




Article

Detection of COVID-19 from Deep Breathing Sounds Using Sound Spectrum with Image Augmentation and Deep Learning Techniques

Olusola O. Abayomi-Alli ¹, Robertas Damaševičius ^{1,*}, Aaqif Afzaal Abbasi ² and Rytis Maskeliūnas ³¹ Department of Software Engineering, Kaunas University of Technology, 51368 Kaunas, Lithuania² Department of Software Engineering, Foundation University, Islamabad 46000, Pakistan³ Department of Applied Informatics, Vytautas Magnus University, 44404 Kaunas, Lithuania

* Correspondence: robertas.damasevicius@ktu.lt

Abstract: The COVID-19 pandemic is one of the most disruptive outbreaks of the 21st century considering its impacts on our freedoms and social lifestyle. Several methods have been used to monitor and diagnose this virus, which includes the use of RT-PCR test and chest CT/CXR scans. Recent studies have employed various crowdsourced sound data types such as coughing, breathing, sneezing, etc., for the detection of COVID-19. However, the application of artificial intelligence methods and machine learning algorithms on these sound datasets still suffer some limitations such as the poor performance of the test results due to increase of misclassified data, limited datasets resulting in the overfitting of deep learning methods, the high computational cost of some augmentation models, and varying quality feature-extracted images resulting in poor reliability. We propose a simple yet effective deep learning model, called DeepShufNet, for COVID-19 detection. A data augmentation method based on the color transformation and noise addition was used for generating synthetic image datasets from sound data. The efficiencies of the synthetic dataset were evaluated using two feature extraction approaches, namely Mel spectrogram and GFCC. The performance of the proposed DeepShufNet model was evaluated using a deep breathing COSWARA dataset, which shows improved performance with a lower misclassification rate of the minority class. The proposed model achieved an accuracy, precision, recall, specificity, and f-score of 90.1%, 77.1%, 62.7%, 95.98%, and 69.1%, respectively, for positive COVID-19 detection using the Mel COCOA-2 augmented training datasets. The proposed model showed an improved performance compared to some of the state-of-the-art-methods.

Keywords: sound classification; audio processing; small data; data augmentation; transfer learning; deep learning; COVID-19 recognition



Citation: Abayomi-Alli, O.O.; Damaševičius, R.; Abbasi, A.A.; Maskeliūnas, R. Detection of COVID-19 from Deep Breathing Sounds Using Sound Spectrum with Image Augmentation and Deep Learning Techniques. *Electronics* **2022**, *11*, 2520. <https://doi.org/10.3390/electronics11162520>

Academic Editor: Rui Pedro Lopes

Received: 6 July 2022

Accepted: 8 August 2022

Published: 11 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The coronavirus (COVID-19) pandemic can be described as a respiratory infection majorly caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and has infected more than 44 million individuals globally [1]. The effect of this 21st-century pandemic has negatively affected global economic activities such as finance [2], security [3], food security, education, and global peace [4], with some positive results in reducing urban pollution [5]. The influence of this virus from the alpha to the beta variant has affected both the health and the welfare status of citizens around the world [6]. The World Health Organization (WHO) declared it to be a novel coronavirus disease and named it as a Public Health Emergency of International Concern (PHEIC) on 30 January 2020 due to the easy spread and high transmission rate and communicability of this disease [7].

Previous studies have shown that some of the clinical signs of patients infected with COVID-19 are closely related to other viral upper respiratory diseases such as a respiratory syncytial virus (RSV), influenza, and bacterial pneumonia, while other common symptoms

are sore throat, pleurisy, shortness of breath, dry cough, fever, headache, etc. [8]. Different tools and methods have been used for monitoring and diagnosing this virus, such as Real-Time Polymerase Chain Reaction (RT-PCR) [9], medical imaging such as computer tomography (CT) scan images [10,11], chest X-ray [12,13], and lung ultrasonography [14], as well as blood samples [15], urine [16], feces [17], etc. However, some of the limitations of previous studies include inaccuracies of results, cost implications, varying quality and reliability of available SARS-CoV-2 nucleic acid detection kits, and the insufficient number and throughput of laboratories performing the RT-PCR test, etc. [18]. Similarly, the use of medical images for diagnosis has its share of limitations, such as the cost implications of setup, and insufficient machines in hospitals for conducting timely COVID-19 screening [19]. These medical images are processed using various machine learning, deep learning [20], and other artificial intelligence methods [21], making them more effective.

Recently, the use of respiratory sound or human audio samples such as coughing, breathing, counting, and vowel sounds for the detection of COVID-19 [22–24], are being presented as alternative, simple and inexpensive methods for monitoring the disease. Sound or audio classification tasks have continued to increase thanks to their wide span of applications in our everyday lives, including medical diagnostics for cognitive decline [25] and laryngeal cancer [26]. The concept of sound or audio recognition involves recognizing the audio stream, related to various environmental sounds. Thus, the advancement of deep convolutional neural network (CNN) applications in sound classification have shown very impressive performances. This is based on the strong capabilities of deep CNN architectures in identifying key features that are mapping audio spectrograms to relative or different sound labels such as the time and frequency energy modulation patterns over spectrogram data inputs [27]. The need for a deep CNN model in sound classification is due to some of the challenges posed by conventional machine learning methods, which include the inability to effectively identify features in Spectro-temporal patterns for different sounds [28]. The recent adoption of a deep network is basically due to its stronger representational ability, thereby achieving better classification performance [29]. However, the real-life applications of deep neural networks suffer from overfitting, which is always a result of limited datasets (data starvation), class imbalance, and the challenges of proper annotations in many practical scenarios due to the cost and time complexity of carrying out such annotations [30]. In addition to these challenges, there are also some shortcomings in traditional audio features techniques, such as Mel Frequency Cepstral Coefficients (MFCC), which is the problem of identifying important features within different audio/sounds for efficient classification. Therefore, alternative methods such as cochleagrams are sought for audio feature extraction [31].

CNNs are effective at learning from images. Deep CNNs are particularly well suited to the problem of sound classification for two reasons: first, when used with spectrogram-like image inputs, they can capture energy modulation patterns across time and frequency, which has been shown to be a key characteristic for differentiating between different sounds [32]. Deep CNNs are particularly suited for sound classification because they can learn discriminative spectro-temporal patterns [27]. The human body is too complex for performing effective classification, making it difficult to spot data's underlying patterns. The introduction of image-based sound classification allows for the efficient recording of a variety of sound patterns, including those coming from the heart and lungs [33,34]. However, in many situations, data augmentation is required to accomplish generalization [35].

Data augmentation has consistently shown its relevance in improving data generalization based on the application of one or more deformations properties in a set of labelled training samples, thus generating additional training data samples. Some of the most effective data augmentation methods proposed in existing studies for audio/sound dataset include the following: semantics-preserving deformations in music datasets, random time-shifting [36], pitch shifting and time stretching, etc. Some of the traditional data augmentation techniques have proven to be insufficient in other sound datasets with very high time complexity for training, and to have an insignificant impact on the performance

of some state-of-the-art models [27]. Wang et al. [30] applied GAN-based semi-supervised learning using a low-density sample annealing scheme for generating a new fake audio spectrogram with labelled IFER data. Other studies also adopted image augmentation techniques for increasing spectrogram images. Mushtaq et al. [37] applied some of the most widely used image augmentation techniques on the converted audio files to spectrogram images. The authors also applied five of the most popular deformation approach to the audio files which include the pitch shift, time stretch, trim silence, etc. The study concluded that their proposed data augmentation method improved the performance of the DCNN model more than the traditional image augmentation methods with increasing accuracy for training, validation and test datasets. Based on some of the findings deduced from recent studies, we can agree that the combination of appropriate feature extraction methods with deep learning models using suitable data augmentation technique(s) can aid the performance of classifiers in sound classification. Therefore, this paper introduces effective and improved data augmentation schemes on deep learning models for sound record classification in COVID-19 detection.

In summary, the main contributions of our study are as follows: Firstly, applied simple and effective data augmentation schemes for efficient data generalizations for COVID-19 detection. Secondly, a pre-trained CNN architecture called DeepShufNet was analyzed and evaluated. The experimental analysis of the augmented datasets in comparison with baseline results showed significant improvement in performance metrics, better data generalization and enhanced optimal test results. In addition, we compared and investigated the impact of data augmentation on two methods (Mel-spectrograms and GFCC) for the detection of COVID-19 symptomatic cases, positive asymptomatic cases, and fully recovered cases. The results showed an impressive result with near-optimal performance, especially in the rate of recall, precision, and F1-Score. The remaining part of this paper is sectioned as follows.

The related work is presented in Section 2, where we discuss in detail all significant approaches used for data augmentation, and learning classifiers concerning audio/sound classification. In Section 3, an introduction to our proposed methodology is fully discussed with emphasis on the dataset used, as well as our proposed data augmentation and deep learning methods. Detailed results from and discussions on the comparison of the proposed method with others' published results are presented in Section 4. In Section 5, conclusive remarks are given.

2. Related Work

This section discusses in detail some of the state-of-the-art methods used by previous researchers for data augmentation techniques and classification models in sound/audio classification. Research trends in COVID-19 detection include the use of conventional machine learning algorithms in sound datasets, which include but are not limited to coughing, deep breathing, sneezing, etc. Machine learning algorithms have been applied in the detection of COVID-19 with improved results, such as a study by Sharma et al. [22], who analyzed audio texture for COVID-19 detection using datasets with different sound samples and a weighted KNN classifier. Tena et al. [38] conducted COVID-19 detection using five classifiers, namely: Random Forest, SVM, LDA, LR, and Naïve Bayes algorithms. RF classifier outperformed other machine learning methods with significant improvement in the accuracy on five datasets; however, the shortfall is lower specificity rates. Chowdhury et al. [39] presented an ensemble method using the multi-criteria decision making (MCDM) method, and the best performance was obtained with extra tree classifier.

