE of each channel's signal.

In the Coswara dataset, heavy cough data is chosen for diagnosing COVID-19 since coughing is a significant characteristic of such respiratory diseases, and heavier cough sounds can provide more detailed information. The samples labelled as 'healthy' are selected as COVID-negative samples, and samples labelled as 'positive mild' and 'positive moderate' are selected as COVID-positive samples. After data cleaning, the dataset consists of 589 positive samples and 1,405 negative samples, totalling 1,994 samples. The number of positive samples is very small, so resampling the dataset would reduce the total number of samples to only 1,198. Such a small number of data could easily result in inadequate generalisation capabilities of the model.

Figure 3a depicts an example of segmenting an audio signal. After the audio segmentation processing, only the coughing sound is retained, while noises and blank segments are cut off. The total number of data increases to more than

twice the original, 4,432, with 1,091 positive samples and 3,341 negative samples. Finally, data augmentation techniques are applied to each audio segment, doubling the dataset size to a final count of 8,864 samples, which is 4.45 times the original. These processes help improve the model's generalisation ability.

Finally, non-repetitive resampling is conducted on the majority class (Negative), with the number of samples being the data quantity of the minority class (Positive). Since the negative sample is more than three times the positive sample quantity, it is possible to resample three groups of majority subsets equivalent to the minority quantity without repetition. The resampled majority subsets are combined with the minority set separately, forming three balanced datasets. Using three balanced datasets for training allowed us to produce comprehensive outcomes.

Because of diversified devices and formats, the maximum signal amplitude and sampling rates can vary significantly, especially in crowd-sourced datasets. To minimise such sample biases and other biases caused by external factors such as background noises, volume diversity, etc., the histogram of the maximum signal values is investigated. As can be seen in Fig. 3b, the maximum value of most audio is 1, but about 50% of the audio has a maximum value of less than 1. This can increase the convergence time of the model and may affect the final results. Therefore, resampling and normalisation are performed. In this study, sounds are resampled to 48,000 Hz and normalised to range from -1 to 1.

4.3 Feature Extraction

For each sample, multiple time-domain and frequency-domain features are extracted, including Mel spectrogram, MFCC, Chromagram, Centroid, Bandwidth, Flatness, Rolloff and Contrast.

All the features are extracted with the same frame length and hop length, thus concatenating all features as a matrix, namely a feature map. Given the human ear's sound length resolution is 20–50 ms, the frame length is set to 2,048, which is approximately 42 ms. The hop length is set to half the frame length, 1024, which allows overlap between frames to prevent the sound signal from being separated by the frames and the effective features from being extracted. To select the best set of MFCC features during feature selection, we used the Librosa Python package to extract 13, 26, and 39 coefficients as three types of features, which include MFCC features, MFCC differential features, and second-order differential features.

4.4 Data Analysis

First, we conduct PCA analysis on the original time signals, as shown in Fig. 4. It can be seen that the first principal component only accounts for 0.014 of the total proportion, and the boundary between positive and negative samples in the sample graph drawn with the first two components is also unclear. This

