2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10340579

Constrained Synthetic Sampling for Augmentation of Crackle Lung Sounds

Annapurna Kala[1] and Mounya Elhilali[1]

Abstract— **Crackles are explosive breathing patterns caused by lung air sacs filling with fluid and act as an indicator for a plethora of pulmonary diseases. Clinical studies suggest a strong correlation between the presence of these adventitious auscultations and mortality rate, especially in pediatric patients, underscoring the importance of their pathological indication. While clinically important, crackles occur rarely in breathing signals relative to other phases and abnormalities of lung sounds, imposing a considerable class imbalance in developing learning methodologies for automated tracking and diagnosis of lung pathologies. The scarcity and clinical relevance of crackle sounds compel a need for exploring data augmentation techniques to enrich the space of crackle signals. Given their unique nature, the current study proposes a crackle-specific constrained synthetic sampling (CSS) augmentation that captures the geometric properties of crackles across different projected object spaces. We also outline a task-agnostic validation methodology that evaluates different augmentation techniques based on their goodness of fit relative to the space of original crackles. This evaluation considers both the separability of the manifold space generated by augmented data samples as well as a statistical distance space of the synthesized data relative to the original. Compared to a range of augmentation techniques, the proposed constrained-synthetic sampling of crackle sounds is shown to generate the most analogous samples relative to original crackle sounds, highlighting the importance of carefully considering the statistical constraints of the class under study.**

I. Introduction

Recent advances in deep learning methods have enabled unprecedented breakthrough in tackling complex problems from healthcare to autonomous systems and human-computer interactions. But deep learning is data-hungry [1], [2]. Access to large amounts of data enables learning systems to infer abstractions, semantic rules and generalize to different scenarios and contexts. Yet, access to data is a continued challenge due to issues with collection, curation, annotation as well as data imbalance. Data augmentation has been adopted as a way to broaden access to more data by including new artificial data points to the original dataset [3], [4]. This additional data is either a modified form of the original data or an artificially-produced data using generative rules.

Healthcare applications are a prime example of challenging domains where data access and uniformity across scenarios continue to hamper progress to develop accurate automated screening and diagnostics [5], [6]. In case of auscultation data, identifying abnormal sounds from body organs offer a cheap and non-invasive tool for diagnosis of a plethora of diseases (pneumonia, asthma, Bronchiolitis, heart murmurs) [7]. Like other learning systems, Computerized Auscultation Analysis (CAA) develops models to leverage information from auscultation data in order to identify adventitious patterns in body sounds indicative of abnormal conditions[8]. Crackle sounds are specific patterns in lung sounds that are caused by fluid filling lung air sacs and associated with popping or snapping breathing sounds. A number of studies have linked presence of crackle sounds to pneumonia-related mortatility rates, COPD, bronchitis or even heart failure [9], [10], [11]. Despite their pathophysiological importance, crackles are extremely short bursts that are infrequent relative to other breathing sounds.

Given their clinical importance yet rare manifestation in most auscultation datasets, data augmentation for crackles is an important tool that can hugely benefit CAA. Crackles are usually discontinuous along the time axis due to their abrupt nature. The question remains, do existing data augmentation techniques commonly used for image and speech processing translate to specialized sounds like crackle breathing sounds. A number of CAA models have shown promising results with auscultation augmentation methods using generative models to attain synthetic samples [12], [13], [14]. However, such methods are not broadly applicable to the specific case of crackle sounds given the acute class imbalance issue. Moreover, it is important to consider augmentation techniques that not only promote future processing using classification, detection and discrimination methods downstream, but also maintain a close correspondence to the original crackle sounds. Access to as authentic representative sounds as possible enable not only use in learning systems for various purposes but also facilitate clinical training of medical students. A few studies explored adverserial networks that analyze abnormal auscultation similarity for augmentation [15]. However, applicability of these methods for the crackle sound subclass is impractical given their rare occurance.