The authors of [40] applied Gaussian noise augmentation techniques and AUCOREsNet for the detection of COVID-19. Loey and Mirjalili [41] compared six deep learning architectures such as GoogleNet, ResNet 18, 50 and 101, MobileNet and NasNetmobile for detection of COVID-19 using the Coughdataset. The study shows that ResNet-18 outperforms the other models with a significant performance result. Pahar et al. [42] presented three pre-trained deep neural networks CNN, an LSTM and a Resnet50 architecture for

detection of COVID-19 using five datasets. Erdogan and Narin [43] applied deep feature ResNet 50 and MobileNet architecture on support vector machine in the detection of COVID-19 and the feature extraction method used two conventional approaches, which are empirical mode decomposition (EMD) and discrete wavelet transform (DWT). The study shows a high-performance result with ResNet50 deep features. Sait et al. [44] proposed a transfer learning model called CovScanNet for classification of COVID-19 using multimodal datasets. Soltanian and Borna [45] investigated the impact of the lightweight deep learning model on classification of Covid from non-Covid cough Virufy datasets. The authors combined separable kernels in deep neural networks for COVID-19 detection.

Despotovic et al. [46] applied a CNN model based on VGGish in a Cough and Voice Analysis (CDCVA dataset) and the study gave an improved performance of 88.52% accuracy, while Mohammed et al. [47] presented shallow machine learning, Convolutional Neural Network (CNN), and pre-trained CNN models on Virufy and Coswara datasets with performance metrics showing 77% accuracy. Brown et al. [48] presented ML algorithms such as Logistic Regression (LR), Gradient Boosting Trees, and Support Vector Machines in the detection of COVID-19.

Some of the data augmentation techniques presented by previous researchers include studies by Lella and Pja [49], which applied traditional audio augmentation methods on a one-dimensional CNN for diagnosing respiratory diseases of COVID-19 using human-generated sounds such as voice/speech, cough and breath datasets. Salamon and Bello [27] examined the impact of different data augmentation methods on the CNN model. Authors concluded that there is a need for class-conditional data augmentation for improved performance of deep learning models. Leng et al. [29] proposed a Latent Dirichlet Allocation (LDA) approach for augmentation of audio events from audio recordings. The authors compared the performance of the proposed LDA algorithm to other data augmentation techniques such as time and pitch shifting and Gaussian noise. Based on this thorough literature review, we can agree that to a great extent, existing data augmentation and classification methods in COVID-19 using sound/audio datasets still suffer from setbacks in identifying an appropriate and lightweight data augmentation method to overcome the problem of limited training data and data imbalance. The issue of background noise on sound datasets affects effective feature extraction; therefore, creating synthetic datasets from such noisy datasets would also affect the efficiency of the classification of deep learning models. There is a need to collect more quality data and thereby improve the performance of the learning models [38,50]. Therefore, this study proposed a simple and efficient deep learning architecture referred to as DeepShufNet model for improved classification of COVID-19. In addition, we applied effective data augmentation techniques using noise and color transformation methods in generating better synthetic datasets, thus improving data generalization and COVID-19 detection.

3. Methodology

3.1. Dataset

This experimental study was conducted using the publicly available Coswara dataset generated by Sharma et al. [51] which consists of nine different audio/sound samples collected from 2130 recordings. The different audio/sound samples include the following: breathing (two types: deep and shallow), cough (two types: heavy or shallow), digit counting (two types: fast and normal), and finally vowel phonation (three types: a, e, and o), respectively. The audio recordings from the Coswara dataset consist of seven categories which are as follows: healthy (1372), positive_moderate (72), positive_mild (231), positive_asymp (42), recovered_full (99), respiratory_illness_not_identified (RINI) (150), and no_respiratory_illness_exposed (NRIE) (164). The summary of each category of audio samples for the entire Coswara dataset is described in Figure 1 and Table 1 summarizes the selected classes used for this study. However, in this study, our experiment is majorly focused on deep breathing audio samples (coined as COCOA-DB). The architectural framework for our proposed model is presented in Figure 2.

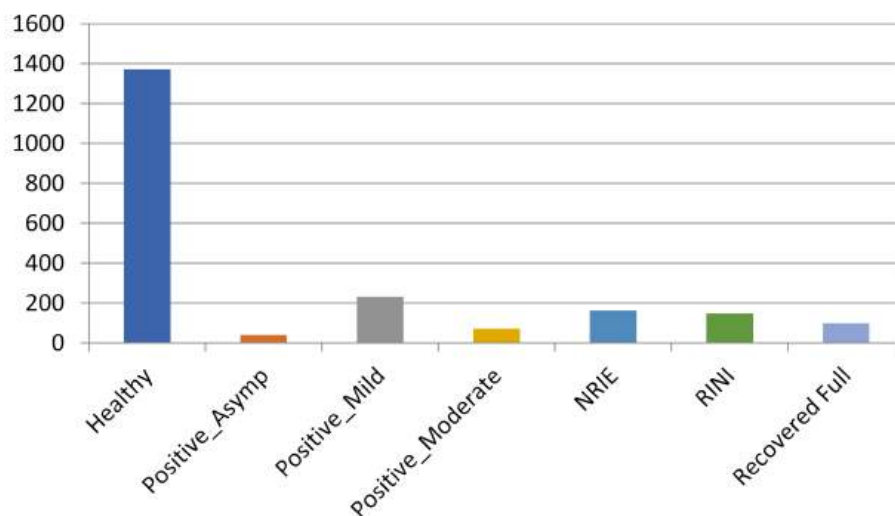


Figure 1. Summary of classes in Coswara Dataset.

Table 1. Summary of the selected audio recordings for this study.

Classes	Size	# of Audio Samples
Positive Asymptotic	48 kHz	42
Positive Mild	48 kHz	231
Positive Moderate	48 kHz	72
Healthy	48 kHz	1372

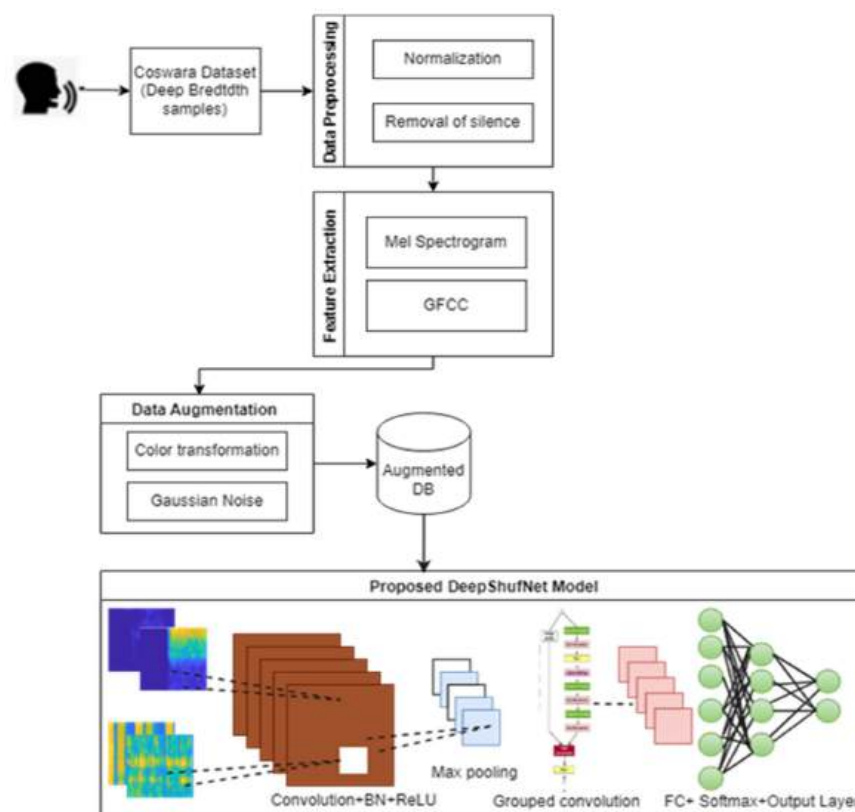


Figure 2. Proposed Architectural framework.

For this study, we merged some classes, as we will see in the next section. The reason for this merger is due to the similarity in the names and the audio spectrum; therefore, later in this study, classes such as positive mild and positive moderate were merged and

represented as a COVID-19 positive class. The information of our proposed architecture model is stated in the remaining subsection and the specifics or key blocks in the architecture are discussed in detail.

3.2. Data Pre-Processing

For each audio recording file within the Coswara dataset, there is an uneven time of audio signal, and to determine the different time duration for each file, we used the expression $L = (N(Y) ./ f(s))$, where $N(Y)$ is the sample length and $f(s)$ is the sampling frequency for each audio sample, which is 48 kHz. Based on the mathematical expression to calculate L s, the minimum and maximum length of audio files are 4 and 29 s, respectively. To ensure that all relevant features are captured during the analysis, we applied a simple pre-processing and normalization method by scaling speech by its peak value with an amplitude maximum value of 1 [52]. Secondly, we applied a silent region deletion method which allows the elimination of the silent part of the signal speech and uses only the voiced portion. Progress in previous studies using silent region elimination has shown its complete usefulness in improving the performance of the system and reducing processing time.

