A Lightweight CNN Model for Detecting Respiratory Diseases from Lung Auscultation Sounds using EMD-CWT-based Hybrid Scalogram

Preprint *in* IEEE Journal of Biomedical and Health Informatics · September 2020

CITATIONS
0

READS
1,459

5 authors, including:

Samiul Based Shuvo
Bangladesh University of Engineering and Technology
9 PUBLICATIONS **11** CITATIONS
SEE PROFILE

Shams Nafisa Ali
Bangladesh University of Engineering and Technology
17 PUBLICATIONS **8** CITATIONS
SEE PROFILE

Soham Irtiza Swapnil
Bangladesh University of Engineering and Technology
8 PUBLICATIONS **7** CITATIONS
SEE PROFILE

Taufiq Hasan
Bangladesh University of Engineering and Technology
56 PUBLICATIONS **1,188** CITATIONS
SEE PROFILE

A Lightweight CNN Model for Detecting Respiratory Diseases from Lung Auscultation Sounds using EMD-CWT-based Hybrid Scalogram

Samiul Based Shuvo[1, ‡], Shams Nafisa Ali[1, ‡], Soham Irtiza Swapnil[1, ‡],
Taufiq Hasan[1], *Member, IEEE* and Mohammed Imamul Hassan Bhuiyan[2], *Senior Member, IEEE*

***Abstract*—Listening to lung sounds through auscultation is vital in examining the respiratory system for abnormalities. Automated analysis of lung auscultation sounds can be beneficial to the health systems in low-resource settings where there is a lack of skilled physicians. In this work, we propose a lightweight convolutional neural network (CNN) architecture to classify respiratory diseases from individual breath cycles using hybrid scalogram-based features of lung sounds. The proposed feature-set utilizes the empirical mode decomposition (EMD) and the continuous wavelet transform (CWT). The performance of the proposed scheme is studied using a patient independent train-validation-test set from the publicly available ICBHI 2017 lung sound dataset. Employing the proposed framework, weighted accuracy scores of 98.92% for three-class chronic classification and 98.70% for six-class pathological classification are achieved, which outperform well-known and much larger VGG16 in terms of accuracy by absolute margins of 1.10% and 1.11%, respectively. The proposed CNN model also outperforms other contemporary lightweight models while being computationally comparable.**

***Index Terms*—Lung auscultation sound, respiratory disease detection, lightweight convolutional neural networks, empirical mode decomposition, continuous wavelet transform, scalogram.**

I. Introduction

LUNG diseases are the third largest cause of death in the world [1]. According to the World Health Organization (WHO), the five major respiratory diseases [2], namely chronic obstructive pulmonary disease (COPD), tuberculosis, acute lower respiratory tract infection (LRTI), asthma, and lung cancer, cause the death of more than 3 million people each year worldwide [3], [4]. These respiratory diseases severely affect the overall healthcare system and adversely affect the lives of the general population. Prevention, early diagnosis and treatment are considered key factors for limiting the negative impact of these deadly diseases.

Auscultation of the lung with a stethoscope is the traditional and most popular diagnostic method used by specialists and general practitioners for performing the initial investigation of the respiratory system. Although physicians use various other investigation strategies such as plethysmography, spirometry, and arterial blood gas analysis, lung sound auscultation remains vital for physicians due to its simplicity and low-cost [5]. The primary classification of these non-periodic and non-stationary sounds consists of two groups: normal (vesicular) and abnormal (adventitious) [6]. The first group is observed when there are no respiratory diseases, while the latter group indicates complications in the lungs or airways [7]. Crackle, wheeze, rhonchus, squawk, stridor, and pleural rub are the commonly known abnormal lung sounds. These anomalies can be differentiated from the normal lung sounds on the basis of frequency, pitch, energy, intensity, timbre, and musicality [8], [9]. Therefore, lung sounds are of particular importance for recognizing specific respiratory diseases and assessing its chronic and non-chronic characteristics. However, identifying the subtle differences between some of the adventitious lung sound classes can be a strenuous task even for a specialist and may introduce subjectivity in the diagnostic interpretation [10]. In this scenario, artificial intelligence (AI)-empowered algorithms can be of great benefit in automatically interpreting respiratory diseases from lung sounds, especially in underdeveloped regions of the world, with a scarcity of skilled physicians.

In the past decade, a number of research approaches have been considered and evaluated for automatic identification of respiratory anomalies from lung auscultation sounds. Numerous feature extraction techniques including statistical features [11], entropy-based features [12], wavelet coefficients [13], Mel Frequency Cepstral Coefficients (MFCC) [10], spectrograms [14], scalograms [15] etc. have been adopted in conjunction with a diverse set of machine learning (ML) algorithms [10]–[23].

With the advent of deep learning (DL), new developments have been made in recent times, demonstrating promising results in various clinical applications [24]–[29]. With automatic feature learning, deep learning approaches are more generic and can mitigate the limitations of traditional ML-based methods. In the same vein, DL-based paradigms that are employed in recent years for the identification of respiratory anomalies and pathologies from lung auscultation data have exhibited promising results [5], [30]–[41]. However, for attaining proper functionality, the deep networks require to undergo an extensive training scheme with a large training dataset that subsequently calls for a considerable amount of

This work was supported by Bangladesh University of Engineering and Technology (BUET), Dhaka-1205, Bangladesh.

[1]Samiul Based Shuvo, Shams Nafisa Ali, Soham Irtiza Swapnil and Taufiq Hasan are with Department of Biomedical Engineering, BUET, Email: {sbshuvo.bme.buet, snafisa.bme.buet, swapnil.buetbme}@gmail.com, taufiq@bme.buet.ac.bd.

[2]Mohammed Imamul Hassan Bhuiyan is with Department of Electrical and Electronic Engineering, BUET, Email: imamul@eee.buet.ac.bd

[‡]These authors share first authorship on, and contributed equally to, this work.

