



Accumulated bispectral image-based respiratory sound signal classification using deep learning

Sandeep B. Sangle¹ · Chandrakant J. Gaikwad¹

Received: 31 January 2023 / Revised: 6 April 2023 / Accepted: 7 April 2023 / Published online: 5 May 2023
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2023

Abstract

The COVID-19 virus is increasingly crucial to human health since new variants appear frequently. Detection of COVID-19 through respiratory sound has been an important area of research. This study analyzes respiratory sounds using novel accumulated bi-spectral features. The principal domain bispectrum is used for computing accumulated bispectrum. The resulting magnitude bispectrum is used in forming the bispectral image. In this work, a convolutional neural network (CNN) and ResNet-50 algorithms are designed to classify respiratory sounds as either COVID-19 or healthy. The performance of the proposed method is compared with the state-of-the-art methods. The proposed CNN-based method achieves the highest accuracy of 87.68% for shallow breath sounds, and ResNet-50 achieves the highest accuracy of 87.62% for deep breath sounds. Similarly, proposed methods gives the improved performance for other respiratory sounds.

Keywords Accumulated bispectrum · Bispectral image · CNN · COVID-19

1 Introduction

In February 2020, the World Health Organization (WHO) declared Corona Virus Disease of 2019 (COVID-19) a worldwide pandemic [1]. As the COVID-19 virus has spread worldwide, it is crucial to provide an accurate method for determining whether or not an individual is infected. Fever, dry cough, and breathing difficulties caused by acute respiratory infections are some of the signs of COVID-19. It is crucial to recognize COVID-19 quickly and prevent harm to various organs.

Researchers are searching for bio-indicators to detect COVID-19. These studies have either been based on the analysis of speech sounds [2–4], CT images, or chest X-rays. Also this is receiving significant attention from the signal processing and artificial intelligence researchers.

In [5], the Environmental Sound Classification 50 (ESC-50) dataset is used to find COVID-19. In this work, the author

used a fusion of MFCC and its principal component analysis (PCA) for support vector machine (SVM). They used dataset with a small number of samples and achieved 95.6% accuracy using 247 healthy samples and 70 COVID-19 samples. For detecting COVID-19 cough shallow (CS) samples [6], authors used data samples from the Coswara and the Virufy datasets. The entropy of Energy, MFCC, Zero Crossing Rate (ZCR), and spectral-based features are used in SVM, KNN and RNN models. Their RNN model reported 81.25% accuracy testing 200 samples of coswara dataset. In [7], authors used empirical mode decomposition (EMD) based 45 features and discrete wavelet transform-(DWT) based 54 features. After applying the Relief feature selection method and SVM linear classifier 98.4% accuracy and 98.6% F1-score values were reported using balanced 595 COVID-19 positive and 592 COVID-19 negative VIRUFY dataset.

In [8, 9] authors reported transfer learning-based deep neural network (DNN) classification methods for covid-19 identification using speech, cough, and breath samples with a ROC-AUC of 0.923, 0.982, and 0.942, respectively. A total of 1171 respiratory sound samples are used for both classes to extract the MFCC, ZCR, and kurtosis features from the audio signal. SMOTE oversampling was utilized to balance the training set. In [10], a fractal-dimension (FD) image is proposed for COVID-19 cough detection. The Coughvid, Virufy, and Coswara datasets were used in this study and reported

✉ Sandeep B. Sangle
sandeepsangle12@gmail.com

Chandrakant J. Gaikwad
cjgaikwad@gmail.com

¹ Department of Electronics and Telecommunication Engineering, Ramrao Adik Institute of Technology, D.Y. Patil Deemed to be University, Nerul, Navi Mumbai, Maharashtra 400706, India

98.45%, 98.15%, and 97.59% accuracy, respectively. In this author used two networks sequentially, the first to detect the cough and the second to identify the COVID-19 samples, which increased the system complexity. The STFT and MFCC features are utilized in [11] for COVID-19 cough analysis and classification using SVM. Authors used 121 cough samples from the VIRUFY dataset.