3.3. Feature Extraction

This study considered three categories of audio signal features, which are based on Mel-spectrograms, and the Gammatone Frequency Cepstral Coefficient (GFCC) image described below as well as a sample of generated images for each class using the two feature extraction methods is depicted in Figures 3 and 4.

3.3.1. Mel Spectrogram

One of the most widely used time–frequency spectra in sound classification is the Mel spectrogram [53]. This input representation has continuously shown its effectiveness and importance when compared to other structures such as short-time Fourier transform (STFT). Based on this knowledge, we transformed all our selected Coswara audio recordings into spectrograms using the default Mel spectrograms function in the MATLAB toolbox. The Mel spectrogram images were created with an FFT window, frequency range up to 2.0×10^4 Hz, and an average length of audio files varies ranging from 10 s to 25 s. The samples of the Mel spectrogram created are depicted in Figure 3, showing the time–frequency spectrogram for each class category in the Coswara dataset. We can also agree that the power spectral energy density $P(f, t)$ for each audio file differs with increasing power for healthy samples in comparison with other class samples to the number of points around the spaced times t and frequencies f .

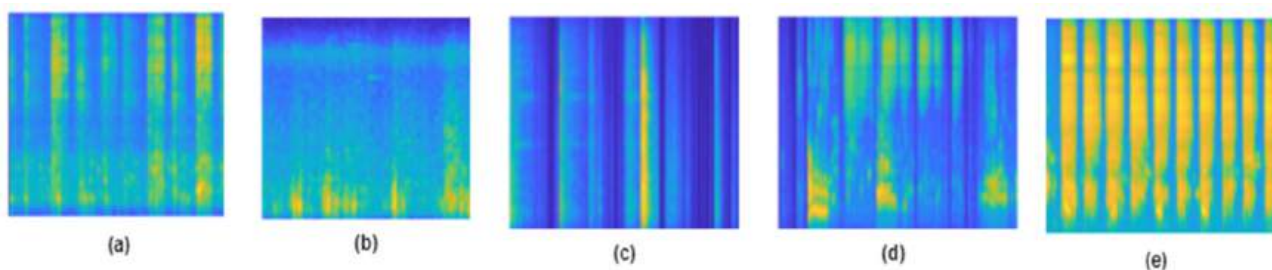


Figure 3. Samples of Mel Spectrogram images generated from the audio files in each class (a) Positive Asymptotic; (b) Positive mild; (c) Positive Moderate; (d) Recovered full; (e) healthy.

3.3.2. Gammatone Frequency Cepstral Coefficients (GFCC)

Gammatone Frequency Cepstral Coefficients (GFCC) was developed by Patterson et al. using Gammatone filter banks which model the auditory system of humans as an overlapping band-pass filter [54]. In the GFCC feature extraction process, the speech signal is expanded to the Gammatone filterbanks in the frequency domain. The output of the Gammatone filterbanks is used in achieving the cochleagram, which is a representation of

a frequency–time signal. Therefore, the impulse response for each gammatone filter can be expressed mathematically in Equation (1).

$$g(t) = mt^{y-1}e^{-2\pi nt} \cos(2\pi f_c t + \varnothing), \quad (1)$$

where: m is constant (mostly equal to 1), controlling the gain and the order of the filters and is defined by the value of y , which is mostly set to a value less than 4; the bandwidth is represented as n and can be expressed in Equation (2); \varnothing is the phase but is generally set to zero. The samples of the GFCC created are depicted in Figure 4, showing the time–frequency spectrum for each class category in the Coswara dataset.

$$n = 25.17 \left(\frac{4.37f_c}{1000} + 1 \right), \quad (2)$$

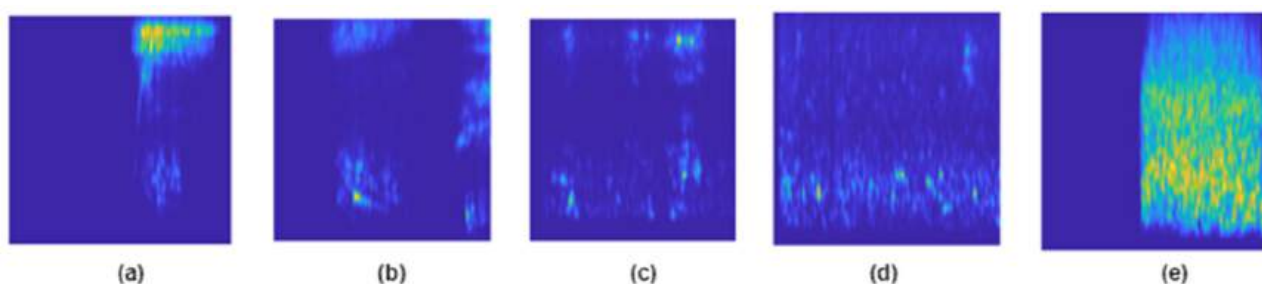


Figure 4. Samples of GFCC images generated from the audio files in each class (a) Positive Asymptomatic; (b) Positive mild; (c) Positive Moderate; (d) Recovered full (e) healthy.

3.4. Data Augmentation Scheme

After the feature extraction steps, there is a problem with the data distribution of each class, as shown in Figure 1, with a huge factor of class imbalance among the seven different classes. In the worst case, the number of samples of the majority class is an average of 10 times more than the minority classes. This factor plays a crucial role in the difficulty of the classification task and thus influences the performance of our model. Therefore, the application of data augmentation will not only provide more training data samples, or reduce the overfitting of models during the training, but it will also improve the accuracy and overall performance of the models [37]. Basically, for this study, we applied the two categories of data augmentation methods to increase the training images of the minority class as depicted in Figure 1, and as a result, we achieved a newly synthetic dataset referred to as COCOA (Table 2), COswara-COvid-Augmented datasets, which are as follows:

- **Color transformation method:** In this category of data transformation, there are three popularly used color models in the literature; however, in this study, we adopted rgb2lab and grayscale transformation methods. These transformer methods are also referred to as monochrome simply because they are made of 256 shades of grey and have a brightness value between 0 (black) to 1 (white). In this research, several types of color transformation techniques were applied, namely, brightness, contrast, rgb2gray, and rgb2lab. Horizontal flip, zoom, and shear transforms were applied to each image in the dataset to generate a new dataset called COCOA-1.
- **Noise Addition:** We applied Gaussian noise and salt–pepper noise with different parameters to each image in the dataset to generate a new synthetic dataset, called COCOA-2.

In addition to these two categories of data augmentation techniques, we also applied some of the traditional data augmentation methods such as horizontal flip, vertical flip, and random reflection to each image in the datasets. Table 3 shows the summary of the total data samples used in this study with the number of augmented samples per class. The number of synthetic data generated by each of the transformation methods using the training datasets are 1098, 760, and 760 synthetic samples for All positive COVID-19, positive asymptomatic, and recovered full classes, respectively.

Table 2. Summary of Augmented dataset and reference names of image augmentations used.

Augmented Dataset	Description
COCOA-1	Color transformation: Grayscale and Rgb2lab
COCOA-2	Noise: Gaussian, salt and pepper
COCOA-3	Combo (combination of color transformation, and Noise)

Table 3. Summary of sample sizes used for each classes.

Category	Total Samples	Train/Validation (80%)	Augmented Data	Test (20%)
Healthy	1372	1098	-	274
All positive COVID-19	303	243	1098	60
Positive Asymp	42	34	760	8
Recovered Full	99	79	760	20

3.5. Structure of The Proposed DeepShufNet Model

This study proposed the DeepShufNet model, which is a lightweight deep CNN model as shown in Figure 5. Our choice of proposed pretrained ShuffleNet architecture in this study is based on the concept of pointwise Group Convolution, which has been described in recent studies as a light-weighted network that assigns models over two GPUs and uses repeated building blocks and channel shuffle. In addition, the use of pointwise group convolution and channel shuffling has helped in minimizing computational cost and still improving overall accuracy. The network has been initially pre-trained using ImageNet. As audio recording is a one-dimensional time series, we train a one-dimensional convolutional neural network for binary classification.

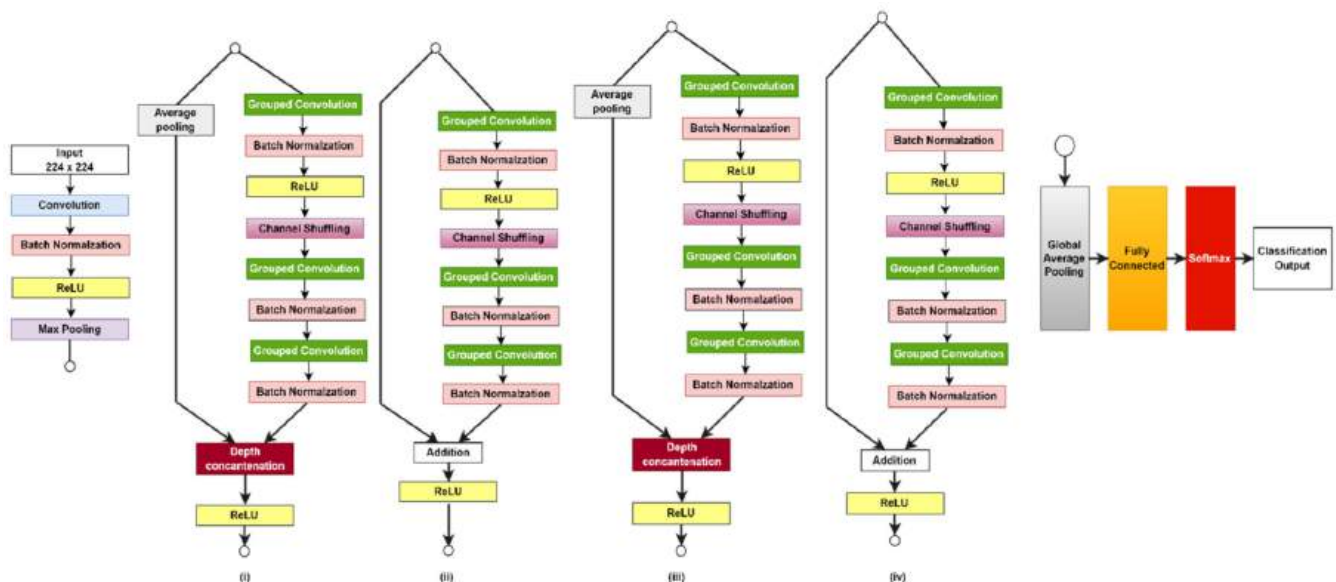


Figure 5. The proposed deepShuffleNet architecture consists of several building blocks.

