IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 8 December 2022, accepted 16 December 2022, date of publication 22 December 2022, date of current version 29 December 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3231640

RESEARCH ARTICLE

Productive and Non-Productive Cough Classification Using Biologically Inspired Techniques

RONEEL V. SHARAN, (Senior Member, IEEE)
Australian Institute of Health Innovation, Macquarie University, Sydney, NSW 2109, Australia
e-mail: roneel.sharan@mq.edu.au

This work was supported by the Google Research Scholar Program and Macquarie University.

ABSTRACT Cough is a common symptom of respiratory diseases and the type of cough, in particular, productive (wet) or non-productive (dry) cough, is an important indicator of the condition of the respiratory system. It is useful in differential diagnosis and in understanding disease progression. However, determining the cough type in clinical practice can be subjective and sometimes unfeasible. This work, therefore, aims to develop an objective assessment method of the cough type. The proposed approach emulates the sound recognition process of humans. In particular, it uses the human auditory model to reveal the frequency characteristics of the cough sound signals and convolutional neural networks for decision-making. It is validated on a dataset of smartphone recordings of 396 cough samples from 88 subjects annotated as wet or dry by up to four expert pulmonologists. The cough signals are automatically segmented and time-frequency image data augmentation is performed during training using the synthetic minority oversampling technique to prevent model overfitting. A sensitivity of 93.13% and specificity of 91.42% (AUC=0.9700) is achieved in segmentation of cough and non-cough sounds and a sensitivity of 100% and specificity of 82.50% (AUC=0.9234) is achieved in detecting subjects with wet and dry cough. The proposed fully automated system in detecting subjects with wet and dry cough demonstrates strong classification performance. It has the potential to provide objective assessment of cough type using smartphone technology, such as in virtual healthcare which has seen an increased uptake during the ongoing pandemic.

INDEX TERMS Cochleagram, convolutional neural networks, cough sound, data augmentation.

I. INTRODUCTION

Globally, more than a billion people suffer from respiratory diseases [1]. Chronic obstructive pulmonary disease, asthma, acute lower respiratory tract infections, tuberculosis, and lung cancer are among the most common causes of severe illness and death [1], [2]. More recently, coronavirus disease 19 (COVID-19), a respiratory infection, was declared a pandemic in March 2020 [3]. As of December 2022, there have been more than 647 million confirmed cases of COVID-19 with more than 6.6 million deaths worldwide [4], although these numbers are likely to be significantly underreported [5].

The associate editor coordinating the review of this manuscript and approving it for publication was Stavros Ntalampiras.

Cough is a common symptom of respiratory diseases [6] and one of the most common presenting conditions in primary care globally [7]. It is a natural reflex of the body to clear the irritants in the respiratory system [8]. The sound of the cough is a result of the turbulent flow in expiration, causing vibration of the larger airways and laryngeal structures [9]. Different respiratory diseases affect the airways differently, thereby producing different types of cough, broadly categorized as productive (wet) or non-productive (dry) cough.

Wet cough typically produces sputum (mucus or phlegm) and, in the absence of sputum, a dry cough. Causes of wet cough can be infection and inflammation of the lungs and bronchi while other causes include chronic obstructive pulmonary disease, bronchiectasis, cystic fibrosis, and tuberculosis [10], [11]. Dry cough can be caused by diseases such

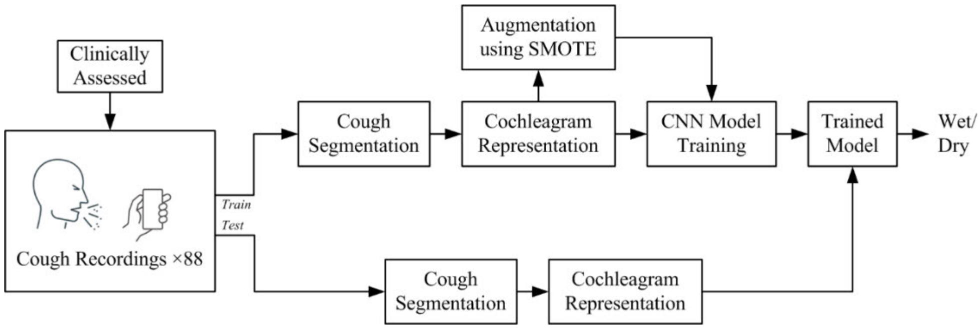

FIGURE 1. An overview of the proposed method in productive (wet) and non-productive (dry) cough classification.

as asthma or following a respiratory infection [11], [12]. In addition, dry cough has been reported in about two-thirds of COVID-19 cases [13]. In some diseases, such as bronchitis, the cough is usually dry in the early stages of the disease but later turns into phlegmy or wet cough as the progression of the disease leads to more secretions in the airways [14]. Therefore, cough is a vital indicator of the condition of the respiratory system and the differentiation of wet and dry cough is useful in differential diagnosis and understanding the progression of the disease. It is also useful in epidemiological studies and clinical research [15], [16].

The differentiation of the cough type is based on the perception of the sounds relating to airway secretions, that is, wet when the sound carries characteristics associated with sputum and dry in the absence of perceived wetness [17]. In clinical practice, the clinician can assess the nature of the cough by asking the patient or their carers and, where feasible, by listening to voluntary coughs. This can be subjective while bronchoscopy, an alternate method to evaluate airway secretions [18], requires a specialized device (bronchoscope), is invasive, and may require recovery time.

Despite the significance of cough type in clinical decision-making of respiratory diseases, objective detection of wet and dry using signal processing and machine learning techniques has received little attention [19], [20], [21], [22], [23], [24]. The temporal and spectral characteristics of wet and dry coughs have been studied in [19] and [20]. The size of the dataset in both works is small, with 20 productive and non-productive cough samples from 5 subjects in [19] and 16 cough samples, taken from publicly available datasets with verification from medical professionals, in [20].

A more comprehensive work in the automatic classification of wet and dry coughs is presented in [21]. Various handcrafted features and logistic regression classifier are used to distinguish between wet and dry coughs annotated by two respiratory physicians. However, their method uses manual segmentation to identify cough samples in the recordings. Also, conventional feature engineering and classification techniques employed in their work have since been superseded by deep learning methods, even on small datasets [25]. The recordings were made in a relatively controlled environment – at a single hospital in Indonesia using two recording devices with predetermined microphone positioning. As such, the dataset lacks diversity in terms of demographics and recording device characteristics and environments. These, together with manual cough segmentation, would make translation into a real-life tool difficult. In addition, their study is limited to the pediatric population but certain respiratory diseases, such as COVID-19, affect adults more than children [26]. Similar shortcomings can also be identified in [22], [23], and [24].

