



Active Learning for Abnormal Lung Sound Data Curation and Detection in Asthma

Shabnam Ghaffarzadegan¹, Luca Bondi¹, Ho-Hsiang Wu¹, Sirajum Munir¹, Kelly J. Shields², Samarjit Das¹, Joseph Aracri³

¹Bosch Center for Artificial Intelligence-USA, ²Highmark Health, ³Allegheny Health Network

{shabnam.ghaffarzadegan, luca.bondi, ho-hsiang.wu, sirajum.munir, samarjit.das}@us.bosch.com, kelly.shields@highmarkhealth.org, joseph.aracri@ahn.org

Abstract

Existing audio-based asthma monitoring solutions rely on feature engineering designs paired with contact-based auscultation which are brittle in practice and do not scale beyond point of care setups. Data-driven methods utilizing contactless microphones have the potential to address such limitations. These solutions are under-explored in healthcare due to high cost of data curation requiring physicians-in-the-loop. Here, we propose an active learning (AL) system to facilitate audio data collection and annotation. It detects lung sound abnormalities in asthma. AL reduces the annotation cost while increasing the model performance under a constrained annotation budget. It automatically extracts interesting audio segments from the continuous recordings, and efficiently annotates and trains anomaly detector model. The experimental results confirm the effectiveness of the proposed system as an enabler for larger scale data curation on a newly collected audio corpus for pediatric asthma.

Index Terms: asthma, audio event detection, active learning, data annotation, healthcare

1. Introduction

Asthma is a chronic respiratory condition affecting more than 300 million children and adults worldwide [1] with a high hospitalization rate during acute episodes. Historically, the definition of asthma has been subjective [2]; however, there is an agreement in the physiological response including a tightening of the muscles around small airways combined with inflammation in the lung leading to airway obstruction. Common asthma symptoms include coughing, wheezing, shortness of breath, and chest tightness. These symptoms may appear intermittently, often worse at night, and are triggered by various conditions such as exercise, allergen and irritant exposure, viral infections (e.g. colds), changes in weather, etc. [3].

Diagnosis of asthma is often based on assessing a patient's medical history, identifying patterns of respiratory symptoms, and more recently blood biomarkers. Additionally, clinicians may utilize a questionnaire, which is simple and convenient, to aid in an asthma diagnosis. World-wide initiatives including the Global Initiative of Asthma (GINA) are using sample questionnaires to collect data on the frequency and severity of lung sound abnormalities (e.g. wheezing and coughing) [3]. Despite health records, questionnaires, and patients' descriptions, a true asthma diagnosis remains challenging due to a lack of accepted universal rules. An asthma diagnosis may vary due to clinicians' experience level, geographic area, and due to the disease and manifestations changing over time. A true asthma diagnosis is especially challenging in young children where questionnaires cannot be used as a reliable tool of choice [2].

Though asthma cannot be cured, a timely diagnosis along

with proper medications and awareness can control the disease [1]. However, untreated asthma lessens a patient's quality of life and can lead to major healthcare costs. Self-awareness and monitoring are an asthma patient's best tools in reducing the impact of the disease on their life. Self-monitoring is subjective, particularly with patients that have caregivers (e.g. children), and can lead to a misdiagnosis and ultimately poor treatment. An automatic, data-driven, at-home asthma condition monitoring system that is non-invasive, cost-effective, and patient-friendly can play an important role in disease control and promote remote medical care which has become increasingly popular since the COVID-19 Pandemic [4].

The usage of data-driven approaches in asthma detection can be categorized into four major groups: screening and diagnosis, patient classification, monitoring, and treatment. Various data resources and modalities have been utilized for each including lung sounds from auscultation, medical history documents, patterns of symptoms, environmental data, and questionnaires. Most studies in this space utilize smaller datasets due to the difficulty in collecting data on the target population and the cost in terms of time and resources. Developing robust traditional machine learning applications or deep-learning based approaches to flag an asthma condition becomes difficult with such sparse data, and poses a critical concern of an experiment's validity and generalizability [5]. Audio data collection can be non-invasive, cost effective, and patient friendly. Hence, a non-contact, automatic lung sound abnormality assessment system can be seamlessly integrated into the patient's life to simplify at-home self-monitoring and reduce human errors.

