Grand Challenge on Respiratory Sound Classification for SPRSound Dataset

Qing Zhang[1,2§], Jing Zhang[2,3,4§], Jiajun Yuan[4,5,6§], Huajie Huang[1], Yuhang Zhang[1], Changyan Chen[1], Jilei Lin [3], Baoqin Zhang[7], Gaomei Lv[8], Shuzhu Lin[9], Na Wang[10], Xin Liu[11], Mingyu Tang[3], Yahua Wang[3], Lu Liu[3], Hui Ma[3], Dan Xie[6], Lihua Wu[3], Haibo Yang[3], Shuhua Yuan[3], Mengjun Chen[12], Bingxue Zhang[13], Hongyuan Zhou[14], Jian Zhao[1,2], Yongfu Li[1,2*], Yong Yin[2,3,4*], Liebin Zhao[4*], Guoxing Wang[1,2], Yong Lian[1,2]

***Abstract*—Globally, respiratory diseases are the leading cause of death, making it essential to develop an automatic respiratory sounds software to speed up diagnosis and reduce physician workload. A recent line of attempts have been proposed to predict accurately, but they have yet been able to provide a satisfactory generalization performance. In this contest, we invited the community to develop more accurate and generalized respiratory sound algorithms. A starter code is provided to standardize the submissions and lower the barrier. New testing set is prepared to evaluate the generalization performance of the submissions. Top 3 teams will present their work at IEEE BioCAS 2023 conference.**

***Index Terms*—Open-source, respiratory sound classification, grand challenge, feature extractions, data augmentation, and machine-learning models, deep-learning models**

This work is supported in part by the National Key Research and Development Program of China under Grant No. 2019YFB2204500in part by the Science, Technology, and Innovation Action Plan of Shanghai Municipality, China under Grant No. 22015821400, in part by the Program for artificial intelligence innovation and development of Shanghai Municipal Commission of Economy and Informatization under Grant No. 2020-RGZN02048, in part by the Translational Medicine Cross Research Fund of Shanghai Jiao Tong University (No. YG20210N109 and No. YG20220N095), in part by the Science and Technology Innovation-Biomedical Supporting Program of Shanghai Science and Technology Committee under Grant No. 19441904400, in part by Special Clinical Research Project of Health Industry of Shanghai Municipal Health Commission (No. 20224Y0180), in part by the Special Research Project of Health Industry of Shanghai Pudong New Area Health Committee (No. PW2021E-06), in part by Supporting Program of Shanghai Science and Technology Committee under Grant No. 20DZ2253900, in part by Hainan Branch of Shanghai Children's Medical Center affiliated to Shanghai Jiao Tong University School of Medcine (No. JYZZ-HQ-202106), in part by Sanya Science and Technology Innovation Program under grant No. 2022KJCX41, and in part by Hainan Provincial Health and Health Commission research program under grant No. 22A200147.

[§]Qing Zhang, Jing Zhang and Jiajun Yuan contributed equally to this work.
*The corresponding authors are Yongfu Li (e-mail: yongfu.li@sjtu.edu.cn), Yong Yin (e-mail: yinyong@scmc.com.cn), and Liebin Zhao (e-mail: zhaoliebin@126.com).
The authors[1] are with the Department of Micro-Nano Electronics and MoE Key Lab of Artificial Intelligence, Shanghai Jiao Tong University, China.
The authors[2] are with Pediatric AI Clinical Application and Research Center, Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, China.
The authors[3] are with Department of Respiratory Medicine, Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, China.
The authors[4] are with the Shanghai Engineering Research Center of Intelligence Pediatrics (SERCIP), Shanghai Children's Medical Center, School of Medicine, Shanghai Jiao Tong University,Shanghai, China.
The author[5] is with School of Computer Engineering and Science, Shanghai University, China.
The authors[6] are with Sanya Maternity and Child Care Hospital, China.
The author[7] is with Department of Paediatrics, Taicang Affiliated Hospital of Soochow University, The First People's Hospital of Taicang, China.
The author[8] is with Linyi City People Hospital, China.
The author[9] is with Department of Paediatrics, Fengcheng Hospital, China.
The author[1]0 is with Linyi City Maternal and Child Health Hospital, China.
The author[11] is with Department of Pediatrics, Fifth Affiliated Hospital of Harbin Medical University, China.
The author[12] is with Fujian Children's Hospital, China.
The author[13] is with the First Affiliated Hospital of Henan University of Traditional Chinese Medicine, China.
The author[14] is with Shanghai Tuoxiao Intelligent Technology Co., Ltd, China.