In [12], the authors used 1040 COVID-19 and non-COVID-19 cough samples from Dicova Dataset. The authors extracted MFCC, ZCR, RMS energy and spectral features and used VGG-13 for classification. They computed AUC values using cross-entropy loss and focal loss methods. In [13], the authors used 755 COVID-19 and 702 cough sound samples to classify COVID-19 cough sounds using six deep transfer models. In this, the authors used scalograms for training the models. The authors reported 94.9% accuracy for ResNet-18 model. In [14], the Cambridge, Coswara, and COUGHVID datasets are combined and authors removed all poor quality and noisy cough sounds. Overall they considered 2049 samples. The authors classified the samples as COVID-19, non COVID-19 and symptomatic. They used MFCC, spectral features, RMS energy, and ZCR are for training the model.

In [15], they evaluated an ensemble model by combining the ICBHI with the Coswara speech, cough, and breathing datasets. Approximately 110 samples for the COVID-19 positive and negative classes are utilized for respiratory analysis. The CRNN, BiLSTM, and BiGRU models are implemented using MFCC features. The accuracy reported for cough, breath, and speech sounds is 98.25%, 90.66%, and 92.40%, respectively. Here, authors have used only 110 samples.

In [16], the author has used a multi-branch input convolution network with MFCC, spectrogram, and chromagram features. The VIRUFY dataset achieved 61% accuracy and 90.4% accuracy in the NoCoCoDa dataset. Here, authors used few samples for testing. In [17], the authors used vowel sounds (/a, /e, /o) from the Coswara dataset. They used 822 training samples and 205 testing samples and reported 97.07% accuracy.

The various methods described above have been proposed in the literature to capture the characteristics of COVID-19 from respiratory sound signals. Most works use features related to MFCC, spectral and basic features. The MFCC and the features based on second-order statistics ignore the signal's nonlinear information. In the case of noisy data, the performance of these methods degrades [18]. Additionally, most of the research mentioned above used small datasets, and the models are complex in many cases.

In the literature, we find many bispectrum-based works that analyse the sound signals. A bispectrum phase reconstruction algorithm has been proposed in [19]. An adaptive system identification method based on bispectrum is described in [20]. The characterization of sleep spindles

using Higher Order Statistics (HOS) and spectra is described in [21]. The authors have used HOS for computing modified AMDF and modified AMDFSoA; further, these functions are used to characterize sleep spindles [21]. In [22], a technique for cross term reduction in wigner-ville distribution is described using modified bispectrum. The key difference between these works and the proposed work is the use of machine learning and deep learning algorithms. These works do not use machine learning or deep learning algorithms for the problems under consideration.

This work explores the use of higher-order spectra (HOS) in COVID-19 detection. Here, we have proposed a 1-D bispectrum obtained from a 2-D bispectrum by using the accumulation of the bispectrum on the frequency variable. Thus, this study uses novel features driven by HOS. The magnitude of the accumulated bispectrum is used in forming the image features that are further used to train the CNN and ResNet-50 networks. Typically, these networks require features in the form of images. We have created image features using the magnitude of the accumulated bispectrum. Thus, for COVID-19 detection using audio signals, accumulated bispectrum-based magnitude features are proposed in this work. A detailed performance analysis is presented in the paper. The proposed method performs better than the methods available in the literature.

The organization of the paper is as follows. The proposed methodology is described in Sect. 2. Experimental results and comparative study is presented in Sect. 3. A summary of the proposed work and the conclusions are given in Sect. 4.

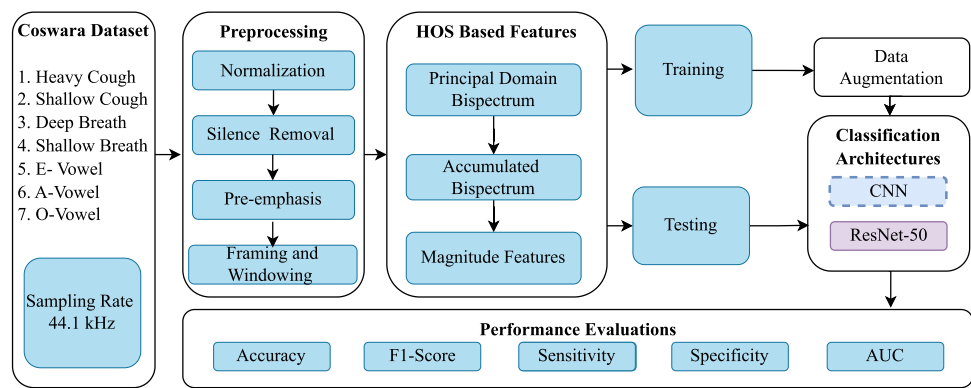
2 Proposed methodology

To analyze the importance of respiratory sound signals in detecting COVID-19, accumulated bispectrum-based features and deep neural networks (DNN) are considered. A description of the accumulated bispectrum based features and classification systems is given below.