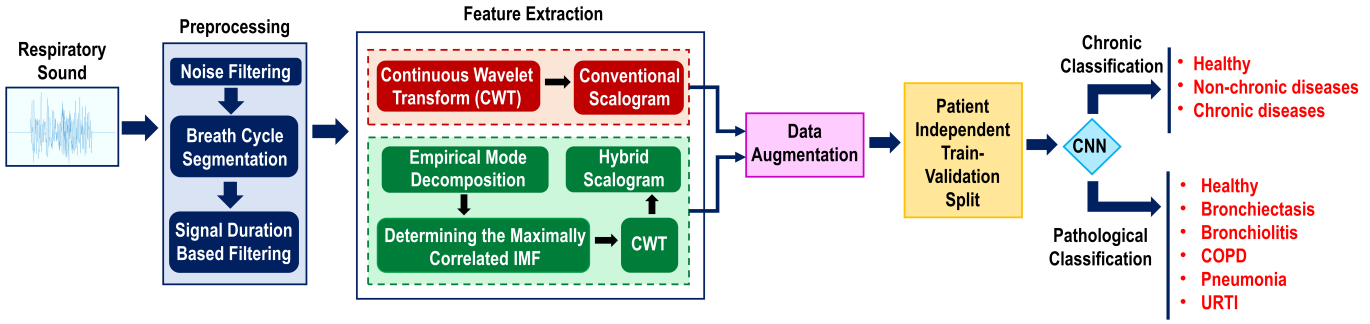

Fig. 1. **A graphical overview of the proposed framework. After several generic pre-processing steps, the lung sound signals are converted into scalograms using both conventional and hybrid approaches. The resulting images are further augmented and fed into the proposed lightweight CNN model to perform experiments using two different tasks: (i) 3-class chronic classification (healthy, chronic, non-chronic) and (ii) 6-class pathology classification (healthy, bronchiectasis, bronchiolitis, COPD, Pneumonia, URTI).**

time and the engagement of powerful computational resources. As a result, it becomes quite challenging to incorporate the deep learning frameworks in the currently available wearable devices and mobile platforms. In order to reduce the number of parameters of these networks, various methods have been investigated, including weight quantization [35], lightweight networks [42] and low precision computation [43].

While constructing AI-assisted automated medical diagnosis systems, patient independent train, validation and test datasets should be utilized in order to achieve reliable performance for unseen patient data [44], [45]. Due to the limited availability of medical data, this factor is often neglected in the existing literature. The random adoption of 80%-20% or any other percentage of the train-test split of the dataset introduces intra-patient dependency, and ultimately the obtained results cannot be generalized in case of a new patient [35].

In this work, a lightweight CNN architecture is proposed to perform respiratory disease classification utilizing the ICBHI 2017 scientific challenge respiratory sound database [46] while maintaining patient independent train-validation-test dataset splits. A hybrid approach for obtaining scalograms from respiratory sound signals is presented wherein CWT is performed only on the maximally correlated intrinsic mode function (IMF) obtained from the EMD of respiratory sound signals. The class discrimination capability of a hybrid scalogram is evaluated with respect to the CWT-based conventional scalogram. Subsequently, along with the proposed CNN model, complex CNN models such as VGG16 [47], AlexNet [48] and several contemporary lightweight architectures including MobileNet V2 [49], NASNet [50] and ShuffleNet V2 [51] are used for the classification of respiratory diseases. A comparative study among the proposed CNN model and existing models is presented in terms of classification performance and model size.

The rest of the paper is organized as follows. Previous studies related to lung sound classification using different AI-based approaches are discussed in Section II. Section III describes the dataset, feature extraction process, and the proposed lightweight CNN model. The experimental setup and results are discussed in Section V. The performance of the proposed method is compared with existing methods in Section VI, and finally, the paper is concluded in Section VII.

II. Related Work

A substantial amount of previous research employing machine learning and deep learning have been reported on automated respiratory sound classification. However, the majority of the methods have focused on respiratory anomaly prediction, i.e., classifying the lung sounds as wheeze, crackles [10]–[23], [30]–[35], in contrast to directly predicting respiratory diseases from lung auscultation recordings. The few recent approaches dealing with pathology classification mostly involve elaborate signal processing or dedicated CNN and RNN networks [36]–[41]. However, at pathology-level, so far, the classification task has been investigated at three different resolutions; the binary classification (healthy, pathological) [36], [37], the three-class chronic classification (healthy, chronic disease, non-chronic disease) [37], [41] and multi-class distinct disease classification [38], [41]. Among the diseases, Upper and Lower Respiratory Tract Infection (URTI and LRTI), bronchiolitis, and pneumonia have been included in the non-chronic disease class while COPD, asthma, and bronchiectasis have been combined to form the chronic class [37].

In [36], a novel CNN based ternary (three-class) classification approach has been implemented and performed considerably well with 82% accuracy and 88% ICBHI score. Later, the same authors proposed a Mel-Frequency Cepstral Coefficient (MFCC) and Long Short-term Memory (LSTM) based framework capable of conducting both binary and ternary classification of respiratory diseases [37] which demonstrated excellent performance with 99% and 98% accuracy, respectively. A separate work involving a complex RNN architecture and extensive pre-processing has reported an accuracy of 95.67% in a six-class pathology classification [38]. However, by employing a CRNN network with a CNN-Mixture-of-Experts (MoE) baseline to learn both spatial and time-sequential features from the spectrograms, another recent work has reported a specificity of 83% and a sensitivity of 96% in the three-class respiratory disease classification [39]. For binary classification, the authors reported specificity and sensitivity values of 83% and 99%, respectively. As an extension of [39], a separate study involving the robust Teacher-Student learning schemes

with knowledge distillation has been conducted, which resulted in a substantially reduced specificity while maintaining the sensitivity [40].

Since the existing heavily imbalanced datasets of lung auscultations further exacerbate the task of respiratory disease classification, a contemporary study has dealt with this issue by experimenting with several data augmentation techniques, such as SMOTE, Adaptive Synthetic Sampling Method (ADASYN), and Variational autoencoder (VAE) [41]. Among the methods, the VAE-based Mel-spectrogram augmentation strategy, in conjunction with a CNN model, has achieved the best results with 98.5% sensitivity and 99.0% specificity in three-class chronic classification. The strategy has also exhibited an equally sophisticated performance with 98.8% sensitivity and 98.6% specificity in the case of the six-class respiratory disease classification [41].

Although the scope of DL methods with spectrogram-image features has been previously investigated for direct classification of respiratory diseases from lung auscultation sounds [39]–[41], to the best of our knowledge, a scalogram based approach has not been considered in this domain. Additionally, no dedicated lightweight, efficient CNN framework has been developed and investigated for this task. Furthermore, none of the studies consider the issue of intra-patient dependency in the train-validation split. Motivated by the above-mentioned factors, in this work, a scalogram based method using a lightweight CNN model is proposed to predict respiratory diseases from lung auscultations while also maintaining patient independence training and test. The proposed framework is schematically represented in Fig. 1.