The DeepShufNet consists of an input layer of a $224 \times 224 \times 3$ image, and multiple hidden layers which include a convolutional layer, batch normalization layer, pooling layer, flatten layer, fully connected layer, and an output layer. However, the original size of each image is 875×656 pixels, but this paper applied an imresize to resize all images to the size 224×224 , which is enough to identify all target ranges. In addition, based on the literature, the use of smaller input aids in improving computational speed, reduction of the

number of parameters, and finally minimizing the possibilities of overfitting. The proposed DeepShuffleNet used in our experiment has a total number of 172 layers and a total number of 1.4 million learnable parameters. We applied 50% dropout layers to the hidden neurons, which helps to prevent overfitting. Despite the larger numbers of layers, the DeepShuffleNet architecture utilizes some interesting operations such as grouped convolution, channel shuffle, and depth concatenation which significantly minimize computational complexity and improve accuracy.

In this study, we utilize the training options with Adam (adaptive momentum algorithm), and a minibatch size of 250 for searching and final training. The learning rate of our optimizer is subject to a warm start ranging from $1 \times e^{-4}$ to 0.001, a total number of epochs of 50, and an L2 regularization parameter of $\lambda = 2 \times e^{-4}$. To ensure the optimal training of our model and to prevent overfitting, which is a major challenge for deep neural network models, we applied a drop-out rate of 50%. The shuffleNet architectural layers are made up of 172 layers and 1.4 million total learnable parameters, as summarized in Table 4.

Table 4. Summary of ShuffleNet architecture layers.

<i>S/n</i>	Type of Layer	# Layers
1	Input Layer	1
2	Convolution	1
3	Batch Normalization	49
4	Activation	32
5	Max/Average Pooling	6
6	Grouped Convolution	49
7	Channel shuffling	16
8	Depth concatenation	2
9	Addition	13
10	Fully Connected	1
11	Softmax	1
12	Output Layer	1
Total		172

3.6. Performance Evaluation

In this paper, we assess the performance of our proposed method using three datasets, which are the original Coswara datasets for deep breathing recordings, COCOA1 (offline data augmentation based on time shift, pitch shift, noise), and COCOA2 (using image augmentations). We investigated the performance of our classification task using some of the state-of-the-art evaluation metrics, namely Accuracy, Recall, Precision, F1-Score, and Confusion matrix. The mathematical expression and the description of the performance metrics used in this study are represented in Table 5.

Table 5. Summary of the Evaluation Metrics.

Metrics	Description	Mathematical Expression
Accuracy	Degree of true values (correctness of COVID-19) measurements against all the evaluated instances.	$Acc = \frac{Tp + Tn}{Tp + Tn + Fp + Fn}$
Sensitivity/Recall	Proportion of true positive (COVID-19) people against the actual number of people with the disease.	$Sen = \frac{Tp}{Tp + Fn}$
Precision	Proportion of true positive (COVID-19) people against the predicted number of people with the disease.	$Prec = \frac{Tp}{Tp + Fp}$

Table 5. Cont.

Metrics	Description	Mathematical Expression
Specificity	Proportion of true negative (non-COVID-19) people against the actual number of people without the disease.	$Spec = \frac{Tn}{Tn + Fp}$
F1-Score	The weighted average of precision and recall.	$F1 - Score = 2 * \frac{Sen * Spec}{Sen + Spec}$

4. Experimental Results and Discussion

This section is based on an extensive experiment and effective investigation of all the different datasets on the proposed DeepShufNet. All experiments were conducted in MATLAB R2020b on a desktop PC built with an Intel(R) core i5 (3.2 GHz) processor, 8 GB of RAM, and an NVIDIA GeForce GTX 1070 GPU server with 120 G memory.

Taking into consideration the condition of the hardware and the issue of out-of-memory errors, we reduced the batch size to 200 for both training and testing. Considering the huge data sparsity within the Coswara dataset classes, the repeated experiments were conducted five times.

4.1. Training and Testing Prediction

The proposed DeepShufNet model was trained and tested on the feature-extracted images combined from all Coswara datasets. Cross-validation method was applied to find the optimal parameter configuration and the model was trained and validated on 80% of the total images extracted from the sound dataset, which consist of 1706 data samples comprising healthy, positive asymptotic, positive mild, positive moderate, recovered full, RINI, and NRIE with 1098, 34, 185, 58, 79, 120, and 132, respectively. The adaptive momentum algorithm ADAM was used as the training algorithm, and different hyperparameter values as summarized in Table 6. The learning rate controls the rate of the weights update, therefore reducing the prediction error, while the batch size helps to determine the number of sample rows processed/time before updating the parameters of the internal network. The baseline experiment was evaluated using the raw feature-extracted images, the training process was with and without fine-tuning. The final DeepShuffleNet model was selected using the model with the least loss in the validation set during training.

Table 6. Training hyperparameter settings.

	Parameters	Values
Optimization	Optimizer name	ADAM
	Loss function	Cross-entropy
	Learning rate	1×10^{-4}
Fitting	Batch size	200
	Dropout	50%
	Epoch	50

The training model for each experiment was analyzed and observations of improvement in the classification results to validation accuracy and losses were noted. The results of the original dataset without augmentation suffer from the increasing misclassification rate of the minority class, especially in the case of classifying positive asymptotic and positive COVID-19 classes with a recall and precision rate of almost NA to less than 10%. However, training the DeepShufNet model with our categories of synthetic dataset gave a near-optimal result with a better performance in detection of COVID-19.

The experimental results are presented in four comparative categories and all results were obtained based on the experiments with the test dataset. The overall performance of the model with each category of dataset is compared using an optimal model in five recorded experiments in this research. In each comparative experiment, the combination of

accuracy, recall and specificity is the main metric to judge the performance of the model in each dataset's categories, since it examines both classes' outcomes and improvement in the classification results for the minority class. The detailed summary of all measures for each category is all stated as follows.

- All Positive vs. Healthy;
- Positive Asymptomatic vs. Healthy;
- Healthy vs. Recovered Full.

4.2. Classification Deep Breath Sound (All Positive COVID-19 vs. Healthy)

This section compares the results of the transfer learning DeepShufNet on 224×224 pixels for binary classification of healthy versus all positive classes. Due to the similarities between the positive mild and moderate classes, we combined these two classes to create a new class called the All-positive-Covid class. A comparison of the detection power of our proposed DeepShufNet on the Mel spectrogram feature images and GFCC features is shown in Table 7. The classification results reflect some improvement and stability of the DeepShuffleNet in the data augmentation datasets.

Table 7. Results Analysis of Positive COVID-19 vs. Healthy on applied feature extraction methods.

Feature Extraction	Data Augmentation	Accuracy (%)	Recall (%)	Precision (%)	Specificity (%)	F1-Score (%)
Mel Spectrogram	No Aug Color	71.2 ± 7.3	60.2 ± 12.6	51.8 ± 15.1	85.22 ± 11.3	53.47 ± 6.1
	Transformation (COCOA-1)	78.7 ± 6.1	57.9 ± 13.5	45.41 ± 9.6	83.19 ± 9.4	49.2 ± 5.8
	Noise Addition (COCOA-2)	85.1 ± 4.2	70.85 ± 7.7	59.64 ± 13.1	88.25 ± 6.14	63.61 ± 6.7
	Combo (COCOA-3)	87.8 ± 1.3	69.49 ± 4.9	64.82 ± 4.7	91.75 ± 1.9	66.9 ± 2.8
GFCC	No Aug Color	74.9 ± 3.8	48.7 ± 14.1	40.1 ± 10.16	86.99 ± 1.55	42.4 ± 6.3
	Transformation (COCOA-1)	76.4 ± 2.5	71.33 ± 2.2	41.23 ± 3.4	77.51 ± 3.3	52.17 ± 2.6
	Noise Addition (COCOA-2)	83.1 ± 0.9	46.7 ± 11.5	53.3 ± 2.32	91.06 ± 2.01	49.27 ± 6.5
	Combo (COCOA-3)	83.1 ± 1.4	38.33 ± 9.3	50.21 ± 1.6	92.21 ± 1.7	43.1 ± 6.5

On the test set, the best performance for DeepShufNet was achieved using the Mel spectrogram image in the COCOA-2 dataset (see Table 6), with an enriching positive COVID-19 detection case summarized as mean accuracy with 85.1 (standard deviation [SD], 4.23), 70.85 (SD, 7.7) for recall/sensitivity, 59.64 (SD, 13.12) for precision, 88.25 (6.14) for specificity, and 63.61 (SD, 6.7) for F1-score. However, the test set results of our proposed model on COCOA-3 show a substantial improvement in accuracy mean of 87.82 (SD, 1.3), 69.49 (SD, 4.9) for recall/sensitivity, 64.82 (SD, 4.7) for precision, 91.75 (1.9) for specificity, and 66.9 (SD, 2.8) for F1-score. Therefore, the test set comparison of the original dataset without augmentation can be said to perform the worst when compared with the outcome of the other datasets. The datasets using the GFCC images with augmentation still outperforms the original datasets with significant comparison result of accuracy as 83.1 (SD, 1.4), 83.05 (SD, 0.9), 76.4 (SD, 2.5), and 74.9 (SD, 3.8) for COCOA-3, COCOA-2, COCOA-1, and raw data (no augmentation), respectively. More interesting is the increasing mean recall for DeepShufNet being 71.33 (SD, 2.2), 48.7 (SD, 14.1), 46.7 (SD, 11.5), and 38.8 (SD, 9.3) for COCOA-1, no aug, COCOA-2, and COCOA-3, respectively.

