An IEEE Societies and Technical Councils Publication
IEEE Data Descriptions

Received 13 August 2024; revised 23 September 2024; accepted 4 October 2024; Date of publication 10 October, 2024; date of current version 15 November, 2024.

Digital Object Identifier 10.1109/IEEEDATA.2024.3477390

Meta: *Data Compression and Event Detection Grand Challenge 2024 With SPRSound Dataset*

QING ZHANG [1,2] (GRADUATE STUDENT MEMBER, IEEE), CHANGYAN CHEN [1,2] (GRADUATE STUDENT MEMBER, IEEE), SHUHUA YUAN [2,3], JING ZHANG[2,3] (MEMBER, IEEE), JIAJUN YUAN [2,3,4,5,6,7], HUAJIE HUANG [1,2], YUHANG ZHANG [1,2] (MEMBER, IEEE), RUI PAN [1,2] (MEMBER, IEEE), XUYA JIANG [1,2,8], JIAN ZHAO [1,2] (SENIOR MEMBER, IEEE), YONGFU LI [1,2] (SENIOR MEMBER, IEEE), YONG YIN [2,3], LIEBIN ZHAO [2], GUOXING WANG [1,2] (SENIOR MEMBER, IEEE), AND YONG LIAN [1,2] (FELLOW, IEEE)

[1]Department of Micro-Nano Electronics and MoE Key Laboratory of Artificial Intelligence, Shanghai Jiao Tong University, Shanghai 200240, China
[2]Shanghai Engineering Research Center of Intelligence Pediatrics, Shanghai 200127, China
[3]The Pediatric AI Clinical Application and Research Center, Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China
[4]Department of Respiratory Medicine, Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China
[5]Child Health Advocacy Institute, China Hospital Development Institute, Shanghai Jiao Tong University, Shanghai 200240, China
[6]School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China
[7]Shanghai Children's Medical Center, School of Medicine, Shanghai Jiao Tong University, Sanya 200240, China
[8]School of Design, Jiangnan University, Wuxi 214122, China

CORRESPONDING AUTHORS: Yongfu Li (e-mail: yongfu.li@sjtu.edu.cn); Yong Yin (e-mail: yinyong@scmc.com.cn); Liebin Zhao (e-mail: zhaoliebin@126.com).

Qing Zhang, Changyan Chen, and Shuhua Yuan contributed equally to this article. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFB2204500; in part by the Medical-Engineering Joint Funds of Shanghai Jiao Tong University under Grant YG2024ZD20, Grant YG2021QN109, and Grant YG2022QN095; in part by the Shanghai Public Health Excellent Talent Project under Grant GWVI-11.2-YQ58; in part by the Science and Technology Committee of Shanghai, China, under Grant 22015821400; in part by the Sanya Science and Technology Special Fund under Grant 2022KJCX41; in part by the Hainan Provincial Health and Health Commission research program under Grant 22A200147; in part by the Hainan Province Science and Technology Special Fund under Grant ZDYF2024SHFZ059; in part by JINYE ZHONGZI Project from SCMC-HN under Grant JYZZ-HQ-202106; and in part by the Science and Technology Committee of Shanghai, China, under Grant 20DZ2253900.

ABSTRACT The development of digital stethoscopes and automatic respiratory sound algorithms is crucial for accelerating diagnosis, reducing physician workload, and lowering mortality rates from respiratory diseases. However, transmitting data from digital stethoscopes to cloud or other storage devices require substantial storage and transmission capacities, especially for wearable devices used for long-term monitoring. Besides, current automatic algorithms often predict labels for segmented sound events, lacking precision in detecting the onsets and offsets of respiratory sound events. To address these challenges, we organized a grand challenge inviting the community to develop data compression and event detection algorithms to reduce the storage and transmission burden and event segmentation workload. A new testing set was prepared to evaluate the performance of the submissions. The top teams presented their work at the 20th IEEE Biomedical Circuits and Systems Conference (BioCAS) 2024.

IEEE SOCIETY/COUNCIL Circuits and Systems (CAS) Society

DATA TYPE/LOCATION Sound and Text; Worldwide

DATA DOI/PID 10.21227/bhw7-2044; 10.21227/nfkk-1x47

INDEX TERMS Data compression, event detection, grand challenge, open-source, respiratory sound.

BACKGROUND

Respiratory diseases rank among the leading causes of death globally, causing significant socioeconomic burdens [1]. Digital stethoscopes are valuable tools to record and transmit respiratory sounds to specialists remotely, aiding in the diagnosis of respiratory diseases. However, digital stethoscope generates large amount of data continuously, facing limitations in storage, computational, and transmission capacities. Therefore, making the development of a respiratory sound compression algorithm is essential.

In recent years, numerous automated respiratory sound algorithms have emerged, aiming to alleviate physicians' workloads [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], [18]. Over the past two years, the IEEE Biomedical Circuits and Systems (BioCAS) conference, flagship conference of the IEEE Circuits and Systems (CAS) Society, has hosted two grand challenges [19], [20], focused on respiratory sound classification using the open-source SJTU pediatric respiratory sound dataset (SPRSound) [21]. These events attracted over 60 research teams dedicated to advancing automatic respiratory sound classification algorithms. As these classification algorithms are based on the specific event/record time on a sound, the detection of onsets and offsets of the respiratory sound event is not required, after that is the assignment of event labels of respiratory events [22], [23], [24], [25], [26], [27], [28], [29], [30]. To complete the entire end-to-end system as shown in Fig. 1, there is a need for event detection that automatically locates and categorizes respiratory sound events (onsets/offsets detection with classifications).