Existing literature in lung sound recognition and characterization includes algorithms, pipelines, and methodologies for cough detection [6, 7], wheeze recognition [8–10], respiration phase characterization [11, 12], onset detection of drug inhalation [13], and asthma detection using speech signals [14, 15]. Most studies propose a combination of normalization, filtering, and pre-processing to emphasize the sound of interest within the recorded respiratory sounds. Mel Frequency Cepstral Coefficients (MFCCs) and Linear Predictive Coding (LPC) coefficients are typically employed as features. Most recent works rely on supervised learning techniques to train linear classifiers [8, 16], Gaussian models [17], or shallow neural networks [18, 19] tailored to the downstream task [5, 10, 13, 20, 21].

Although early studies demonstrate the existence of asthma fingerprints in audio signals, the experiments are typically performed on small datasets, with limited scope, and with specialized equipment (e.g. stethoscopes or wearable devices), which becomes a challenge to scale to an in-home concept that could seamlessly integrate into a patient's life. Moreover, through our preliminary work, we have observed radically different lung sounds when recorded via a non-contact microphone versus a

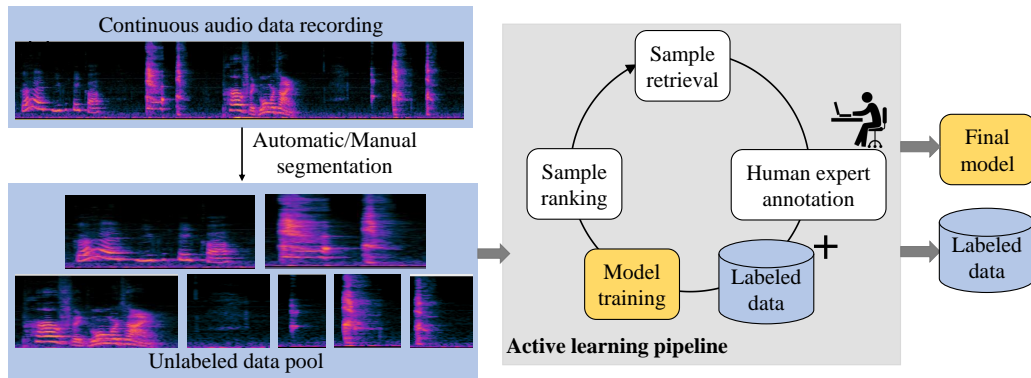


Figure 1: Proposed system for large scale audio data collection and annotation via active learning framework.

stethoscope. In particular, for pediatric patients, soft sounds such as breathing are almost indistinguishable within the background noise when collected with a non-contact microphone, while wheezing sounds are also not as well identifiable when compared to contact stethoscope acquisitions.

Despite the clear advantage of creating an automatic lung sound assessment system for asthma monitoring, the field remains under-explored due to a lack of large scale data availability. Collecting lung sound data from asthmatic patients was limited in the literature due to: 1) the required expertise and complexity of the necessary hardware, often in the form of multiple contact sensors or stethoscope auscultation, 2) the chronic nature of asthma which requires long-term data recordings, and 3) expensive data annotation cost of the rare lung sound abnormalities during long recordings, requiring the expertise of specialized physicians or trained individuals. Increasing the potential of audio-based data-driven approaches to support clinical decision making in healthcare applications hinges on collecting and annotating new datasets in a reasonable time frame from large and diverse populations is critical to train robust and generalizable models.

Motivated by this, we propose a system to enable large audio data collection and annotation based on active learning (AL) for assessing lung sound abnormalities. The proposed system is made of three main components: 1) a pool of unlabeled data from continuous audio recordings – from a controlled or in-the-wild data collection setup; 2) automatic data segmentation to extract audio chunks containing relevant audio events or sounds of interest from the unlabeled data stream; 3) an AL pipeline to efficiently annotate the data while reducing the human effort and increasing the modeling performance under a constrained annotation budget. While we perform the experiments on a pool of audio data collected for a pediatric asthma use case, we expect the proposed framework to enable more scalable data collection and annotation in other audio abnormality detection tasks in the healthcare domain. To the best of our knowledge, this is the first work utilizing active learning framework in audio health care applications.