The summary of DeepShufNet on the Mel spectrogram images is presented in Figure 6, which reflects the best experimental outcome for COCOA-2 with values for accuracy,

recall, specificity, precision, and F1-score being 90.1%, 62.71%, 95.99%, 77.1%, and 69.2%, respectively. The second-best results were achieved with COCOA-3, with an accuracy of 89.5%, 71.2% recall, 93.4% specificity, 70% precision, and 70.6% F1-score. The worst result was achieved by the raw dataset without augmentation, with an accuracy of 79%, 54.23% recall, 84.3% specificity, 42.67% precision, and 47.76% F1-score.

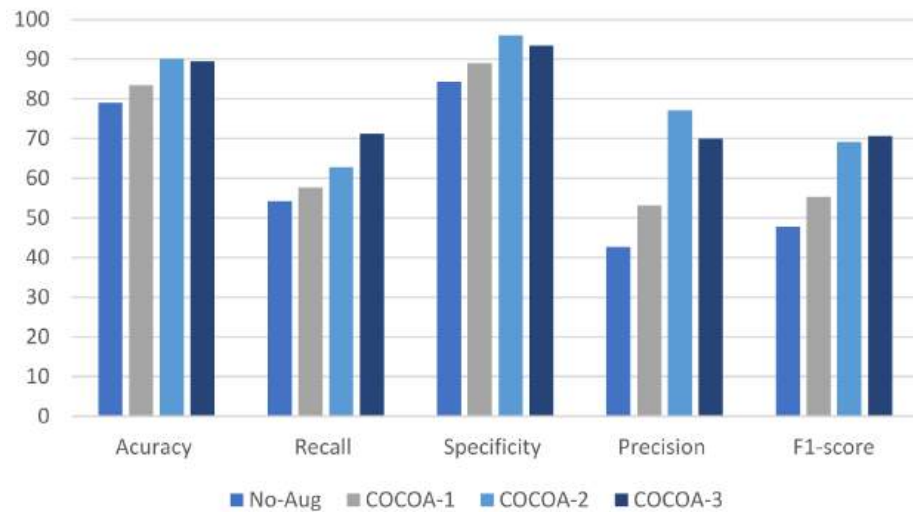


Figure 6. Comparison of performance metrics for Mel-spectrogram (Healthy vs. positive COVID-19).

In the same manner, Figure 7 shows comparison results of DeepShufNet for GFCC images. The application of noise augmentation COCOA-2 and the combo datasets (COCOA-3) show 84.1% and 84.7% accuracy, respectively. The two best recall results were achieved by COCOA-1 and COCOA-2, which depicts that the application of the data augmentation approach helps to improve classification results.

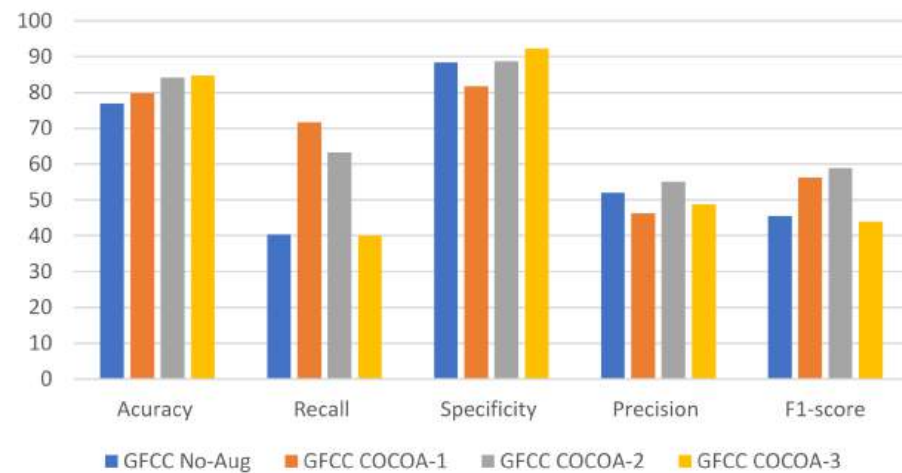


Figure 7. Comparison of performance metrics for GFCC (Healthy against positive COVID-19).

4.3. Experimental Results: Positive Asymptotic vs. Healthy

Aiming to indicate the contribution of our proposed DeepShufNet models, a second experiment was conducted to classify the healthy versus positive asymptotic alone. The wide margin in data sparsity between these two classes could result in serious overfitting of the model. However, the growth in the performance metrics for both Mel-spectrogram and GFCC images has not been continuous for the raw dataset, but the application of data augmentation approach on training data has reduced overfitting with a training accuracy much lower than the accuracy of testing in the last epoch. In summary, the experimental results indicate that the training with augmented datasets has not had a

significant influence on the improvement of classification accuracy; however, training the model with COCOA-1 showed a good classification performance on the test sets in terms of accuracy, but the second worst results for recall rate. On the other hand, training our DeepShufNet with COCOA-2 slightly increases the test classification accuracy, specificity, and F1-score. Considering the efficiency of the data augmentation methods, classification using noise augmentation is more suitable for practical application when the dataset is small, as reflected in Table 7. Figures 8 and 9 show an improvement in the augmentation of Mel-spectrogram images with higher performance results in recall rate, precision, and F1-score. Therefore, we can claim that the impact of data augmentation methods in both feature extraction images achieved a more remarkable improvement in classification results on the proposed DeepShufNet model.

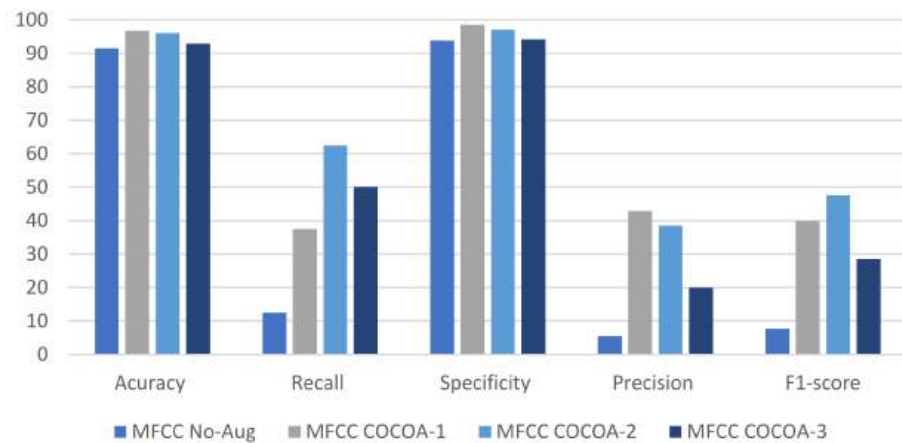


Figure 8. Comparison of metrics for classification of Positive Asymptotic vs. Healthy (Mel-spectrogram images).

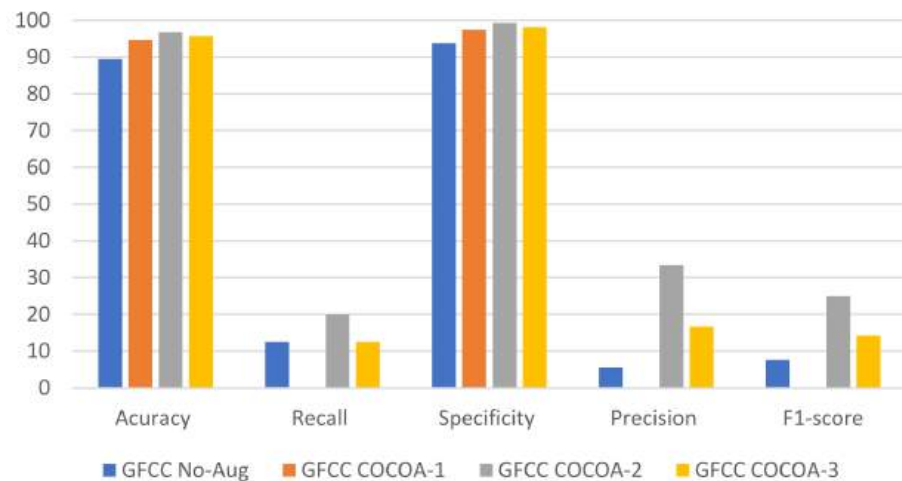


Figure 9. Comparison of metrics for classification of Positive Asymptotic vs. Healthy (GFCC images).