Advanced sound compression and event detection algorithms hold the promise of transforming the diagnosis and management of respiratory diseases globally, ultimately improving patient outcomes and healthcare efficiency. This year, the grand challenge includes the respiratory sound compression and event detection, inviting participants to explore different feature extraction techniques and models to improve the current state-of-the-art works.

PROPOSED METHODOLOGY

Overview

The IEEE BioCAS 2024 has organized a grand challenge on respiratory sound data compression (track 1) and event detection (track 2). The contest began on 1 March 2024, and the submission phase started on 19 May to 9 July 2024. The timeline is shown in Fig. 2. To ensure fairness and transparency among all participating teams, we have provided the following guidelines.

1) A new testing set (Grand Challenge'24) has been prepared to assess the generalization performance of the submissions and will be made open-source after the contest ends.
2) To prevent analysis of correlations between classification models and results, each team is allowed only one submission per day, with scoreboards reflecting each team's best entry.

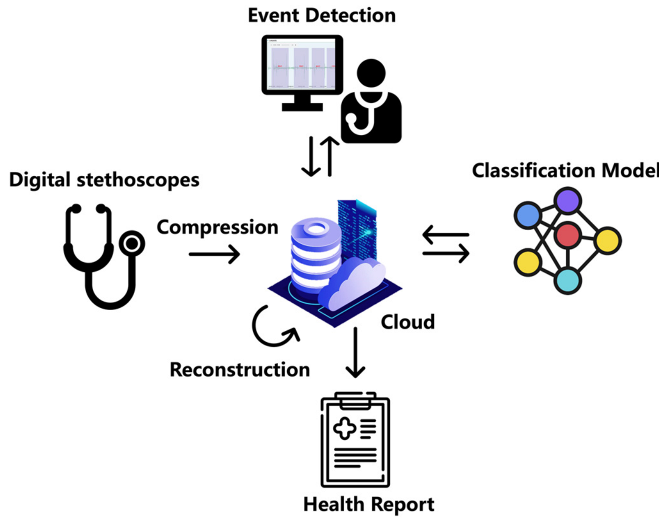

FIG. 1. Overview of respiratory sound system.

To maintain high standards, all submitted papers undergo the same review process as regular submissions. The top winning teams were chosen based on the quality of their papers and the score of their algorithms. They competed for the top prizes at the IEEE BioCAS 2024 conference.

Main Tracks

Task 1 (Respiratory Recording Compression): This track deals with respiratory recordings compression using compressive sensing-based compression methods.

Task 2 (Respiratory Event Detection): This track deals with the detection of onsets and offsets in addition to the assignment of event labels of respiratory events in respiratory recordings using sound event detection methods.

Evaluation Metrics

Track 1 is evaluated with comprise compression ratio (CR), percent root mean square difference (PRD), and correlation coefficient (CC). The detailed definitions are as follows.

Definition 1 (CR): CR measures the reduction in data size achieved by the compression algorithm, which is defined as the ratio of the size of initial signal to the size of the compressed signal

$$\mathrm{CR} = \frac{\text{Size of initial signal}}{\text{Size of compressed signal}}. \tag{1}$$

Definition 2 (CC): CC measures the similarity or linear relationship between the initial and reconstructed signal, which is commonly used to assess the effectiveness of compression algorithms

$$\mathrm{CC} = \frac{\mathrm{Cov}(X, X')}{\sigma(X)\sigma(X')} \tag{2}$$

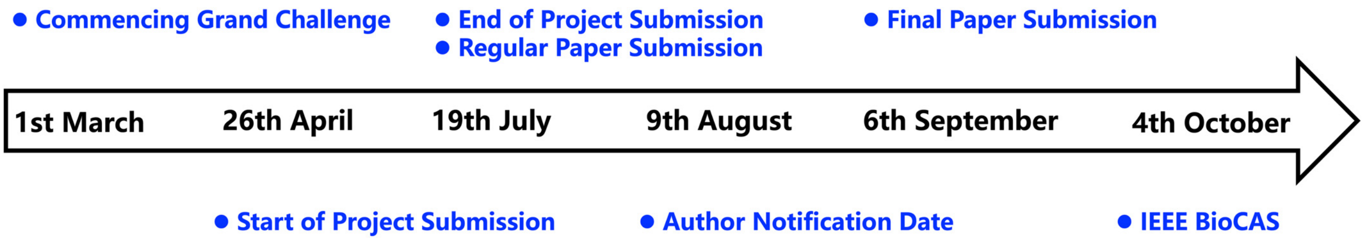

FIG. 2. Timeline of IEEE BioCAS 2024 Grand Challenge.

where $\mathrm{Cov}(X, X')$ refers to the covariance between the original signal X and the reconstructed signal X'. $\sigma(X)$ and $\sigma(X')$ refer to the standard deviations of the original and reconstructed signals, respectively.