The methodology followed in this work can be divided into signal preprocessing and feature extraction parts shown in Fig. 1. The signal preprocessing step involves normalizing, silence removal, framing and windowing the raw data. The feature extraction step computes the novel accumulated bispectrum based features. The classification algorithms CNN and ResNet-50 are trained and tested and the performance evaluation is carried out using metrics, such as accuracy, F1-score, etc. The following subsection explains the detailed methodology.

This study uses the Coswara dataset, which is produced by the Indian Institute of Science Bengaluru, India [23]. This dataset is open and available online. It includes COVID-19 and healthy samples of healthy and shallow coughs, shallow and deep breaths, and vowel a/e/o sounds. This dataset

Fig. 1 Flow diagram of the proposed methods for COVID-19 detection



contains over 2700 COSWARA sound samples, each with a sampling frequency of 44.1 kHz.

This dataset is unbalanced data set, but the data should be balanced to create a precise algorithm. For data augmentation in this study, we used the SMOTE oversampling approach [24]. There are binary classes available in the dataset used in this work. The minority (COVID-19 positive) and majority (COVID-19 negative) samples are available. The SMOTE oversampling method generates an equal amount of minority and majority samples. The above method has been applied in several works [2, 8, 9].

2.1 Preprocessing

Preprocessing consists primarily of three steps. As illustrated in Fig. 1, these steps are normalization, silence removal, and pre-emphasis. The main objective of normalization is to reduce the impact of amplitude fluctuations on the decision. This is achieved using Eq. (1).

$$x_N[n] = 0.9 * \frac{x[n]}{|\max(x[n])|} \quad (1)$$

where $x_N[n]$ represents the normalised signal and $\max(x[n])$ represents the maximum magnitude of $x[n]$. Silence is detected using signal power. An empirical threshold is used to remove the silence from the normalized signal. Signal is grouped as speech or silence, depending on whether it is above or below this threshold. Further, this is given to the pre-emphasis filter with coefficient set to 0.97.

This pre-emphasised signal is further used for windowing and framing. The Hanning window length is 1024 samples, and the hop length is 256 samples. The sampling frequency is 44.1 KHz. These preprocessed sound samples are used for further processing.

2.2 Respiratory sound feature extraction

In this work, we propose accumulated bispectrum-based features for COVID-19 signal detection. These features are used

as image input to train the CNN and ResNet-50 networks. Respiratory sound signals are used to compute the bispectrum as given in (2). This bispectrum is a two-dimensional complex quantity.

$$B(f_1, f_2) = X(f_1)X(f_2)X^*(f_1 + f_2) \quad (2)$$

where $X(f)$ is the Fourier transform and $X^*(f)$ is complex conjugate of $X(f)$. The details of the computation of principal domain bispectrum are given in [25, 26]. This computation only covers the frequency range of the principal domain as shown in Fig. 2. The accumulation of this bispectrum is computed by (3).

$$AB[f_1] = \sum_{f_2=0}^{(N-1)-f_2} B[f_1, f_2] \quad (3)$$

The resulting accumulated bispectrum will be a 1-dimensional quantity representing an accumulated bispectrum (AB) along one frequency variable. The principle domain bispectrum of each frame is shown in Fig. 2. The AB is computed from this principal domain bispectrum.

Now, this AB is used to extract magnitude and phase values. The $AB(f_1)$ is a complex quantity. According to Eq. (4) the magnitude and phase relationship between the bispectrum of the signal is given by

$$AB(f_1) = |AB(f_1)|e^{j\Psi(f_1)} \quad (4)$$

Each frame's AB magnitude values are stored as a row in the magnitude AB image. The magnitude AB images are used for further training and detecting COVID samples using the CNN and ResNet-50 classifiers. The results and performance analysis is presented in the next section.

The first and second column of Fig. 3 shows an analysis of the Healthy Cough (HC) and COVID-19 Cough (CC) sound samples. Healthy Breath (HB) and COVID-19 Breath (CB) sound is shown in the third and fourth column of Fig. 3.