2. Proposed System

Figure 1 shows the overall proposed system for enabling large scale audio data collection and annotation for healthcare applications via an Active Learning (AL) framework. The first step of the system includes continuous audio data collection in-the-wild (e.g. at-home, doctor’s office) using far-field non-contact

microphone(s) to reduce the need for specialized sensors (e.g. stethoscope). Next, depending on the dataset size, we manually segment the continuous data into audio chunks or utilize pre-trained models to automatically extract the desired segments. Finally, we use the generated unlabeled pools of audio samples in the AL pipeline to efficiently annotate the target audio events thus capitalizing on expert annotators’ (e.g. physicians) time with a limited annotation budget. We explain the details of each component below. We detail each component in the next few sections.

2.1. Automatic Segmentation

During automatic segmentation, we use an existing pre-trained model to extract target audio events from a stream of data for later processing steps. For this study, we use pre-trained audio neural networks (PANNs) [22] to propose segments for our desired events. PANNs provide convolutional neural networks (CNNs) based architectures and are trained (supervised) on AudioSet [23], a massive general audio dataset containing 1.9 million audio clips with an ontology of 527 sound classes, including some relevant to clinical biomarkers (e.g. cough, breath, heart sound) We adopt the pre-trained model *Cnn14_DecisionLevelMax*, outputting frame-wise sound event detection (SED) scores for all 527 classes every 10ms. We take the predictions of our target events, apply a threshold for binary predictions, and group consecutive positive predictions to acquire the information of onsets and offsets for each segment of the desired sound events.

2.2. Efficient Data Annotation and Anomaly Detection with Active Learning

Active learning (AL) is a machine learning solution that prioritizes the data to be labeled for optimizing the labeling cost while increasing the model robustness for use cases in which data annotation is costly and time consuming. In this current study, the labeling budget is the number of samples that human annotators can manually label. Hence, AL frameworks develop labeled training data in an iterative manner with less manual annotation effort leading to a more robust model.

AL has been previously used in various audio and speech tasks, from automatic speech and emotion recognition, to sound event classification [24–28]. The applications of AL in the medical domain are more limited to a few works on medical imaging and medical text classification [29–31]. To our knowledge, AL has not been explored in audio-enabled healthcare applications,

where the cost of data annotation is particularly high.

As shown in Fig. 1, an AL framework consists of the following iterative steps: 1) Applying an unsupervised/semi-supervised ranking method to sort unlabeled data samples by assigning higher scores to more informative samples; 2) Retrieving a subset of the ranked samples based on the available annotation budget, B , of the expert human annotator; 3) Annotating the retrieved samples by human expert(s); 4) Training or fine-tuning a classification or anomaly detection model based on the collected annotations and the available unlabeled data; 5) Re-iterating steps 1 to 4 until the desired model performance is achieved. In healthcare applications, we often look for the presence of anomalies in the audio data (e.g., the presence of wheezing during cough, murmur in heart sound). To aid in this type of detection, we integrated a few-shot anomaly detection method in the AL pipeline as detailed below.

2.2.1. Anomaly detection with Deviation network

We use a Deviation Network (DevNet) originally introduced in [32] as a few-shot anomaly detection method for image applications. DevNet is a semi-supervised method that predicts anomaly scores in an end-to-end manner versus previous works that learn representative features to separate normal vs abnormal examples [33], which makes DevNet a great sampling strategy and a strong candidate to be used in an AL pipeline. DevNet deploys a neural anomaly score learner network to assign anomaly scores to each sample. The network is first trained unsupervisedly by assigning normal scores to all the training samples. The experiments in [32] confirm the robustness of DevNet with respect to anomaly contamination in the training set during the initial step. Next, it defines an average anomaly score of the normal data based on prior probabilities (e.g., Gaussian distribution) as a reference score to guide the subsequent anomaly score learner. Finally, DevNet introduces a bias loss to enforce statistically significant deviation of anomaly scores of the abnormal data, while having the scores of the normal data close to the mean of the Gaussian prior.

3. Data collection and experiments

To evaluate the performance of the proposed system, we have collected clinical pediatric audio data for identifying pediatric asthma. The experiments are designed to detect cough instances and the presence of wheezing within the cough as two main characteristics of asthma within the collected respiratory sounds. The experiments assess the efficiency of automatic data segmentation and AL pipeline for the target sound events.

3.1. Data Collection

The audio samples utilized in this study are a subset of the collected data in collaboration with Highmark Health and Allegheny Health Network (AHN) after the institutional review board (IRB) approval. The data is collected in a physician’s office after the patient’s scheduled appointment, under typical noise conditions of a healthcare facility. All participants’ caregivers gave consent prior to enrolling. The corpus contains samples from 8 healthy (non-asthma), 10 non-asthma with a respiratory cold, and 9 with diagnosed asthma between 3 and 11 years old. The participants with asthma are on treatment medications. No other co-morbidities were present in the participants. Additional information such as height, weight, BMI, temperature, medication, and gender are also available for each participant.