Definition 3 (PRD): PRD measures the difference between the initial and the reconstructed signal

$$\mathrm{PRD} = \sqrt{\frac{\sum_{i=1}^{N}(X'(i) - X(i))^2}{\sum_{i=1}^{N} X(i)^2}} \tag{3}$$

where $X(i)$ and $X'(i)$ refer to the initial and reconstructed signal value at the ith sample, respectively. N refers to the number of samples in the signal. A lower PRD value indicates a smaller difference between the original and reconstructed signals, implying better reconstruction quality.

Using these metrics, we can comprehensively assess the effectiveness of compression and reconstruction algorithms. A higher CR indicates more efficient compression, and a lower *PRD* indicates better reconstruction fidelity. The value of CC ranges from -1 to 1, with 1 indicating a perfect positive linear relationship, -1 indicating a perfect negative linear relationship, and 0 indicating no linear relationship.

Based on these evaluation metrics, we define the total score for track 1 (TS_1) as follows:

$$\mathrm{TS}_1 = \frac{\mathrm{CR}}{2 - \mathrm{CC} + \mathrm{PRD}}. \tag{4}$$

Track 2 is evaluated with event-level microaverage F-score (F) and error rate (ER) to offer a comprehensive evaluation of the model's ability to accurately detect the temporal onsets and offsets of events within the complete dataset. A fixed 200 ms collar at onset and a 20% event length collar at offsets are applied [31], [32]. The detailed definitions are as follows.

Definition 4 [F-score (F)]: The F-score measures the model's ability to identify positive classes, which is derived from precision (*P*) and recall (*R*). *P* indicates the precision of the results returned by a system, specifically the ratio of the correct results to the total number of results returned. However, *R* measures the completeness of the results, indicating the ratio of the correct results to the total number of relevant elements in the dataset. The *P*, *R*, and *F* are calculated as follows:

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{5}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{6}$$

$$F = \frac{2 \cdot P \cdot R}{P + R} \tag{7}$$

$$= \frac{2 \cdot \mathrm{TP}}{2 \cdot \mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{8}$$

where TP, FP, and FN denote the counts of true positives, false positives, and false negatives, respectively. These are calculated by aligning the model's detection results with the actual event onsets and offsets based on the collar tolerance.

Definition 5 (ER): The error rate measures the overall proportion of incorrect predictions made by a model, which is derived from substitution (*S*), deletion (*D*), and insertion (*I*). *S* indicates the number of reference events for which a correct event was not output, effectively representing the cases where an event is replaced by an incorrect one. *D* indicates the number of reference events that were not correctly identified by the model, representing the missed detections. *I* indicates the number of events in the model predictions that do not correspond to any reference event, representing the events that were incorrectly detected by the model. The *S*, *D*, *I*, and ER are calculated as

$$S = \min(\mathrm{FN}, \mathrm{FP}) \tag{9}$$

$$D = \max(0, \mathrm{FN} - \mathrm{FP}) \tag{10}$$

$$I = \max(0, \mathrm{FP} - \mathrm{FN}) \tag{11}$$

$$\mathrm{ER} = \frac{S + D + I}{N} \tag{12}$$

where N is the total number of respiratory events marked as active in the annotation, ensuring that the error rate is normalized by the total number of events.

Using these metrics, we can comprehensively assess the effectiveness of respiratory sound event detection models. The *F* ranges from 0 to 1. A higher *F* indicates that the model is more effective in displaying the correct event detections. An *F* of 1 indicates the model's perfect ability to predict all reference events in the dataset. The ER is larger than 0, with no upper limit. Lower ER indicates that the model makes fewer incorrect predictions. An ER of 0 indicates a perfect model that accurately outputs all and only the events present in the dataset without false positives.

Based on these evaluation metrics, we define the total score for track 2 (TS_2) as follows:

$$\mathrm{TS}_2 = F - \mathrm{ER}. \tag{13}$$

Training and Validation Datasets

The IEEE BioCAS 2024 has organized a grand challenge on respiratory sound compression and event detection using the open-sourced SPRSound [21] and Grand Challenge'23 datasets as the training and validation datasets [20]. An additional blind testing (Grand Challenge'24) dataset is reserved to validate the algorithms in this contest.

To enhance model performance, we have combined the publicly available SPRSound [21] and Grand Challenge'23 [20] databases into one large dataset for training and validation purposes. The SPRSound dataset includes 2683 records with 9089 respiratory sound events from 292 participants. The Grand Challenge'23 dataset has 871 records with 3124 events from 95 participants. Both datasets were developed by Shanghai Jiao Tong University and its affiliated hospitals, with labels annotated by over 20 physicians from ten hospitals using a quality assurance process. Specifically, the signal quality of the records was carefully assessed, with poor quality data labeled accordingly to distinguish them from the usable records. Respiratory sound records with good signal quality were independently annotated by 20 experienced pediatric physicians, with each record annotated by at least three. Given the subjective nature of such annotations, labels were considered a "Gold Standard" only when 100% agreement was achieved among the annotators. In cases where discrepancies occurred in the annotations, the data were forwarded to two authoritative physicians for further review. This multilayered process ensured the robustness and reliability of the final dataset. Additionally, the same strict quality assurance methodology has been applied to our new blind testing set, ensuring consistent label quality across all datasets used in this study.

The datasets include both record- and event-level labels. Record-level labels help to identify and exclude noisy data marked as "poor quality" from the dataset for track 2, which lacks valid respiratory sound events and is not suitable for event detection. Event-level labels, provided in JSON format, cater to the needs of tracks 1 and 2.