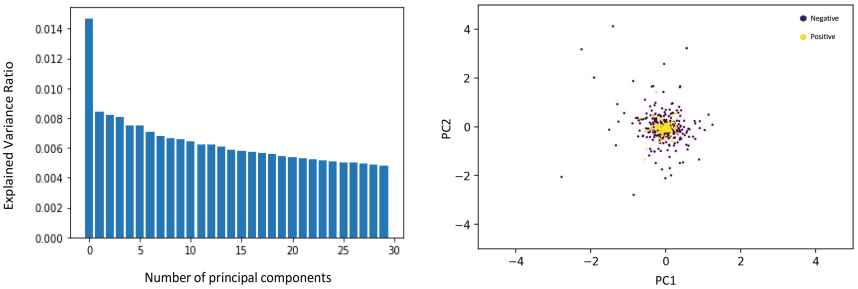
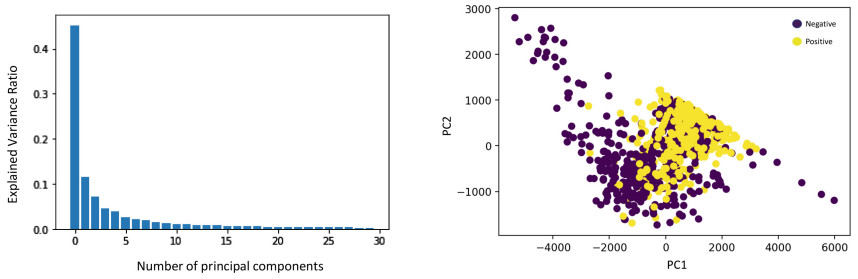
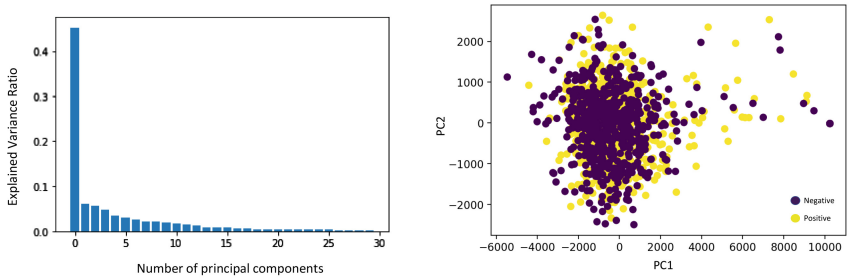


Fig. 4. PCA explained variance ratio (left) and first two principal components (right) of time signals in the Russian dataset. The purple points are negative samples, and the yellow points are positive samples. (Color figure online)



(a) Russian dataset (b) Russian dataset



(c) Coswara dataset (d) Coswara dataset

Fig. 5. PCA explained variance ratio (left) and first two principal components (right) of MFCC features. The purple and the yellow points refer to the negative and positive samples, respectively. (Color figure online)

suggests that it is difficult to obtain features suitable for classification from the time signal series or time domain features. All analyses were done using Python (version 3.8).

Table 2. 5-fold cross-validation accuracies with feature selection. Each entry represents the accuracy obtained using the selected features and a new feature. *Selected* means the feature is selected, and no more cross-validation is performed.

Features	Classification accuracy				
	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5
39 MFCCs	0.699	*Selected*	*Selected*	*Selected*	*Selected*
26 MFCCs	0.683	0.687	0.698	0.705	0.716
13 MFCCs	0.683	0.687	0.698	0.704	0.714
Chromagram	0.579	0.698	0.702	0.708	0.720
Mel spectrogram	0.601	0.663	0.673	0.681	0.792
Centroid	-	0.697	0.701	**0.723**	*Selected*
Bandwidth	-	**0.705**	*Selected*	*Selected*	*Selected*
Flatness	-	0.701	**0.712**	*Selected*	*Selected*
Onset	-	0.693	0.703	0.698	0.716
ZCR	-	0.685	0.702	0.710	0.718
Rolloff	-	0.700	0.701	0.709	0.711
Contrast	0.643	0.687	0.700	0.709	0.713

After the feature extraction, PCA performs quite differently. In Fig. 5 (a, b), the first two components account for over 0.5. In the sample location graph drawn using the first two components, the distribution of positive and negative samples is significantly different, with negative samples distributed more towards the left and positive samples distributed more towards the right. This validates the effectiveness of MFCC and corroborates our research direction.

However, interestingly, the MFCC feature map extracted from the Coswara dataset does not perform the same, as shown in Fig. 5 (c, d). Although the first two components also account for over 0.5, it is difficult to see the difference in the distribution of positive and negative samples from the drawn position map. This could be due to greater noise in the Coswara dataset due to crowdsourcing. The distribution of sample points can be another reason. Despite the difficulty in differentiating, the successful COVID detection reflects the reliability of our model.

4.5 Feature Selection

The test accuracies in each feature selection step are recorded in Table 2. As Centroid, Bandwidth, Flatness, Onset, ZCR, and Rolloff provide less information and perform poorly as independent training features, only the remaining features were considered in the first step to optimising run time. Therefore, 39 MFCCs, centroid, bandwidth and flatness are selected as the optimal feature combination.