III. Data Resources

This work utilizes the International Conference on Biomedical Health Informatics (ICBHI) 2017 dataset that is a publicly available benchmark dataset of lung auscultation sounds [46]. It is collected by two independent research teams of Portugal and Greece. The dataset contains 5.5 hours of audio recordings sampled at different frequencies (4 kHz, 10 kHz, and 44.1 kHz), ranging from 10s to 90s, in 920 audio samples of 126 subjects from different anatomical positions using heterogeneous equipment [52].

The samples are professionally annotated considering two schemes: (i) according to the corresponding patient's pathological condition, i.e., healthy and seven distinct disease classes, namely Pneumonia, Bronchiectasis, COPD, URTI, LRTI, Bronchiolitis, Asthma and (ii) according to the presence of respiratory anomalies, i.e., crackles and wheezes in each respiratory cycle. Further details about the dataset and data collection method can be found in [52].

IV. Proposed Method

A. Data Pre-processing

1) Bandpass filtering: Since lung auscultation signals generally reside in the frequency range 50 – 2500 Hz [7], the audio samples are first processed with a 6th order Butterworth bandpass filter with upper and lower cut-off frequencies of 50 and 2500 Hz, respectively. All the sample audio signals are resampled at 8 kHz [7] to ensure consistency while avoiding the loss of important lung sound components and lowering the computational cost. The signals are also amplitude normalized to reduce the effect of device/sensor variation.

2) Segmentation: Each of the lung sound recordings is segmented according to the annotated respiratory cycle timing. Samples with a minimum respiratory cycle duration of 3s are taken into account to extract useful respiratory sound information [39]. The selected audio samples are converted into homogeneous signals of 6s duration, padding zeros if necessary. If any respiratory cycle is longer than 6s, only the first 6s duration is used. Two of the disease classes, namely asthma and LRTI, are found to have an inadequate number of segmented samples for meaningful classification experiments and thus are not considered for this study. Finally, a total of 87 out of 126 lung auscultation recordings of unique patients are found to be usable. Table I summarizes the data distribution at several levels of processing corresponding to the disease classes considered in this study.

B. Feature Extraction

1) Empirical Mode Decomposition (EMD): EMD is a powerful self-adaptive signal decomposition method highly suitable for analysis and processing of non-linear and non-stationary signals such as lung sounds and heart sounds [53]. It decomposes a given signal $x(t)$ into a finite set (N) of intrinsic mode functions, $\text{IMF}_1(t), \text{IMF}_2(t), ..., \text{IMF}_N(t)$, depending on the local characteristic time scale of the signal, with a view to expressing the original signal as the sum of all its IMF plus a final trend either monotonic or constant called residue, $r(t)$, such that: $x(t) = \sum_{i=1}^{N} \text{IMF}_i(t) + r(t)$ [54]. An IMF is an oscillatory function with an equal number of extrema and zero crossings, while having envelopes symmetrical with respect to zero. Thus, the EMD detrends a signal and elicits its underlying spectral patterns [53].

2) Continuous Wavelet Transform (CWT): Wavelet transform is defined as a signal processing method that can decompose a signal into an orthonormal wavelet basis or into a set of independent frequency channels [15], [28]. Using a basis function, i.e. the mother wavelet $g(t)$, and its scaled and dilated versions, the Continuous Wavelet Transform (CWT) can be used to decompose a finite-energy signal, $x(t)$ as [29]:

$$Z(a,b) = \frac{1}{\sqrt{b}} \int x(t) g\left(\frac{t-a}{b}\right) dt \tag{1}$$

where b and a are the scale and translation factors. Larger scale values reveal low-frequency information while the smaller scale values reveal high-frequency information [28]. The squared-modulus of the CWT coefficients Z is known as the scalogram [15].

C. Scalogram Representations

1) Conventional Scalogram: Scalogram is defined as the time-frequency representation of a signal that depicts the obtained energy density using CWT [5], [55]. The segmented and filtered lung sound samples are decomposed into the wavelet domain using Morse analytic wavelet as the mother-wavelet

TABLE I
DISTRIBUTION OF DATA AT DIFFERENT LEVELS OF PROCESSING CORRESPONDING TO THE DISEASE CLASSES

Disease class	# Unsegmented audio files	# Segmented and filtered samples	# Unique patients	# Augmented images in training	# Augmented images in validation	# Images in test
Bronchiectasis	16	55	6	152	24	11
Bronchiolitis	13	65	6	160	20	20
COPD	793	1,963	51	1,363	196	404
Healthy	35	42	13	92	16	15
Pneumonia	37	41	3	116	20	7
URTI	23	21	8	52	8	6
Total	917	2,187	87	1,953	284	463

with the time-bandwidth product and symmetry parameter set to 60 and 3, respectively. According to the range of energy of the wavelet in frequency and time the minimum and maximum scales are automatically determined using 10 voices per octave [56]. Points out of the cone of influence have been handled by the approximation [57] used in MathWorks MATLAB *cwt* function. Scalogram plots are generated with a resolution of 224×224 using these coefficients. Fig. 2 shows the scalograms of lung sounds in different disease categories.

2) Hybrid Approach for Scalogram: For each segmented and filtered sample under each pathological class, 9 IMFs are generated using the *emd* function in MATLAB 2020a. Based on the cross-correlation between the source signal and the IMFs, the most physically significant IMF output with the highest correlation coefficient is determined [58], [59]. The reason for using 9 IMFs is that in most cases the 4th or 5th IMF shows the maximal correlation for the respiratory sound signals. As a precaution, the correlation process was continued up to the 9th IMF. Subsequently, the squared-modulus of the CWT of the corresponding IMF is calculated to obtain the scalogram.

The diverse frequency bands varying from the maximum to the minimum range give the IMFs the capability to extract the temporal and spectral information [54] effectively. Hence, when this IMF based scheme is combined with CWT-oriented scalogram representation, the newly formed hybrid scalograms can be expected to yield more discriminative and significant features. Thus, it has the potential to provide better classification performance by a CNN model. The box plots of the scalograms of lung sounds for various respiratory diseases are shown in Fig. 3. The distinction among the plots is more evident when using the hybrid approach than those of the conventional scalograms obtained using only CWT.

It should be mentioned that the proposed scalogram is distinctly different from that of [5], [59] in that the CWT modulus is computed from the maximally correlated IMFs and thus, providing a better representation of the underlying information. Note that the works of [5] and [59] are on detecting respiratory anomalies such as crackle and wheeze, and analysis and segmentation of heart sounds, respectively, whereas our objective is to detect respiratory diseases from the lung auscultation sounds.