The experimental results from Table 8 show an improvement using the data augmentation method as compared to the baseline experiment with the best accuracy being achieved by COCOA-1 with an accuracy of 97.15% (SD, 0.5); 95.8% (SD, 1.1) for COCOA-2; 92.7% (SD, 0.17) for COCOA-3; and 92.2% (SD, 0.9) for no aug data.

Table 8. Results Analysis of Positive Asymptotic vs. Healthy on applied feature extraction methods.

Features Extraction	Data Augmentation	Accuracy (%)	Recall (%)	Precision (%)	Specificity (%)	F1-Score (%)
Mel Spectrogram	No Augmentation	92.2 ± 0.9	18.75 ± 8.8	9.03 ± 4.9	94.3 ± 0.8	10.73 ± 4.1
	Color Transformation (COCOA-1)	97.15 ± 0.5	25 ± 17.7	71.42 ± 20.4	99.27 ± 1.0	31.11 ± 12.6
	Noise Addition (COCOA-2)	95.8 ± 1.1	37.5 ± 15.3	30.43 ± 8.3	97.37 ± 1.1	32.45 ± 8.9
	Combo (COCOA-3)	92.7 ± 0.17	41.67 ± 7.2	17.19 ± 2.43	94.16 ± 0.0	24.34 ± 3.7
GFCC	No Augmentation	88.54 ± 0.9	12.5 ± 3.3	12.5 ± 1.2	93.79 ± 0.77	12.18 ± 6.35
	Color Transformation (COCOA-1)	91.15 ± 5.4	NA	NA	93.79 ± 5.4	NA
	Noise Addition (COCOA-2)	92.86 ± 2.1	19.17 ± 6.3	17.84 ± 13.63	96.59 ± 2.35	17.06 ± 7.7
	Combo (COCOA-3)	89.0 ± 5.93	25.0 ± 12.5	9.96 ± 5.8	90.88 ± 6.5	12.01 ± 2.3

4.4. Experimental Results: Healthy vs. Recovered-Full

In this experiment, we tried to validate the effectiveness of our proposed model by analyzing the detection rate of the DeepShufNet model in classifying healthy against recovered. This experimental results of the applied model on the four datasets based on MFCC feature-extracted images, namely raw data (no aug), COCOA-1, COCOA-2, and COCOA-3, which gave the following performance results for accuracy: 93.45 (SD, 0.41) for COCOA-2; 93.33 (SD, 0.51) for COCOA-1; 91.68 (SD, 4.0) for COCOA-3; and 91.03 (SD, 0.8) for no augmentation (see Table 9). Figures 10 and 11 show the best results of all the four datasets on the DeepShufNet model, and it reflects that the combination of the two data augmentation techniques (COCOA-3) gave the best results.

Table 9. Results Analysis of Positive COVID-19 vs. Healthy on applied feature extraction methods.

Feature Extraction	Data Augmentation	Accuracy (%)	Recall (%)	Precision (%)	Specificity (%)	F1-Score (%)
Mel Spectrogram	No Augmentation	91.03 ± 0.8	96.96 ± 2.7	94.56 ± 0.6	26.32 ± 0.0	95.72 ± 0.9
	Color Transformation (COCOA-1)	93.33 ± 0.51	98.78 ± 1.1	94.31 ± 0.4	18.33 ± 7.6	96.49 ± 0.3
	Noise Addition (COCOA-2)	93.45 ± 0.41	96.42 ± 1.2	96.5 ± 1.0	52.5 ± 14.4	96.45 ± 0.2
	Combo (COCOA-3)	91.68 ± 4.0	92.89 ± 5.6	97.14 ± 1.4	61.25 ± 20.6	95.39 ± 2.4
GFCC	No Augmentation	90.5 ± 1.9	95.86 ± 2.8	93.16 ± 0.6	18.33 ± 11.54	94.98 ± 1.1
	Color Transformation (COCOA-1)	77.96 ± 12.9	81.47 ± 14.66	93.90 ± 0.5	27.0 ± 16.05	86.66 ± 8.5
	Noise Addition (COCOA-2)	91.15 ± 2.8	97.54 ± 3.1	93.28 ± 0.22	5.0 ± 0.0	95.34 ± 1.6
	Combo (COCOA-3)	91.04 ± 4.0	97.08 ± 4.3	93.53 ± 0.32	8.0 ± 4.5	95.23 ± 2.2

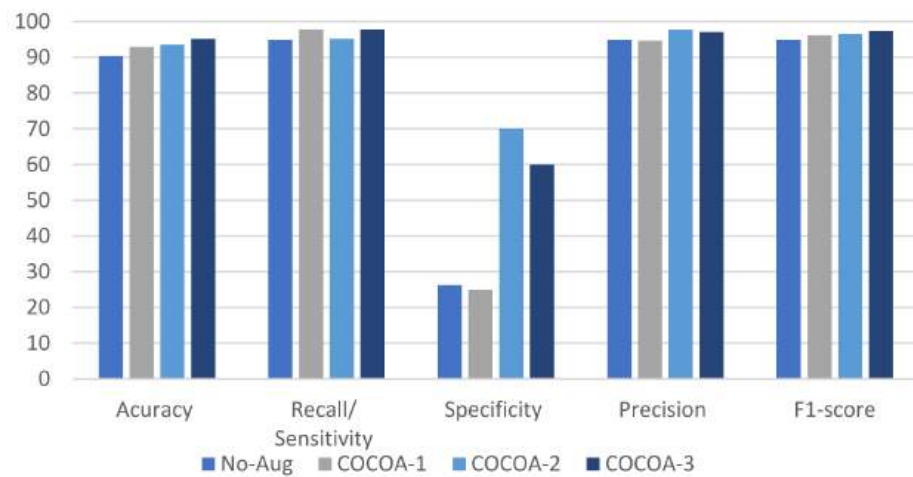


Figure 10. Comparison of metrics for classification of Healthy vs. Recovered Full (Mel-spectrogram images).

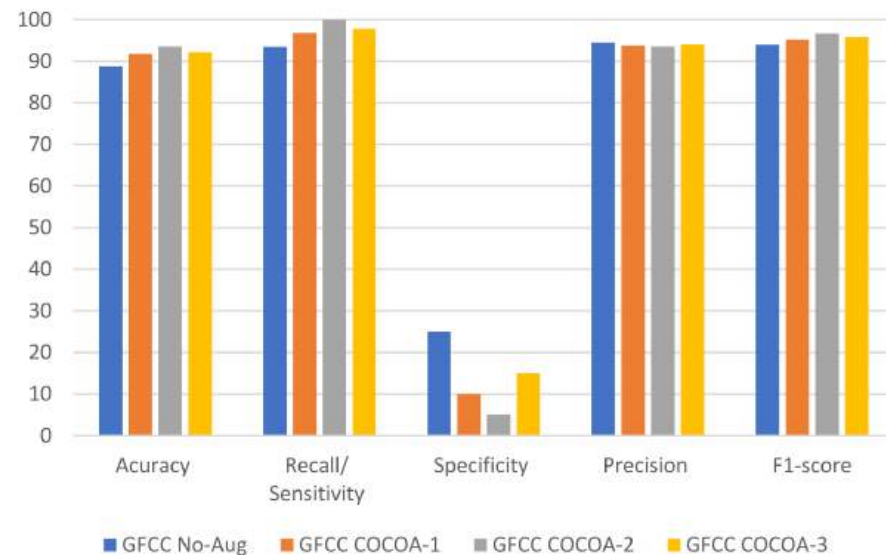


Figure 11. Comparison Metrics for classification of Healthy vs. Recovered Full (GFCC images).

4.5. Limitations

One of the major issues faced in this study is the problem of misclassification errors associated with the poor generalization of some noisy images. As expected, the majority of the error in misclassification can be attributed to a serious imbalance of classes and limited data samples. The differences between each class of sound, when represented as either Mel-spectrogram images or GFCC feature images, are almost similar to power representation and this could impact the ability of the model to generalize the data efficiently. The generated spectrogram for each audio file is a two-dimensional array of intensity values that is majorly noisy because of environmental noises connected to audio signals [55]. Therefore, it is important to equalize values distribution to enhance feature learning.

The proposed model is designed based on existing data augmentation techniques (color transformation and noise) and the features in the frequency domain, which makes the model simple and intuitive with low space cost. On the one hand, image spectra for sound signals could be a complex system, since some of the images cannot fully reflect the characteristic information of sound signals, although the frequency-domain feature has been used by previous researchers in sound classification tasks.

Regardless of these limitations, the proposed DeepShufNet model has proven to be effective in terms of the detection of COVID-19, despite the gross imbalance in classes and

the limited dataset. Moreover, it has low computational complexity in terms of resources and time. In the future, there is still a need to explore more complex data augmentation methods to overcome some of the errors due to the misclassification of the images by generating a cleaner dataset for proper generalization.

4.6. Comparison to Related Work on COVID-Sound Databases and Discussion

Further comparison in terms of accuracy, recall, and precision was carried out between our proposed system and other existing COVID-19 sound database systems. Despite applying different experimental conditions to each classification task, the proposed DeepShufNet model shows improved and promising results with respect to COVID-19 detection compared to the existing studies. The summary of the comparison table with related work is presented in Table 10.

Table 10. Comparison with other works.