4.6 Train and Test Models

Split Dataset. As samples have been segmented, it is possible that different samples come from the same audio file, making samples in the training set and

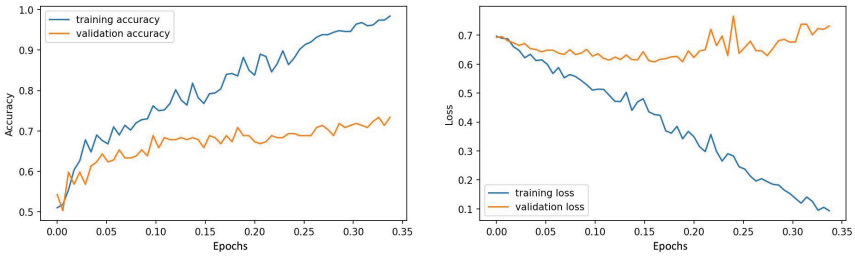


Fig. 6. Classification accuracy (left) and loss (right) during training.

test set similar to each other. To address this, the data are split specifically based on the source of the samples to ensure that samples from the same source are not distributed to different datasets.

Hyperparameters. We use Adam optimiser, whose learning rate is set at 1e-4 with β at (0.9, 0.999). The batch size is set to 10, running 20 epochs each time to find the optimal early stopping scheme, which is usually less than 10 epochs.

Cross Validation. We use 5-fold cross-validation to evaluate the performance of the model. As shown in Fig. 6, this model shows serious overfitting in later epochs; we thus impose early stopping techniques to obtain the optimal model. Therefore, unlike ordinary cross-validation, this paper further divides each fold's training set into a training subset and a validation subset. The training subset is used to train the model, and the validation subset is used to decide the epoch number to stop training based on the validation result. At last, the final result of this fold is tested on the test set using the early-stopped model. Here, the position of early stopping is actually treated as a hyperparameter. Such a nested structure ensures that the choice of hyperparameters has no correlation with the test set.

4.7 Results

As shown in Table 3, the proposed model's test result shows an AUC-ROC of 80.0% and an accuracy of 72.3% on the Coswara dataset. As Coswara is a continuously updated crowdsourced dataset, articles using Coswara may not necessarily use the same dataset. Therefore, we compare with the results from the second DiCOVA challenge [22], which also used the Coswara dataset. The proposed model ranked 4^{th} in all results, with a gap of 1.9% from the top AUC-ROC, indicating there is still room for improvement.

On the Russian dataset, our proposed model, **C3-PO**, achieves an AUC-ROC of 98.1%, and an accuracy of 92.7%, outperforming the Covid-Envelope model [8] by 22%.

Table 3. Performance comparison with other methods, including the top five teams from the DiCOVA challenge.

Dataset	Reference	AUC-ROC	Accuracy
Coswara	West Lab†	**0.819**	-
	WhyNot†	0.812	-
	USTCer†	0.801	-
	Team SMILE†	0.790	-
	ProPTIT†	0.778	-
	DiCOVA Baseline [22]	0.749	-
	C3-PO (Ours)	0.800	0.723
Russian	Covid-Envelope [8]	0.890	0.683
	C3-PO (Ours)	**0.981**	**0.927**

† Participants in the 2nd DiCOVA challenge [22]. Participants' results are available at https://competitions.codalab.org/competitions/34801#results (Track-2).

The difference in performance between the two datasets could be due to various factors. Firstly, the quality of the Coswara dataset is not as high as that of the Russian dataset; the voice information may be mixed with noise, and differences in recording devices and methods can also influence the recording results. Secondly, it is challenging to verify the annotation of crowdsourced datasets due to false information by crowdsource participants [23], which means that there is a possibility that the COVID status uploaded by users does not correspond with reality.

4.8 Ablation Study

Audio segmentation, data augmentation and the ensemble model are three primary methods used to handle small data volumes and data imbalance in our proposed **C3-PO** framework. To demonstrate the effectiveness of these three techniques, we perform ablation studies around them. By removing each technique one by one from the main workflow, we can see from the results the impact each technique has on model performance. The results recorded in Table 4 indicate that all three techniques contribute to a certain extent to the final performance of the model on both the Russian dataset and the Coswara dataset.