This work proposes a method for fully automated classification of wet and dry cough for the adolescent and adult populations. It is inspired by advancements in audio signal classification using artificial neural networks, in particular, convolutional neural networks (CNN) [27], [28]. CNN is originally an image classification technique and, in this work, time-frequency analysis is used to generate an image-like representation of the cough signals, referred to as cochleagram. This is achieved using gammatone filterbanks, which model the human auditory system, to analyze the frequency characteristics of the cough signals. The nonlinear gammatone filters offer finer frequency characterization at low frequency than at high frequency and have shown to be effective in cough sound classification [25], [29].

The method is validated on a dataset of 396 cough samples from 88 subjects annotated by up to four expert pulmonologists. The dataset is crowdsourced from many different countries on different continents with possibly varying recording devices and environments. In this work, the cough signals are automatically segmented and the relatively small dataset is augmented during training to prevent model overfitting. The performance of the proposed method is compared against three baseline methods to demonstrate its effectiveness in wet and dry cough classification.

II. MATERIALS AND METHODS

An overview of the proposed method for classifying the cough type is illustrated in Fig. 1 with details in the following subsections.

A. DATASET

The dataset of cough sound recordings was collected as part of COVID-19 research, to create a dataset for screening

TABLE 1. Overview of the dataset used in this work.

		Wet	Dry	Overall
Number of recordings		8	80	88
Number of coughs		34	362	396
Age	Range (Years)	16-46	14-60	14-60
	Average (Years)	36.88±10.51	33.42±12.06	33.79±11.88
COVID-19 Status	COVID-19	3	21	24
	Symptomatic	2	26	28
	Healthy	3	20	23
	Unknown	0	13	13

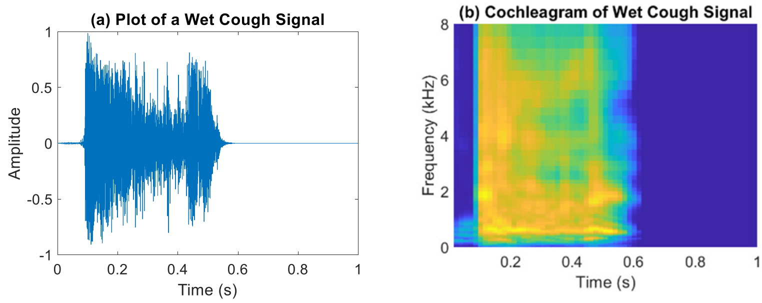

FIGURE 2. Illustration of (a) a wet cough signal and (b) its cochleagram representation.

COVID-19 using cough sound analysis [30]. The recordings have been made using smartphones and sampled at 16 kHz. The crowdsourced data is from subjects from around the world, likely from different make and model of smartphones, and recorded in varying environments, making this a very diverse dataset.

This work uses a subset of the dataset annotated by up to four pulmonologists as having *wet* or *dry* cough. This work utilizes recordings where the majority (three or more) pulmonologists agreed on the cough type. This results in a total of 88 recordings an overview of which is provided in Table 1, including the self-reported COVID-19 status. 8 of these recordings are labeled as wet with a total of 34 cough samples, and the remaining 80 recordings are labeled as dry with a total of 362 cough samples. This gives a total of 396 cough samples with the number of cough(s) per recording varying from 1 to 12. The age range of the subjects is 14-60 years (33.79±11.88 years).

B. COCHLEAGRAM REPRESENTATION

The cochleagram is a time-frequency representation of the cough signal, created using a filterbank based on the frequency selectivity property of the human cochlea and which decomposes the cough signal into different frequency bands. The filterbank is modeled using a gammatone filter, the impulse response of which is given as

$$g(t) = at^{n-1}e^{-2\pi bt}\cos(2\pi f_c t + \phi) \quad (1)$$

where a is the amplitude factor, t is the time in seconds, n is the filter order (set to four to model human hearing), f_c is the center frequency, b is the bandwidth of the filter, and ϕ is the phase factor [31]. The relationship between the center frequency and bandwidth is given by the equivalent rectangular bandwidth, a psychoacoustics measure approximating the bandwidths of the human auditory filters [31]. The gammatone filter was implemented as a cascade of four second-order digital filters [32].

In forming the cochleagram representation, all segmented cough signals are zero-padded or cropped to the same length (1 second). The signal is then decomposed using 64 gammatone filters. Each decomposed cough signal is divided into 64 equally sized frames with 50% overlap and the energy in each frame is computed to form a cochleagram representation of size 64 × 64, the input size of the CNN used in this work. An illustration of a wet cough signal and its cochleagram representation, with values in dB, is shown in Fig. 2.

C. CNN

1) CNN ARCHITECTURE

An important consideration in the design of CNN is network size. While many popular large networks are available today, large networks can suffer from overfitting when trained on small datasets [33]. Research shows that shallow networks can be more effective than deep networks on small datasets [34]. In [35], for example, a shallow CNN is trained successfully for the recognition of 50 sound event classes with only 50 training samples per class, without data augmentation. Due to the relatively small training data in this work, a relatively shallow network is utilized making it less prone to overfitting.

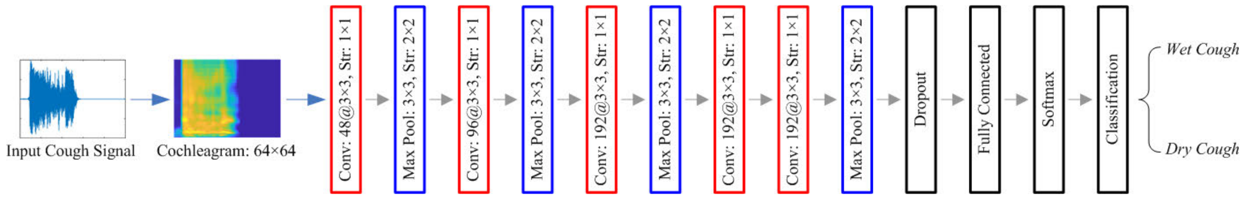

FIGURE 3. Architecture of the CNN used in wet and dry cough classification and cough segmentation.

The cochleagram images of dimension 64 × 64 are normalized at the input of the CNN network [25], illustrated in Fig. 3. The network has five convolutional layers, each with a filter size of 3 × 3. The first convolutional layer has 48 filters, 96 filters in the second convolutional layer, and 192 filters in the remaining three convolutional layers. Each convolutional layer is followed by a batch normalization layer [36] and rectified linear unit (ReLU) [37]. Each ReLU layer, except the fourth, is followed by a max pooling layer [38] with a pool size of 3 × 3 and a stride of 2 × 2. The final layers include a dropout layer [39], a fully connected layer, a softmax layer [40], and a classification layer.

2) DATA AUGMENTATION: SMOTE FOR COCHLEAGRAM IMAGES

Standard CNNs have a balanced focus on misclassification error making them unsuitable for datasets with severely imbalanced class distributions, as in this work. Two common approaches to address this problem are undersampling and oversampling. In undersampling, the majority class is undersampled by removing observations until the dataset is balanced while in oversampling the minority class is oversampled by adding observations. In this work, the minority class has only 34 samples, therefore, oversampling is preferred. Conventional oversampling works by duplicating the existing observations. While this increases the amount of data, it does not present new information to the classification model.