Reference	Classifier	Acc (%)	Spec (%)	Sens (%)
Chaudhari et al. [56]	Ensemble deep learning model	77.1	NA	NA
Bagad et al. [57]	VGGish shallow	72	NA	NA
Pahar et al. [42]	ResNet 50, LR, KNN, SVM, LSTM	91	92	90
Our work	CNN based on DeepShufNet	90.1	95.98	62.3

5. Conclusion

The increasing popularity of the application of different deep neural network models in sound classification tasks is quite impressive. However, there has been some research work on COVID-19 detection based on different CNN architectures and some of the publicly available datasets still suffer from huge data imbalance, limited datasets, and poor classification of some of the machine learning models. Therefore, this work aims to apply a deep learning model, called DeepShufNet, to different categories of data augmentation techniques. The main contributions of this work include:

1. Covering the gap between limited datasets and class imbalance by creating a larger corpus of synthetic datasets using some simple and effective data augmentation techniques. Additionally, three different synthetic datasets were created, namely COCOA-1, COCOA-2, COCOA-3.
2. Deep learning based on pre-trained Shufflenet architecture, called the DeepShufNet model, was trained and evaluated on the analyzed datasets for comparison. The experimental analysis of the augmented datasets in comparison with baseline results showed significant improvement in performance metrics, better data generalization and enhanced optimal test results.

We compared and analyzed the effects of the two different feature extraction methods, namely Mel-spectrogram and GFCC imaging, on the DeepShufNet model. This study investigated the effects of augmented images in the detection of COVID-19, including positive asymptomatic cases, and fully recovered cases. The results showed that the DeepShufNet model had the highest accuracy on COCOA-2 Mel-spectrogram images for almost all the comparison cases. The proposed DeepShufNet models showed an improved performance, especially in the recall rate, precision, and F1-Score rate for all three types of augmented images. The proposed model showed the highest test results, with scores for accuracy, precision, recall, specificity, and f-score being 90.1%, 77.1%, 62.7%, 95.98%, and 69.1%, respectively, for positive COVID-19 detection using the Mel COCOA-2 training datasets. In the same manner, the experimental result for the detection of positive asymptomatic achieved the best recall rate of 62.5% and specificity rate of 97.1%, and a 48% F1-score.

In the future, we will explore advanced data augmentation techniques such as the application of generative adversarial networks (GANs) to train and test the model. Furthermore, more deep learning architectures will be implemented to improve and enhance COVID-19 recognition performance. In addition, the proposed DeepShufNet deep learning

model could also be applied and evaluated with the combination of all the different sound datasets.

Author Contributions: All authors have contributed equally to this manuscript. Conceptualization, R.D.; methodology, R.D.; software, O.O.A.-A.; validation, O.O.A.-A., R.D. and R.M.; formal analysis, O.O.A.-A., R.D., A.A.A. and R.M.; investigation, O.O.A.-A., R.D., A.A.A. and R.M.; data curation, O.O.A.-A.; writing—original draft preparation, O.O.A.-A. and R.D.; writing—review and editing, A.A.A. and R.M.; visualization, O.O.A.-A. and R.D.; supervision, R.D.; funding acquisition, R.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The COSWARA dataset used in this study is available at <https://github.com/iiscleap/Coswara-Data> (accessed on 16 June 2022).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Anand, S.; Sharma, V.; Pourush, R.; Jaiswal, S. A comprehensive survey on the biomedical signal processing methods for the detection of COVID-19. *Ann. Med. Surg.* **2022**, *76*, 103519. [[CrossRef](#)] [[PubMed](#)]
2. Dunbar, K. Impact of the COVID-19 event on US banks' financial soundness. *Res. Int. Bus. Financ.* **2022**, *59*, 101520. [[CrossRef](#)]
3. Nielson, K.; Zhang, Y.; Ingram, J. The impact of COVID-19 on police officer activities. *J. Crim. Justice* **2022**, *82*, 101943. [[CrossRef](#)]
4. Rai, P.K.; Sonne, C.; Song, H.; Kim, K.H. The effects of COVID-19 transmission on environmental sustainability and human health: Paving the way to ensure its sustainable management. *Sci. Total Environ.* **2022**, *838*, 156039. [[CrossRef](#)]
5. Girdhar, A.; Kapur, H.; Kumar, V.; Kaur, M.; Singh, D.; Damasevicius, R. Effect of COVID-19 outbreak on urban health and environment. *Air Qual. Atmos. Health* **2021**, *14*, 389–397. [[CrossRef](#)]
6. Salih, S.Q.; Abdulla, H.K.; Ahmed, Z.S.; Surameery, N.M.S.; Rashid, R.D. Modified AlexNet Convolution Neural Network for Covid-19 Detection Using Chest X-ray Images. *Kurd. J. Appl. Res.* **2020**, *5*, 119–130. [[CrossRef](#)]
7. World Health Organization. *Infection Prevention and Control during Health Care When Novel Coronavirus (nCoV) Infection Is Suspected: Interim Guidance, 25 January 2020 (No. WHO/2019-nCoV/IPC/2020.2)*; World Health Organization: Geneva, Switzerland, 2020.
8. Shaikh, F.; Anderson, M.; Sohail, M.R.; Mulero, F.; Awan, O.; Dupont-Roettger, D.; Kubassova, O.; Dehmsehki, J.; Bisdas, S. Current landscape of Imaging and the potential role for Artificial intelligence in the management of COVID-19. *Curr. Probl. Diagn. Radiol.* **2020**, *50*, 430–435. [[CrossRef](#)]
9. Corman, V.M.; Landt, O.; Kaiser, M.; Molenkamp, R.; Meijer, A.; Chu, D.K.W.; Bleicker, T.; Brünink, S.; Schneider, J.; Schmidt, M.L.; et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* **2020**, *25*, 2000045. [[CrossRef](#)]
10. Elaziz, M.A.; Al-Qaness, M.A.A.; Zaid, E.O.A.; Lu, S.; Ibrahim, R.A.; Ewees, A.A. Automatic clustering method to segment COVID-19 CT images. *PLoS ONE* **2021**, *16*, e0244416. [[CrossRef](#)]
11. Alliou, H.; Mohammed, M.A.; Benameur, N.; Al-Khateeb, B.; Abdulkareem, K.H.; Garcia-Zapirain, B.; Damaševičius, R.; Maskeliūnas, R. A multi-agent deep reinforcement learning approach for enhancement of COVID-19 CT image segmentation. *J. Pers. Med.* **2022**, *12*, 309. [[CrossRef](#)]
12. Yousri, D.; Abd Elaziz, M.; Abualigah, L.; Oliva, D.; Al-qaness, M.A.A.; Ewees, A.A. COVID-19 X-ray images classification based on enhanced fractional-order cuckoo search optimizer using heavy-tailed distributions. *Appl. Soft Comput.* **2021**, *101*, 107052. [[CrossRef](#)]
13. Rehman, N.-U.; Zia, M.S.; Meraj, T.; Rauf, H.T.; Damaševičius, R.; El-Sherbeeney, A.M.; El-Meligy, M.A. A self-activated cnn approach for multi-class chest-related covid-19 detection. *Appl. Sci.* **2021**, *11*, 9023. [[CrossRef](#)]
14. Diaz-Escobar, J.; Ordóñez-Guillén, N.E.; Villarreal-Reyes, S.; Galaviz-Mosqueda, A.; Kober, V.; Rivera-Rodriguez, R.; Lozano Rizk, J.E. Deep-learning based detection of COVID-19 using lung ultrasound imagery. *PLoS ONE* **2021**, *16*, e0255886. [[CrossRef](#)]
15. Abayomi-Alli, O.O.; Damaševičius, R.; Maskeliūnas, R.; Misra, S. An ensemble learning model for COVID-19 detection from blood test samples. *Sensors* **2022**, *22*, 2224. [[CrossRef](#)]
16. Morell-Garcia, D.; Ramos-Chavarino, D.; Bauça, J.M.; del Castillo, P.A.; Ballesteros-Vizoso, M.A.; de Guadiana-Romualdo, L.G.; Gómez-Cobo, C.; Pou, J.A.; Amezaga-Menéndez, R.; Alonso-Fernández, A.; et al. Urine biomarkers for the prediction of mortality in COVID-19 hospitalized patients. *Sci. Rep.* **2021**, *11*, 11134. [[CrossRef](#)]
17. Yuan, C.; Wang, H.; Li, K.; Tang, A.; Dai, Y.; Wu, B.; Zhang, H.; Chen, J.; Liu, J.; Wu, W.; et al. SARS-CoV-2 viral shedding characteristics and potential evidence for the priority for faecal specimen testing in diagnosis. *PLoS ONE* **2021**, *16*, e0247367. [[CrossRef](#)]
18. Jędrusik, P.; Gaciong, Z.; Sklinda, K.; Sierpiński, R.; Walecki, J.; Gujski, M. Diagnostic role of chest computed tomography in coronavirus disease 2019. *Pol. Arch. Intern. Med.* **2020**, *130*, 520–528. [[CrossRef](#)]

