INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

AsthmaSCELNet: A Lightweight Supervised Contrastive Embedding Learning Framework for Asthma Classification Using Lung Sounds

Arka Roy, Udit Satija

Department of Electrical Engineering, Indian Institute of Technology Patna, Bihar-801106, India
arka_2121ee34@iitp.ac.in, udit@iitp.ac.in

Abstract

Asthma is one of the most prevalent respiratory disorders, which can be identified by different modalities such as speech, wheezing of lung sounds (LSs), spirometric measures, etc. In this paper, we propose AsthmaSCELNet, a lightweight supervised contrastive embedding learning framework, to classify asthmatic LSs by providing adequate classification margin across the embeddings of healthy and asthma LS, in contrast to vanilla supervised learning. Our proposed framework consists of three steps: pre-processing, melspectrogram extraction, and classification. The AsthmaSCELNet consists of two stages: embedding learning using a lightweight embedding extraction backbone module that extracts compact embedding from the melspectrogram, and classification by the learnt embeddings using multi-layer perceptrons. The proposed framework achieves an accuracy, sensitivity, and specificity of 98.54%, 98.27%, and 98.73% respectively, that outperforms existing methods based on LSs and other modalities.

Index Terms: lung sounds, wheeze, asthma classification.

1. Introduction

Asthma is one of the severe chronic respiratory diseases and has been listed as one of the members of the "Big Five Respiratory Diseases" by the world health organization [1]. According to the global asthma report, more than 339 million (M) individuals have been affected from asthma worldwide. Spirometry-based measurements are often used to diagnose and monitor the asthmatic condition [2], which evaluates how quickly and how much air a person can exhale. In spirometry test, a patient has to inhale deeply and exhale forcefully into a mouthpiece with clipped nose. To evaluate the level of severity in asthma, several spirometric measures are used, such as forced expiration capacity, forced expiration volume of 1 second (FEV1), and the ratio of both of these quantities [2]. However, spirometry is highly dependent on patient efforts as it is a highly laborious procedure, especially for elders and children [3]. One of the techniques for asthma monitoring is peak flow meter [4] which measures the airflow rate of major airways through which air reaches to lungs. However, the main drawback of this modality is its inability to measure the airflow rate through the smaller airway paths which gets affected by asthma [4]. Hence, wheezing events of LSs can be exploited for asthma detection [5], as these sounds are associated with many structural faults that occur in the lungs as a result of respiratory diseases [6], [7]. Thereby, developing artificial intelligence (AI)-based automated algorithms using LSs will be extremely beneficial in the detection of asthma.

In recent years, several medical diagnostic modalities have been used to identify asthma, such as Yadav et al. [8] used pathological speech signals from healthy and asthma subjects. They used mel-frequency cepstral coefficient (MFCC) features and support vector machine (SVM) based machine learning classifier and achieved an accuracy of 77.8% in classifying asthma patients. Altan et al. [9] proposed an asthma classification framework with 84.61% accuracy rate using LS recordings. To identify the asthma, the LSs were segmented into 15-second frames and passed through a 1st-order high pass filter, then decomposed using the Hilbert Huang transform (HHT), followed by statistical feature extraction, and classification using a deep belief network (DBN). Tripathy et al. [10] recently developed an asthmatic LS classification system based on empirical wavelet transform (EWT) and feature extraction, followed by a variety of machine learning (ML) classifiers such as SVM, random forest, K-nearest neighbor (KNN), etc., and achieved a classification accuracy of 80.35% [10]. Existing methods uses traditional ML classifiers with handcrafted features which fails to derive the accurate representation of the highly varying time-frequency content of LSs,leading to poor classification performance. Therefore, there is a need to develop a novel deep learning network that can provide accurate distinct feature representation from LSs and achieve higher classification rate.