Fig. 2 Flow of extraction of accumulated bispectrum

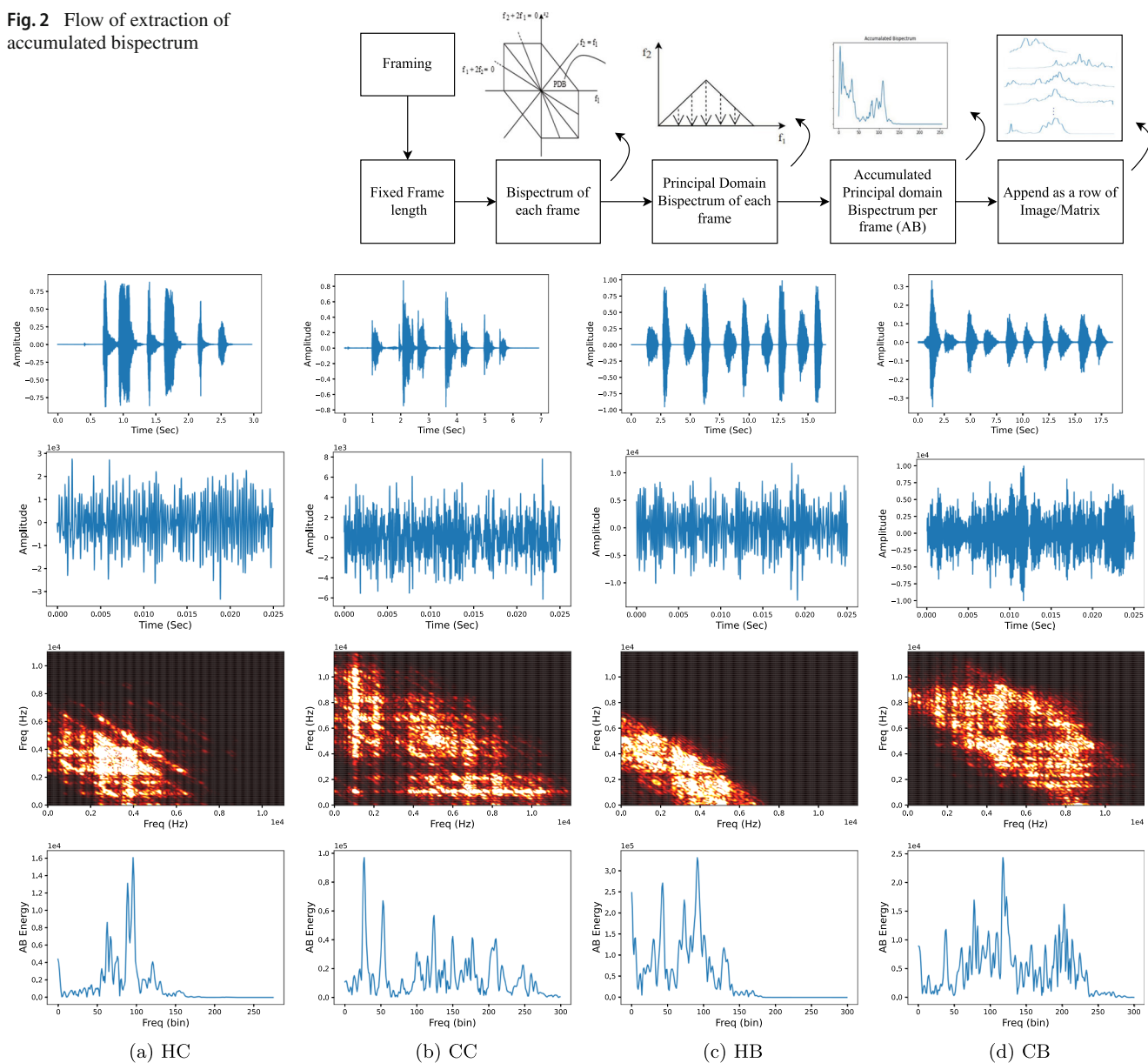


Fig. 3 Analysis of accumulated bispectrum

The audio signals are shown in the first row in Fig. 3. The sample frames are shown in row two. The bispectrum and accumulated bispectrum are shown in rows three and four respectively. We can observe AB for healthy and COVID-19 coughs and note that AB is more concentrated in healthy coughs, whereas AB energy is distributed in COVID-19 coughs.