1) Track 1: No annotations file is provided in the respiratory sound compression track due to the unnecessary of labels.
2) Track 2: Detailed event-level labels are provided to allow participants to detect and classify respiratory sound events accurately. The labels include event types of "normal," "rhonchi," "wheeze," "stridor," "coarse crackle," "fine crackle," and "wheeze & crackle." The annotation files also provide the exact start and end times for each event. An example of the annotation file ('_3.9_0_p1_18151') is shown in Fig. 3(a).

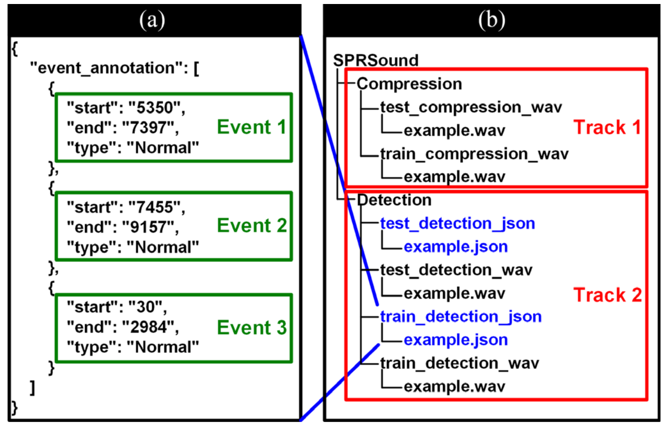

FIG. 3. (a) Example of the annotation file ('_3.9_0_p1_18151') in JSON format. (b) Dataset folder structure for each track.

TABLE I. Detailed Statistics of Datasets for Challenge

Level	Type	Training	Validation	Grand Challenge'24
Event	Normal	7430	1915	7252
	Rhonchi	119	29	62
	Wheeze	871	188	141
	Stridor	44	5	17
	Coarse Crackle	85	15	12
	Fine Crackle	1204	272	169
	Wheeze & Crackle	32	4	6
	Total	9785	2428	7659
Record		2660	664	1704

An 80/20 split is used to divide the dataset into training and validation datasets. In total, the contest's training and validation datasets consist of 3324 records and 12 213 events, with the exact counts of each event type detailed in Table I. The dataset folder structure for each track is shown in Fig. 3(b).

Grand Challenge'24 Dataset

To further ensure the robustness of the contest, we have developed an additional dataset, Grand Challenge'24, specifically for blind testing. The Grand Challenge'24 dataset was collected and annotated using the same rigorous data collection and quality assurance processes by the same team of physicians as the SPRSound [21] and Grand Challenge'23 [20] datasets, ensuring consistency in label accuracy and robustness. The primary difference lies in the participants involved in the datasets, which allows us to further validate the generalization ability of models. This ensures that Grand Challenge'24 serves as a reliable blind test for models trained on the other datasets. Table I summarizes the detailed information. The Grand Challenge'24 database consists of 1704 respiratory sound records and 7659 respiratory sound events from 324 participants. The numbers of "normal," "rhonchi," "wheeze," "stridor," "coarse crackle," "fine crackle," and "wheeze & crackle" are 7252, 62, 141, 17, 12, 169, and 6, respectively.

The training, validation, and Grand Challenge'24 databases for respiratory sound analysis show different average durations. The mean duration of "normal," "rhonchi," "wheeze," "stridor," "coarse crackle," "fine crackle," and "wheeze & crackle" for the training set are 1540, 1444, 697, 1647, 1135, 1039, and 1090 ms, respectively. For the validation set, the mean durations are 1560, 1339, 732, 1287, 952, 997, and 1600 ms, respectively. For the Grand Challenge'24 set, the mean durations are 2120, 2375, 1992, 2601, 2114, 1898, and 1534 ms, respectively.

The datasets used in this study are licensed under the Creative Commons Attribution 4.0 International license, which permits the use, adaptation, and distribution of the data provided that appropriate credit is given and a license notice is included. To ensure participant privacy, all data were anonymized following strict ethical guidelines and received ethical approval from the SCMC Institutional Review Board (Approval Number: SCMCIRB-K2019056-1). Specifically, personal identifiers were removed, and participants were assigned randomized IDs that cannot be traced back to any individual. Sensitive information, such as names and contact details, was excluded from the dataset to protect confidentiality. Researchers are encouraged to reuse the dataset in publications and presentations while adhering to these ethical standards and citing the original data collection references as detailed in the dataset repository. Comprehensive details on data sharing, licensing, and reuse policies can be found within the repository, ensuring that the dataset can be reused responsibly in line with legal and ethical considerations.

APPLIED ANALYSIS

Summary of Participants

A total of 20 teams took part in the IEEE BioCAS grand challenge for respiratory sound compression (track 1) and detection of respiratory sound events (track 2). In track 1, 16 submissions and 3 papers were submitted. In track 2, 16 submissions and 3 papers were submitted. Following two rounds of peer review, the top teams of each track were chosen based on the excellence of their paper submissions and solutions.

Summary of the Top Winning Teams of IEEE BioCAS 2024 Grand Challenge for Respiratory Sound Compression (Track 1)

To achieve an optimal compression performance, it is necessary to explore the compatible compression and reconstruction algorithms along with postprocessing techniques. We have summarized the key techniques used by the top winning teams of this track. The statistical testing results in the Grand Challenge'24 dataset are shown in Table II.