I. Organization of the Grand Challenge

Respiratory diseases are among the leading causes of death worldwide, causing huge socioeconomic burdens [1]. There have been numerous automated respiratory research proposals in recent years to speed diagnosis and reduce physicians' workloads [2]–[18]. In 2022, the IEEE BioCAS conference organized the first grand challenge on respiratory sound classification using the open-source SJTU paediatric respiratory sound dataset (SPRSound) [19], attracting more than 50 research teams to continuously develop automatic respiratory sound analysis algorithms [20]–[28]. However, the generalization performance of these algorithms has yet to be improved.

The IEEE BioCAS 2023 has organized the second grand challenge on respiratory sound classification using SPRSound dataset [19] with additional dataset for blind testing in this contest, which provides a platform where participants only focus on further improving the model accuracy and generalization performance and let the organizers handle the rest with their model evaluation framework. The contest begins at 1st March 2023, and the submission phase starts from 19th May to 9th June 2023. Fig. 1 illustrates the timeline of the grand challenge. To ensure fairness and transparency among all participating teams, we have provided the following guidelines. (i) To standardize submissions in the expected format and lower the entry barrier and encourage participation in general, a starter code is provided for all participating teams; (ii) To evaluate the generalization performance of the submissions, a brand new testing set (Grand Challenge'23) is prepared, which will be open-sourced after the contest concludes; (iii) To avoid analyzing correlations between the classification models and results, teams are only permitted to submit one entry per day. Scoreboards update based on each team's best entry.

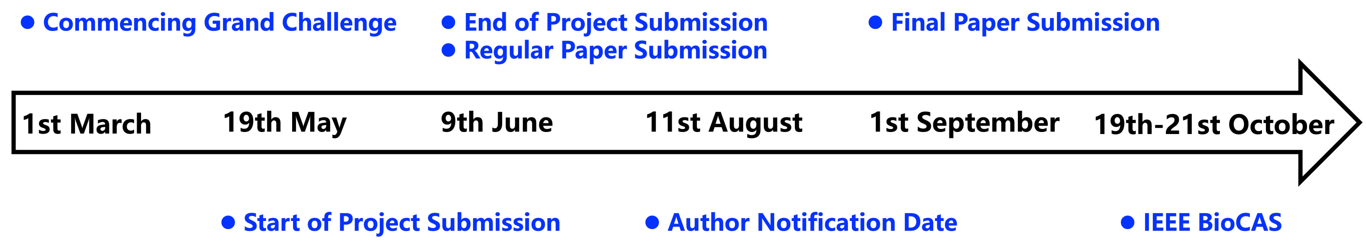

Figure 1. Timeline of grand challenge on respiratory sound classification.