In conventional machine learning, the synthetic minority oversampling technique (SMOTE) [41] is commonly used for oversampling. SMOTE works by randomly selecting data from the minority class and then selecting its k-nearest neighbors, such as using distance measures between the data points. Synthetic data is generated by randomly selecting a point between the chosen data and a randomly selected neighboring data from the k neighbors. This procedure is repeated to generate as many synthetic samples as required.

While SMOTE was originally developed to generate synthetic data in the feature space, this work utilizes SMOTE to generate synthetic cochleagram time-frequency representation. SMOTE for the generation of synthetic cochleagram images works the same as conventional SMOTE, except that structural similarity (SSIM) [42] is used to select the nearest neighbors. Unlike distance measures, which are absolute, SSIM measures the spatial interdependencies between images making it more suitable for the given task.

3) CNN TRAINING

Adaptive moment estimation [43] is used for training the network and the training parameters are tuned using a simple grid search. The final setting for the initial learning rate is 0.001, mini batch size is 16, the maximum number of epochs is 10, learn rate drop factor is 0.5, learn rate drop period is 2, and L_2 regularization of 0.2. The model implementation is in MATLAB R2021a. The model is trained using a single NVIDIA V100 Tensor Core GPU and the training stops after the maximum number of epochs is reached.

4) SUBJECT/RECORDING PREDICTION

The pulmonologists annotated the recording from each subject as wet/dry. Each recording has one or more cough samples which are assigned the same label as the recording. While CNN determines the probability of the cough sample as wet/dry, the average of all the cough probability values from a recording is used to estimate the probability of the subject as having wet/dry cough.

D. AUTO SEGMENTATION

Each recording has one or more cough samples along with silence and other acoustic events, speech and non-speech, which are collectively referred to as non-cough samples. This work uses a supervised method for the automatic segmentation of cough samples in the recordings. In auto segmenting, firstly, the boundary (start and end points) of all sound events in all the recordings are detected using the envelope of the signal [44]. The cough and non-cough sounds are converted to cochleagram image, as described in Section II-B, and CNN, the same method as described in Section II-C, is trained to classify the cough and non-cough sounds. For supervised classification, for the training recordings, the detected sound events are labeled as cough if it has 50% or more overlap with manually segmented cough sounds and non-cough otherwise, as in [45]. The sound events in the test recordings are also labeled similarly to compute the classification results of the proposed auto segmentation method.

E. BASELINE METHODS

Along with the proposed cochleagram-CNN method, the performance of the *cough vs non-cough* and *wet vs dry* cough classification tasks is also evaluated on three baseline feature sets. *Baseline feature set 1* is based on earlier work in classifying wet/dry coughs [21], *baseline feature set 2* is descriptors of the cochleagram representation described in

Section II-B and is inspired by work in time-frequency image feature extraction for sound classification described in [46], and *baseline feature set 3* is inspired from work in detecting COVID-19 using cough sounds [47].

Baseline feature set 2 is the cochleagram image features (CIF) [48], analogous to the spectrogram image features of [46]. In computing the CIF, the 64 $\times$ 64 cochleagram representation described in Section II-B is divided into 8 $\times$ 8 blocks, that is, 8 vertical and 8 horizontal blocks, for a total of 64 blocks. The second and third central moments [46] are computed in each block. These values are then concatenated into a 128-dimensional feature vector.

In baseline feature set 3, two sets of handcrafted features and a set of transfer learning-driven features [47] are engineered and populated in each segmented signal. In computing the handcrafted features, each signal is divided into frames of 32 milliseconds with a 50% overlap. The first handcrafted feature set is based on mel-frequency cepstral coefficients (MFCCs) [49], a widely used feature in audio classification tasks that utilizes frequency scales based on the auditory perception. 13 MFCCs and the first and second derivatives of these coefficients [50] are computed in each frame.

The second handcrafted feature set has 4 features capturing temporal and spectral characteristics of the signal which are once again computed in each frame. These are the zero-crossing rate, short-time energy, spectral centroid, and spectral roll-off point [51], [52]. For both the handcrafted feature sets, the raw features are represented using the following 11 statistical features across all the frames: *mean*, *median*, *root mean square*, *maximum*, *minimum*, *1st and 3rd quartile*, *interquartile range*, *standard deviation*, *skewness*, and *kurtosis*. These result in a 429-dimensional MFCC feature set (including the first and second derivatives) and a 44-dimensional temporal and spectral feature set for each segmented signal.

In addition, 128 *VGGish* features are computed for each segmented signal using a pretrained convolutional neural network for audio classification [53]. VGGish is inspired by the popular VGG networks in image classification. It has been trained on a large YouTube audio dataset of 128-dimensional embeddings. In computing the VGGish features, each segmented signal is zero-padded or cropped to 0.975 seconds and transformed into a 94 $\times$ 64 log mel-spectrogram [28]. This time-frequency representation forms the input to the VGGish network for extracting the feature embeddings. The combined feature vector is, therefore, 601-dimensional (429 MFCC features, 44 temporal and spectral features, and 128 VGGish features).

The features in all three baseline feature sets are standardized using a z-score and the discriminative features are identified using a t-test with a p-value threshold of 0.05. While only logistic regression (LR) is used for binary (wet/dry) classification in [21], in this work, random forest (RF), support vector machine (SVM), and feedforward, fully connected neural network (FNN) classifiers are also experimented with on all three baseline feature sets. The FNN has two fully connected layers, of size 16, each of which is followed by ReLU activation. The final layers include a fully connected layer of size 2, a softmax layer, and a classification layer. The network is trained using the limited-memory Broyden-Fletcher-Goldfarb-Shanno quasi-Newton algorithm [54].

F. EVALUATION METRICS

The performance of the proposed cough type classification method is evaluated using sensitivity and specificity, where sensitivity is the fraction of wet coughs that are correctly classified, and specificity is the fraction of dry coughs that are correctly classified. The area under the curve (AUC) of the receiver operating characteristic (ROC) curve is also used as a single measure of classification performance. In auto segmentation, sensitivity refers to the proportion of cough samples that are correctly identified, and specificity the proportion of non-cough samples that are correctly identified. The optimal threshold on the ROC curve in all cases is determined as the point on the ROC curve that minimizes the distance to the point (0,1). In addition, the agreement between the pulmonologists is evaluated using percent agreement, and Cohen's kappa (κ) [55] when measuring the agreement between two pulmonologists and Fleiss' kappa [56] between more than two pulmonologists.