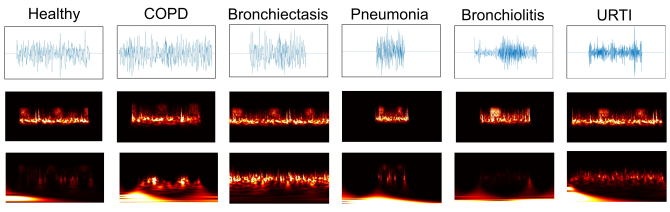

Fig. 2. **Scalograms of the lung auscultation sounds for 6 disease classes; lung sound recordings (1st row), conventional scalogram (2nd row) and scalogram using the proposed hybrid approach (3rd row).**

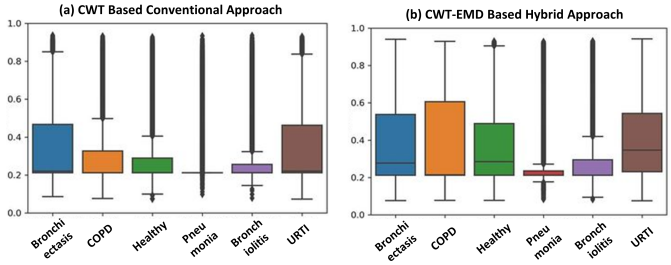

Fig. 3. **Box plot. (a) Scalogram using the conventional CWT approach; (b) Scalogram using the hybrid approach.**

D. Augmentation

The ICBHI 2017 dataset is highly imbalanced, with around 86% of the data belonging to COPD. Image augmentation using different color mapping schemes is employed to oversample the less represented classes and address the data imbalance issue [60]. Colormaps are three-column arrays containing RGB triplets where each row defines a distinct color. Scalogram representation using different color maps helps generalize the produced images.

From each of the audio samples of the less represented data classes, four scalograms are generated for each segmented sample using four different color mapping schemes: Bone, Copper, Winter, and Hot, which are predefined in MATLAB 2020a while for the most represented class, COPD, only one image is produced from each audio sample. Nevertheless, for ensuring generalization and homogeneity of the augmented data, all four-color mapping schemes are randomly utilized for COPD. Since appropriate colormap selection is a complex issue [61], the four colormaps that have been utilized here are selected after extensive experimentation for achieving both the purposes of data oversampling and data augmentation with a single action. A summary of segmented audio files and final augmented scalogram images with corresponding diseases classes are presented in Table I.

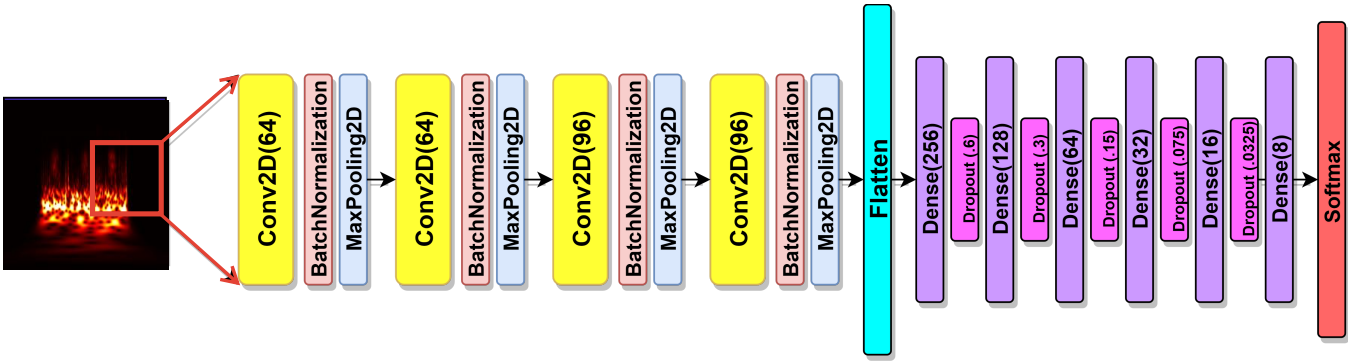

Fig. 4. **The detailed architecture of the proposed lightweight CNN model.**

V. Proposed Lightweight CNN Architecture

Various CNN architectures are most commonly utilized for classifying images while recent methods have considered CNN models for classifying 2D images produced from audio signals [5], [30], [32]. However, due to memory constraints, a regular deep CNN model is computationally expensive with its large number of learnable parameters and arithmetic operations. Thus, it is not suitable for embedded devices as they cannot afford the processing complexity and storage space for parameters and weight values of filters [35]. Cloud computing methodology requires a higher RAM for this computationally intensive training and hence are outsourced [62]. For this reason, lightweight CNN models are gaining popularity among researchers for their faster performance and compact size without compromising the much-needed accuracy performance compared to the well-known deep learning networks [63].

The architecture of the proposed CNN model consists of an input layer corresponding to the 3-channel input of 224×224 images. The architecture of the proposed model is illustrated in Fig. 4. The 1st convolutional (Conv) layer uses 64 output filters with a 5×5-pixel kernel followed by a 2×2-pixel max-pooling layer. Three additional Conv blocks are stacked over the first block, each having a 3×3-pixel kernel with 64, 96, and 96 filters sequentially with corresponding batch-normalization and max-pooling layers with 2×2 pooling window. Outputs from all these layers are flattened and connected with five pairs of fully connected and dropout layers, followed by a SoftMax output layer with probability nodes for each class. Rectified linear unit (ReLU) activation is used for the Conv blocks and fully connected layers. Max pooling is used after the ReLU activation, which reduces the spatial dimensionality of the extracted feature maps. It also extracts the most important features and is unaffected from locational bias [34].

The batch normalization (BN) operation normalizes the extracted features in each layer and reduces the problem of diversity in data variance. This provides the network a representative power with a small number of parameters and faster training capability.

VI. Experimental Results

A. Evaluation Criteria

The segmented audio files are split into three sets with the approximate proportion of $70 : 10 : 20$ maintaining patient independence, and they are respectively grouped for creating the training, validation, and testing set as shown in Table I. For training and validation, scalograms were augmented while the test set contains only a single scalogram for each segmented and filtered audio sample. Patient uniqueness, a critical aspect in the real-world applications, is maintained while dividing into training and validation parts as speaker dependency results in biased accuracy [45].

The performance of the classifier model is evaluated based on the evaluation matrices, accuracy, recall (sensitivity), precision and F1-score. Additionally, specificity and ICBHI-score [37], [52], a dedicated metric involving both sensitivity and specificity to assess the performance of the frameworks using the ICBHI dataset, is used to evaluate the performance of our method.