TABLE II. Performance Comparison of the Top Winning Teams of IEEE BioCAS 2024 Grand Challenge on Respiratory Sound Compression on Grand Challenge'24 Dataset

Rank	Team	CR	PRD	CC	TS_1
1	TJU	256.001	0.147	0.990	221.415
2	Bio_ICDS	222.141	0.223	0.972	178.301

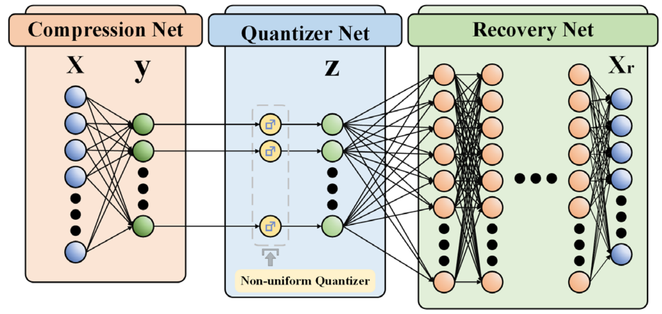

FIG. 4. Proposed three-stage framework for respiratory sound compression and reconstruction by Team TJU [33].

1) *Team TJU* [33] from Tian Jin University (Jinglei Zhai et al.) achieved the top prize in track 1 with the highest total score. They proposed a three-stage respiratory sound compression and reconstruction framework, as shown in Fig. 4. In the compression stage, they used a variational auto-encoder (VAE) to reduce the sound data size while keeping the key features. In the quantization stage, a nonuniform quantizer was applied to make the data more compact, while the nonuniform concept helped to maintain the sound quality and the model's accuracy. In the reconstruction stage, they used a Transformer network to improve the reconstruction performance and understand the connections between sounds over time, making the reconstruction more precise. To further enhance the quality of the reconstructed signals, they used a combination of wavelet transformation, a low-pass filter, and a median filter to clean up the sounds, removing any noise and smoothing them out. They also ensured that their model did not just memorize the training data by using techniques such as normalization and dropout, which helped to avoid overfitting. Their efforts resulted in a total score of 221.4145, with a CR of 256.0009, a PRD of 0.1466, and a CC of 0.9903.
2) *Team Bio_ICDS* [34] from Nanyang Technological University (Shuailin Tao et al.) achieved 2nd place in track 1. They developed a novel framework grounded in the auto-encoder concept for respiratory sound compression and reconstruction tasks. This framework features a residual convolutional encoder (Fig. 5) for efficient signal compression and a corresponding residual transpose convolutional decoder for high-fidelity signal reconstruction. To address variable-length signals, the team integrated 1-D convolutional layers, eliminating

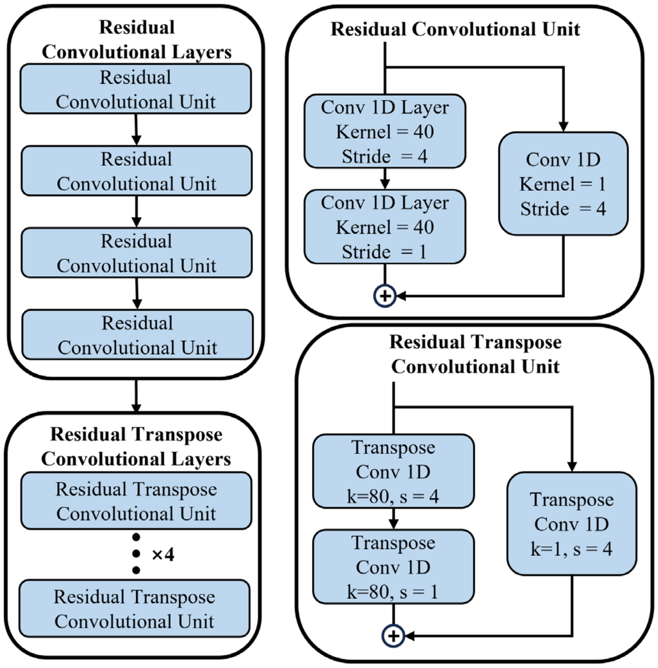

FIG. 5. Proposed residual convolutional encoder by Team Bio_ICDS [34].

the need to standardize input signal dimensions. They also delved into the latent space, opting not to use a variational distribution to ensure the integrity and continuity of respiratory sounds throughout the compression process. During model training, they implemented a sophisticated composite loss function that bridges the gap between domains, focusing on both time- and frequency-domain reconstruction to enhance the signal's frequency accuracy. Additionally, they conducted a thorough exploration to identify the optimal kernel size, striking a balance between CR and reconstruction quality. Their efforts culminated in a commendable total score of 178.3007, with standout metrics: a CR of 222.1410, a PRD of 0.2235, and a CC of 0.9719, demonstrating their model's exceptional performance in the competition.

Summary of the Top Winning Teams of IEEE BioCAS 2024 Grand Challenge for Respiratory Sound Event Detection (Track 2)

Besides selecting the appropriate machine-learning/deep-learning sound events timestamp detection and events classification models, there is a need to explore the compatible signal preprocessing techniques, feature extraction techniques, data augmentation methods, and loss functions to achieve their best performance. We have summarized the key techniques used by the top winning teams of this track. The statistical testing results in SPRSound and Grand Challenge'24 dataset are shown in Table III.