III. RESULTS

A. PULMONOLOGIST AGREEMENT

The agreement between the four pulmonologists (referred to as P1, P2, P3, and P4) in annotating the recordings as wet or dry cough is presented in Table 2. Pulmonologists 1 and 2 annotated 83 recordings in common, agreeing to the annotation of 71 recordings. This gives an overall percent agreement of 85.54% (κ = 0.3261). Similarly, pulmonologists 1 and 3 have an overall percent agreement of 67.09% (κ = 0.2596), pulmonologists 1 and 4 have an overall percent agreement of 92.94% (κ = 0.5860), pulmonologists 2 and 3 have an overall percent agreement of 67.33% (κ = 0.2821), pulmonologists 2 and 4 have an overall percent agreement of 82.14% (κ = 0.3103), and pulmonologists 3 and 4 have an overall percent agreement of 68.57% (κ = 0.2979).

While Table 2 analyzes the agreement between two pulmonologists at a time, the agreement between three and four pulmonologists is analyzed in Table 3. Pulmonologists 1, 2, and 3 annotated 77 recordings in common and agreed to the annotation of 45 of these recordings for an overall percent agreement of 58.44% (κ = 0.2165). Similarly, pulmonologists 1, 2, and 4 have an overall percent agreement of 81.71% (κ = 0.3760), pulmonologists 1, 3, and 4 have an overall percent agreement of 62.03% (κ = 0.2282), and pulmonologists 2, 3, and 4 have an overall percent agreement of 58.42% (κ = 0.2558). In addition, all four pulmonologists agreed on the cough type in 44 of the possible 77 recordings for an overall percent agreement of 57.14% (κ = 0.2383).

TABLE 2. Agreement between two pulmonologists in annotating wet and dry cough recordings.

	P1 & P2	P1 & P3	P1 & P4	P2 & P3	P2 & P4	P3 & P4
Number of common annotations	83	79	85	101	112	105
Number of agreements	71	53	79	68	92	72
% agreement	85.54%	67.09%	92.94	67.33%	82.14%	68.57%
Kappa (κ)	0.3261	0.2596	0.5860	0.2821	0.3103	0.2979

TABLE 3. Agreement between three and four pulmonologists in annotating wet and dry cough recordings.

	P1, P2 & P3	P1, P2 & P4	P1, P3 & P4	P2, P3 & P4	All Four
Number of common annotations	77	82	79	101	77
Number of agreements	45	67	49	59	44
% agreement	58.44%	81.71%	62.03	58.42%	57.14%
Kappa (κ)	0.2165	0.3760	0.2282	0.2558	0.2383

B. WET VS DRY COUGH CLASSIFICATION RESULTS USING MANUALLY SEGMENTED COUGHS

The performance of the cochleagram-CNN classification model described in Section II is evaluated in 8-fold stratified cross-validation at the subject level, whereby cough samples from 77 recordings (7 wet cough recordings and 70 dry cough recordings) are used to train the model and cough samples from the remaining 11 recordings (1 wet cough recording and 10 dry cough recordings) are used for validation. In each fold, the training data (cochleagram image) is balanced using SMOTE before training the classification model. A similar validation strategy is also used for the baseline methods where SMOTE, in its conventional form [41], is used to balance the training data by generating synthetic feature vectors.

The wet vs dry cough classification results using manually segmented cough samples are given in Table 4. An AUC of 0.8169 is achieved using SVM in classifying wet and dry cough samples using the first baseline feature set. This is a relative improvement of 8.83%, 16.62%, and 4.87% over the AUC values using LR, RF, and FNN classifiers, respectively. Similarly, an AUC of 0.8453 is achieved in classifying the individual subject recordings as wet or dry using SVM, a relative improvement of 9.07%, 9.95%, and 5.37% over the LR, RF, and FNN classifiers, respectively.

Interestingly, on the second baseline feature set, the FNN classifier achieves the highest AUC when compared to the LR, RF, and SVM classifiers. An AUC of 0.7457 and 0.8601 is achieved in cough and recording classification, respectively. On the final baseline feature set, an AUC of 0.7801 is achieved using SVM in classifying the cough samples as wet or dry, better than what could be achieved using LR, RF, and FNN. An AUC of 0.8609 is achieved using SVM in classifying the recordings as wet or dry which is once again better than what could be achieved using LR, RF, and FNN.

Overall, the classification of wet and dry cough recordings using the third baseline feature set yields slightly better AUC than the other two baseline feature sets. The highest AUC in classifying the recordings is 0.8609, which is 1.85% better than the best AUC of 0.8453 using baseline feature set 1 and 0.09% better than the best AUC of 0.8601 using baseline feature set 2.

The boxplot of the most significant features (lowest p-values) from each of the three feature sets in baseline feature set 3 is shown in Fig. 4. The first derivative of the 7^{th} mel-frequency cepstral coefficient is determined to be the most significant feature (p-value$=2.92 \times 10^{-13}$) from the MFCC feature set. The energy in the cough signals is also determined to be important (p-value$= 1.23 \times 10^{-7}$) together with VGGish feature embedding 26 (p-value$= 1.24 \times 10^{-8}$).

The proposed cochleagram-CNN approach achieves an AUC of 0.8376 in classifying the cough samples, an improvement of 2.53% over the best baseline method, and an AUC of 0.9313 in subject classification, an improvement of 8.18% over the best baseline method. With a sensitivity of 0.7647 and specificity of 0.7680 in cough sample classification and a sensitivity of 1.0000 and specificity of 0.8250 in the classification of subject recordings, these are the best overall classification results of all the methods considered in this work.

Next, the predictions of the CNN are investigated on a validation cochleagram image from a wet cough recording. This is performed using occlusion sensitivity [57], a deep learning visualization method. Occlusion sensitivity perturbs small areas of the cochleagram image by replacing it with an occluding mask and then measuring the change in probability score for the given class as the mask moves along the cochleagram image. The occlusion sensitivity map, Fig. 5(a), for the wet cough cochleagram image of Fig. 2(b) suggests that it is focusing on the area of high frequency for the wet class.

C. AUTO SEGMENTATION RESULTS

Results using the automatic cough segmentation algorithm are presented in Table 5 using the same training and validation procedure described in Section III-B. The cough and non-cough classification results using baseline feature set 3 are better than those using baseline feature set 1 and 2. The

TABLE 4. Classification results for wet vs dry cough using manually segmented cough samples.