2.3 Classification architecture

In this experiment, CNN and ResNet-50 architectures are used to categorise COVID-19 or healthy sounds. When the input layer receives an image in CNN, the convolution

and pooling layers extract the features through the layers. Therefore, features are modified from low-level to high-level features and classified through fully connected layers. Similarly, transfer learning is defined as training a neural network on a new task using the weights of another similar task. ResNet50 is also a version of the ResNet transfer learning model that contains 48 convolution layers, one max pool and one average pool layer.

In this network, we used two convolutions and max-pooling layers. The three fully connected layers have 32, 125, and 2 neurons respectively. All convolution layers used 3×3 kernel size and did not use zero padding. The convolution processes in both CNN architectures employ the Relu

Table 1 CNN parameters used for magnitude AB image features

Layer (type)	Output shape	No. of parameters
Input layer	(100, 256, 1)	0
Conv2D	(98,254, 24)	240
MaxPooling	(49, 127, 24)	0
Conv2D	(47, 125, 36)	7812
MaxPooling	(23, 62, 36)	0
Flatten1	(None, 51,336)	0
Dense1	(None, 32)	1642784
Dropout	(None, 32)	0
Dense2	(None, 125)	4125
Dense3	(None, 2)	252

activation function, while the output layer uses the softmax activation function. After the convolutional layer, a 2×2 size max-pooling layer is used. A dropout layer has been added to prevent overfitting since some neurons were disabled during the data training. The output shape and the number of trainable parameters with respective layers are shown in Table 1. We used 100 epochs and 125 batch sizes to train the CNN architecture.

ResNet50 requires a 3-channel input; thus, the features are modified to 3 channels by replicating the same single-channel to all three channels to overcome this difficulty. The shape of the input layer is (100, 256, 1); therefore, it also needs to be resized into (100, 256, 3). Imagenet's trainable parameter has been set to include_top = False refers to the activation of feature extraction, which removes the last dense layer. Furthermore, ResNet-50 utilizes a total of four dense layers. The last two layers contained 64 and 2 neurons, respectively, while the first two used 1024 and 256 neurons. Furthermore, this experiment has used regularisation methods to prevent over-fitting as drop-out between the first and second dense layers. There are 50 epochs used during ResNet-50 training.

The ADAM optimizer and binary cross-entropy are used for the loss function in both classification architectures. Low learning rates are preferred to preserve a large portion of previous learning. The learning rate in this study is kept at 0.0001. Softmax activation function is used in the last dense layer.

3 Results and discussion

In this section, we analyze and discuss the results. The augmented data is divided into 80 and 20% for training and testing purposes. The CNN and ResNet-50 classification architectures used in this work are explained in the previous section. The performance is shown using 2×2 confusion

matrices. Details of performance evaluation parameters are described in [2].

The simulation results of the method given in [9] with the updated larger dataset are shown in Table 2. The MFCC, ZCR, and kurtosis features are used in [9]. As indicated in [9], 13 MFCC and 26 features using first and second-order time derivatives of MFCC are used. The ZCR and kurtosis of the respiratory sound signal are used to capture additional information about the characteristics of the sound. The 39 MFCC features are combined with the ZCR and kurtosis of the respiratory sound signal and these features are applied to the CNN and ResNet-50 model.

The results are shown in Table 2. For CNN model, the F1 score is 82.22% and the accuracy is 84.15% for shallow cough. For shallow breathing, the F1 score is 81.55%, and the accuracy is 83.49%. The accuracy for vowel sounds is relatively less than cough and breath sounds. For ResNet-50 model, all sound types gives similar accuracy of around 83.2%.

The results of the method given in [13] are shown in Table 3 for comparison purposes. In this method, scalogram images are extracted using the Continuous Wavelet Transform (CWT). The CWT coefficients provide a time-frequency representation of the signal. CWT allows for frequency content analysis over time. The Morlet filter is used in wavelets for capturing a signal's high and low-frequency components. The CWT is computed using the equation (5) at a scale of ($x > 0$).

$$\text{CWT}_{(x,y)} = \frac{1}{\sqrt{x}} \int_{-\infty}^{\infty} f(s) \vartheta \left(\frac{s-y}{x} \right) ds \quad (5)$$

where $\vartheta(s)$ is a continuous function in both the time and frequency domains and is called as the father signal. x is a scale parameter and y is a position parameter.