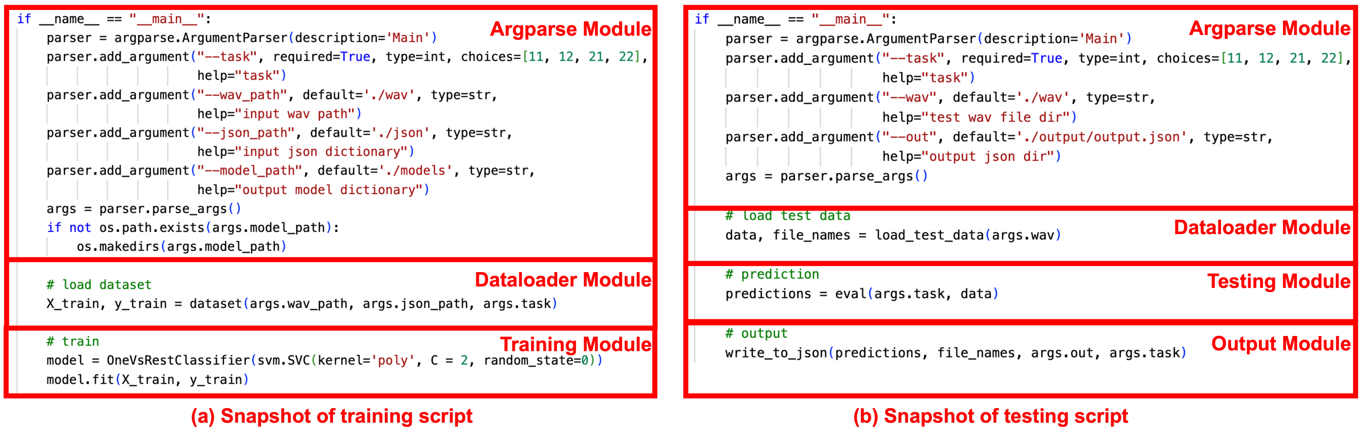

Figure 2. Snapshot of the starter code, including (a) training and (b) testing scripts.

The participating teams are able to submit their papers to a special session at the IEEE BioCAS conference. Submitted papers are reviewed in the same way as regular submissions to ensure high paper quality. Based on the quality of their papers and the efficacy of their models, the top 3 teams are selected to compete for the top 3 prizes at the IEEE BioCAS 2023 conference.

II. Summary of Participants

A total of 15 teams have been participated in the IEEE BioCAS grand challenge. A total of 36 solutions and 4 papers have been submitted for this contest. The top 3 teams were selected based on the quality of their paper submissions and solutions after two rounds of peer review.

III. Grand Challenge

The grand challenge focuses on the classification of respiratory sounds at both an event and record level [29]. Each entry's *total score* is the weighted sum of its individual tasks' scores and its runtime bonus, where the runtime bonus is calculated by taking the logarithm scale of the execution time difference over the worst-case solution and normalizing it to the best result [29].

A. Started Code

We have released a starter code to standardize submissions in the expected format. Specifically, the starter code contains training and testing scripts with the following modules: (1) argparse module, (2) data loader module, (3) training module, (4) testing module, and (5) output module. The argparse module parses command-line arguments, checking for the required parameters and required input files. The data loader module is specifically designed for respiratory sound classification tasks. It automatically converts respiratory sound audios into feature tensors for the subsequent training/testing process. Training and testing modules are provided to establish a baseline for data. The output module standardizes the output format, simplifying the subsequent results evaluation process. Fig. 2 illustrates a snapshot of the starter code. The release of the starter code lowers the entry barrier and encourages participation in general. In addition, participating teams will have more time to polish core functionality.

IV. Grand Challenge'23 Dataset

The SPRSound database [19] is the first open-access pediatric respiratory sound database, which includes 2,683 records and 9,089 respiratory sound events from 292 participants. Specifically, it consists of one training set and two testings sets for intra-patient and inter-patient evaluation. However, since the dataset has been made public, it is challenging for us to assess the model's performance in this years' grand challenge using the provided testing set. Hence, developing an additional dataset (Grand Challenge'23) for blind testing becomes imperative in the context of this competition.

The Grand Challenge'23 dataset is developed with the same data collection and label annotation process as SPRSound [19]. To further expand the dataset resources, such as the sources of participants and physicians, we have invited more

Table I
THE DETAILED STATISTICS OF THE SPRSOUND DATASET AND THE GRAND CHALLENGE'23 DATASET.

Level	Type	SPRSound Training	SPRSound Testing	Grand Challenge'23
Record	Normal	1303	482	539
	CAS	126	107	135
	DAS	248	99	133
	CAS & DAS	95	36	21
	Poor Quality	177	10	43
	Total	1,949	734	871
Event	Normal	5,159	1728	2458
	Rhonchi	39	14	95
	Wheeze	452	413	194
	Stridor	15	2	32
	Coarse Crackle	49	17	34
	Fine Crackle	912	255	309
	Wheeze & Crackle	30	4	2
	Total	6,656	2433	3124