Feature/Input	Classifier	Cough Classification Results			Subject Classification Results		
		Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC
Baseline Feature Set 1	LR	0.7059	0.7735	0.7506	0.8750	0.7375	0.7750
	RF	0.6176	0.7569	0.7005	0.7500	0.7875	0.7688
	SVM	0.7353	0.8370	0.8169	0.8750	0.8875	0.8453
	FNN	0.7647	0.6878	0.7790	0.8750	0.7875	0.8022
Baseline Feature Set 2	LR	0.7353	0.5331	0.6728	0.7500	0.6750	0.7281
	RF	0.7941	0.5580	0.7064	0.7500	0.7375	0.7500
	SVM	0.7059	0.6077	0.7089	0.8750	0.7125	0.8109
	FNN	0.7941	0.6215	0.7457	0.8750	0.7375	0.8601
Baseline Feature Set 3	LR	0.5294	0.8370	0.6840	0.8750	0.7000	0.8188
	RF	0.7353	0.7099	0.7465	0.7500	0.8500	0.7875
	SVM	0.7647	0.7762	0.7801	0.8750	0.7375	0.8609
	FNN	0.6765	0.7293	0.7522	0.8750	0.7125	0.8219
Cochleagram	CNN	0.7647	0.7680	**0.8376**	1.0000	0.8250	**0.9313**

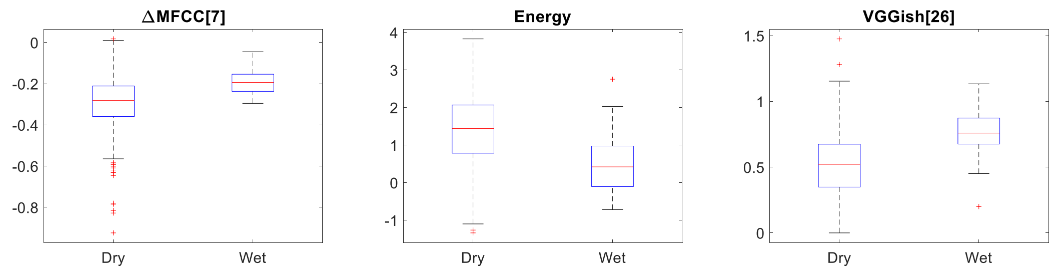

FIGURE 4. Boxplot of the most significant features from each feature group in baseline feature set 3.

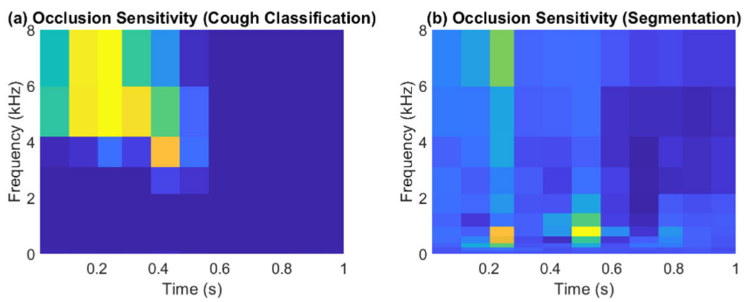

FIGURE 5. Occlusion sensitivity map for the wet cough cochleagram image of Fig. 2(b) in (a) wet vs dry cough classification task and (b) cough vs non-cough classification task.

highest AUC for all three feature sets is achieved using RF, the AUC of 0.8902 on baseline feature set 1, 0.9143 on baseline feature set 2, and 0.9600 on baseline feature set 3. However, with a sensitivity of 0.9313, specificity of 0.9142, and AUC of 0.9700, the best cough vs non-cough classification results are achieved using the proposed cochleagram-CNN method. The occlusion sensitivity map, Fig. 5(b), for the wet cough cochleagram image of Fig. 2(b) highlights different regions possibly because it contains many small features, including the region of high intensity and low frequency.

D. WET VS DRY COUGH CLASSIFICATION RESULTS USING AUTOMATICALLY SEGMENTED COUGHS

The results for the classification of wet/dry cough samples and subject recordings using automatically segmented coughs are given in Table 6. The experimental setup used here is exactly the same as that for manually segmented cough samples except that the cough samples are now automatically segmented. There are some differences in the AUC values in the classification of cough samples and subjects from manually segmented cough samples (Table 4) to automatically segmented cough samples (Table 6). However, the general

TABLE 5. Classification results for cough and non-cough segmentation.

Feature/Input	Classifier	Cough Classification Results		
		Sensitivity	Specificity	AUC
Baseline Feature Set 1	LR	0.7172	0.7657	0.8065
	RF	0.8263	0.8218	0.8902
	SVM	0.7293	0.7756	0.8150
	FNN	0.6485	0.6832	0.7233
Baseline Feature Set 2	LR	0.6949	0.8251	0.7484
	RF	0.8222	0.8416	0.9143
	SVM	0.8343	0.8284	0.9014
	FNN	0.8081	0.8482	0.8915
Baseline Feature Set 3	LR	0.8747	0.8086	0.8626
	RF	0.9051	0.8878	0.9600
	SVM	0.8747	0.9076	0.9563
	FNN	0.8667	0.8680	0.9395
Cochleagram	CNN	0.9313	0.9142	**0.9700**

trend is similar with the AUC using baseline feature set 3 generally higher than baseline feature sets 1 and 2, and the proposed cochleagram-CNN approach outperforms all three baseline feature sets. The AUC in the classification of cough samples and subjects using the cochleagram-CNN approach is 0.8071 and 0.9234, respectively. These are only marginally lower than what is achieved using manually segmented coughs. The sensitivity and specificity in subject classification are 1.0000 and 0.8250, respectively, the same as what is achieved using manually segmented coughs.

IV. DISCUSSION

This work proposes a method for the fully automated classification of wet and dry coughs. The proposed method first performs automatic cough segmentation. The segmented coughs are then classified as wet or dry coughs. To the best of my knowledge, this is the first work which performs automatic segmentation and classification of wet and dry coughs.

The proposed method for segmentation and classification is based on time-frequency (cochleagram) image classification using CNN with data augmentation during training using SMOTE. Conventionally, SMOTE is applied to feature vectors to generate new feature vectors. In this work, SMOTE is utilized to generate new time-frequency (cochleagram) representations. When using SMOTE with feature vectors, the nearest neighbors are often computed using the Euclidean distance measure, which is not suited for capturing the spatial relationship between images. When using SMOTE for time-frequency images, this work proposes to select the nearest neighbors using structural similarity, a technique used to compare similarity between images in image classification tasks. In addition, while gammatone filters (cochleagram) have been used for audio signal analysis before, to my knowledge, this is the first work where gammatone filters are used to analyze the frequency characteristics of wet and dry coughs.

In this work, the agreement (κ) between the pulmonologists in annotating the cough recordings as wet/dry is in the range of 0.21–0.40 which indicates fair agreement [21], except for between pulmonologists 1 and 4 which is 0.5860 indicating moderate agreement. The agreement (κ) between the two scorers is 0.5520 in [21] (moderate agreement) and 0.89 in [58] (almost perfect agreement). However, there are some differences in this work compared to [21] and [58], such as this work has data from the adolescent and adult population compared to children in their work. In particular, the recordings in [21] and [58] are made in the same environment (hospital) with the same recording device and experimental setup. The dataset in this work is crowd-sourced, likely coming from devices with different audio recording characteristics (hardware and software), different environments, and different microphone positioning. All of these can affect the quality of the sound and thereby make cough sound interpretation difficult, potentially highlighting the subjectiveness in determining cough type in this manner and the need for objective methods such as those presented in this paper.