1) *Team IESOC* [35] from National Yang Ming Chiao Tung University (Chiayu Yep et al.) won first place on track 2, earning the highest total score. They crafted a three-stage respiratory sound event detection model, including the preprocessing stage, the event timestamp detection stage, and the event classification stage, as shown in Fig. 6. In the first stage, they separated respiratory sounds into segments and obtained spectrograms of the segments using the fast Fourier transform (FFT) for feature extraction. In the second stage, an accurate voice activity detection (VAD) model based on convolutional and recurrent layers, silero VAD, was introduced to precisely find the onset and offset of each respiratory sound event. For the third stage, they pulled out more details from the detected events using techniques such as mel-frequency cepstral coefficients (MFCCs), short-time Fourier transform (STFT), and FFT. They put these details into a streamlined version of a network that connects in a dense way, which helped them to figure out the type of event with a model that is not too complex and does not take up too much computing power. Due to the imbalance of the dataset, a waveform augmentation method, time-shift, was utilized to generate more events of the less

TABLE III. Performance Comparison of the Top Winning Teams of IEEE BioCAS 2024 Grand Challenge on Respiratory Sound Event Detection on Grand Challenge'24 Dataset

Rank	Team	F	ER	TS_2
1	IESOC	0.330	1.362	-1.027
2	Oxy	0.0341	1.563	-1.527

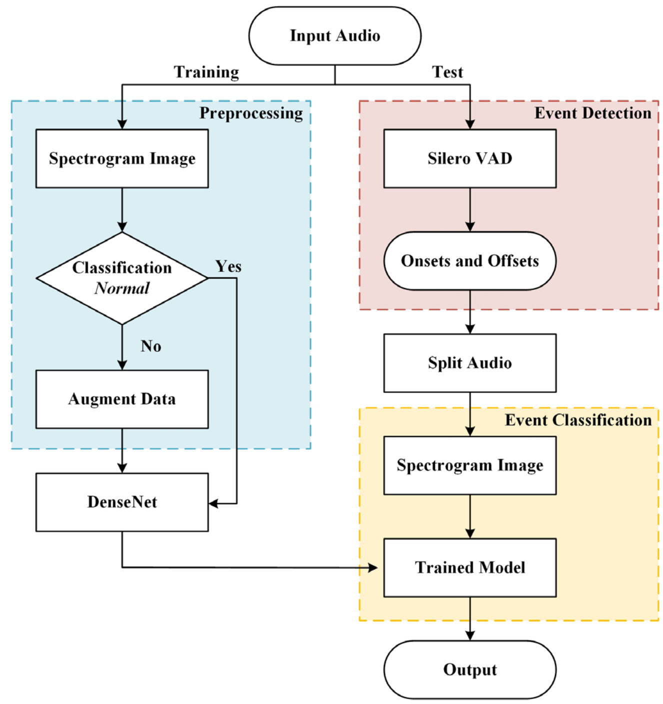

FIG. 6. Proposed three-stage framework for respiratory sound events detection by team IESOC [35].

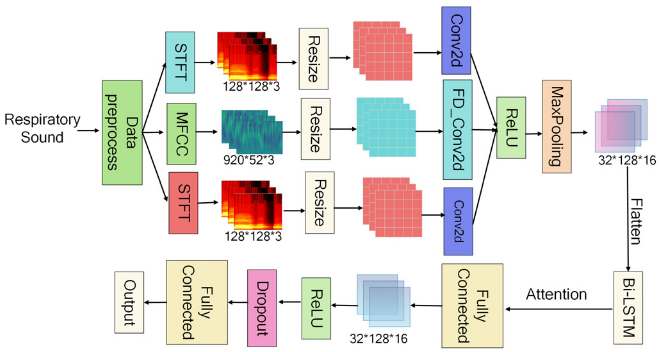

FIG. 7. Proposed two-stage framework for respiratory sound events detection by team Oxy [36].

common events, which helped to improve the model's classification accuracy. Among all teams, they have achieved a total score of -1.027, with an F1 score and an error rate of 0.330 and 1.362, respectively.

2) *Team Oxy* [36] from Xidian University (Chenyang Xu et al.) took second place in track 2 by introducing a novel two-stage framework for detecting respiratory sound events, including the feature extraction and classification stages, as shown in Fig. 7. The efficiency of their feature extraction process lies in the use of a high-pass filter to enhance high-frequency audio components, followed by generating spectrograms with MFCC and STFT that reflect the nuances of human auditory perception. The core of the feature classification stage is a dynamic convolutional recurrent neural network (DCRNN) that adeptly maintains frequency features while processing sound data. They further enhanced the model's predictive performance with attention-based bidirectional long short-term memory (BiLSTM) layers. Furthermore, during training, they introduced the dynamic triple center loss (DTCL) to tackle the class imbalance problem and refine classification accuracy. Evaluated on the blind dataset, their framework achieved a total score of -1.527, with an F1-score and *ER* of 1.563 and 0.034.