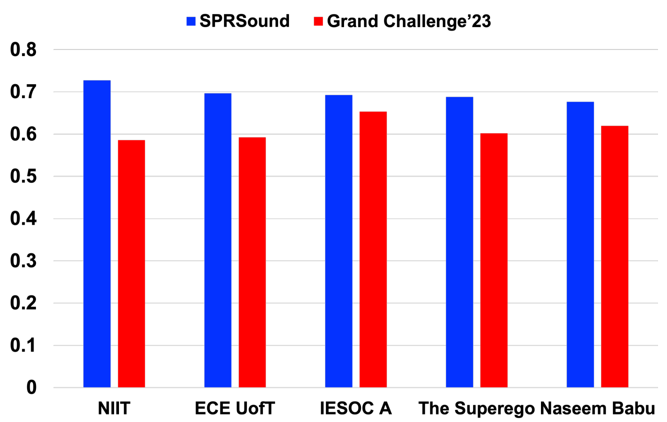

Figure 3. Performance comparison of top-5 teams of IEEE BioCAS 2022 grand challenge on the SPRSound dataset and the Grand Challenge'23 dataset.

hospitals and physicians to participate in the project. In total, 20 physicians from 10 hospitals have participated in the study so far. The Grand Challenge'23 dataset consists of 871 respiratory sound records and 3124 respiratory sound events from 95 participants. At the event level, the numbers of "Normal", "Rhonchi", "Wheeze", "Stridor", "Coarse Crackle", "Fine Crackle", and "Wheeze & Crackle" are 2,458, 95, 194, 32, 34, 309, and 2, respectively. At the record level, the number of "Normal", "CAS", "DAS", "CAS & DAS", and "Poor Quality" records is 539, 135, 133, 21, and 43, respectively. Table I summarized the details of the SPRSound dataset and the Grand Challenge'23 dataset.

V. SUMMARY

A. Summary of the Top-5 of IEEE BioCAS 2022 Grand Challenge on Grand Challenge'23 Dataset

"NIIT", "The Superego", "IESOC A", "ECE_UofT", and "Naseem Babu" are the top-5 of IEEE BioCAS 2022 grand challenge [20]–[23]. We validated the accuracy of their model submitted last year using Grand Challenge'23 dataset, enabling us to delve deeper into its generalization performance analysis. Table II presents a performance comparison between the top-5 teams on both the SPRSound dataset and the Grand Challenge'23 dataset. We observe that "NIIT" achieves the highest total score on the SPRSound dataset, while achieves the lowest total score on the Grand Challenge'23 dataset, indicating a lack of robustness. "IESOC A" boasts a relatively high total score on the SPRSound dataset, and this performance extends to the Grand Challenge'23 dataset as well, showcasing commendable generalization ability. Fig. 3 illustrates the performance comparison, where the blue and red represent the score on the SPRSound dataset and Grand Challenge'23 dataset, respectively. Overall, the total scores of all teams on the new testing dataset are lower compared to those on the SPRSound dataset. Therefore, we conclude that the robustness and generalization of respiratory sound classification algorithms can be further improved.

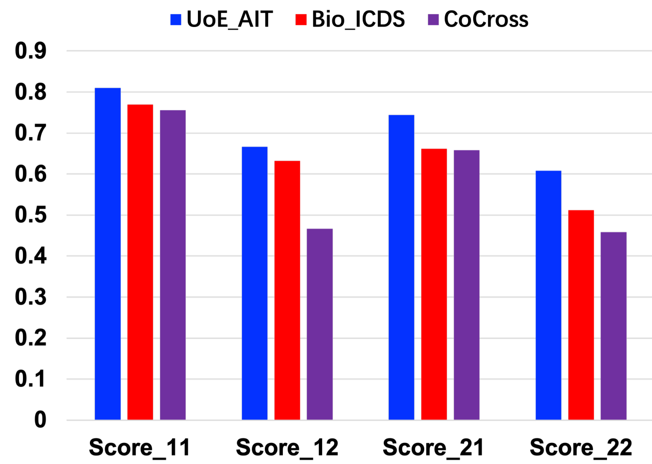

Figure 4. Performance comparison of the top-3 winning teams of IEEE BioCAS 2023 grand challenge on the Grand Challenge'23 dataset.