B. Experimental Setup

The proposed CNN model is constructed using Keras and TensorFlow backend, and trained using NVidia K80 GPUs provided by Kaggle notebooks. The balanced mini-batch training scheme is employed while feeding the image data into a model for tackling the class imbalance issue. This technique performs oversampling of the minority classes while randomly undersampling a majority class. This strategy ensures that the CNN model takes an equal number of samples from each class during each of the training epochs and thereby forms a balanced training set [25].

The adaptive learning rate optimizer (Adam) with a learning rate of 0.00001 and categorical cross-entropy as a cost function are used for training the model. The batch size needs to be a multiple of 6 since an equal number of samples from each of the classes (3 or 6 depending on the experiment) needs to be included in each mini-batch for balancing [25]. In this study, a batch size of 6 has been selected for both of the classification schemes.

As stated earlier, both the 3-class chronic classification (chronic, non-chronic, healthy) and six class (Bronchiectasis, Bronchiolitis, COPD, Healthy, Pneumonia, and URTI) pathological classification are carried out in this work. The classification performance of the proposed CNN model is compared with that of VGG16, a well-known CNN architecture commonly used for image classification [47]. It should be

noted that the experiments are performed using both the conventional CWT-based scalogram and hybrid scalogram images. In addition, the performance of our proposed CNN model is compared with a number of well-known DL architectures such as AlexNet and several other lightweight networks in terms of computational complexity and accuracy.

C. Sensitivity Analysis of the Proposed Network

The parameters used in the proposed lightweight model are determined after performing rigorous experimental evaluation on the 6-class pathological classification scheme. The experiments were carried out considering two aspects i.e., optimization of the number of filters in the Conv block and optimization of the depth of the Conv block.

1) No. of Filters in Conv block optimization: Keeping other parameters unchanged, 5 different combinations i.e., $(32, 32, 64, 64)$, $(64, 64, 128, 128)$, $(64, 64, 96, 96)$, $(32, 32, 96, 96)$ and $(128, 128, 256, 256)$ have been experimented for evaluating the optimum no. of filters in the Conv blocks. As shown in Table II, the proposed network with 4 convolutional blocks yields the best result on the validation set.

TABLE II
SENSITIVITY OF PARAMETER: NO. OF FILTER IN CONV BLOCK

No. of filters in the Conv block	Parameters	Accuracy
$32, 32, 64, 64$	$2,471,374$	97.588%
$64, 64, 128, 128$	$5,027,566$	97.807%
$64, 64, 96, 96$ **(Proposed)**	$3,764,654$	$\mathbf{98.026}\%$
$32, 32, 96, 96$	$3,706,638$	97.800%
$128, 128, 256, 256$	$10,527,022$	78.509%

TABLE III
SENSITIVITY OF PARAMETER: CONV BLOCK DEPTH

No. of Conv blocks	Model Description (No. of filters)	Accuracy
2	$64, 96$	93.640%
3	$64, 64, 96$	97.588%
4 **(Proposed)**	$64, 64, 96, 96$	$\mathbf{98.026}\%$
5	$64, 64, 64, 96, 96$	97.680%
6	$64, 64, 64, 96, 96, 96$	97.807%

2) Conv Block Depth Optimization: Keeping other parameters unaltered, 5 different combinations i.e., 2, 3, 4, 5 and 6 Conv blocks have been experimented for identifying the optimum no. of Conv blocks. As depicted in Table III, the proposed network with 4 convolution blocks has reported the best results.

D. Classification Performance of the Proposed Framework

1) Chronic Classification: From Table IV, it can be seen that using the hybrid scalogram method in conjunction with the proposed CNN model shows the best accuracy of 98.92%. The corresponding accuracy obtained by using VGG16 is similar (97.84%). Despite being a heavy model, the slightly lower accuracy of VGG16 can be attributed to the overfitting issue due to the limited number of samples in different classes. When comparing conventional CWT scalogram to the proposed hybrid scalogram, considerable improvement in accuracy is obtained for the latter using VGG16 and our proposed CNN model (10.89%-12.46%). The corresponding confusion matrices for both the models' best results are illustrated in Fig. 5. The results depict that the proposed method is better in the 3-class chronic classification than the VGG16 model.

2) Pathological Classification: For the six-class pathological classification task, the proposed method involving the hybrid scalogram and CNN model yields the best accuracy, 98.70%, as seen in Table IV. Similar to the case in the 3-class chronic classification scheme, the accuracy of VGG16 is slightly lower. However, since the dataset is more segregated, being divided into six different disease classes, the accuracy drop is larger. Nevertheless, the proposed hybrid scalogram outperforms the conventional CWT scalogram with a larger margin (13.17%-13.77%) for both VGG16 and our proposed model. The overall superiority of the proposed method is evident from the confusion matrices shown in Fig. 6.

Since the ultimate goal of the proposed framework is to determine if a given person is afflicted by a particular respiratory condition or not, the pathological classification performance of the segmented breath sounds was further investigated to present an additional classification result per patient. Among the six classes, in Bronchiectasis 2 patients out of 2, in Bronchiolitis 2 patients out of 2, in COPD 9 patients out of 9, in Healthy 4 subjects out of 4, in Pneumonia 1 patient out of 1, and in URTI 2 patients out of 4 have been correctly identified based on the predicted classification of their segmented breath sounds by the proposed model.

E. Comparison with Existing Methods

1) Respiratory Disease Classification: As discussed in Section II, none of the existing methods for respiratory disease classification explore the domain of patient-specific prediction. Some of the studies address the issues regarding class imbalance [38], [41]. Nevertheless, the extensive pre-processing, coupled with the ambiguous undersampling of the COPD disease class while oversampling all other disease classes, can complicate the reproducibility of [38]. Furthermore, in [41], Fast Fourier Transform (FFT) is applied to the entire respiratory sound signals, whereas our work focuses on segmented breath sounds. In our work, complete patient independence has been maintained in the train, validation and test set, which is not possible while using the entire lung auscultation signal due to the low number of samples. Therefore, our work aims to overcome the limitations present in the existing methods. A comparison among the various methods, including the Proposed method, is provided in Table V.

It is observed that our proposed CNN model with the hybrid scalogram can perform on par with the existing state-of-the-art CNN and RNN models for classification tasks while maintaining a patient independent train-validation scheme.