Audio data were recorded at a sampling rate of 44.1 kHz

for a range of tasks with a portable hardware setup of a far-field microphone connected to a laptop and placed on a desk ~ 0.2 m away from each participant. The data collection is divided into 5 tasks: (1) reciting ABCs, (2) reciting numbers one to ten, (3) pronouncing the vowel /ah/, (4) naming two favorite colors, (5) naming two favorite foods, (6) coughing, (7) deep breathing, and (8) sitting silent. During deep breathing and sitting silent, a digital stethoscope, EKO¹, is used to collect ground truth data accompanying microphone data. Doctors performed each recording by instructing each participant to perform each task while the participant was accompanied by a caregiver.

The audio recordings are segmented and labeled for each task by two human annotators. The quality of the segments is then verified by an independent annotator. The obtained segments have variable lengths depending on the nature of the task and the individual participants. In a second round of data annotation, a pediatric specialist labeled the cough samples for the presence of wheezing. After data labeling and validation, in total we obtain 152 individual cough segments from which 18 segments contain wheezing, 41 individual /ah/ segments, 23 breathing segments, 25 ABC segments, 26 counting to ten segments, 26 naming two favorite food segments, 26 naming two favorite color segments, and 22 silent segments.

3.2. Evaluation of the Automatic Segmentation Method

In this section, we evaluate the effectiveness of the automatic segmentation method using the PANNs pre-trained model on collected data as well as a synthesized cough dataset to simulate the use case of health monitoring in-the-wild.

Table 1: *SED results for the collected data and synthesized soundscapes in-the-wild.*

Data	Event	Precision	Recall	F1-score
Collected data	Cough	0.71	0.89	0.79
Soundscapes		0.94	0.41	0.57

To generate the synthesized data, we obtained cough samples from the COUGHVID dataset [34], taking those shorter than 5 seconds with confidence (*cough_detected*) higher than 0.85, resulting in ~ 2 k samples as foreground events. We also utilized the Domestic Environment Sound Event Detection (DESED) evaluation dataset² [35] for background scenes. The DESED data contains mixtures of sounds present in daily life (e.g. alarms, dishes, vacuums, etc.) to mimic the domestic environments in which cough events may take place when actually monitoring. We generate a total of 2,000 soundscape mixtures, 1,000 mixtures for validation and 1000 mixtures for evaluation. We use separate foreground and background sounds to generate each set. We utilize *scaper*³ with default configurations as provided by the tutorial so that each soundscape contains 1 to 3 cough events. We then compute segment-based sound event detection metrics for cough sound prediction using *sed_eval*⁴ with *time_resolution* = 1.0.

To determine the prediction probability threshold, we use the synthesized validation set, and sweep between threshold values from 0.05 to 0.9, with a step size 0.05. Subsequently, we

¹<https://www.ekohealth.com/>

²<https://zenodo.org/record/5524373>

³<https://github.com/justinsalamon/scaper>

⁴https://github.com/TUT-ARG/sed_eval

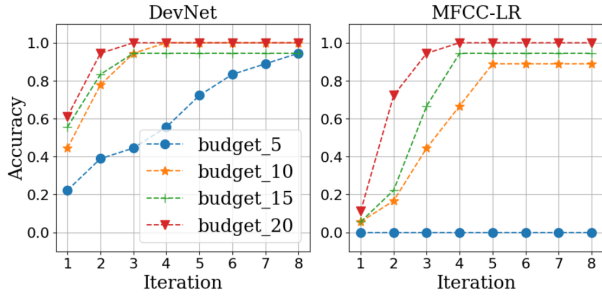


Figure 2: Accumulated accuracy rate of the learned anomaly detection model as function of AL iteration and annotation budget.

find that $threshold = 0.1$ yields the best F1-score and use this threshold for both the collected and the synthesized evaluation data.

As per Table 1, the selected pre-trained audio model performs well (both precision and recall) on the collected data which was collected in a controlled environment. This indicates that our proposed model is a good candidate for the initial segmentation of the collected cough events. Looking at the cough events detected in the synthesized soundscapes in a domestic environment, we see that the precision is still good. However, the recall drops significantly. This is mostly caused by the challenging scenario of overlapping sounds occurring simultaneously in daily life, as well as other objects sounding similar to the impulse characteristics of coughing. Nonetheless, the analysis provides a clear path forward to addressing this challenge in future work.