The present work tackles both these goals: We propose a novel augmentation method specifically tailored for crackle sounds and propose an evaluation technique that appraises the use of different augmentation methods relative to the statistical space and examplars in the dataset. Specifically, this study makes two contributions: (1) We propose a new method based on constrained synthetic sampling (CSS). The proposed method is an extension of existing Synthetic Minority Over-Sampling Technique (SMOTE)[16] that is more faithful to the statistical distribution of crackle sounds, instead of randomly filling the space of the original data as adopted by SMOTE. The proposed CSS method is evaluated against a wide range of augmentation methods and shown to maintain a close fidelity of the original crackles. (2) We

[1]Johns Hopkins University, Department of Electrical and Computer Engineering

979-8-3503-2447-1/23/$31.00 ©2023 IEEE

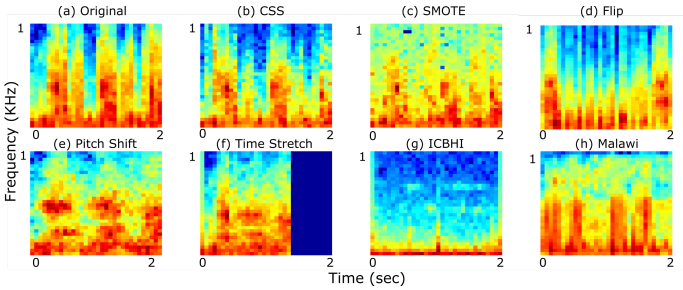

Fig. 1: Comparison of Original Crackle spectrogram (panel a) against various augmentation methods. The proposed CSS is shown in panel b, panel c shows example from the SMOTE technique, panel d shows example using spectrogram flipping, panel e shows pitch shifting using 3.5 semitones, panel f shows time stretching using a rate of 1.2. Panels g and h show example crackles from two different databases (ICBHI and Malawi respectively).

propose an evaluation method for augmentation techniques based on the inherent structure of the newly augmented dataset in a task-agnostic fashion. The proposed evaluation appraises the new augmented space both as a statistical space and based on individual samples within the space. As such, the method remains impartial to any backend classifier for discrimination, detection or other tasks, which impose different constraints based on the learning goal.

II. Crackle Augmentation Methods

A. constrained synthetic sampling (CSS) for crackle sounds

Given the unique nature of crackle sounds, the proposed method takes into account the geometry of original crackle spectrogram space. This approach builds on the established Synthetic Minority Over-sampling Technique (SMOTE) [16] and imposes constraints on the new data samples to maintain the fidelity of the original crackle space.

Starting with an original crackle spectrogram $\vec{\mathbf{A}}$ of dimensions 32x64, a synthetic variation $\vec{\mathbf{A}}_N$ is obtained using the SMOTE technique based on neighbor sampling. The methodology uses K nearest neighbors of $\vec{\mathbf{A}}$ amongst original crackles and introduces perturbation in each time-frequency bin of $\vec{\mathbf{A}}$ proportional to the gap from the corresponding bin of a random k^{th} nearest neighbor $\vec{\mathbf{A}}_K$ (Eq. 1).

$$\vec{\mathbf{A}}_N(t,f) = \vec{\mathbf{A}}(t,f) + U(0,1) * \left(\vec{\mathbf{A}}(t,f) - \vec{\mathbf{A}}_K[k](t,f)\right) \quad (1)$$

The new constrained synthetic sampling (CSS) method takes into account the fact that crackle sounds tend to be very transient, resulting in a rather discrete time axis. A synthetic variation $\vec{\mathbf{A}}_N$ generated using the classic SMOTE method averages the distance of K nearest neighbors inducing unwanted time frequency content across the spectrogram (as observed in 1(c)). Instead, the proposed CSS method implements a Principal Component Analysis on the original crackle spectrograms and further constraints the new crackle sample by projecting the synthetic sample along the principle components of the original space.