In this paper, we first introduce a melspectrogram time-frequency representation driven supervised contrastive embedding learning (SCEL) framework for asthma classification based on LSs, that mitigates two main drawbacks of the traditional cross-entropy loss-based supervised deep learning training procedure: inadequate classification margins across samples of various classes [11], and susceptibility to noisy labels [12], by focusing on inter-class dissimilarity and intra-class similarity via triplet loss-based contrastive learning mechanism. To the best of our knowledge, this is the first deep learning-based framework for asthma classification using LSs. The major contributions of the paper are summarized as follows:

- Investigating the potential of mel-spectrogram representation for the first time in asthma classification.
- Designing a novel contrastive triplet loss-based SCEL framework to provide better classification margin across the healthy and asthmatic LS class by surpassing the vanilla supervised learning methods.
- Designing a lightweight embedding extraction backbone (LEEB) to extract compact embedding representation from LS by exploiting the paradigm of lightweight neural network architecture that reduces the number of trainable parameters.
- Extensive evaluation of the proposed framework using the publicly available database through several performance parameters.

The rest of the paper is organized as: Section 2 describes the publicly available database, and Section 3 includes a detailed

10.21437/Interspeech.2023-428

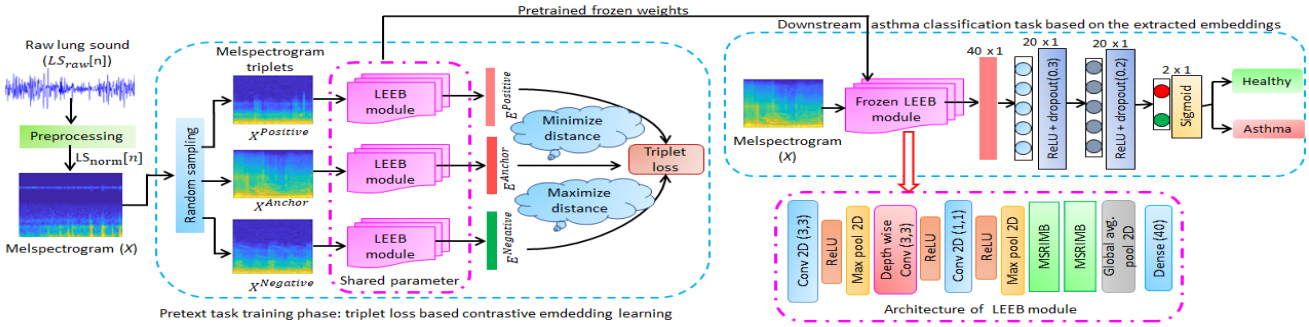

Figure 1: *Block diagram of proposed AsthmaSCELNet framework for asthmatic LS classification.*

discussion of the proposed framework. Section 4 evaluates the proposed framework, and finally, Section 5 concludes the work.

2. Database Description

Chest wall lung sound database (CWLSD) [13] is a recently published publicly available database for respiratory disease identification based on LS recordings. The LS recordings are collected using Littmann 3200 digital stethoscope with a sampling rate of 4000 Hz. The database includes a total of 336 LSs which have been collected from 112 human subjects at King Abdullah University Hospital in Jordan. The database contains a total of 96, 9, 14, 56, 38, 12, and 6 records for each class of respiratory disorders, including asthma, bronchiectasis (BRON), pneumonia, heart failure, chronic obstructive pulmonary disease (COPD), fibrosis, and pleural effusion. The audio signal's duration is irregular, varying from 10 seconds to 50 seconds. In this study, we have considered a total of 105 LS recordings from the healthy class, and 96 recordings from the asthma class to evaluate the potential of our framework.

3. Proposed Framework

In this study, we introduce a novel lightweight supervised contrastive embedding learning framework to distinguish between asthmatic and healthy LS signal. Our proposed framework consists of three main stages: (a) pre-processing, (b) melspectrogram extraction, and (c) classification using the proposed AsthmaSCELNet which exploits the potential of a supervised contrastive embedding learning approach. The individual stages are covered with details in the following subsections.