3.3. Evaluation of the Active Learning Pipeline

Baseline As a baseline system for ranking and training steps of the AL pipeline, we extracted MFCCs from the raw audio signal. The audio is downsampled to 16 kHz and MFCCs are extracted with a Hamming window size of 512, a hop size of 160 samples, 20 MFCC coefficients, and 50 Hz and 8 kHz lower and upper cut-off frequencies. We expose outliers with a K-Means clustering algorithm for unsupervised ranking in the first iteration of the AL pipeline, and a Linear Regression (LR) model as supervised training and ranking method in later iterations, when labeled samples are available.

Devnet The implementation of DevNet is taken from its official release⁵ with the variation of substituting the backbone architecture with an audio pre-trained model. We use PANNs backbone [22] without the last fully connected layer as a feature extractor, specifically the *CNN14* model with a 16 kHz sampling rate. The input of PANNs is a log Mel spectrogram extracted from the raw audio signal with a Hamming window size of 512, a hop size of 160 samples, 64 Mel filter banks, and 50 Hz and 8 kHz lower and upper cut-off frequencies. DevNet learns the anomaly scores by passing the feature map to a 1×1 convolutional layer, followed by a top-K Multiple Instance Learning (MIL)-based anomaly score optimization using the proposed deviation loss [32]. As a result of the deviation loss, during the inference phase the final model assigns large anomaly scores to the samples as long as their features deviate significantly from the normal samples' features.

⁵<https://github.com/Choubo/deviation-network-image>

The results of the AL pipeline, illustrated in Fig. 2, show the accuracy rate of retrieving wheezing events, which comprises 11% of the collected cough samples, as a function of the AL iteration and annotation budget for two training methods based on DevNet and MFCC-LR. Annotation "budget_x" in the plots indicates the human annotator only labels "x" samples in each AL iteration. Note that in the MFCC-LR plot, the reported numbers for iteration 1 are based on outlier exposure via the K-Means clustering approach with MFCC features, due to a lack of available annotated data. Both DevNet and MFCC-LR based methods can fully retrieve wheezing samples without annotating the entire dataset. Though, DevNet can reach a higher retrieval accuracy (defined as the accumulative retrieved wheeze samples/total number of wheeze samples) in fewer iterations and, thus, a lower total annotation cost.

For the collected dataset, the MFCC-LR method has an annotation cost of around 40%, meaning that we need to annotate approximately 40% of the entire dataset to retrieve each of the existing wheeze samples. Conversely, the DevNet method has an annotation cost of approximately 26% of the data size for complete retrieval. Note that the accuracy line of the MFCC-LR method with "budget_5" remains 0 throughout the iterations, indicating the model was not able to retrieve any wheeze data in the first iteration and hence could not improve the performance in later AL iterations. For comparison, we calculate the annotation cost of random sampling with the annotation budgets of 5, 10, 15, and 20 samples at each AL iteration. The random sampling method requires a minimum labeling of 98% of the data to retrieve each wheezing instance. The random sampling experiments are repeated 20 times.

The above experiments show the effectiveness of the automatic audio segmentation and AL learning pipeline in efficiently labeling and detecting anomalous events in the data, hence enabling larger scale audio data collection for healthcare applications moving forward.

Next Steps We plan to expand our work by enhancing the automatic segmentation approach in the presence of background noise. We would like to investigate the presence of asthma audio fingerprints in the other tasks of the collected data such as speech and breathing. Finally, we would like to incorporate the proposed system for a large-scale audio data collection in-the-wild beyond the controlled environment of a physician's office.

4. Conclusion

Timely diagnosis of asthma combined with continuous monitoring may have a significant impact on disease management and lessen the burden on a patient's life. Audio-based monitoring solutions are great candidates to be integrated into a general monitoring system considering the abnormal, symptomatic respiratory sounds in asthma. However, these solutions are still under-explored due to a lack of available datasets to train robust and generalizable data-driven models. Here, we proposed a data collection and annotation strategy to enable large scale, clinical audio data curation using an Active Learning framework based on non-contact microphone(s) recordings. Two main components of this system are automatic data segmentation to extract target segments from the recorded audio stream and an Active Learning pipeline for efficient data annotation and anomaly detection. We evaluated the system on collected audio data for pediatric asthma. The experimental results show the effectiveness of the proposed approach in reducing human annotation effort while retrieving target anomalous sounds.