Specifically, we formulate a matrix $\mathbb{A}$ with vectorized original crackle spectrograms as columns. A new projected matrix $\mathbb{A}_p$ is then derived using the eigenvectors of the covariance matrix $Cov[\mathbb{A}]$. Any new synthetic sample generated is projected onto the columns of $\mathbb{A}_p$ to achieve a constrained synthetic sample that is more faithful to the structure of original crackles to obtain the final CSS augmented sample $\vec{\mathbf{A}}_S$. Equation 2 summarizes this operation.

$$\vec{\mathbf{A}}_S = \mathbb{A}_p \mathbb{A}_p^T \vec{\mathbf{A}}_N \quad (2)$$

B. Alternative Augmentation Methods

We compare the proposed crackle augmentation $\vec{\mathbf{A}}_S$ with alternative augmentation techniques that have been adopted in many audio learning systems including auscultation sounds [17]. We present an example spectrogram for each of these augmentation techniques in Fig 1.

1) Sound Flip: A commonly used auscultation augmentation technique is to flip the spectrogram along time axis and consider the time reversal version.

2) Pitch Shifting: The pitch of auscultation recordings are shifted by a nominal number of semitones. Pitch shifts of scale $-3.5, -2.5, 2.5, 3.5$ are considered in this study.

3) Time Stretching: The audio data is scaled horizontally by stretching the time at different rates 0.5, 0.7, 1.2, 1.5. The final spectrogram is either truncated or padded to ensure the spectrogram is 2 seconds long.

C. Alternative Datasets

One of the common augmentation considerations in case of data scarcity is to combine data across different datasets in order to increase the sample size as well as diversify the examplars of a particular class. In the case of healthcare data, this approach of data augmentation using dataset merging can raise its own sets of challenges. For auscultation signals, different datasets introduce very large uncontrolled variability that can be induced by sensors differences (different auscultation devices reflect different sensitivities and transfer functions [18], [19], surround conditions, clinical populations, in addition to annotation standards. To further examine this issue, we compare our proposed augmentation technique with crackle data from two other datasets.

1) ICBHI Dataset: This publicly available dataset [20] contains 1864 crackles collected from across 179 patients whose demographics cover a wide range across age, gender, and BMI. The data collection is done using three different stethoscopes. The subjects are diagnosed for COPD, lower and upper respiratory tract infections.

2) Malawi Dataset: A second smaller dataset [21] is collected following a similar protocol as the PERCH study [22] with a focus on severe pneumonia in 100 children in Lilongwe, Malawi under non-ideal clinical conditions in a low-resource setting. Data was collected using a Feelix stethoscope [23] and data annotation and crackle identification followed similar procedures as the PERCH study.

III. Augmentation Evaluation Methods

The paper proposes a method to evaluate data generated using any augmentation method (Fig. 2). The evaluation

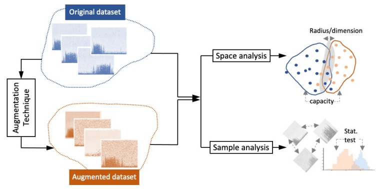

Fig. 2: Methods - Proposed augmentation method is evaluated by a two-level validation analysis at manifold level and sample level as depicted in the figure. This methodology is upper bounded when the analyses are performed on two Original Data Subsets and lower bounded on comparing original data subset against another randomly shuffled subset

follows two prongs: (i) the geometric properties of the Original and Augmented Dataset in a separable manifold space and (ii) statistical analysis on pair-wise distances on the optimally flattened spectrograms.

Both evaluation metrics are obtained by comparing a subset of the original crackle data against a corresponding augmentation technique. To better calibrate the metrics obtained from these comparisons, we define an upper and lower bound for each of the metrics. The upper bound is defined as an ideal case where the 'new' augmented data is *identical* to the original crackle data. This is achieved by comparing different subsets of the original crackle data against each other in a five-fold cross-validation. The lower bound is defined using an arbitrary augmentation technique whereby the augmented spectrogram is obtained by randomly shuffling the time-frequency content of the original crackle hence completely shattering the time-frequency signature of what a crackle is. These random samples are then compared to original crackles also in a cross-fold validation to determine a lower bound of what an augmentation technique can achieve.