By using CWT coefficients, a wavelet matrix with scale and position information is generated. The results of this method using the updated Coswara dataset are shown in Table 3. These scalogram images are applied to the CNN model. For shallow cough sounds, the F1-score is 81.63%, and the accuracy is 83.08%. For shallow breath, the F1 score is 80.97%, and the accuracy is 82.46%.

As can be seen in Tables 2 and 3, all respiratory sound types in Table 2 have the higher accuracy values except for vowel a and o. This shows that the method [9] performs better than the method in [13].

The results of the proposed method based on the magnitude of AB are shown in Table 4. A five-time mean is performed for all of the experiments. The proposed CNN network performs better for shallow and deep breath sounds. The 87.68% in accuracy (Acc) and F1-score (F1), 86.07% for the shallow breath sound, is visible in Table 4. ResNet-50 performs relatively better than CNN for the deep breath sound

Table 2 Classification performance using features proposed in [9] with updated dataset

Sound	Accuracy	F1-score	Sensitivity	Sepecificity	AUC
<i>CNN</i>					
Cough heavy	84.15	82.22	73.26	95.04	0.841
Cough shallow	83.2	81.11	72.09	94.32	0.832
Breath deep	82.51	80.27	71.18	93.84	0.825
Breath shallow	83.49	81.55	72.95	94.04	0.834
Vowel a	81.6	79.39	70.86	92.34	0.816
Vowel e	82.63	79.94	69.21	96.05	82.63
Vowel o	80.32	79.32	75.49	85.14	80.32
<i>ResNet-50</i>					
Cough heavy	82.45	78.96	65.65	99.04	0.824
Cough shallow	83.08	79.7	66.41	99.75	0.830
Breath deep	83.25	80.05	67.24	99.26	0.832
Breath shallow	83	79.64	66.5	99.5	0.830
Vowel a	83.2	79.94	66.91	99.5	0.832
Vowel e	83.25	79.94	66.74	99.75	0.832
Vowel o	83.16	79.94	67.07	99.25	0.831

Table 3 Classification performance using scalogram image features proposed in [13] with the updated dataset

Sound	Accuracy	F1-score	Sensitivity	Sepecificity	AUC
<i>CNN</i>					
Cough heavy	83.08	81.63	72.46	94.53	0.831
Cough shallow	80.16	80.00	68.97	95.33	0.810
Breath deep	81.85	80.32	70.42	94.53	0.818
Breath shallow	82.46	80.97	71.43	94.53	0.823
Vowel a	82.16	80.65	71.94	93.08	0.830
Vowel e	81.25	79.68	71.94	90.98	0.822
Vowel o	82.77	81.30	74.07	91.67	0.833

Table 4 Performance of magnitude AB image features

Sound	Accuracy	F1-score	Sensitivity	Sepecificity	AUC
<i>CNN</i>					
Cough heavy	84.27	81.54	69.49	99.05	0.842
Cough shallow	84.34	81.53	69.1	99.59	0.843
Breath deep	87.45	85.81	75.89	99.02	0.874
Breath shallow	87.68	86.07	76.1	99.26	0.876
Vowel a	83.51	80.58	68.42	98.61	0.835
Vowel e	85.00	82.51	70.81	99.18	0.850
Vowel o	84.82	82.27	70.46	99.18	0.842
<i>ResNet-50</i>					
Cough heavy	83.17	80.36	68.86	97.48	0.831
Cough shallow	84.14	82.43	74.39	93.9	0.841
Breath deep	87.62	86.02	76.22	99.02	0.876
Breath shallow	86.76	85.18	76.1	97.42	0.867
Vowel a	83.37	80.45	68.42	98.33	0.833
Vowel e	84.72	82.26	70.81	98.64	0.847
Vowel o	85.23	82.88	71.54	98.91	0.852

type and its accuracy and F1-score are 87.62% and 86.02%, respectively. The CNN outperformed ResNet50, according to our results shown in Table 4 except in the case of deep breath sound type.

The CNN extracted features faster as compared to Resnet-50 from AB images. The CNN model has significantly fewer parameters than the huge ResNet50 model. This is happening due to ResNet-50 being pre-trained on the three-channel dataset ImageNet, its input layer has been modified from one channel to three.