Furthermore, a comparison of the method proposed in this work against earlier works is provided in Table 7. Simple temporal and spectral analysis of wet and dry coughs is performed on small datasets in [19] and [20]. In [21], the cough signals are manually segmented from the audio recordings. Leave-one-out cross-validation is performed on 310 cough sounds, from 60 subjects. The training and validation are performed on cough data from the same subjects. Sensitivity and specificity of 81% and 83% are reported in cross-validation. In addition, they report test results on a separate dataset of 18 recordings containing 117 cough sounds, achieving a sensitivity and specificity of 84% and 76%, respectively.

The dataset of [22] contains 352 wet and dry coughs but the number of subjects and the cough segmentation method could not be determined. They achieve an accuracy of 56% in detecting wet cough and 91% in detecting dry cough using Gabor filterbank features and the Gaussian mixture model. However, their dataset contains an additional third class of non-cough sounds, which were removed in my work at the auto segmentation stage.

While sensitivity and specificity of 93.00% and 95.93%, respectively, are reported in [23], there are two important differences compared to my work. Firstly, in [24], evaluation is performed with coughs from the same subjects present in both the training and validation folds. This makes it difficult to gauge how their proposed method would generalize on cough sounds from unseen subjects, which has greater practical value. In my work, coughs from test subjects are not used for training. Secondly, the cough signals in their work are manually segmented, once again limiting the practical use of their method.

A wet vs dry cough classification method using various cepstral features and multilayer perceptron on 1,554 recordings from the COUGHVID dataset is proposed in [24], achieving an AUC of 0.8570. However, the wet and dry

TABLE 6. Classification results for wet vs dry cough using automatically segmented cough samples.

Feature/Input	Classifier	Cough Classification Results			Subject Classification Results		
		Sensitivity	Specificity	AUC	Sensitivity	Specificity	AUC
Baseline Feature Set 1	LR	0.6154	0.7768	0.7146	0.8750	0.8875	0.8969
	RF	0.7179	0.6362	0.7140	0.8750	0.7875	0.8625
	SVM	0.7179	0.7121	0.7862	0.7500	0.9125	0.8688
	FNN	0.7179	0.7254	0.8003	0.8750	0.8000	0.8828
Baseline Feature Set 2	LR	0.4872	0.6272	0.5317	0.5000	0.7875	0.5953
	RF	0.6410	0.7076	0.7144	0.8750	0.8625	0.9047
	SVM	0.5897	0.8080	0.6956	0.7500	0.8125	0.8484
	FNN	0.7692	0.6786	0.7992	0.8750	0.8750	0.9016
Baseline Feature Set 3	LR	0.5385	0.8393	0.6869	0.8750	0.7375	0.8750
	RF	0.7179	0.7545	0.8031	0.8750	0.7875	0.8938
	SVM	0.8205	0.7188	0.8058	0.7500	0.9250	0.9109
	FNN	0.8574	0.6652	0.8067	0.8750	0.8500	0.9150
Cochleagram	CNN	0.7692	0.7679	**0.8071**	1.0000	0.8250	**0.9234**

TABLE 7. Comparison of different works in wet and dry cough classification.

Study	No. of subjects (coughs)	Method	Wet vs Dry Cough Classification Results		
			Sensitivity	Specificity	AUC
Murata *et al.* (1998) [19]	5 (20)	Analysis of time lapse, duration, wave pressure, sound spectrogram, frequency range, and time expanded waveform	–	–	–
Chatrzarrin *et al.* (2011) [20]	– (16)	Time and frequency domain analysis of wet and dry cough sounds	–	–	–
Swarnkar *et al.* (2013) [21]	60 (310)	Cough segmentation method: manual Cough features: temporal, spectral, cepstral features Feature selection: *p*-values of LR Classifier: LR Validation: leave-one-out (subject mix)	81%	83%	–
Schröder *et al.* (2016) [22]	– (352)	Cough segmentation method: unknown Cough features: Gabor filterbank features Classifier: Gaussian mixture model Validation: 5-fold (unknown)	56%	91%	–
Renjini *et al.* (2021) [23]	– (115)	Cough segmentation method: manual Cough features: wavelet and complex network features Classifier: feedforward neural network Validation: 15-fold (subject mix)	93.00%	95.93%	–
Pande *et al.* (2022) [24]	1,554 (–)	Cough segmentation method: none (silence removal) Cough features: cepstral features Data augmentation: adaptive synthetic oversampling Classifier: Multilayer perceptron Validation: partition (subject independent)	89.85%	–	0.8570
This work	88 (396)	Cough segmentation method: automatic Cough analysis: gammatone filters (cochleagram) Data augmentation: SMOTE Classifier: CNN Validation: 8-fold (subject independent)	100%	82.50%	0.9234

cough recordings in the COUGHVID dataset have been annotated by up to four pulmonologists, with majority agreement (where at least three pulmonologists agreed to the annotation) on only 88 recordings. As such, the inclusion and exclusion criteria for the recordings used in their work are unclear. Also, they only remove silence from their recordings but, as per the analysis in [30], the recordings in the COUGHVID dataset can contain various non-cough sounds as well which need to be removed, hence the need for auto segmentation.

In all earlier works in wet vs dry cough classification analyzed in Table 7, various feature engineering methods are employed. However, the method proposed in my work only requires the transformation of the signal to a cochleagram representation. The CNN then learns the signal characteristics directly from the cochleagram. The proposed method demonstrates strong classification results in differentiating between cough and non-cough sound events in auto segmentation (AUC=0.9700) and in differentiating between wet and dry coughs (AUC=0.9234).

While the method proposed in this work offers various advantages against earlier works, there are limitations. In particular, due to the small dataset, a shallow CNN is used as large networks tend to suffer from overfitting when trained on small datasets. The CNN has only 879.6k total learnable parameters, which is a much smaller number of parameters compared to the millions of parameters in popular deep CNN networks, such as ResNet, which has shown to be useful in COVID-19 cough classification [59]. It is possible that training a deep CNN network on a larger dataset will yield even better results in wet vs dry cough classification.

In addition, the audio recordings used in this work are annotated as wet or dry. Several audio recordings in the dataset have two or more coughs and these are assigned the same label as the recording. However, wet and dry coughs can be present in the same recording. This could explain the slightly lower classification results in cough classification than subject recording classification. As such, in the future, it would be preferred to annotate each cough as wet or dry, similar to [21].

V. CONCLUSION

A method for the classification of wet and dry coughs using signal processing and deep learning techniques is presented in this paper. The proposed cochleagram-CNN classification approach with automatically segmented cough samples is seen to yield a similar classification performance to manually segmented cough samples. The proposed method offers various advantages compared to earlier works (Table 7), such as the use of automatically segmented cough samples and variability in the dataset which can be expected in real-life applications. This also makes for a challenging dataset which is probably why the handcrafted features from [21] (baseline feature set 1), which are proposed for a relatively controlled environment, achieved lower classification performance compared to the proposed deep learning based methods.