3.1. Pre-processing

The raw LS signals $(LS_{raw}[n])$ are initially framed into 5sec window keeping 50% overlap with the next adjacent window. Thereafter, low-frequency baseline wandering (BW) component is removed by employing a discrete Fourier transform (DFT) based filtering approach as presented in [14], by extracting the DFT coefficients with frequency values below 1Hz [14].

3.2. Melspectrogram extraction

LSs are highly nonstationary signal. Therefore, to extract more information from the LS signal, it is beneficial to transform the signal from time domain to time-frequency domain which helps to capture the variation of changing frequency over time [9]. In this paper, we have extracted melspectrogram representation from the normalized LS signal by using mel-scale mapping of basic frequency bins [15]. Firstly, we compute short-time Fourier transform (STFT) of BW-free LS $(LS_{bwf}[n])$ as:

$$S[m,k] = \sum_{n=0}^{N-1} LS_{bwf}[n] \cdot W[n-mH] \cdot e^{-j\frac{2\pi nk}{N}} \quad (1)$$

where, $W[n]$ is taken as Hanning window of 1024 samples, with hop-length (H) of 512 samples. Thereafter, we project the Hertz frequency (f) to mel-scale frequency (f_{mel}) to construct the mel-filter banks. In this work, we have considered 64 triangular overlapping mel-filters. This mel-scale conversion is formulated as [15]:

$$f_{mel} = 2595\log(1 + f/700) \quad (2)$$

To extract the melspectrogram, the mel-filters are multiplied with each frame of STFT $(S[m,k])$ [15], [16]. Lastly, a log transform is used on the amplitudes of the melspectrogram. Then, these 2D melspectrograms are converted to 3-channel images by using the 'jet' colormap [15] and reshaped into a size of $224 \times 224 \times 3$. Fig. 2 (a), Fig. 2 (c) illustrate the temporal visualization and melspectrogram representation of asthmatic LS, and Fig. 2 (b), and Fig. 2 (d) illustrates the same for healthy LS signal.

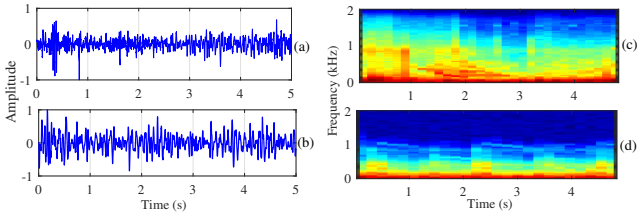

Figure 2: *(a), (c) Illustrates the temporal visualization and melspectrogram representation of asthmatic LS, (b), (d) temporal visualization and melspectrogram representation of healthy LS.*

3.3. Model architecture of proposed AsthmaSCELNet

In this subsection, we discuss the proposed AsthmaSCELNet which consists of two steps. In the first step, we train our proposed LEEB module using triplet loss-based supervised contrastive learning, which facilitates the embedding representation learning and extracts compact representation from the input melspectrogram (X). In the second step, we use the pretrained lightweight embedding extraction backbone (LEEB) module to extract embedding and train a multi-layer perceptron (MLP)-based classifier using the extracted embedding to classify asthmatic LS signals.

3.3.1. Contrastive embedding learning based on Triplet loss

Here, first we introduce the proposed lightweight neural network architecture, namely the LEEB module to extract embeddings from the given melspectrograms. The detailed architecture of the LEEB module is provided in Fig. 1. This LEEB module consists of stacked layers of standard convolution block (SCB), depthwise and pointwise convolution block (DPCB), two multiscale residual inception mobile blocks (MSRIMBs), global average pooling (GAP) and dense layer with 40 neurons. SCB computes convolution using (3×3) kernel and 16

filters. While DPCB uses (3×3) depthwise and (1×1) pointwise convolution layer with 32 filters and helps to reduce the parameter size [17]. Both SCB and DPCB contains ReLU activation and maxpooling (MP) layers with (2×2) kernel and stride of 2. Two MSRIMBs facilitates multiscale feature extraction from same input tensor. The detailed architecture of MSRIMB is provided in Fig. 3, which contains inception like structure [18], however, modified by DPCBs [19] and consists of a residual skip connection [20] which helps to mitigate the problem of overfitting [20]. A detailed ablation study regarding the total trainable parameters of the proposed LEEB module is discussed in Table 1, which tabulates the number of parameters required to construct each layer of the LEEB module. From Table 1, we can observe that our proposed LEEB module requires only 18856 trainable parameters to extract efficient embedding representation from the given melspectrograms. This indicates the lightweight nature of the proposed architecture. Finally, the LEEB module provides embedding size of $E \in \mathscr{R}^{40\times1}$ after the dense layer with 40 neurons. As the LEEB module is trained