As indicated in Tables 2, 3 and 4, the proposed method using magnitude accumulated bispectral image features outperforms the methods in [9] and [13]. For a shallow breath sound signal, the improvement in accuracy is 4.19% and in F1-score is 4.52%. The results of the scalogram method are given in Table 3. The proposed method provides an accuracy improvement of 5.22% and the F1-score improvement is 5.1%. Similarly, the proposed method also gives improved results for other respiratory sounds, as shown in Tables 2 and 3. These results suggest that the accumulated bispectrum is a valuable tool in the analysis of the respiratory sound signals.

As discussed in the introduction, most of the authors have used MFCC, ZCR, spectral features and scalogram-based image features. These features are based on second-order statistics like power spectrum, mel-spectrum, etc. The proposed method based on the accumulated bispectral features gives improved performance. This could be due to the fact that the HOS has the ability to detect nonlinear properties of the signals. Additionally, HOS are better at handling the noise in the signals. The performance of the proposed method is specifically better for the cases of coughing and breathing sounds. The accuracy, F1-score, sensitivity and specificity values are improved. The state-of-the-art methods are implemented and tested for comparison purposes on the same dataset.

4 Conclusion

A novel accumulated bispectrum magnitude image features are proposed in this paper for detecting COVID-19 sound signals. The proposed method captures the COVID-19-specific characteristics of the respiratory sound signals. The accumulated bispectrum image-based features perform better and achieve the highest accuracy using a CNN classifier compared to the state-of-the-art methods. The results for shallow breath sound outperform the remaining respiratory sound types and achieves 87.68% classification accuracy, and the F1 score of 86.07%.

In this work, we have explored the use of the magnitude of the accumulated bispectrum. In our future work, we will

explore use of accumulated bispectrum with various slices and phase information for biomedical applications.

Author Contributions All authors have made contributions to conception and design, analysis and interpretation of data and they have been involved in drafting the manuscript. All authors read and approved the final manuscript.

Funding There is no funding for the submitted work from any funding institutions.

Data Availability The datasets used and analyzed in this study are described in COSWARA—A Database of Breathing, Cough, and Voice Sounds for COVID-19 Diagnosis and available at <https://doi.org/10.48550/arXiv.2005.10548>.

Declarations

Ethical approval Not applicable.

Conflict of interest The authors declare that they have no competing interests.

References

1. WHO Coronavirus (COVID-19) Dashboard. <https://covid19.who.int/>. Accessed 14 Jan 2023
2. Sangle, S., Gaikwad, C.: Covid-19 detection using spectral and statistical features of cough and breath sounds. In: 2021 International Conference on Decision Aid Sciences and Application (DASA) (2021). <https://doi.org/10.1109/dasa53625.2021.9682308>
3. Deshpande, G., Schuller, B. W.: COVID-19 biomarkers in speech: on source and filter components. In: 2021 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (2021). <https://doi.org/10.1109/embc46164.2021.9629831>
4. Al-khassaweneh, M., Bani Abdelrahman, R.: A signal processing approach for the diagnosis of asthma from cough sounds. *J. Med. Eng. Technol.* **37**(3), 165–171 (2013). <https://doi.org/10.3109/03091902.2012.758322>
5. Imran, A.: 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>
6. Feng, K., He, F., Steinmann, J., Demirkiran, I.: Deep learning Based Approach to Identify Covid-19. *SoutheastCon 2021* (2021). <https://doi.org/10.1109/southeastcon45413.2021.9401826>
7. Erdoğan, Y.E., Narin, A.: COVID-19 detection with traditional and deep features on cough acoustic signals. *Comput. Biol. Med.* **136**, 104765 (2021). <https://doi.org/10.1016/j.combiomed.2021.104765>
8. Pahar, M., Klopper, M., Warren, R., Niesler, T.: COVID-19 cough classification using machine learning and global smartphone recordings. *Comput. Biol. Med.* **135**, 104572 (2021). <https://doi.org/10.1016/j.combiomed.2021.104572>
9. Pahar, M., Klopper, M., Warren, R., Niesler, T.: COVID-19 detection in cough, breath and speech using deep transfer learning and bottleneck features. *Comput. Biol. Med.* **141**, 105153 (2022). <https://doi.org/10.1016/j.combiomed.2021.105153>
10. Sobahi, N., Atila, O., Deniz, E., Sengur, A., Acharya, U.R.: Explainable COVID-19 detection using fractal dimension and vision transformer with Grad-CAM on cough sounds. *Biocybern. Biomed. Eng.* **42**(3), 1066–1080 (2022). <https://doi.org/10.1016/j.bbe.2022.08.005>