In the Coswara dataset, the model's accuracy is improved by 7.5% through these three techniques, whereas the improvement on the Russian dataset is only 2.2%. This is because the distribution difference between positive and negative samples in the Russian dataset is more significant. As we can see, the model can achieve a 90.5% accuracy rate without additional data preprocessing techniques, so the remaining room for improvement is limited. In contrast, the distribution of positive and negative sample data in the Coswara dataset is hard to distinguish linearly. More samples can help the model better identify the differences between

Table 4. Ablation study results. Three primary methods are added one by one in different experiments to find their significance.

Data	Method	Exp. 1	Exp. 2	Exp. 3	Exp. 4
	Cough Segmentation		✓	✓	✓
	Data Augmentation			✓	✓
	Ensemble Models				✓
Russian	AUC-ROC	0.959	0.970	0.975	0.981
	Accuracy	0.905	0.918	0.921	0.927
Coswara	AUC-ROC	0.731	0.762	0.781	0.800
	Accuracy	0.648	0.682	0.707	0.723

positive and negative samples. Notably, the most improvement was seen after the cough audio segmentation, making it the most effective method among the three.

5 Conclusion

This study proposes **C3-PO**, a novel framework to diagnose COVID-19 through cough sounds. Utilising data augmentation and segmentation techniques, the volume of data is increased to more than three times the original size to mitigate the impacts of a small dataset and imbalances within the dataset.

We tested the model on the crowdsourced Coswara dataset and the Russian dataset. On the Russian dataset, we achieved a state-of-the-art accuracy of 92.7% and an AUC-ROC of 98.1%, with 22% higher accuracy than the other work in the literature. On the Coswara dataset, we achieved an accuracy rate of 72.3% and an AUC-ROC of 80.0%, an AUC-ROC difference of 1.9% compared to the top-ranked models from the DiCOVA competition. Ablation studies are conducted to verify the influence of data augmentation, segmentation and ensemble models on the model's performance. The results indicate that all three methods positively contribute to the model's accuracy.

A limitation of this work is the use of crowdsourced datasets like Coswara, which run the risk of false or unverified data. A future direction for experimentation could be the application of transfer learning, transferring experience gained in other tasks, such as cough detection and speech recognition, into diagnosing COVID-19.

References

1. Andreu-Perez, J., et al.: A generic deep learning based cough analysis system from clinically validated samples for point-of-need Covid-19 test and severity levels. IEEE Trans. Serv. Comput. **15**(3), 1220–1232 (2021). https://doi.org/10.1109/TSC.2021.3061402