19. Zhang, J.; Xie, Y.; Li, Y.; Shen, C.; Xia, Y. Covid-19 screening on chest X-ray images using deep learning-based anomaly detection. *arXiv* **2020**, arXiv:2003.12338.
20. Alyasseri, Z.A.A.; Al-Betar, M.A.; Abu Doush, I.; Awadallah, M.A.; Abasi, A.K.; Makhadmeh, S.N.; Alomari, O.A.; Abdulkareem, K.H.; Adam, A.; Damasevicius, R.; et al. Review on COVID-19 diagnosis models based on machine learning and deep learning approaches. *Expert Syst.* **2022**, *39*, e12759. [[CrossRef](#)]
21. Kumar, V.; Singh, D.; Kaur, M.; Damaševičius, R. Overview of current state of research on the application of artificial intelligence techniques for COVID-19. *PeerJ Comput. Sci.* **2021**, *7*, e564. [[CrossRef](#)]
22. Sharma, G.; Umopathy, K.; Krishnan, S. Audio texture analysis of COVID-19 cough, breath, and speech sounds. *Biomed. Signal Process. Control.* **2022**, *76*, 103703. [[CrossRef](#)]
23. Xia, T.; Han, J.; Qendro, L.; Dang, T.; Mascolo, C. Uncertainty-Aware COVID-19 Detection from Imbalanced Sound Data. *arXiv* **2021**, arXiv:2104.02005.
24. Ramasubramanian, C. Diagnosing the Stage of COVID-19 using Machine Learning on Breath Sounds. In Proceedings of the PHM Society European Conference, Virtual, 28 June–2 July 2021.
25. Lauraitis, A.; Maskeliunas, R.; Damaševičius, R.; Krilavičius, T. Detection of speech impairments using cepstrum, auditory spectrogram and wavelet time scattering domain features. *IEEE Access* **2020**, *8*, 96162–96172. [[CrossRef](#)]
26. Maskeliūnas, R.; Kulikajėvas, A.; Damaševičius, R.; Pribušis, K.; Ulozaitė-Stanienė, N.; Uloza, V. Lightweight deep learning model for assessment of substitution voicing and speech after laryngeal carcinoma surgery. *Cancers* **2022**, *14*, 2366. [[CrossRef](#)]
27. Salamon, J.; Bello, J.P. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Process. Lett.* **2017**, *24*, 279–283. [[CrossRef](#)]
28. Cotton, C.V.; Ellis, D.P. Spectral vs. spectro-temporal features for acoustic event detection. In Proceedings of the 2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), New Paltz, NY, USA, 16–19 October 2011; pp. 69–72.
29. Leng, Y.; Zhao, W.; Lin, C.; Sun, C.; Wang, R.; Yuan, Q.; Li, D. LDA-based data augmentation algorithm for acoustic scene classification. *Knowl. Based Syst.* **2020**, *195*, 105600. [[CrossRef](#)]
30. Wang, S.; Wu, Z.; He, G.; Wang, S.; Sun, H.; Fan, F. Semi-supervised classification-aware cross-modal deep adversarial data augmentation. *Future Gener. Comput. Syst.* **2021**, *125*, 194–205. [[CrossRef](#)]
31. Maskeliunas, R.; Raudonis, V.; Damasevicius, R. Recognition of emotional vocalizations of canine. *Acta Acust. United Acust.* **2018**, *104*, 304–314. [[CrossRef](#)]
32. Khamparia, A.; Gupta, D.; Nguyen, N.G.; Khanna, A.; Pandey, B.; Tiwari, P. Sound Classification Using Convolutional Neural Network and Tensor Deep Stacking Network. *IEEE Access* **2019**, *7*, 7717–7727. [[CrossRef](#)]
33. Tariq, Z.; Shah, S.K.; Lee, Y. Feature-Based Fusion Using CNN for Lung and Heart Sound Classification. *Sensors* **2022**, *22*, 1521. [[CrossRef](#)]
34. Barua, P.D.; Karasu, M.; Kobat, M.A.; Balik, Y.; Kivrak, T.; Baygin, M.; Dogan, S.; Demir, F.B.; Tuncer, T.; Tan, R.-S.; et al. An accurate valvular heart disorders detection model based on a new dual symmetric tree pattern using stethoscope sounds. *Comput. Biol. Med.* **2022**, *146*, 105599. [[CrossRef](#)] [[PubMed](#)]
35. Zhu, H.; Luo, M.-D.; Wang, R.; Zheng, A.-H.; He, R. Deep audio-visual learning: A survey. *Int. J. Autom. Comput.* **2021**, *18*, 351–376. [[CrossRef](#)]
36. Piczak, K.J. The details that matter: Frequency resolution of spectrograms in acoustic scene classification. In Proceedings of the Detection and Classification of Acoustic Scenes and Events 2017 Workshop, Munich, Germany, 16–17 November 2017; pp. 103–107.
37. Mushtaq, Z.; Su, S.F.; Tran, Q.V. Spectral images based environmental sound classification using CNN with meaningful data augmentation. *Appl. Acoust.* **2021**, *172*, 107581. [[CrossRef](#)]
38. Tena, A.; Clarià, F.; Solsona, F. Automated detection of COVID-19 cough. *Biomed. Signal Processing Control.* **2022**, *71*, 103175. [[CrossRef](#)]
39. Chowdhury, N.K.; Kabir, M.A.; Rahman, M.M.; Islam, S.M.S. Machine learning for detecting COVID-19 from cough sounds: An ensemble-based MCDM method. *Comput. Biol. Med.* **2022**, *145*, 105405. [[CrossRef](#)]
40. Dentamaro, V.; Giglio, P.; Impedovo, D.; Moretti, L.; Pirlo, G. AUCO ResNet: An end-to-end network for Covid-19 pre-screening from cough and breath. *Pattern Recognit.* **2022**, *127*, 108656. [[CrossRef](#)]
41. Loey, M.; Mirjalili, S. COVID-19 cough sound symptoms classification from scalogram image representation using deep learning models. *Comput. Biol. Med.* **2021**, *139*, 105020. [[CrossRef](#)]
42. 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.* **2022**, *141*, 105153. [[CrossRef](#)]
43. Erdoğan, Y.E.; Narin, A. COVID-19 detection with traditional and deep features on cough acoustic signals. *Comput. Biol. Med.* **2021**, *136*, 104765. [[CrossRef](#)]
44. Sait, U.; Gokul, L.K.V.; Shivakumar, S.; Kumar, T.; Bhaumik, R.; Prajapati, S.; Bhalla, K.; Chakrapani, A. A deep-learning based multimodal system for Covid-19 diagnosis using breathing sounds and chest X-ray images. *Appl. Soft Comput.* **2021**, *109*, 107522. [[CrossRef](#)]
45. Soltanian, M.; Borna, K. Covid-19 recognition from cough sounds using lightweight separable-quadratic convolutional network. *Biomed. Signal Processing Control.* **2022**, *72*, 103333. [[CrossRef](#)]

46. Despotovic, V.; Ismael, M.; Cornil, M.; Mc Call, R.; Fagherazzi, G. Detection of COVID-19 from voice, cough and breathing patterns: Dataset and preliminary results. *Comput. Biol. Med.* **2021**, *138*, 104944. [CrossRef]
47. 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.* **2021**, *11*, 15404. [CrossRef]
48. Brown, C.; Chauhan, J.; Grammenos, A.; Han, J.; Hasthanasombat, A.; Spathis, D.; Xia, T.; Cicuta, P.; Mascolo, C. Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data. *arXiv* **2020**, arXiv:2006.05919. [CrossRef]
49. Lella, K.K.; Pja, A. Automatic COVID-19 disease diagnosis using 1D convolutional neural network and augmentation with human respiratory sound based on parameters: Cough, breath, and voice. *AIMS Public Health* **2021**, *8*, 240–264. [CrossRef]
50. Melek, M. Diagnosis of COVID-19 and non-COVID-19 patients by classifying only a single cough sound. *Neural Comput. Appl.* **2021**, *33*, 17621–17632. [CrossRef]
51. Sharma, N.; Krishnan, P.; Kumar, R.; Ramoji, S.; Chetupalli, S.R.; Ghosh, P.K.; Ganapathy, S. Coswara—A database of breathing, cough, and voice sounds for COVID-19 diagnosis. *arXiv* **2020**, arXiv:2005.10548.
52. Chen, Y.-W. Audio Normalization by MATLAB. 2021. Available online: <https://www.mathworks.com/matlabcentral/fileexchange/69958-audio-normalization-by-matlab> (accessed on 24 November 2021).
53. Choi, K.; Fazekas, G.; Sandler, M.; Cho, K. A comparison of audio signal preprocessing methods for deep neural networks on music tagging. In Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 3–7 September 2018; pp. 1870–1874.
54. Patterson, R.; Nimmo-Smith, I.; Holdsworth, J.; Rice, P. An efficient auditory filterbank based on the gammatone function. In Proceedings of the Paper Meeting of the IOC Speech Group on Auditory Modelling at RSRE, Malvern, UK, 14–15 December 1987.
55. Esmaeilpour, M.; Cardinal, P.; Koerich, A.L. Unsupervised feature learning for environmental sound classification using weighted cycle-consistent generative adversarial network. *Appl. Soft Comput.* **2020**, *86*, 105912. [CrossRef]
56. Chaudhari, G.; Jiang, X.; Fakhry, A.; Han, A.; Xiao, J.; Shen, S.; Khanzada, A. Virufy: Global applicability of crowdsourced and clinical datasets for AI detection of COVID-19 from cough. *arXiv* **2020**, arXiv:2011.13320. [CrossRef]
57. Bagad, P.; Dalmia, A.; Doshi, J.; Nagrani, A.; Bhamare, P.; Mahale, A.; Rane, S.; Agarwal, N.; Panicker, R. Cough against Covid: Evidence of Covid-19 signature in cough sounds. *arXiv* **2020**, arXiv:2009.08790. [CrossRef]