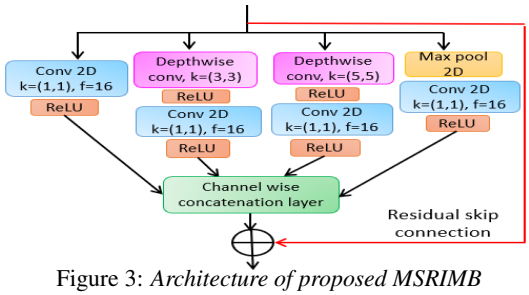

Figure 3: *Architecture of proposed MSRIMB*

using triplet loss [21], it is configured as a parameter shared triplet neural network [22]. We have trained the LEEB module to transform the melspectrograms to an embedding space using three sets of data: anchor data (X^{Anchor}), positive data $(X^{Positive})$, and negative data $(X^{Negative})$, as demonstrated in Fig 1. These three sets of data are randomly sampled from the training set where, $\{X^{Anchor}, X^{Positive}\} \in$ same class, and $X^{Negative} \in$ some other class. The two main reasons why the proposed framework is referred to as supervised contrastive embedding learning are: **firstly**, the triplet training subsets are sampled based on their labels [23] i.e., sampled with supervision of labels [23]. And **secondly**, the use of triplet contrastive loss [21], [23] in the context of learning faithful embedding representation from the melspectrogram representations. The LEEB module is trained independently to learn the embedding using one of the traditionally used contrastive loss functions; the triplet loss function [23], which increases the distance between embeddings of two distinct classes and reduces the distance between embeddings of the same classes [21]. The triplet loss can be formulated as:

$$\mathscr{L}_{triplet} = \max\{0, \gamma + \mathscr{D}(E^{Anc}, E^{Pos}) - \mathscr{D}(E^{Anc}, E^{Neg})\} \quad (3)$$

where, $\mathscr{D}(E^{Anc}, E^{Pos})$ indicates the Euclidean distance between anchor-embedding (E^{Anc}), and positive-embedding (E^{Pos}). Similarly, $\mathscr{D}(E^{Anc}, E^{Neg})$ indicates the Euclidean distance between anchor-embedding (E^{Anc}), and negative-embedding (E^{Neg}). The margin value γ is used to create adequate margin in the embedding space among the embeddings of several classes to obtain discriminant embedding from each class. While experimenting, we have found that $\gamma = 0.2$ provides the best result for this problem.

3.3.2. Downstream asthma classification stage

In this stage, we use the pretrained frozen LEEB module to extract embeddings from the melspectrograms (X). Thereafter, these embeddings are passed through two dense layers or MLP layers with 20 neurons in each layer. The activation function used for these layers is ReLU, and dropout rate of 0.3, 0.2 are used in the two dense layers respectively, to alleviate overfitting problem. Finally, output of the dense layer is classified using sigmoid activation function. The operation for the classification layer can be expressed as:

$$\mathscr{O} = \sigma(< \mathscr{U}, \mathscr{W}_0 > + \mathscr{B}_0) \quad (4)$$

where, $< \mathscr{U}, \mathscr{W}_0 >$ denotes the dot product between weight vector $(\mathscr{W}_0)$ and the output of the dense layer $(\mathscr{U})$, $\mathscr{B}_0$ denotes the bias, and σ refers to the sigmoid activation function [24] which is used for binary classification. Finally, the classifier is trained using a binary cross entropy loss function [24], and a gradient descent-based weight update approach. The optimal simulation parameters used to train the AsthmaSCELNet are shown in Table 2 which have been selected using the GridSearchCV-based KerasTuner hyperparameter optimization framework [25].