A. *Manifold Analysis*

To capture the geometric similarity of different sub-populations of discrete representations (crackle spectrograms), we analyse the stochastic geometric properties of their separable manifolds. Mean Field Theoretic Manifold Analysis (MFTMA) [24], [25] implemented in the neuronal population of deep networks is used for this purpose. This is done by first projecting crackle spectrogram samples and corresponding crackle augmentation samples onto a low dimensional subspace where both the classes are linearly separable. Multiple random realizations of such separable subspaces are taken into consideration. Each such subspace mapping realization generates two anchor points, i.e vectors from the center of each class in the low-dimensional subspace to the classifying plane in that subspace. The stochastic properties of these anchor points capture the geometric separability characteristics of the manifolds. For an augmentation technique to be ideal, we want both samples in the comparison to be as non-separable as possible and increasing separability indicates a step away from the desired properties of the original crackle population. The following three properties from the MFTMA are considered as our augmentation validation metrics.

1) Capacity (ϕ): This capacity quantizes the linear separability of the two samples in the manifold subspace. This quantity serves as a measure of the linearly decodable information per time-frequency mapping about object separability.

2) Manifold Radius (ρ): The variance of random anchor points indicate the compactness of the object manifold in this separable space. The smaller the radius, higher is the compactness when the spectrogram comparison samples are mapped onto this space with maximum discrimination indicating better separability.

3) Manifold Dimension (δ): The angular spread of the anchor points gives the Dimension of the pairwise comparison of crackle and augmentation sample pair in the manifold subspace. A narrower spread indicates the presence of a stable subspace where the comparing pair are separable while a broader spread indicates lower separability.

For this analysis, original crackles are divided into 5 subsets and manifold parameters are generated for each original & augmented subset. Mean values of obtained geometric properties across different subsets are reported in the results.

B. *Sample Analysis*

To emphasize the differences between Constrained Synthetic Sample augmentation and a basic SMOTE augmentation, we analyse the surface properties of spectral content across time-frequency bins. This distance measure is adapted from a feature engineering technique used for auscultation classification [26]. The 2-dimensional spectrogram is first re-visualized as a 3-dimensional mesh with time-frequency-spectral content as the three axes. This 3D mesh is sampled and triangulated by Delaunay triangulation [27]. Given the sparse nature of the original spectrogram mesh, there is a need for transforming the original mesh space to a flattened mesh space constraining the properties of the non-sparse triangles from the original triangulated mesh. This is done by Adaptive Block Coordinate Descent Algorithm[28]. Each spectrogram mesh has to be separately optimized to reach an ideal representation in the flattened mesh space over a constraint. We analyze the L2-Distance Measures between random sample pairs of Original and Augmented crackle in the optimally flattened space.

To ensure that the randomness in the optimization is accounted for, we perform L2-distance analysis in the flattened mesh space on 500 Monte-Carlo random sample pairs (sample size = 100) of original crackle and the augmentation method in the flattened space.

IV. Auscultation Data

The recordings collected by the Pneumonia Etiology Research for Child Health (PERCH) study group [9] are used in the current study. A diverse set of 742 interpretable patient recordings from subjects of age 1-59 months across 7 different countries were collected with a Thinklabs digital stethoscope. 7-8 second recordings were collected from two frontal, two back and two axial body positions. All signals were originally sampled at 44.1KHz, pre-processed by applying a low-pass filtered with a fourth-order Butterworth filter at 4 kHz cutoff, downsampled to 8 kHz, and normalized.