TABLE IV
SUMMARY OF THE CLASSIFICATION PERFORMANCE. ALL RESULTS ARE SHOWN IN PERCENTAGE (%)
THE RED AND BLUE MARKED VALUES REPRESENT THE HIGHEST ACCURACY OBTAINED WITH OUR PROPOSED CNN MODEL AND VGG16

Network	Chronic Classification								Pathological Classification							
	Scalogram using CWT				Scalogram using EMD and CWT				Scalogram using CWT				Scalogram using EMD and CWT			
	Prec.	Recall	Acc.	F1	Prec.	Recall	Acc.	F1	Prec.	Recall	Acc.	F1	Prec.	Recall	Acc.	F1
Proposed	88.00	88.50	89.20	88.25	98.90	98.90	98.92	98.90	85.30	86.61	87.21	86.00	98.68	98.27	98.70	98.47
VGG16	87.00	87.00	87.05	87.00	97.95	97.83	97.84	97.89	85.00	85.00	85.80	85.00	97.60	97.86	97.62	97.01

TABLE V
COMPARISON OF THE PROPOSED FRAMEWORK WITH EXISTING METHODS USING THE ICBHI 2017 DATASET

Pre-processing	Classifier Network(s)	Number of Classes Considered	Acc.	Spec.	Sen.	ICBHI Score
Gammatone Spectrogram [39], [40]	C-RNN	3 (Healthy, Chronic, Non-chronic)	-	0.570	0.940	0.760
	CNN-MoE		-	0.860	0.960	0.910
	Ensemble		-	0.710	0.950	0.830
MFCC [37]	CNN	3 (Healthy, Chronic, Non-chronic)	0.820	0.760	0.890	0.830
	LSTM		0.980	0.820	0.980	0.900
MFCC [38]	RNN	6 classes (excluding Asthma, LRTI)	0.957	-	0.957	-
Mel-spectrogram- VAE [41]	CNN	3 (Healthy, Chronic, Non-chronic)	0.990	0.990	0.985	0.988
		6 classes (excluding Asthma, LRTI)	0.990	0.986	0.988	0.987
Hybrid Scalogram (Proposed)	Lightweight CNN	3 (Healthy, Chronic, Non-chronic)	**0.989**	1.00	0.989	0.994
		6 classes (excluding Asthma, LRTI)	**0.987**	1.00	0.986	0.993

2) Computational Performance as a Lightweight Network: A detailed comparison is presented in Table VI among VGG16 [47], our proposed CNN model, AlexNet [48] and the existing state-of-the-art lightweight models such as MobileNetV2 [49], NASNet [50], ShuffleNetV2 [51] in terms of size, trainable parameters, the number of operations measured by multiply-add (MAdd) and accuracy on both chronic and pathological approach. In terms of accuracy, our proposed CNN model shows better results than VGG16 while requiring only 3% of the parameters. The proposed CNN model also outperforms the contemporary lightweight models, ShuffleNet V2, MobileNet V2, and NASNet relatively by 0.43%, 0.02%, and 1.14% in chronic classification, respectively, while obtaining better trade-off between the number of parameters, requiring significantly lower storage space and computational power. This makes our proposed lightweight model more suitable for real-time wearable devices with faster and less resource-intensive training. We have calculated the time required for end-to-end classification of an auscultation sound using the proposed framework. For this experiment, we only performed the pre-processing and inference step using all of our test data and calculated the mean and standard deviation of the required CPU time. We found that the pre-processing time for EMD+CWT is $7.5s \pm 0.5s$., and only CWT is $6.8s \pm 0.5s$. These algorithms are run on a Core i7 7500 processor with clock speed specifications of 2.70-2.90GHz. Time required for the classification of a scalogram image using the proposed network is $0.07s \pm 0.01s$, while the MobileNetV2 takes $0.085s \pm 0.01s$. Thus, the proposed CNN architecture is faster in classifying a sound image as compared to MobileNetV2.

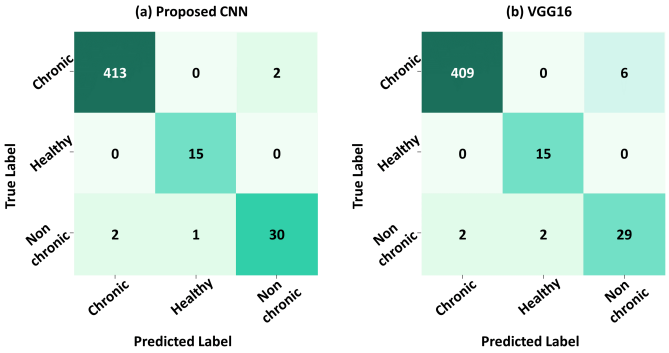

Fig. 5. **Confusion matrices for the best results obtained in 3-class chronic classification. (a) Proposed CNN model with batch size 6; (b) VGG16 with batch size 6.**

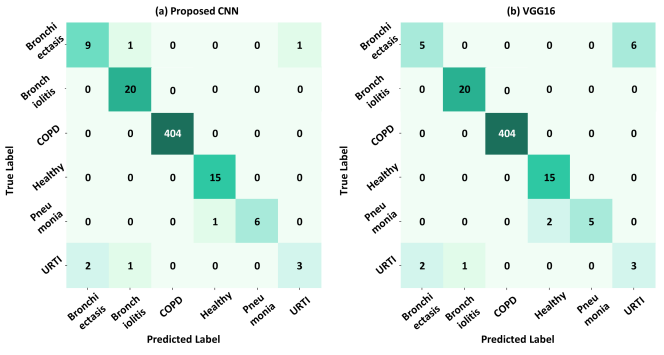

Fig. 6. **Confusion matrices for the best results obtained in six-class Pathological classification. (a) Proposed CNN model with batch size 6; (b) VGG16 with batch size 6.**

TABLE VI
COMPARISON OF MODELS WITH RESPECT TO SIZE AND PERFORMANCE

Parameter	Network					
	VGG16	AlexNet	**Proposed model**	Mobile-Net (V2)	Shuffle-Net (V2)	NASNet
Model size	1.5GB	294MB	44.85MB	49MB	46.9MB	64MB
Trainable parameters	138M	25.704M	3.7674M	4.2M	5.4M	4.2M
MAdd	154.7G	725M	371.93M	575M	564M	567M
Accuracy (6-class)	97.60%	98.237%	98.70%	98.68%	98.27%	97.58%
Accuracy (3-class)	97.60%	99.519%	98.92%	98.47%	99.06%	98.03%