11. Melek Manshouri, N.: Identifying COVID-19 by using spectral analysis of cough recordings: a distinctive classification study. *Cogn. Neurodyn.* **16**, 239–253 (2022). <https://doi.org/10.1007/s11571-021-09695-w>
12. Rao, S., Narayanaswamy, V., Esposito, M., Thiagarajan, J.J., Spanias, A.: COVID-19 detection using cough sound analysis and deep learning algorithms. *Intell. Decis. Technol.* **15**(4), 655–665 (2021)
13. Loey, M., Mirjalili, S.: COVID-19 cough sound symptoms classification from scalogram image representation using deep learning models. *Comput. Biol. Med.* **139**, 105020 (2021). <https://doi.org/10.1016/j.combiomed.2021.105020>
14. Kim, S., Baek, J.Y., Lee, S.-P.: COVID-19 detection model with acoustic features from cough sound and its application. *Appl. Sci.* **13**, 2378 (2023). <https://doi.org/10.3390/app13042378>
15. Wall, C., Zhang, L., Yu, Y., Kumar, A., Gao, R.: A deep ensemble neural network with attention mechanisms for lung abnormality classification using audio inputs. *Sensors* **22**(15), 5566 (2022). <https://doi.org/10.3390/s22155566>
16. Ulukaya, S., Sarica, A.A., Erdem, O., et al.: MSCCov19Net: multi-branch deep learning model for COVID-19 detection from cough sounds. *Med. Biol. Eng. Comput.* (2023). <https://doi.org/10.1007/s11517-023-02803-4>
17. Verde, L., De Pietro, G., Ghoneim, A., Alrashoud, M., Al-Mutib, K.N., Sannino, G.: Exploring the use of artificial intelligence techniques to detect the presence of coronavirus Covid-19 through speech and voice analysis. *IEEE Access* **9**, 65750–65757 (2021). <https://doi.org/10.1109/ACCESS.2021.3075571>
18. Shahnawazuddin, S., Kumar, A., Kumar, S., Ahmad, W.: Enhancing robustness of zero resource children's speech recognition system through bispectrum based front-end acoustic features. *Digit. Signal Process.* **118**, 103226 (2021). <https://doi.org/10.1016/j.dsp.2021.103226>
19. Alshebeili, S., Cetin, A.E.: A phase reconstruction algorithm from bispectrum (seismic reflection data). *IEEE Trans. Geosci. Remote Sens.* **28**(2), 166–170 (1990). <https://doi.org/10.1109/36.46695>
20. Alshebeili, S.A., Cetin, A.E., Venetsanopoulos, A.N.: An adaptive system identification method based on bispectrum. *IEEE Trans. Circuits Syst.* **38**(8), 967–969 (1991)
21. Akgül, T., Sun, M., Scلابassi, R.J., Cetin, A.E.: Characterization of sleep spindles using higher order statistics and spectra. *IEEE Trans. Biomed. Eng.* **47**(8), 997–1009 (2000). <https://doi.org/10.1109/10.855926>. (PMID: 10943047)
22. Gaikwad, C.J., Sircar, P.: Bispectrum-based technique to remove cross terms in quadratic systems and Wigner Ville distribution. *SIViP* **12**(4), 703–710 (2017). <https://doi.org/10.1007/s11760-017-1211-7>
23. Sharma, N.: Coswara A database of breathing, cough, and voice sounds for COVID-19 diagnosis. *Interspeech 2020* (2020). <https://doi.org/10.21437/interspeech.2020-2768>
24. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002). <https://doi.org/10.1613/jair.953>
25. Mendel, J.: Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications. *Proc. IEEE* **79**(3), 278–305 (1991). <https://doi.org/10.1109/5.75086>
26. Chandran, V., Elgar, S.: A general procedure for the derivation of principal domains of higher-order spectra. *IEEE Trans. Signal Process.* **42**(1), 229–233 (1994). <https://doi.org/10.1109/78.258147>

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.