Augmentation Category.	Comparison	ϕ (x10^{-3})	ρ	δ
Upper Bound	Original	2.89	2.32	410.2
Synthetic Sampling	**CSS**	**4.44**	**1.66**	**305.1**
	SMOTE	4.45	1.65	302.3
Spectrogram Variation	Flip	9.79	1.17	174.9
Pitch Shift	Scale = -2.5	9.29	1.21	180.6
	Scale = -1.5	8.15	1.24	201.3
	Scale =1.5	8.69	1.19	196.0
	Scale =2.5	8.91	1.72	192.8
Time Stretch	Rate = 0.5	7.92	1.27	203.22
	Rate = 0.7	11.5	1.15	154.1
	Rate = 1.2	12.5	1.13	142.8
	Rate = 1.5	13.2	1.10	136.3
Alterative Datasets	Malawi	7.36	1.43	191.0
	ICBHI	12.31	1.35	129.3
Lower Bound	Random	16.49	1.15	117.5

TABLE I: Manifold Analysis Geometric Properties Capacity (ϕ), Radius (ρ), and Dimension (δ) across different Crackle Augmentation Techniques.

The data collected in the PERCH study was annotated by 9 expert reviewers (pediatricians or pediatric-experienced physicians). They indicated the presence of adventitious breathing patterns like wheezes and crackles. In this study, we focus on the recordings in which at least two reviewers agree with confidence that they observe a presence of crackle. Out of 13.3 hours of collected data, only 48 minutes of recordings indicate the presence of crackles. Moreover, within each recording, crackles occur very briefly. In this study, we identify the peak of crackles within the onset and offset given by the expert reviewers and center a truncating window of 2 seconds around it obtaining 1607 samples.

V. Results

1) Manifold Analysis: In this analysis, we mapped pairwise subsets of two crackle representations into a manifold space such that they are optimally separable. If the data augmentation technique achieves its objective of being a good representative of a crackle sound , it will not be differentiable from an actual crackle even in this separable space. Based on this intuition, we look at the geometric properties in the manifold space that depict good separability. Having high classification capacity, as it nominally suggests, indicate good separability. Additionally, smaller radius and dimension of individual manifolds imply compact representations in the classifying space. As noted in the methods section, we also consider an upper-bound and lower bound of each metric by considering an Original-Original comparison (best case) and Original-Random (worst case). The overall evaluation of all augmentation methods is reported in TABLE I.

The results show that the proposed CSS results in the lowest classification capacity (ϕ) and highest radius (ρ) & dimension (δ) only next to the best case scenario. The classic SMOTE augmentation also yields reasonable manifold parameters though individual samples tend to be a bit noisier in time-frequency bins where little energy is expected in a crackle (see example in Fig 1, panel c). The manifold measure is rather insensitive to these minor variations in the spectrogram space, hence the need for a more refined sample-based comparison - as outlined next-.

In contrast, we note that basic spectrogram manipulations (flip, shift, stretch), often successful in a number of fields such as computer vision and speech systems yield quite drastically different spaces relative to original crackles, occasionally an order of magnitude away from the best case scenario.

Interestingly, we note that using crackles from two different datasets also result in drastically different statistical space than crackles from the PERCH dataset. The ICBHI dataset comprises of crackles that are drastically different from the PERCH signals. Fig 1, panel g shows this drastic contrast which can be due to a number of factors including sensing technology (which greatly shapes the signal profile) as well as patient population and clinical conditions. This contrast is confirmed in the manifold analysis whereby the radius, dimension and capacity are closer to the worse case scenario than they are to the original crackle data. In addition, crackles from the Malawi dataset are also quite different from the PERCH original despite a greater overlap in clinical protocol and patient population. Nevertheless, differences in auscultation devices as well as clinical settings appear to also cause a drastic contrast in the crackle samples between the originl PERCH and Malawi datasets.

2) Sample Analysis: While the CSS Augmentation results are the closest match when compared to the upper bound (based on the original crackle space), it is not dramatically different than SMOTE augmentation in terms of manifold parameters. However, we can clearly observe the presence of random artifacts all across the SMOTE crackle in Fig 1 (c) when compared against the Original Crackle 1 (a) and CSS Crackle 1 (b).

To better quantize the likeness of the augmentation techniques at a signal level, we perform a statistical analysis on random pairwise L2-distances of original crackle and each of the synthetic augmentation techniques. As L2-distance does not weigh each time-frequency bin differently, it is not a best fit to capture the distance on crackle spectrograms that are ususally discrete along time axis. We instead work with an optimally flattened new space to compute our pairwise difference samples.