DISCUSSION OF RESULTS

The respiratory sound datasets in our contest, including SPRSound, Grand Challenge'23, and Grand Challenge'24 datasets, serve as a benchmark for evaluating the performance of automatic algorithms under various conditions. Participants in the data compression track developed advanced compression algorithms based on auto-encoder that achieved high CRs, significantly reducing the data size of respiratory sound recordings, which, in turn, reduces transmission and storage burdens. In addition, their recognition algorithms were carefully designed to recover the sound signal with minimal differences, ensuring that the integrity of respiratory sounds is preserved after compression. However, there is a tradeoff between achieving higher CRs and maintaining the quality of the reconstructed signal, which poses a challenge in balancing efficiency and accuracy. Besides, these compression algorithms involve numerous layers and complex neuron operations, which poses significant challenges when implementing these algorithms in edge devices, where computational power and resources are often limited. Therefore, future research should focus on developing simplified compression algorithms that are feasible for hardware implementation coupled with highly accurate reconstruction algorithms. In addition, hybrid approaches that combine lossless and lossless compression should be explored to optimize the tradeoff between compression efficiency, computational complexity, and signal fidelity. In addition, developing techniques that integrate domain-specific knowledge, such as the specific features of respiratory sounds that are most crucial for diagnosis, could help to maintain essential signal quality even with high compression.

In the grand challenge on the detection of respiratory sound events, participants introduced diverse innovative solutions. Prominent among these was the use of MFCC and STFT for feature extraction, emphasizing the transformation of audio signals into spectrograms that capture key time-frequency attributes of respiratory sounds. Additionally, the use of VAD for the initial identification of events and advanced deep learning models, including DenseNet and DCRNN, was notably effective for the precise detection of respiratory sound events.

Despite the sophistication of the proposed solutions, two technical constraints were observed in the algorithm design. The algorithms' utilization of a uniform segmentation approach resulted in conspicuous discrepancies in timestamp alignment, failing to accurately reflect the variable durations inherent in respiratory sounds. Consequently, the predicted event lengths were significantly shorter than those of the ground truth, underscoring a critical limitation in the algorithms' temporal resolution. Moreover, the application of feature extraction techniques that are more appropriate for speech analysis introduced inconsistencies within the dataset, detracting from the algorithms' ability to discern the nuanced time-frequency characteristics of respiratory sounds. This mismatch in methodology led to an impaired representation of the acoustic data, further exacerbating the inaccuracy of the classification results.

Future research should focus on refining feature extraction methods to more effectively capture time-domain relationships, potentially enhancing the accuracy of timestamp detection in respiratory sound events. The development of concise, efficient models is essential, ensuring they can be integrated into real-world diagnostic processes without sacrificing accuracy. In addition, exploring hybrid models that integrate various techniques may yield a more robust approach to sound analysis, leading to the creation of more precise and effective diagnostic tools.

CONCLUSION

The Data Compression and Event Detection Grand Challenge 2024 with SPRSound represents another significant step forward in the development of efficient, precise algorithms for digital stethoscopes. By addressing the dual challenges of data storage and transmission, as well as improving the accuracy of respiratory event detection, this initiative fosters innovation that can directly impact clinical practices. The successful participation and advancements presented at the IEEE BioCAS 2024 underline the importance of collaborative efforts in pushing the boundaries of biomedical signal processing, ultimately contributing to better healthcare outcomes.

SOURCE CODE AND SCRIPTS

The IEEE BioCAS 2024 has organized a grand challenge on respiratory sound compression and event detection using the open-sourced SPRSound [21] and Grand Challenge'23 [20] datasets as the training and validation datasets. An additional blind testing (Grand Challenge'24) dataset is reserved to validate the algorithms in this contest. The DataPort links and GitHub repository host the scripts and datasets used in the competition. The scripts serve as foundational frameworks for developing respiratory sound algorithms, providing starter code that supports algorithmic development and validation.

DataPort (Competition and Scripts): https://ieee-dataport.org/competitions/respiratory-sound-track-grand-challenge-2024-respiratory-sound-compression-and-event.

DataPort (Dataset): https://ieee-dataport.org/documents/sprsound-open-source-sjtu-paediatric-respiratory-sound-database.

Github (Dataset and Scripts): https://github.com/SJTU-YONGFU-RESEARCH-GRP/SPRSound.

ACKNOWLEDGMENT

Q. Zhang, C. Chen, and S. Yuan are the main authors of this work with data collection, data cleaning, and preparing the data for publication. Q. Zhang, C. Chen, S. Yuan, J. Zhang, J. Yuan, Y. Zhang, R. Pan, X. Jiang, J. Zhao, Y. Li, Y. Yin, L. Zhao, G. Wang, and Y. Lian have designed the grand challenge. All authors reviewed the manuscript.

The authors have declared no conflicts of interest.