Table 1: *Description of Total Parameter Size of LEEB Module*

Layer	Kernel size (k)	Stride	Filter number	Output size	Trainable parameters
Input layer	–	–	–	224*224*3	0
Standard conv2D	(3,3)	2	32	112*112*32	896
ReLU		–	–	112*112*32	0
MP	(2,2)	2	–	56*56*32	0
Depthwise conv2D	(3,3)	1	–	56*56*32	320
ReLU	–	–	–	56*56*32	0
Pointwise conv2D	(1,1)	–	64	56*56*64	2112
ReLU	–	–	–	56*56*64	0
MP	(2,2)	2	–	28*28*64	0
MSRIMB	–	–	16,16,16,16	28*28*64	6464
MSRIMB	–	–	16,16,16,16	28*28*64	6464
GAP	–	–	–	64*1	0
Dense	–	–	–	40*1	2600
Total Parameters					**18856**

Table 2: *Optimal Simulation Parameters Used to Train LEEB Module and Classifier of the Proposed AsthmaSCELNet*

Parameter	LEEB module	Classifier
Input shape	$224 \times 224 \times 3$	40×1
Trainable parameters	18856	1282
Optimizer	Adam	Adam
Learning rate	0.008	0.008
Batch size	64	64
Epochs	400	100

4. Result and discussion

In this section, we compare the quantitative and qualitative performance of the proposed AsthmaSCELNet framework with some notable prior works on asthma classification.

4.1. Evaluation metrics

To evaluate the efficacy of the proposed framework, we have used the following matrices: accuracy (acc) [26], [27] sensitivity (sen) or recall (rec) [28], [27] specificity (spe) [27], precision (prc) [28], F1-score, and ICBHI score [29], [7], an average of specificity and sensitivity, is one of the most widely used performance metrics in LS classification tasks.

4.2. Performance evaluation

The performance of the proposed architecture is evaluated based on training and testing of the network using a 5-fold cross-validation method. Initially, the whole LS data from asthma patients and healthy subjects are splitted into 80% - 20% training-testing set. Further, we take 10% data from the training set to create the validation set. The testing data was not involved

in any training and fine-tuning process of the proposed Asthma SCELNet. Fig. 4 illustrates the performance of the LEEB module in terms of efficient embedding extraction. The embeddings extracted from the LEEB module are visualized in a 2D feature plane by t-distributed stochastic neighbor embedding (t-SNE) [30] shown in Fig 4. In Fig. 4 (a), the initial raw LSs were dispersed randomly in the 2D feature plane. After applying the LEEB module, the embeddings of the same LSs extracted from corresponding melspectrograms are well separated in the 2D feature plane, that can be observed in Fig. 4 (b). As these embeddings are highly discriminative by themselves, it becomes relatively simple for the classifier to categorize asthmatic LSs.

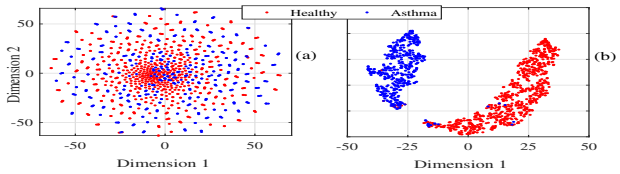

Figure 4: *2D t-SNE visualization of (a) raw LS signals, and (b) embedding of the LS signal extracted from LEEB module.*

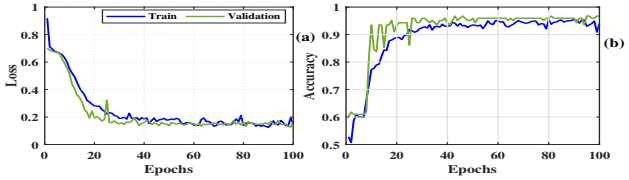

Figure 5: *(a-b) Classification loss and accuracy curve obtained while training the classifier using the embeddings extracted from the frozen LEEB module.*