B. Summary of the Top-3 Winning Teams of IEEE BioCAS 2023 Grand Challenge

(1) Team Bio_ICDS [30] from Nanyang Technological University (Jinhai Hu *et al.*) won 1*st* place in this contest with the highest *total score*. At the pre-processing stage, they have applied pre-normalization for scaling respiratory sound and Mel-Cepstral Frequency Coefficients (MFCCs) for feature extraction. They have also employed a balanced sampler to address the problem of dataset imbalance. Besides, random flipping and random cropping are applied to enhance the diversity and variation of the dataset. Then, the extracted spectrograms are fed into a pre-trained supervised contrastive model (SupCon) and encoded into high-dimensional embedding vectors. Mixup optimization method is adopted in the fine-tuning process to further enhance model performance. Moreover, they have explored optimal hyperparameters (including network parameters and preprocessing parameters) using the bayesian optimization, which enables to maximize the model performance. To implement the respiratory sound classification model in real-time inference, they have proposed a sliding-window approach for classification at recording level. Among all teams, their algorithm has the slowest runtime of 201s. Eventually, they

Table II
PERFORM COMPARISON OF TOP-5 TEAMS OF IEEE BIOCAS 2022 GRAND CHALLENGE ON SPRSOUND DATASET AND GRAND CHALLENGE'23 DATASET.

Team	Testing Set	$Score_{1-1}$	$Score_{1-2}$	$Score_{2-1}$	$Score_{2-2}$	Total Score (W/O bonus)
NIIT	SPRSound	0.8886	**0.8203**	**0.7179**	0.5331	0.7267
	Grand Challenge'23	**0.7848**	**0.6479**	0.5471	0.4167	0.5858
The Superego	SPRSound	0.8196	0.7425	0.7114	0.5314	0.6894
	Grand Challenge'23	0.6677	0.5551	0.7228	0.5241	0.6019
IESOC A	SPRSound	0.8491	0.7473	0.7013	0.5285	0.6930
	Grand Challenge'23	0.7329	0.6460	**0.7587**	**0.5384**	**0.6536**
ECE UofT	SPRSound	**0.8926**	0.7970	0.7151	0.4543	0.6971
	Grand Challenge'23	0.7482	0.5987	0.6986	0.4109	0.5923
Naseem Babu	SPRSound	0.8356	0.7335	0.6696	0.5176	0.6765
	Grand Challenge'23	0.7199	0.5926	0.6653	0.5494	0.6196

Table III
PERFORM COMPARISON OF TOP-3 TEAMS OF IEEE 2023 GRAND CHALLENGE ON GRAND CHALLENGE'23 DATASET.

Rank	Team	$Score_{1-1}$	$Score_{1-2}$	$Score_{2-1}$	$Score_{2-2}$	Runtime (s)	Bonus	Total Score
1	Bio ICDS	0.7693	0.6318	0.6615	0.5121	201	0.1	0.7294
2	UoE AIT	0.8097	0.6666	0.7443	0.6079	7902	0	0.6931
3	CoCross	0.756	0.4666	0.6581	0.4583	2623	0.03	0.5903

have achieved a *total score* of 0.7294 with *bonus* of 0.1 and *score* of 0.7693, 0.6318, 0.6615, and 0.5121 for each task.

(2) Team UoE_AIT [31] from University of Essex (Dat Ngo *et al.*), achieving the highest score in each task while having the longest runtime, won 2*nd* place in this contest overall. At the feature extraction phase, they have re-sampled and duplicated the respiratory events and recordings to consistent lengths of 10s and 30s, respectively. Then, they have employed a 60-2000Hz band-pass filter to suppress noise. Finally, they have applied continuous wavelet transformation to audio feature extraction, which presents both temporal and spectral features. Notice that they have explored multiple feature spectrograms with different mother waves (Amor, Morse, and Bump), different spectrogram sizes, and different temporal dimensions. To deal with the imbalanced problem, they have proposed 3 data augmentation techniques, including (1) spectrogram oversampling, (2) spectrogram masking, and (3) spectrogram mixup. In this way, data diversity and Fisher's criteria are enhanced. At the classification phase, they proposed an inception-residual-based network-based classification model. In particular, they have employed spatio-temporal focusing and multi-head mechanisms to further improve model performance in respiratory sound classification tasks. Kullback-Leibler (KL) divergence loss is used as the loss function. Eventually, they have achieved a *total score* of 0.6931 with *score* of 0.8097, 0.6666, 0.7443, and 0.6079 for each task.