5. References

- [1] "World health organization: Asthma," <https://www.who.int/news-room/fact-sheets/detail/asthma>, accessed: 2023-02-20.
- [2] M. I. Asher, L. García-Marcos, N. E. Pearce, and D. P. Strachan, "Trends in worldwide asthma prevalence," *European Respiratory Journal*, vol. 56, no. 6, 2020.
- [3] "Global initiative for asthma (2020). global strategy for asthma management and prevention." https://ginasthma.org/wp-content/uploads/2020/06/GINA-2020-report_20.06.04-1-wms.pdf, accessed: 2023-02-20.
- [4] J. Shaver, "The state of telehealth before and after the covid-19 pandemic," *Prim Care*, 2022.
- [5] K. P. Exarchos, M. Beltsiou, C.-A. Votti, and K. Kostikas, "Artificial intelligence techniques in asthma: a systematic review and critical appraisal of the existing literature," *European Respiratory Journal*, vol. 56, no. 3, 2020.
- [6] U. Koehler, O. Hildebrandt, A. Weissflog, A. Zacharasiewicz, K. Sohrabi, N. Koehler, and V. Gross, "Leosound - a new device for long-term recording of wheezing and cough in pediatric and adult patients with asthma (during sleep)," *Clinical Investigation*, vol. 08, 01 2018.
- [7] H. I. Hee, B. Balamurali, A. Karunakaran, D. Herremans, O. H. Teoh, K. P. Lee, S. S. Teng, S. Lui, and J. M. Chen, "Development of machine learning for asthmatic and healthy voluntary cough sounds: A proof of concept study," *Applied Sciences*, vol. 9, no. 14, 2019.
- [8] S. Z. H. Naqvi, M. Arooj, S. Aziz, M. U. Khan, M. A. Choudhary, and M. N. ul. Hassan, "Spectral analysis of lungs sounds for classification of asthma and pneumonia wheezing," in *2020 International Conference on Electrical, Communication, and Computer Engineering (ICEECE)*, 2020, pp. 1–6.
- [9] Y. Shabtai-Musih, J. B. Grotberg, and N. Gavriely, "Spectral content of forced expiratory wheezes during air, he, and sf6 breathing in normal humans." *Journal of applied physiology*, vol. 72 2, pp. 629–35, 1992.
- [10] R. X. A. Pramono, S. A. Imtiaz, and E. Rodriguez-Villegas, "Evaluation of features for classification of wheezes and normal respiratory sounds," *PLOS ONE*, vol. 14, no. 3, pp. 1–21, 03 2019.
- [11] D. Skalicky, V. Koucky, D. Hadraba, M. Vitezniak, M. Dub, and F. Lopot, "Detection of respiratory phases in a breath sound and their subsequent utilization in a diagnosis," *Applied Sciences*, vol. 11, no. 14, 2021.
- [12] S.-H. Li, B.-S. Lin, C.-H. Tsai, C.-T. Yang, and B.-S. Lin, "Design of wearable breathing sound monitoring system for real-time wheeze detection," *Sensors*, vol. 17, no. 1, 2017.
- [13] D. Nikos Fakotakis, S. Nousias, G. Arvanitis, E. I. Zacharaki, and K. Moustakas, "AI sound recognition on asthma medication adherence: Evaluation with the RDA benchmark suite," *IEEE Access*, vol. 11, pp. 13 810–13 829, 2023.
- [14] M. Z. Alam, A. Simonetti, R. Brillantino, N. Tayler, C. Grainge, P. Siribaddana, S. A. R. Nouraei, J. Batchelor, M. S. Rahman, E. V. Mancuzo, J. W. Holloway, J. A. Holloway, and F. I. Rezwani, "Predicting pulmonary function from the analysis of voice: A machine learning approach." *Front Digit Health*, vol. 4, p. 750226, 2022.
- [15] M. A. Iqbal, K. Devarajan, and S. M. Ahmed, "Real time detection and forecasting technique for asthma disease using speech signal and denn classifier," *Biomedical Signal Processing and Control*, vol. 76, p. 103637, 2022.
- [16] D. Emmanouilidou, E. D. McCollum, D. E. Park, and M. Elhilali, "Computerized Lung Sound Screening for Pediatric Auscultation in Noisy Field Environments," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 7, pp. 1564–1574, Jul. 2018.
- [17] B.-S. Lin, H.-D. Wu, and S.-J. Chen, "Automatic wheezing detection based on signal processing of spectrogram and back-propagation neural network." *Journal of healthcare engineering*, vol. 6 4, pp. 649–72, 2015.