Fig. 5 (a-b) illustrates the training-validation loss, and accuracy plots of the classifier trained using the embeddings extracted from the frozen LEEB module. It can be observed from Fig. 5 that the classifier achieves a decent amount of accuracy within a few numbers of epochs. Table 3 indicates the classification performance of the proposed AsthmaSCELNet in terms of the aforementioned evaluation metrics. It can be observed that using the proposed framework, we have achieved higher accuracy as compared to the state-of-the-art results on asthma classification by using LS signals. To prove the efficacy of our proposed framework, we have shown the receiver operating characteristics (ROC) curve of both asthma, and healthy LS classes in Fig 6 (a). It can also be observed that for each of the classes, we have achieved a high area under the curve (AUC) value using the proposed AsthmaSCELNet. Additionally, Fig. 6 (b) illustrates the confusion matrix obtained from the AsthmaSCELNet. From the confusion matrix, we can observe that our proposed framework achieves very less misclassification rate for both classes. We had implemented our framework with keras in Python and tested on Windows 10, 32GB RAM desktop consisting of Intel Xeon(R) W-1350 3.30 - 3.31 GHz processor, where, it takes nearly 3.62 sec to classify an entire lung sound signal at the inference time.

Table 3: *Classification Performance Using AsthmaSCELNet*

Performance metrics in percentage (%)					
acc	**sen**	**spe**	**prc**	**F1-score**	**ICBHI score**
98.54	98.27	98.73	98.27	98.27	98.50

4.3. Performance comparison

In this subsection, we compare our proposed AsthmaSCELNet with other existing research works on asthma classification. Table 4 shows the comparative results for the asthma classification task using various diagnostic modalities. From Table 4, it is clear that LS outperforms other modalities, such as pathological speech used by Yadav et al. [8], in diagnosing asthma.

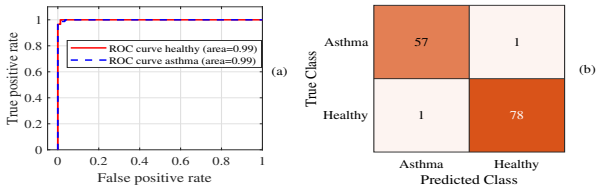

Figure 6: *Illustrates the (a) ROC curve for both asthma and healthy class, (b) confusion matrix obtained from the test data.*

However, it is evident that our suggested framework performs better than all prior studies based on LS-based asthma detection by a margin of 13% and 18%, respectively, when compared to those published by Altan et al. [9] and Tripathy et al. [10]. To the best of our knowledge, our proposed framework presents a deep learning-based method for the first time for asthma classification using LS, which effectively reduces the burden of the hand-crafted feature extraction approach used in [8], [9], [10].

Table 4: *Performance Comparison of AsthmaSCELNet with Other Existing Methodologies*

Reference	**Data type (database)**	**Methodology**	**Results (%)**			
			acc	**sen**	**spe**	**ICBHI score**
Yadav et al. [8]	Speech (own)	MFCC features, SVM	75.40	–	–	–
Altan et al. [9]	LS (own)	Filtering, HHT, statistical features, DBN	84.61	85.83	77.11	81.47
Tripathy et al. [10]	LS (CWLSD)	EWT, temporal-spectral features, ML classifiers	80.35	84.88	75.23	80.05
Proposed framework	**LS (CWLSD)**	**Melspectrogram driven AsthmaSCELNet**	**98.54**	**98.27**	**98.73**	**98.50**

Additionally, to demonstrate the lightweight nature of the proposed AsthmaSCELNet, a rigorous comparative study is given in Table 5 in terms of the following model evaluation parameters such as total trainable parameters, size of the model, and performance rates. From Table 5, our proposed AsthmaSCELNet clearly surpasses the current lightweight deep learning models by attaining the greatest performance metrics for asthmatic LS categorization and also drastically reducing the total trainable parameter size.