It is estimated that there are hundreds of thousands of seasonal influenza-associated respiratory deaths every year [60]. In the United States alone, there were an estimated 18 million medical visits during the 2019-2020 influenza season [61]. Globally, there has been an increased uptake of virtual healthcare during COVID-19 and this is largely expected to continue [62]. There is a potential for the proposed approach to be integrated into existing virtual healthcare systems to provide physicians with an objective assessment of the patient's airway secretions during a consultation using smartphone technology.

REFERENCES

[1] *Forum of International Respiratory Societies, The Global Impact of Respiratory Disease*, 2nd ed., European Respiratory Society, Sheffield, U.K., 2017.

[2] H. Wang, "Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: A systematic analysis for the Global Burden of Disease Study 2015," *Lancet*, vol. 388, no. 10053, pp. 1459–1544, 2016.

[3] *WHO Director-General's Opening Remarks at the Media Briefing on COVID-19–11 March 2020*, World Health Organization, Geneva, Switzerland, 2020.

[4] Johns Hopkins Coronavirus Resource Center. *COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)*. Accessed: Dec. 8, 2022. [Online]. Available: https://coronavirus.jhu.edu/map.html

[5] A. Karlinsky and D. Kobak, "Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset," *eLife*, vol. 10, p. e69336, 2021, doi: 10.7554/eLife.69336.

[6] W. T. Goldsmith, A. M. Mahmoud, J. S. Reynolds, W. G. McKinney, A. A. Afshari, A. A. Abaza, and D. G. Frazer, "A system for recording high fidelity cough sound and airflow characteristics," *Ann. Biomed. Eng.*, vol. 38, no. 2, pp. 469–477, Feb. 2010.

[7] C. R. Finley, "What are the most common conditions in primary care," *Can. Family Physician*, vol. 64, no. 11, pp. 832–840, 2018.

[8] M. Malvè, A. P. del Palomar, J. L. López-Villalobos, A. Ginel, and M. Doblaré, "FSI analysis of the coughing mechanism in a human trachea," *Ann. Biomed. Eng.*, vol. 38, no. 4, pp. 1556–1565, Apr. 2010.

[9] A. B. Chang, G. J. Redding, and M. L. Everard, "Chronic wet cough: Protracted bronchitis, chronic suppurative lung disease and bronchiectasis," *Pediatric Pulmonol.*, vol. 43, no. 6, pp. 519–531, 2008.

[10] K. Lai, H. Shen, X. Zhou, Z. Qiu, S. Cai, K. Huang, Q. Wang, C. Wang, J. Lin, C. Hao, L. Kong, S. Zhang, Y. Chen, W. Luo, M. Jiang, J. Xie, and N. Zhong, "Clinical practice guidelines for diagnosis and management of cough—Chinese thoracic society (CTS) asthma consortium," *J. Thoracic Disease*, vol. 10, no. 11, pp. 6314–6351, Nov. 2018.

[11] P. Kardos, "Management of cough in adults," *Breathe*, vol. 7, no. 2, p. 122, 2010.

[12] R. A. Haque, O. S. Usmani, and P. J. Barnes, "Chronic idiopathic cough: A discrete clinical entity," *Chest*, vol. 127, no. 5, pp. 1710–1713, 2005.

[13] *Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)*, World Health Organization, Geneva, Switzerland, 2020.

[14] M. D. Shields and S. Thavagnanam, "The difficult coughing child: Prolonged acute cough in children," *Cough*, vol. 9, no. 1, p. 11, 2013.

[15] A. H. Morice, L. McGarvey, and I. Pavord, "Recommendations for the management of cough in adults," *Thorax*, vol. 61, no. 1, pp. i1–i24, 2006.

[16] L. P. McGarvey, P. Forsythe, L. G. Heaney, J. MacMahon, and M. Ennis, "Bronchoalveolar lavage findings in patients with chronic nonproductive cough," *Eur. Respiratory J.*, vol. 13, no. 1, p. 59, 1999.

[17] M. Weinberger and M. Hurvitz, "Diagnosis and management of chronic cough: Similarities and differences between children and adults," *F1000Res.*, vol. 9, no. 757, pp. 1–10, 2020.

[18] B. A. Chang, J. T. Gaffney, M. M. Eastburn, J. Faoagali, N. C. Cox, and I. B. Masters, "Cough quality in children: A comparison of subjective vs. bronchoscopic findings," *Respiratory Res.*, vol. 6, no. 1, pp. 1–8, Dec. 2005.

[19] A. Murata, Y. Taniguchi, Y. Hashimoto, Y. Kaneko, Y. Takasaki, and S. Kudoh, "Discrimination of productive and non-productive cough by sound analysis," *Internal Med.*, vol. 37, no. 9, pp. 732–735, 1998.

[20] H. Chatrzarrin, A. Arcelus, R. Goubran, and F. Knoefel, "Feature extraction for the differentiation of dry and wet cough sounds," in *Proc. IEEE Int. Symp. Med. Meas. Appl.*, May 2011, pp. 162–166.

[21] V. Swarnkar, U. R. Abeyratne, A. B. Chang, Y. A. Amrulloh, A. Setyati, and R. Triasih, "Automatic identification of wet and dry cough in pediatric patients with respiratory diseases," *Ann. Biomed. Eng.*, vol. 41, no. 5, pp. 1016–1028, May 2013.

[22] J. Schröder, J. Anemuller, and S. Goetze, "Classification of human cough signals using spectro–temporal Gabor filterbank features," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 6455–6459.

[23] A. Renjini, M. S. Swapna, V. Raj, and S. Sankararaman, "Graph-based feature extraction and classification of wet and dry cough signals: A machine learning approach," *J. Complex Netw.*, vol. 9, no. 6, pp. 1–11, Oct. 2021.

[24] S. Pande, A. Patil, and S. Petkar, "Dry and wet cough detection using fusion of cepstral base statistical features," in *Proc. Int. Conf. Decis. Aid Sci. Appl. (DASA)*, Mar. 2022, pp. 874–878.

[25] R. V. Sharan, S. Berkovsky, D. F. Navarro, H. Xiong, and A. Jaffe, "Detecting pertussis in the pediatric population using respiratory sound events and CNN," *Biomed. Signal Process. Control*, vol. 68, Jul. 2021, Art. no. 102722.

[26] N. Dhochak, T. Singhal, S. K. Kabra, and R. Lodha, ''Pathophysiology of COVID-19: Why children fare better than adults?'' *Indian J. Pediatrics*, vol. 87, no. 7, pp. 537–546, Jul. 2020.

[27] O. Abdel-Hamid, A. R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, ''Convolutional neural networks for speech recognition,'' *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 22, no. 10, pp. 1533–1545, Oct. 2014.