- [18] Y. Kim, Y. Hyon, S. S. Jung, S. Lee, G. Yoo, C. Chung, and T. Ha, "Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning," *Scientific Reports*, vol. 11, no. 1, p. 17186, Aug. 2021.
- [19] H.-C. Kuo, B.-S. Lin, Y.-D. Wang, and B.-S. Lin, "Development of Automatic Wheeze Detection Algorithm for Children With Asthma," *IEEE Access*, vol. 9, pp. 126 882–126 890, 2021.
- [20] H. Hafke-Dys, B. Kuźnar-Kamińska, T. Grzywalski, A. Maciaszek, K. Szarzynski, and J. Kocinski, "Artificial intelligence approach to the monitoring of respiratory sounds in asthmatic patients," in *Frontiers in Physiology*, 2021.
- [21] S. M. Shaharum, K. Sundaraj, S. Aniza, R. Palaniappan, and K. Helmy, "Classification of asthma severity levels by wheeze sound analysis," in *2016 IEEE Conference on Systems, Process and Control (ICSPC)*, 2016, pp. 172–176.
- [22] Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. D. Plumbley, "Panns: Large-scale pretrained audio neural networks for audio pattern recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2880–2894, 2020.
- [23] J. F. Gemmeke, D. P. Ellis, D. Freedman, A. Jansen, W. Lawrence, R. C. Moore, M. Plakal, and M. Ritter, "Audio set: An ontology and human-labeled dataset for audio events," in *2017 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2017, pp. 776–780.
- [24] Y. Wang, A. E. Mendez Mendez, M. Cartwright, and J. P. Bello, "Active learning for efficient audio annotation and classification with a large amount of unlabeled data," in *ICASSP 2019*, 2019, pp. 880–884.
- [25] M. Kholghi, Y. Phillips, M. Towsey, L. Sitbon, and P. Roe, "Active learning for classifying long-duration audio recordings of the environment," *Methods in Ecology and Evolution*, vol. 9, no. 9, pp. 1948–1958, 2018.
- [26] Z. Shuyang, T. Heittola, and T. Virtanen, "Active learning for sound event detection," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2895–2905, 2020.
- [27] D. Hakkani-Tür, G. Riccardi, and A. Gorin, "Active learning for automatic speech recognition," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 4, 2002, pp. IV–3904–IV–3907.
- [28] M. Abdelwahab and C. Busso, "Active learning for speech emotion recognition using deep neural network," in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, 2019, pp. 1–7.
- [29] S. Budd, E. C. Robinson, and B. Kainz, "A survey on active learning and human-in-the-loop deep learning for medical image analysis," *Medical Image Analysis*, vol. 71, p. 102062, 2021.
- [30] M. Kholghi, L. Sitbon, G. Zuccon, and A. Nguyen, "Active learning: a step towards automating medical concept extraction," *Journal of the American Medical Informatics Association*, vol. 23, no. 2, pp. 289–296, 08 2015.
- [31] Y. Chen, S. Mani, and H. Xu, "Applying active learning to assertion classification of concepts in clinical text," *Journal of Biomedical Informatics*, vol. 45, no. 2, pp. 265–272, 2012.
- [32] G. Pang, C. Shen, and A. van den Hengel, "Deep anomaly detection with deviation networks," in *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, 2019, pp. 353–362.
- [33] G. Pang, L. Cao, L. Chen, and H. Liu, "Learning representations of ultrahigh-dimensional data for random distance-based outlier detection," in *24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, ser. KDD '18, 2018, p. 2041–2050.
- [34] L. Orlandic, T. Teijeiro, and D. Atienza, "The coughvid crowd-sourcing dataset, a corpus for the study of large-scale cough analysis algorithms," *Scientific Data*, vol. 8, no. 1, p. 156, 2021.
- [35] N. Turpault, R. Serizel, J. Salamon, and A. P. Shah, "Sound event detection in domestic environments with weakly labeled data and soundscape synthesis," *DCASE*, 2019.