Table 5: *Comparative Performance of Proposed AsthmaCCELNet with Other Existing Lightweight Deep Learning Models*

Model	**Model evaluation factors**					
	No. of trainable parameter in million (M)	**Size of model**	**Performance metrics (%)**			
			acc	**sen**	**spe**	**ICBHI score**
ResNet-50 [20]	25.6 M	298MB	90.82	88.93	94.11	91.52
Mobilenet [19]	4.2 M	46.9MB	92.66	87.23	96.77	92.00
ShufflenetV2 [31]	5.4 M	49MB	87.15	80.35	94.33	87.34
Lightweight CNN [32]	3.8 M	44.8MB	95.74	93.65	92.34	92.99
AsthmaSCELNet	**0.018 M**	**498KB**	**98.54**	**98.27**	**98.73**	**98.50**

5. Conclusion

In this paper, we have investigated the potential of supervised contrastive embedding learning framework to identify the asthmatic LSs using a novel LEEB module and MLP classifier. This study makes use of the melspectrogram's potential for identifying asthma and exploits lightweight neural network architecture that reduces the computational load. Employing the proposed framework, we have outperformed the state-of-the-art asthma classification approaches, by achieving highest classification accuracy of 98.54%. Additionally, we believe that our framework will allow us to develop an on-device system that can diagnose asthma from lung auscultations in real-world clinical situations.

6. Acknowledgement

This research is supported by the Ministry of Education (MoE), Government of India, through Prime Minister Research Fellowship (PMRF) program (PMRF ID: 2702854).

7. References

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

[2] A. C. Ayuk, S. N. Uwaezuoke, C. I. Ndukwu, I. K. Ndu, K. K. Iloh, and C. V. Okoli, "Spirometry in asthma care: a review of the trends and challenges in pediatric practice," *Clinical Medicine Insights: Pediatrics*, vol. 11, p. 1179556517720675, 2017.

[3] K. G. Fan, J. Mandel, P. Agnihotri, and M. Tai-Seale, "Remote patient monitoring technologies for predicting chronic obstructive pulmonary disease exacerbations: review and comparison," *JMIR mHealth and uHealth*, vol. 8, no. 5, p. e16147, 2020.

[4] B. Adeniyi and G. Erhabor, "The peak flow meter and its use in clinical practice," *Afr J Respir Med*, vol. 6, no. 2, pp. 5–7, 2011.

[5] S. Yadav, K. NK, D. Gope, U. M. Krishnaswamy, and P. K. Ghosh, "Comparison of cough, wheeze and sustained phonations for automatic classification between healthy subjects and asthmatic patients," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 1400–1403.

[6] R. Zulfiqar, F. Majeed, R. Irfan, H. T. Rauf, E. Benkhelifa, and A. N. Belkacem, "Abnormal respiratory sounds classification using deep cnn through artificial noise addition," *Frontiers in Medicine*, vol. 8, 2021.

[7] Z. Yang, S. Liu, M. Song, E. Parada-Cabaleiro, and B. W. Schuller, "Adventitious Respiratory Classification Using Attentive Residual Neural Networks," in *Proc. Interspeech*, 2020, pp. 2912–2916.

[8] S. Yadav, M. Keerthana, D. Gope, P. K. Ghosh *et al.*, "Analysis of acoustic features for speech sound based classification of asthmatic and healthy subjects," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 6789–6793.

[9] G. Altan, Y. Kutlu, A. Pekmezci, and S. Nural, "The diagnosis of asthma using hilbert–huang transform and deep learning on lung sounds," *Akilli Sistemlerde Yenilikler ve Uygulamalari (ASYU), Antalya*, p. 82, 2017.

[10] R. K. Tripathy, S. Dash, A. Rath, G. Panda, and R. B. Pachori, "Automated detection of pulmonary diseases from lung sound signals using fixed-boundary-based empirical wavelet transform," *IEEE Sensors Letters*, vol. 6, no. 5, pp. 1–4, 2022.