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 *Proc. Int. Conf. Biomed. Health Inform.*, 2017, pp. 33–37.
[3] B. M. Rocha et al., "An open access database for the evaluation of respiratory sound classification algorithms," *Physiol. Meas.*, vol. 40, no. 3, 2019, Art. no. 035001.
[4] Y. Ma et al., "Live Demo: LungSys—Automatic digital stethoscope system for adventitious respiratory sound detection," in *Proc. IEEE Biomed. Circuits Syst. Conf.*, 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, pp. 1–19, 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 *Proc. 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 Biomed. Circuits Syst. Conf.*, 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 Trans. Biomed. Circuits Syst.*, vol. 14, no. 3, pp. 535–544, Jun. 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, H. Phan, R. Palaniappan, A. Mertins, and I. McLoughlin, "CNN-MoE based framework for classification of respiratory anomalies and lung disease detection," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 8, pp. 2938–2947, Aug. 2021.
[11] T. Nguyen and F. Pernkopf, "Crackle detection in lung sounds using transfer learning and multi-input convolutional neural networks," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2021, pp. 80–83.
[12] S. Ntalampiras and I. Potamitis, "Automatic acoustic identification of respiratory diseases," *Evolving Syst.*, 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," *J. Ambient Intell. Humanized Comput.*, vol. 13, no. 10, pp. 1–13, 2021.
[14] Mohammad Fraiwan et al., "A dataset of lung sounds recorded from the chest wall using an electronic stethoscope," *Data Brief*, vol. 35, 2021, Art. no. 106913.
[15] J. Li et al., "LungAttn: Advanced lung sound classification using attention mechanism with dual TQWT and triple STFT spectrogram," *Physiol. Meas.*, vol. 42, no. 10, 2021, Art. no. 105006.
[16] H. Zhu et al., "Automatic pulmonary auscultation grading diagnosis of Coronavirus Disease 2019 in China with artificial intelligence algorithms: A cohort study," *Comput. Methods Programs Biomed.*, vol. 213, 2022, Art. no. 106500.
[17] J. Li et al., "Explainable CNN with fuzzy tree regularization for respiratory sound analysis," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 6, pp. 1516–1528, Jun. 2022.
[18] S. B. Shuvo, S. N. Ali, S. I. Swapnil, T. Hasan, and M. I. H. Bhuiyan, "A lightweight CNN model for detecting respiratory diseases from lung auscultation sounds using EMD-CWT-based hybrid scalogram," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 7, pp. 2595–2603, Jul. 2021.
[19] Q. Zhang et al., "Grand challenge on respiratory sound classification for SPRSound dataset," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Taipei, Taiwan, 2022, pp. 213–217.
[20] Q. Zhang et al., "Grand challenge on respiratory sound classification for SPRSound dataset," in *IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Toronto, ON, Canada. Piscataway, NJ, USA: IEEE, 2023, pp. 1–5.
[21] Q. Zhang et al., "SPRSound: Open-source SJTU paediatric respiratory sound database," *IEEE Trans. Biomed. Circuits Syst.*, vol. 16, no. 5, pp. 867–881, Oct. 2022.
[22] N. Babu, J. Kumari, J. Mathew, U. Satija, and A. Mondal, "Multiclass categorisation of respiratory sound signals using neural network," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, 2022, pp. 228–232.
[23] L. Zhang, Y. Zhu, S. Tu, and L. Xu, "A feature polymerized based two-level ensemble model for respiratory sound classification," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, 2022, pp. 238–242.
[24] J. Li, X. Wang, X. Wang, S. Qiao, and Y. Zhou, "Improving the ResNet-based respiratory sound classification systems with focal loss," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, 2022, pp. 223–227.
[25] W.-B. Ma, X.-Y. Deng, Y. Yang, and W.-C. Fang, "An effective lung sound classification system for respiratory disease diagnosis using

DenseNet CNN model with sound pre-processing engine," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, 2022, pp. 218–222.

[26] D. Pessoa et al., "BRACETS: Bimodal repository of auscultation coupled with electrical impedance thoracic signals," *Comput. Methods Programs Biomed.*, vol. 240, 2023, Art. no. 107720.

[27] S. Ntalampiras, "Explainable siamese neural network for classifying pediatric respiratory sounds," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 10, pp. 4728–4735, Oct. 2023.

[28] C. Chen, Q. Zhang, S. Sheng, H. Huang, Y. Zhang, and Y. Li, "LungHeart-AtMe: Adventitious cardiopulmonary sounds classification using MMoE with STFT and MFCCs spectrograms," in *Proc. IEEE Int. Conf. Artif. Intell. Circuits Syst. (AICAS)*, 2023, pp. 1–5.

[29] B. TaghiBeyglou et al., "Fusion of manual and deep learning analyses for automatic lung respiratory sounds identification in youth," *CMBES Proc.*, vol. 45, 2023.

[30] F.-S. Hsu et al., "A dual-purpose deep learning model for auscultated lung and tracheal sound analysis based on mixed set training," *Biomed. Signal Process. Control*, vol. 86, 2023, Art. no. 105222.

[31] J. Ebbers, R. Haeb-Umbach, and R. Serizel, "Threshold independent evaluation of sound event detection scores," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, 2022, pp. 1021–1025.

[32] G. Ferroni et al., "Improving sound event detection metrics: Insights from DCASE 2020," 2020, *arXiv:2010.13648*.

[33] J. Zhai et al., "Compression of respiratory sound signals via TranscoderQVAE," in *Proc. IEEE Biomed. Circuits Syst. Conf.*, 2024, pp. 1–5.

[34] S. Tao et al., "Convolutional auto-encoder for variable length respiratory sound compression and reconstruction," in *Proc. IEEE Biomed. Circuits Syst. Conf.*, 2024, pp. 1–5.

[35] C. Yeh et al., "Paediatric respiratory sound event detection based on voice activity detection," in *Proc. IEEE Biomed. Circuits Syst. Conf.*, 2024, pp. 1–5.

[36] C. Xu et al., "Enhancing respiratory sound detection through integrated learning and dynamic convolutional recurrent neural networks," in *Proc. IEEE Biomed. Circuits Syst. Conf.*, 2024, pp. 1–5.