(2) Team CoCross [32] from Aristotle University of Thessaloniki (Diogo Pessoa *et al.*) won 3*rd* place in this contest. At the preprocessing and feature extraction stage, they resample the respiratory sounds to 4000Hz. They have also fixed the respiratory sound events and recordings to be of uniform lengths, which are 7.152s and 15.36s, respectively. Afterwards, they apply the short-time Fourier transform (STFT) as the feature extraction method and normalize the spectrograms to [0, 1] They have designed a dual input model architecture for better classification performance, which are STFT spectorgrams from the time-frequency domain and raw audios from time domain. Specifically, the model consists of two branches with three CNN blocks and one CNN block for processing spectorgrams and raw audios, respectively. At the model training stage, they have employed the hold-out method for model validation. Categorical cross-entropy is set as their loss function. Eventually, they have achieved a *total score* of 0.5903 with *bonus* of 0.03 and *score* of 0.756, 0.4666, 0.6581, and 0.4583 for each task.

VI. CONCLUSIONS

A special session is organized for the grand challenge at the IEEE BioCAS 2023 conference. This provides an opportunity for the top-3 teams to present their algorithms and classification results. We hope that the grand challenge will increase researchers' interest in developing more accurate algorithms and push forward this medical field.

REFERENCES

[1] A. Marques, A. Oliveira, and C. Jácome, "Computerized adventitious respiratory sounds as outcome measures for respiratory therapy: a systematic review," *Respiratory Care*, vol. 59, no. 5, pp. 765–776, 2014.

[2] B. M. Rocha *et al.*, "A respiratory sound database for the development of automated classification," in *International Conference on Biomedical and Health Informatics*, 2017, pp. 33–37.

[3] B. M. Rocha *et al.*, "An open access database for the evaluation of respiratory sound classification algorithms," *Physiological Measurement*, vol. 40, no. 3, p. 035001, 2019.

[4] Y. Ma *et al.*, "Live Demo: LungSys-Automatic Digital Stethoscope System For Adventitious Respiratory Sound Detection," in *Proc. IEEE Biomedical Circuits and Systems Conference*, 2019, pp. 1–1.

[5] B. M. Rocha *et al.*, "Automatic Classification of Adventitious Respiratory Sounds: A (Un)Solved Problem?" *Sensors*, vol. 21, no. 1, 2021.

[6] 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 *Interspeech*, 2020, pp. 2902–2906.

[7] Y. Ma *et al.*, "LungBRN: A smart digital stethoscope for detecting respiratory disease using bi-resnet deep learning algorithm," in *Proc. IEEE Biomedical Circuits and Systems Conference*, 2019, pp. 1–4.

[8] 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.

[9] M. T. García-Ordás *et al.*, "Detecting respiratory pathologies using convolutional neural networks and variational autoencoders for unbalancing data," *Sensors*, vol. 20, no. 4, p. 1214, 2020.

[10] L. Pham *et al.*, "CNN-MoE Based Framework for Classification of Respiratory Anomalies and Lung Disease Detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 8, pp. 2938–2947, 2021.

[11] T. Nguyen and F. Pernkopf, "Crackle Detection In Lung Sounds Using Transfer Learning And Multi-Input Convolutional Neural Networks," in *Proc. International Conference of the IEEE Engineering in Medicine & Biology Society*, 2021, pp. 80–83.

[12] S. Ntalampiras and I. Potamitis, "Automatic acoustic identification of respiratory diseases," *Evolving Systems*, vol. 12, no. 1, pp. 69–77, 2021.

[13] M. Fraiwan *et al.*, "Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–13, 2021.