[11] K. Cao, C. Wei, A. Gaidon, N. Arechiga, and T. Ma, "Learning imbalanced datasets with label-distribution-aware margin loss," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[12] Z. Zhang and M. Sabuncu, "Generalized cross entropy loss for training deep neural networks with noisy labels," *Advances in Neural Information Processing Systems*, vol. 31, 2018.

[13] M. Fraiwan, L. Fraiwan, B. Khassawneh, and A. Ibnian, "A dataset of lung sounds recorded from the chest wall using an electronic stethoscope," *Data in Brief*, vol. 35, p. 106913, 2021.

[14] U. Satija, B. Ramkumar, and M. S. Manikandan, "Real-time signal quality-aware ecg telemetry system for iot-based health care monitoring," *IEEE Internet of Things Journal*, vol. 4, no. 3, pp. 815–823, 2017.

[15] O. Ilyas, "Pseudo-colored rate map representation for speech emotion recognition," *Biomedical Signal Processing and Control*, vol. 66, p. 102502, 2021.

[16] K. W. Cheuk, H. Anderson, K. Agres, and D. Herremans, "nnaudio: An on-the-fly gpu audio to spectrogram conversion toolbox using 1d convolutional neural networks," *IEEE Access*, vol. 8, pp. 161 981–162 003, 2020.

[17] A. Koumparoulis and G. Potamianos, "MobiLipNet: Resource-Efficient Deep Learning Based Lipreading," in *Proc. Interspeech*, 2019, pp. 2763–2767.

[18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[19] W. Sae-Lim, W. Wettayaprasit, and P. Aiyarak, "Convolutional neural networks using mobilenet for skin lesion classification," in *2019 16th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 2019, pp. 242–247.

[20] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[21] M. Schultz and T. Joachims, "Learning a distance metric from relative comparisons," *Advances in Neural Information Processing Systems*, vol. 16, 2003.

[22] E. Hoffer and N. Ailon, "Deep metric learning using triplet network," in *International workshop on similarity-based pattern recognition*. Springer, 2015, pp. 84–92.

[23] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," *Advances in Neural Information Processing Systems*, vol. 33, pp. 18 661–18 673, 2020.

[24] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[25] M. Saini, U. Satija, and M. D. Upadhayay, "Dscnn-cau: Deep learning-based mental activity classification for iot implementation towards portable bci," *IEEE Internet of Things Journal*, pp. 1–1, 2022.

[26] J. Mallela, A. Illa, Y. Belur, N. Atchayaram, R. Yadav, P. Reddy, D. Gope, and P. K. Ghosh, "Raw Speech Waveform Based Classification of Patients with ALS, Parkinson's Disease and Healthy Controls Using CNN-BLSTM," in *Proc. Interspeech 2020*, 2020, pp. 4586–4590.

[27] A. K. Dubey, S. M. Prasanna, and S. Dandapat, "Pitch-adaptive front-end feature for hypernasality detection." in *Proc. Interspeech*, 2018, pp. 372–376.

[28] S. Chaudhary, S. Sadbhawna, V. Jakhetiya, B. N. Subudhi, U. Baid, and S. C. Guntuku, "Detecting covid-19 and community acquired pneumonia using chest ct scan images with deep learning," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 8583–8587.

[29] Y. Ma, X. Xu, and Y. Li, "LungRN+NL: An Improved Adventitious Lung Sound Classification Using Non-Local Block ResNet Neural Network with Mixup Data Augmentation," in *Proc. Interspeech*, 2020, pp. 2902–2906.

[30] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of Machine Learning Research*, vol. 9, no. 11, 2008.

[31] N. Ma, X. Zhang, H.-T. Zheng, and J. Sun, "Shufflenet v2: Practical guidelines for efficient cnn architecture design," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 116–131.

[32] M. Chakraborty, S. V. Dhavale, and J. Ingole, "Corona-nidaan: lightweight deep convolutional neural network for chest x-ray based covid-19 infection detection," *Applied Intelligence*, vol. 51, no. 5, pp. 3026–3043, 2021.