VII. CONCLUSION

In this work, we have proposed a lightweight CNN model to classify respiratory diseases using scalogram images of lung sounds. A hybrid approach employing both EMD and CWT is presented to generate the scalogram images. The publicly available ICBHI 2017 challenge dataset has been used for a 3-class and a 6-class classification of respiratory diseases. The proposed method has provided a considerable accuracy of 98.92% for the 3-class chronic classification task. In pathological classification among six disease classes, an accuracy of 98.70% has been achieved. The obtained accuracies were found to be higher than the much larger VGG16 model. In addition, for classification tasks, the proposed framework has provided better or a comparable performance with respect to the existing state-of-the-art methods in terms of precision, recall, F1-score, sensitivity, specificity, and ICBHI score. It is worthwhile to mention that unlike most of these methods, the classification performance of the proposed technique has been assessed, keeping the training and testing data-independent in terms of patients. The computational complexity of the proposed classifier has also been compared with a number of well-known CNN models and state-of-the-art lightweight networks. The proposed network has been shown to achieve high accuracy in classification while being a deep lightweight architecture. We believe that these attributes can enable the development of the automatic classification of respiratory diseases from lung auscultations in real-world clinical applications.

REFERENCES

[1] "The Global Impact of Respiratory Disease – Second Edition — CHEST Physician," 2017, [Online]. Available: https://www.mdedge.com/chestphysician/article/140055/society-news/global-impact-respiratory-disease-second-edition.

[2] A. A. Cruz, *Global surveillance, prevention and control of chronic respiratory diseases: a comprehensive approach*. World Health Organization, 2007.

[3] C. D. Mathers and D. Loncar, "Projections of global mortality and burden of disease from 2002 to 2030," *PLoS medicine*, vol. 3, no. 11, p. e442, 2006.

[4] "WHO — Global tuberculosis report 2019," 2019, [Online]. Available: https://www.who.int/tb/publications/global_report/en/.

[5] S. Jayalakshmy and G. F. Sudha, "Scalogram based prediction model for respiratory disorders using optimized convolutional neural networks," *Artificial Intelligence in Medicine*, vol. 103, p. 101809, 2020.

[6] A. Abbas and A. Fahim, "An automated computerized auscultation and diagnostic system for pulmonary diseases," *Journal of Medical Systems*, vol. 34, no. 6, pp. 1149–1155, 2010.

[7] S. Reichert, R. Gass, C. Brandt, and E. Andrès, "Analysis of respiratory sounds: state of the art," *Clinical Medicine. Circulatory, Respiratory and Pulmonary Medicine*, vol. 2, pp. CCRPM–S530, 2008.

[8] M. Sarkar, I. Madabhavi, N. Niranjan, and M. Dogra, "Auscultation of the respiratory system," *Annals of Thoracic Medicine*, vol. 10, no. 3, p. 158, 2015.

[9] A. Bohadana, G. Izbicki, and S. S. Kraman, "Fundamentals of lung auscultation," *New England Journal of Medicine*, vol. 370, no. 8, pp. 744–751, 2014.

[10] M. Bahoura and C. Pelletier, "New parameters for respiratory sound classification," in *Canadian Conference on Electrical and Computer Engineering*, vol. 3. IEEE, 2003, pp. 1457–1460.

[11] R. Palaniappan, K. Sundaraj, and N. U. Ahamed, "Machine learning in lung sound analysis: a systematic review," *Biocybernetics and Biomedical Engineering*, vol. 33, no. 3, pp. 129–135, 2013.

[12] J. Zhang, W. Ser, J. Yu, and T. Zhang, "A novel wheeze detection method for wearable monitoring systems," in *2009 International Symposium on Intelligent Ubiquitous Computing and Education*. IEEE, 2009, pp. 331–334.

[13] M. Bahoura, "Pattern recognition methods applied to respiratory sounds classification into normal and wheeze classes," *Computers in Biology and Medicine*, vol. 39, no. 9, pp. 824–843, 2009.

[14] J. Acharya, A. Basu, and W. Ser, "Feature extraction techniques for low-power ambulatory wheeze detection wearables," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2017, pp. 4574–4577.

[15] N. Gautam and S. B. Pokle, "Wavelet scalogram analysis of phonopulmonographic signals," *International Journal of Medical Engineering and Informatics*, vol. 5, no. 3, pp. 245–252, 2013.

[16] G. Serbes, C. O. Sakar, Y. P. Kahya, and N. Aydin, "Pulmonary crackle detection using time–frequency and time–scale analysis," *Digital Signal Processing*, vol. 23, no. 3, pp. 1012–1021, 2013.

[17] S. İçer and Ş. Gengeç, "Classification and analysis of non-stationary characteristics of crackle and rhonchus lung adventitious sounds," *Digital Signal Processing*, vol. 28, pp. 18–27, 2014.

[18] F. Jin, F. Sattar, and D. Y. Goh, "New approaches for spectro-temporal feature extraction with applications to respiratory sound classification," *Neurocomputing*, vol. 123, pp. 362–371, 2014.

[19] P. Bokov, B. Mahut, P. Flaud, and C. Delclaux, "Wheezing recognition algorithm using recordings of respiratory sounds at the mouth in a pediatric population," *Computers in Biology and Medicine*, vol. 70, pp. 40–50, 2016.

[20] P. Mayorga, C. Druzgalski, R. Morelos, O. Gonzalez, and J. Vidales, "Acoustics based assessment of respiratory diseases using gmm classification," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, 2010, pp. 6312–6316.

[21] T. R. Fenton, H. Pasterkamp, A. Tal, and V. Chernick, "Automated spectral characterization of wheezing in asthmatic children," *IEEE Transactions on Biomedical Engineering*, vol. 32, no. 1, pp. 50–55, 1985.

[22] H. Pasterkamp, S. S. Kraman, and G. R. Wodicka, "Respiratory sounds: advances beyond the stethoscope," *American Journal of Respiratory and Critical Care Medicine*, vol. 156, no. 3, pp. 974–987, 1997.

[23] Z. Dokur, "Respiratory sound classification by using an incremental supervised neural network," *Pattern Analysis and Applications*, vol. 12, no. 4, p. 309, 2009.

[24] B. Bozkurt, I. Germanakis, and Y. Stylianou, "A study of time-frequency features for cnn-based automatic heart sound classification for pathology detection," *Computers in Biology and Medicine*, vol. 100, pp. 132–143, 2018.

[25] A. I. Humayun, S. Ghaffarzadegan, M. I. Ansari, Z. Feng, and T. Hasan, "Towards domain invariant heart sound abnormality detection using learnable filterbanks," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 8, pp. 2189–2198, 2020.