[14] Mohammad Fraiwan *et al.*, "A dataset of lung sounds recorded from the chest wall using an electronic stethoscope," *Data in Brief*, vol. 35, p. 106913, 2021.

[15] J. Li *et al.*, "LungAttn: advanced lung sound classification using attention mechanism with dual TQWT and triple STFT spectrogram," *Physiological Measurement*, vol. 42, no. 10, p. 105006, 2021.

[16] H. Zhu *et al.*, "Automatic pulmonary auscultation grading diagnosis of Coronavirus Disease 2019 in China with artificial intelligence algorithms: A cohort study," *Computer Methods and Programs in Biomedicine*, vol. 213, p. 106500, 2022.

[17] J. Li *et al.*, "Explainable cnn with fuzzy tree regularization for respiratory sound analysis," *IEEE Transactions on Fuzzy Systems*, vol. 30, no. 6, pp. 1516–1528, 2022.

[18] S. B. Shuvo *et al.*, "A Lightweight CNN Model for Detecting Respiratory Diseases From Lung Auscultation Sounds Using EMD-CWT-Based Hybrid Scalogram," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 7, pp. 2595–2603, 2021.

[19] Q. Zhang *et al.*, "SPRSound: Open-Source SJTU Paediatric Respiratory Sound Database," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 16, no. 5, pp. 867–881, 2022.

[20] N. Babu *et al.*, "Multiclass Categorisation of Respiratory Sound Signals using Neural Network," in *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2022, pp. 228–232.

[21] L. Zhang *et al.*, "A Feature Polymerized Based Two-Level Ensemble Model for Respiratory Sound Classification," in *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2022, pp. 238–242.

[22] J. Li *et al.*, "Improving The ResNet-based Respiratory Sound Classification Systems With Focal Loss," in *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2022, pp. 223–227.

[23] W.-B. Ma *et al.*, "An Effective Lung Sound Classification System for Respiratory Disease Diagnosis Using DenseNet CNN Model with Sound Pre-processing Engine," in *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2022, pp. 218–222.

[24] D. Pessoa *et al.*, "BRACETS: Bimodal repository of auscultation coupled with electrical impedance thoracic signals," *Computer Methods and Programs in Biomedicine*, vol. 240, p. 107720, 2023.

[25] S. Ntalampiras, "Explainable Siamese Neural Network for Classifying Pediatric Respiratory Sounds," *IEEE Journal of Biomedical and Health Informatics*, 2023.

[26] C. Chen *et al.*, "LungHeart-AtMe: Adventitious Cardiopulmonary Sounds Classification Using MMoE with STFT and MFCCs Spectrograms," in *IEEE International Conference on Artificial Intelligence Circuits and Systems (AICAS)*, 2023, pp. 1–5.

[27] B. TaghiBeyglou *et al.*, "Fusion of Manual and Deep Learning Analyses for Automatic Lung Respiratory Sounds Identification in Youth," *CMBES Proceedings*, vol. 45, 2023.

[28] F.-S. Hsu *et al.*, "A dual-purpose deep learning model for auscultated lung and tracheal sound analysis based on mixed set training," *Biomedical Signal Processing and Control*, vol. 86, p. 105222, 2023.

[29] Q. Zhang *et al.*, "Grand Challenge on Respiratory Sound Classification for SPRSound Dataset," in *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 2022, pp. 213–217.

[30] J. Hu *et al.*, "Real-time Classification Using Supervised Contrastive Learning and Mixup with Limited Paediatric Respiratory Sound Data," in *Proc. IEEE Biomedical Circuits and Systems Conference*, 2023, pp. 1–5.

[31] D. Ngo *et al.*, "A Deep Learning Architecture with Spatio-Temporal Focusing for Detecting Respiratory Anomalies," in *Proc. IEEE Biomedical Circuits and Systems Conference*, 2023, pp. 1–5.

[32] D. Pessoa *et al.*, "Pediatric Respiratory Respiratory Sound Classification Using a Dual Input Deep Learning Architecture: The IEEE BioCAS 2023 Grand Challenge," in *Proc. IEEE Biomedical Circuits and Systems Conference*, 2023, pp. 1–5.