The results of this analysis are presented in Fig 3. We notice that on looking at 500 Monte Carlo runs of pair wise distance measures obtained from original-original, original-CSS, and original-SMOTE, the population means of the new space L2-differences are 1.45, 1.55, and 1.73 respectively. Moreover, on performing pairwise t-tests, we observe that the Original-Original L2 distance sample in the flattened space is statistically similar to the Original-CSS L2 distance sample with a p-value of 0.09. However, the Original-CSS and the Original-SMOTE distance samples are statistically different with a p-value of 0.0099. This analysis further underscores the conceptual understanding and the visual representations (Fig 1 (a)-(b)-(c)) that the proposed CSS augmentation achieves the best 'similarity measure' to original crackle when compared against all the stated alternatives including SMOTE. While the lower bound of this distance measure is obtained by the best case scenario, we have also presented

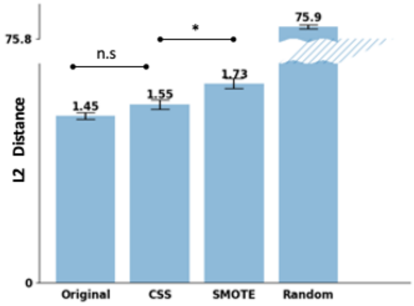

Fig. 3: Monte Carlo L2 Distances on Constrained Flattened Mesh Space

the the results for upper bound with a mean if 75.9.

VI. Conclusion

We propose a synthetic sampling based constrained augmentation for crackles, crucial pathological indicators amongst high-risk infants suffering from pulmonary diseases and a minority subclass of adventitious auscultations. Considering the discrete nature of crackle spectrograms along time axis, we propose two task-independent measures of augmentation validity: (i) Geometric properties of the object manifolds in a two-class (original crackle - comparison method) separable space, (ii) Statistical analysis on random pairwise distances from the original crackle on optimally flattened spectrograms. The proposed CSS augmentation performs significantly better than all the other methods analysed in this work and it is also statistically indistinguishable from original crackles across certain measures. Further analysis could be done by extending the evaluation of augmentation performance across different CAA downstream tasks to cross-verify the task agnosticism of the proposed augmentation across any extreme crackle class-imbalance setting.

Acknowledgment

The authors would like to thank the PERCH study group for guidance throughout the completion of this work, and to the patients and families enrolled in this study. We also acknowledge support from NIH R01HL163439 and ONR N00014-23-1-2086.

References

[1] K. Y. Ngiam and I. W. Khor, "Big data and machine learning algorithms for health-care delivery," *The Lancet Oncology*, vol. 20, no. 5, pp. e262–e273, 5 2019.

[2] Y. Roh, G. Heo, and S. Euijong Whang, "A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective; A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, 2021.

[3] F. H. Kiyoiti *et al.*, "Data Augmentation Using GANs," *Proceedings of Machine Learning Research*, vol. XXX, pp. 1–16, 2019.

[4] K. M. Rashid and J. Louis, "Times-series data augmentation and deep learning for construction equipment activity recognition," *Advanced Engineering Informatics*, vol. 42, p. 100944, 10 2019.

[5] M. J. Willemink *et al.*, "Preparing medical imaging data for machine learning," *Radiology*, vol. 295, no. 1, pp. 4–15, 2 2020.

[6] W. N. Price and I. G. Cohen, "Privacy in the age of medical big data," *Nature Medicine*, vol. 25, no. 1, pp. 37–43, 1 2019.

[7] S. A. WAHLS, "Causes and Evaluation of Chronic Dyspnea," *American Family Physician*, vol. 86, no. 2, pp. 173–180, 7 2012.

[8] G. Petmezas *et al.*, "Automated Lung Sound Classification Using a Hybrid CNN-LSTM Network and Focal Loss Function," *Sensors 2022, Vol. 22, Page 1232*, vol. 22, no. 3, p. 1232, 2 2022.