2. Ashby, A.E., Meister, J.A., Soldar, G., Nguyen, K.A.: A novel cough audio segmentation framework for covid-19 detection. In: Proceedings of the Symposium on Open Data and Knowledge for a Post-Pandemic Era ODAK22, UK, pp. 1–8 (2022). https://doi.org/10.14236/ewic/ODAK22.1
3. Bhattacharya, D., et al.: Coswara: a respiratory sounds and symptoms dataset for remote screening of SARS-COV-2 infection. Sci. Data **10**(1), 397 (2023)
4. Duan, H., Wei, Y., Liu, P., Yin, H.: A novel ensemble framework based on k-means and resampling for imbalanced data. Appl. Sci. **10**(5), 1684 (2020)
5. Furtado, A., da Purificação, C.A.C., Badaró, R., Nascimento, E.G.S.: A light deep learning algorithm for CT diagnosis of COVID-19 Pneumonia. Diagnostics **12**(7), 1527 (2022). https://doi.org/10.3390/diagnostics12071527
6. Geertsen, A., Chmelyuk, V.: Dataset of recordings of induced cough (Dec 2020). https://github.com/covid19-cough/dataset
7. Hasan, M.R., Hasan, M.M., Hossain, M.Z.: How many mel-frequency cepstral coefficients to be utilized in speech recognition? a study with the Bengali language. J. Eng. **2021**(12), 817–827 (2021). https://doi.org/10.1049/tje2.12082
8. Hossain, M.Z., Uddin, M.B., Yang, Y., Ahmed, K.A.: CovidEnvelope: an automated fast approach to diagnose Covid-19 from cough signals. In: 2021 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), pp. 1–6. IEEE, Brisbane, Australia (Dec 2021). https://doi.org/10.1109/CSDE53843.2021.9718501
9. Imran, A., et al.: AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. Inform. Med. Unlocked **20**, 100378 (2020). https://doi.org/10.1016/j.imu.2020.100378
10. Kassania, S.H., Kassanib, P.H., Wesolowskic, M.J., Schneidera, K.A., Detersa, R.: Automatic detection of coronavirus disease (Covid-19) in x-ray and CT images: a machine learning based approach. Biocybernet. Biomed. Eng. **41**(3), 867–879 (2021)
11. Kranthi Kumar, L., Alphonse, P.: Covid-19 disease diagnosis with light-weight CNN using modified MFCC and enhanced GFCC from human respiratory sounds. Europ. Phys. J. Special Topics **231**(18), 3329–3346 (2022)
12. Laguarta, J., Hueto, F., Subirana, B.: COVID-19 artificial intelligence diagnosis using only cough recordings. IEEE Open J. Eng. Med. Biol. **1**, 275–281 (2020). https://doi.org/10.1109/OJEMB.2020.3026928
13. Liu, X., Hasan, M.R., Ahmed, K.A., Hossain, M.Z.: Machine learning to analyse omic-data for COVID-19 diagnosis and prognosis. BMC Bioinform. **24**(7), 1–20 (2023). https://doi.org/10.1186/s12859-022-05127-6
14. Mahanta, S.K., Kaushik, D., Jain, S., Van Truong, H., Guha, K.: COVID-19 diagnosis from cough acoustics using convnets and data augmentation (May 2022). arXiv:2110.06123
15. Mohammed, E.A., Keyhani, M., Sanati-Nezhad, A., Hejazi, S.H., Far, B.H.: An ensemble learning approach to digital corona virus preliminary screening from cough sounds. Sci. Rep. **11**(1), 15404 (2021)
16. Muguli, A., et al.: Dicova challenge: dataset, task, and baseline system for covid-19 diagnosis using acoustics. arXiv preprint arXiv:2103.09148 (2021). 10.48550/arXiv.2103.09148
17. Orlandic, L., Teijeiro, T., Atienza, D.: The coughvid crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms. Sci. Data **8**(1), 156 (2021). https://doi.org/10.1038/s41597-021-00937-4

18. Pramono, R.X.A., Imtiaz, S.A., Rodriguez-Villegas, E.: A cough-based algorithm for automatic diagnosis of pertussis. PLoS ONE **11**(9), e0162128 (2016). https://doi.org/10.1371/journal.pone.0162128
19. Richards, R.: Evidence on the accuracy of the number of reported covid-19 infections and deaths in lower-middle income countries. K4D Helpdesk Report 856 (2020). https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/15576
20. Schuller, B.W., Coppock, H., Gaskell, A.: Detecting covid-19 from breathing and coughing sounds using deep neural networks. arXiv preprint arXiv:2012.14553 (2020)
21. Sharma, N., et al.: Coswara - a database of breathing, cough, and voice sounds for Covid-19 diagnosis. In: Interspeech 2020, pp. 4811–4815 (Oct 2020). https://doi.org/10.21437/Interspeech.2020–2768
22. Sharma, N.K., Chetupalli, S.R., Bhattacharya, D., Dutta, D., Mote, P., Ganapathy, S.: The second dicova challenge: dataset and performance analysis for diagnosis of covid-19 using acoustics. In: ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 556–560 (2022). https://doi.org/10.1109/ICASSP43922.2022.9747188
23. Sheehan, K.B.: Crowdsourcing research: data collection with amazon's mechanical turk. Commun. Monogr. **85**(1), 140–156 (2018). https://doi.org/10.1080/03637751.2017.1342043
24. Tian, S., Hu, W., Niu, L., Liu, H., Xu, H., Xiao, S.Y.: Pulmonary pathology of early-phase 2019 novel coronavirus (covid-19) pneumonia in two patients with Lung Cancer. J. Thorac. Oncol. **15**(5), 700–704 (2020). https://doi.org/10.1016/j.jtho.2020.02.010
25. World Health Organization: WHO coronavirus (COVID-19) dashboard (2023). https://covid19.who.int/