[26] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals," *Computers in Biology and Medicine*, vol. 100, pp. 270–278, 2018.

[27] S. Hershey, S. Chaudhuri, D. P. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, *et al.*, "Cnn architectures for large-scale audio classification," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 131–135.

[28] S. Debbal and F. Bereksi-Reguig, "Analysis of the second heart sound using continuous wavelet transform," *Journal of Medical Engineering & Technology*, vol. 28, no. 4, pp. 151–156, 2004.

[29] A. Meintjes, A. Lowe, and M. Legget, "Fundamental heart sound classification using the continuous wavelet transform and convolutional neural networks," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 409–412.
[30] K. Minami, H. Lu, H. Kim, S. Mabu, Y. Hirano, and S. Kido, "Automatic classification of large-scale respiratory sound dataset based on convolutional neural network," in *2019 19th International Conference on Control, Automation and Systems (ICCAS)*. IEEE, 2019, pp. 804–807.
[31] M. Aykanat, Ö. Kılıç, B. Kurt, and S. Saryal, "Classification of lung sounds using convolutional neural networks," *EURASIP Journal on Image and Video Processing*, vol. 2017, no. 1, p. 65, 2017.
[32] F. Demir, A. Sengur, and V. Bajaj, "Convolutional neural networks based efficient approach for classification of lung diseases," *Health Information Science and Systems*, vol. 8, no. 1, p. 4, 2020.
[33] R. Liu, S. Cai, K. Zhang, and N. Hu, "Detection of adventitious respiratory sounds based on convolutional neural network," in *2019 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*. IEEE, 2019, pp. 298–303.
[34] D. Bardou, K. Zhang, and S. M. Ahmad, "Lung sounds classification using convolutional neural networks," *Artificial Intelligence in Medicine*, vol. 88, pp. 58–69, 2018.
[35] J. Acharya and A. Basu, "Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 14, no. 3, pp. 535–544, 2020.
[36] D. Perna, "Convolutional neural networks learning from respiratory data," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2018, pp. 2109–2113.
[37] D. Perna and A. Tagarelli, "Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks," in *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2019, pp. 50–55.
[38] V. Basu and S. Rana, "Respiratory diseases recognition through respiratory sound with the help of deep neural network," in *2020 4th International Conference on Computational Intelligence and Networks (CINE)*. IEEE, 2020, pp. 1–6.
[39] L. Pham, I. McLoughlin, H. Phan, M. Tran, T. Nguyen, and R. Palaniappan, "Robust deep learning framework for predicting respiratory anomalies and diseases," *arXiv preprint arXiv:2002.03894*, 2020.
[40] L. Pham, "Predicting respiratory anomalies and diseases using deep learning models," *arXiv preprint arXiv:2004.04072*, 2020.
[41] M. T. García-Ordás, J. A. Benítez-Andrades, I. García-Rodríguez, C. Benavides, and H. Alaiz-Moretón, "Detecting respiratory pathologies using convolutional neural networks and variational autoencoders for unbalancing data," *Sensors*, vol. 20, no. 4, p. 1214, 2020.
[42] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.
[43] I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio, "Quantized neural networks: Training neural networks with low precision weights and activations," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 6869–6898, 2017.
[44] S. Kiranyaz, T. Ince, R. Hamila, and M. Gabbouj, "Convolutional neural networks for patient-specific ecg classification," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 2608–2611.
[45] N. U. Maheswari, A. Kabilan, and R. Venkatesh, "Speaker independent speech recognition system based on phoneme identification," in *2008 International Conference on Computing, Communication and Networking*. IEEE, 2008, pp. 1–6.
[46] "ICBHI 2017 Challenge," 2017, [Online]. Available: https://bhichallenge.med.auth.gr/.
[47] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
[48] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.
[49] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.
[50] X. Qin and Z. Wang, "Nasnet: A neuron attention stage-by-stage net for single image deraining," *arXiv preprint arXiv:1912.03151*, 2019.
[51] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in *European conference on Computer Vision (ECCV)*, 2018, pp. 116–131.
[52] B. Rocha *et al.*, "A respiratory sound database for the development of automated classification," in *International Conference on Biomedical and Health Informatics*. Springer, 2017, pp. 33–37.
[53] N. Ibtehaz, M. S. Rahman, and M. S. Rahman, "Vfpred: A fusion of signal processing and machine learning techniques in detecting ventricular fibrillation from ecg signals," *Biomedical Signal Processing and Control*, vol. 49, pp. 349–359, 2019.
[54] M. Altuve, L. Suárez, and J. Ardila, "Fundamental heart sounds analysis using improved complete ensemble emd with adaptive noise," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 426–439, 2020.
[55] Z. Ren, K. Qian, Y. Wang, Z. Zhang, V. Pandit, A. Baird, and B. Schuller, "Deep scalogram representations for acoustic scene classification," *IEEE/CAA Journal of Automatica Sinica*, vol. 5, no. 3, pp. 662–669, 2018.
[56] "Continuous 1-D wavelet transform - MATLAB cwt," [Online]. Available: https://www.mathworks.com/help/wavelet/ref/cwt.html.
[57] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of the American Meteorological society*, vol. 79, no. 1, pp. 61–78, 1998.
[58] R. Fontugne, J. Ortiz, D. Culler, and H. Esaki, "Empirical mode decomposition for intrinsic-relationship extraction in large sensor deployments," in *Workshop on Internet of Things Applications, IoT-App*, vol. 12, 2012.
[59] D. Boutana, M. Benidir, and B. Barkat, "Segmentation and time-frequency analysis of pathological heart sound signals using the emd method," in *2014 22nd European Signal Processing Conference (EUSIPCO)*. IEEE, 2014, pp. 1437–1441.
[60] F. Y. Shih and H. Patel, "Deep learning classification on optical coherence tomography retina images," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 34, no. 08, p. 2052002, 2019.
[61] L. Zhou and C. D. Hansen, "A survey of colormaps in visualization," *IEEE transactions on visualization and computer graphics*, vol. 22, no. 8, pp. 2051–2069, 2015.
[62] S. Y. Nikouei, Y. Chen, S. Song, R. Xu, B.-Y. Choi, and T. R. Faughnan, "Real-time human detection as an edge service enabled by a lightweight cnn," in *2018 IEEE International Conference on Edge Computing (EDGE)*. IEEE, 2018, pp. 125–129.
[63] B. Lim, B. Yang, and H. Kim, "Real-time lightweight cnn for detecting road object of various size," in *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE, 2018, pp. 202–203.