[9] E. McCollum *et al.*, "Digitally-recorded lung sounds and mortality among children 1-59 months old with pneumonia in the Pneumonia Etiology research for Child Health study," Tech. Rep., 2017. [Online]. Available: https://www.atsjournals.org/doi/abs/10.1164/ajrccm-conference.2017.195.1_MeetingAbstracts.A1195

[10] Z. MOUSSAVI, "Respiratory sound analysis [Introduction for the Special Issue]," *IEEE Engineering in Medicine and Biology Magazine*, vol. 26, no. 1, p. 15, 2007. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4069350

[11] H. Melbye *et al.*, "Inspiratory crackles-early and late-revisited: identifying COPD by crackle characteristics," *BMJ Open Resp Res*, vol. 8, p. 852, 2021. [Online]. Available: http://bmjopenrespres.bmj.com/

[12] 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*. ISCA: ISCA, 10 2020, pp. 2902–2906.

[13] M. Seibold *et al.*, "Conditional Generative Data Augmentation for Clinical Audio Datasets," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 13437 LNCS, pp. 345–354, 2022.

[14] K. Kochetov and A. Filchenkov, "Generative Adversarial Networks for Respiratory Sound Augmentation," 2020.

[15] S. Jayalakshmy and G. F. Sudha, "Conditional GAN based augmentation for predictive modeling of respiratory signals," *Computers in Biology and Medicine*, vol. 138, p. 104930, 11 2021.

[16] N. V. Chawla *et al.*, "SMOTE: Synthetic Minority Over-sampling Technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 6 2002.

[17] Z. Tariq, S. K. Shah, and Y. Lee, "Feature-Based Fusion Using CNN for Lung and Heart Sound Classification," *Sensors 2022, Vol. 22, Page 1521*, vol. 22, no. 4, p. 1521, 2 2022.

[18] I. McLane *et al.*, "Design and Comparative Performance of a Robust Lung Auscultation System for Noisy Clinical Settings," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 7, pp. 2583–2594, 7 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9345968/

[19] V. Rennoll *et al.*, "Electronic Stethoscope Filtering Mimics the Perceived Sound Characteristics of Acoustic Stethoscope," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1542–1549, 5 2021. [Online]. Available: https://ieeexplore.ieee.org/document/9184043/

[20] 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, 3 2019.

[21] Z. Smith *et al.*, "Evaluation of a novel digital lung sound recording and denoising platform for use in low-resource settings: expert listening panel agreement with traditional acoustic auscultation in a cohort of Malawian children 2-59 months with WHO-defined severe pneumonia," Washington DC, Tech. Rep., 2023.

[22] "PERCH; The Pneumonia Etiology Research for Child Health project." [Online]. Available: http://perchresults.org

[23] E. West *et al.*, "Introducing Feelix, a digital stethoscope incorporating active noise control and automatic detection of lung sound abnormalities," *The Journal of the Acoustical Society of America*, vol. 145, no. 3, p. 1923, 4 2019. [Online]. Available: https://asa.scitation.org/doi/abs/10.1121/1.5101984

[24] C. Stephenson *et al.*, "Untangling in Invariant Speech Recognition," 3 2020.

[25] U. Cohen *et al.*, "Separability and geometry of object manifolds in deep neural networks," *Nature Communications 2020 11:1*, vol. 11, no. 1, pp. 1–13, 2 2020.

[26] J. Levy, A. Naitsat, and Y. Y. Zeevi, "Classification of audio signals using spectrogram surfaces and extrinsic distortion measures," *Eurasip Journal on Advances in Signal Processing*, vol. 2022, no. 1, pp. 1–23, 12 2022.

[27] H. Edelsbrunner, "Geometry and Topology for Mesh Generation," *Geometry and Topology for Mesh Generation*, 5 2001.

[28] A. Naitsat, Y. Zhu, and Y. Y. Zeevi, "Adaptive Block Coordinate Descent for Distortion Optimization," *Computer Graphics Forum*, vol. 39, no. 6, pp. 360–376, 9 2020.