[28] R. V. Sharan and T. J. Moir, ''Acoustic event recognition using cochleagram image and convolutional neural networks,'' *Appl. Acoust.*, vol. 148, pp. 62–66, May 2019.

[29] R. V. Sharan, U. R. Abeyratne, V. R. Swarnkar, and P. Porter, ''Automatic croup diagnosis using cough sound recognition,'' *IEEE Trans. Biomed. Eng.*, vol. 66, no. 2, pp. 485–495, Feb. 2019.

[30] L. Orlandic, T. Teijeiro, and D. Atienza, ''The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms,'' *Sci. Data*, vol. 8, no. 1, p. 156, Jun. 2021.

[31] R. D. Patterson, K. Robinson, J. Holdsworth, D. McKeown, C. Zhang, and M. Allerhand, ''Complex sounds and auditory images,'' in *Auditory Physiology and Perception*, Y. Cazals, K. Horner, and L. Demany, Eds. Bergama, Turkey: Pergamon, 1992, pp. 429–446.

[32] M. Slaney, ''An efficient implementation of the Patterson–Holdsworth auditory filter bank,'' Apple Comput., Inc., Cupertino, CA, USA, Tech. Rep. 35, 1993.

[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, ''Imagenet classification with deep convolutional neural networks,'' in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[34] K. Pasupa and W. Sunhem, ''A comparison between shallow and deep architecture classifiers on small dataset,'' in *Proc. 8th Int. Conf. Inf. Technol. Electr. Eng. (ICITEE)*, Oct. 2016, pp. 1–6.

[35] H. Zhang, I. McLoughlin, and Y. Song, ''Robust sound event recognition using convolutional neural networks,'' in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 559–563.

[36] S. Ioffe and C. Szegedy, ''Batch normalization: Accelerating deep network training by reducing internal covariate shift,'' 2015, *arXiv:1502.03167*.

[37] V. Nair and G. Hinton, ''Rectified linear units improve restricted Boltzmann machines,'' in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.

[38] K. Jarrett, K. Kavukcuoglu, M. A. Ranzato, and Y. LeCun, ''What is the best multi-stage architecture for object recognition?'' in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep. 2009, pp. 2146–2153.

[39] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, ''Dropout: A simple way to prevent neural networks from overfitting,'' *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[40] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[41] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, ''Smote: Synthetic minority over-sampling technique,'' *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Dec. 2002.

[42] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, ''Image quality assessment: From error visibility to structural similarity,'' *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[43] D. P. Kingma and J. Ba, ''Adam: A method for stochastic optimization,'' 2014, *arXiv:1412.6980*.

[44] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.

[45] Y. A. Amrulloh, U. R. Abeyratne, V. Swarnkar, R. Triasih, and A. Setyati, ''Automatic cough segmentation from non-contact sound recordings in pediatric wards,'' *Biomed. Signal Process. Control*, vol. 21, pp. 126–136, Aug. 2015.

[46] J. Dennis, H. D. Tran, and H. Li, ''Spectrogram image feature for sound event classification in mismatched conditions,'' *IEEE Signal Process. Lett.*, vol. 18, no. 2, pp. 130–133, Feb. 2011.

[47] C. Brown, J. Chauhan, A. Grammenos, J. Han, A. Hasthanasombat, D. Spathis, T. Xia, P. Cicuta, and C. Mascolo, ''Exploring automatic diagnosis of COVID-19 from crowdsourced respiratory sound data,'' in *Proc. 26th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2020, pp. 3474–3484.

[48] R. V. Sharan and T. J. Moir, ''Cochleagram image feature for improved robustness in sound recognition,'' in *Proc. IEEE Int. Conf. Digit. Signal Process. (DSP)*, Jul. 2015, pp. 441–444.

[49] S. B. Davis and P. Mermelstein, ''Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences,'' *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, no. 4, pp. 357–366, Aug. 1980.

[50] S. Young, *The HTK Book (for HTK Version 3.4)*. Cambridge, U.K.: Cambridge Univ. Engineering Department, 2009.

[51] G. Peeters, ''A large set of audio features for sound description (similarity and classification) in the CUIDADO project,'' IRCAM, Paris, France, Tech. Rep., 2004.

[52] E. Scheirer and M. Slaney, ''Construction and evaluation of a robust multifeature speech/music discriminator,'' in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1997, pp. 1331–1334.

[53] S. Hershey, S. Chaudhuri, D. P. W. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, M. Slaney, R. J. Weiss, and K. Wilson, ''CNN architectures for large-scale audio classification,'' in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 131–135.

[54] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. New York, NY, USA: Springer, 2006.

[55] J. Cohen, ''A coefficient of agreement for nominal scales,'' *Educ. Psychol. Meas.*, vol. 20, no. 1, pp. 37–46, 1960.

[56] J. L. Fleiss, ''Measuring nominal scale agreement among many raters,'' *Psychol. Bull.*, vol. 76, no. 5, pp. 378–382, 1971.

[57] M. D. Zeiler and R. Fergus, ''Visualizing and understanding convolutional networks,'' in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Zurich, Switzerland, 2014, pp. 818–833.

[58] N. Bisballe-Müller, A. B. Chang, E. J. Plumb, V. M. Oguoma, S. Halken, and G. B. McCallum, ''Can acute cough characteristics from sound recordings differentiate common respiratory illnesses in children?'' *Chest*, vol. 159, no. 1, pp. 259–269, Jan. 2021.

[59] H. Coppock, A. Gaskell, P. Tzirakis, A. Baird, L. Jones, and B. Schuller, ''End-to-end convolutional neural network enables COVID-19 detection from breath and cough audio: A pilot study,'' *BMJ Innov.*, vol. 7, no. 2, pp. 356–362, Apr. 2021.

[60] A. D. Iuliano, ''Estimates of global seasonal influenza-associated respiratory mortality: A modelling study,'' *Lancet*, vol. 391, no. 10127, pp. 1285–1300, 2018.

[61] Centers for Disease Control and Prevention. (2020). *Estimated Influenza Illnesses, Medical Visits, Hospitalizations, and Deaths in the United States 2019–2020 Influenza Season*. Accessed: Jun. 3, 2022. [Online]. Available: https://www.cdc.gov/flu/about/burden/2019-2020.html

[62] P. Webster, ''Virtual health care in the era of COVID-19,'' *Lancet*, vol. 395, no. 10231, pp. 1180–1181, Apr. 2020.

RONEEL V. SHARAN (Senior Member, IEEE) received the Ph.D. degree in engineering from the Auckland University of Technology, Auckland, New Zealand, in 2016.

From 2016 to 2019, he was a Postdoctoral Research Fellow at the School of Information Technology and Electrical Engineering, University of Queensland, Brisbane, Australia. Since 2019, he has been a Research Fellow with the Australian Institute of Health Innovation, Macquarie University, Sydney, Australia. His research interests include biomedical signal processing and machine learning.