Research Respiratory Disease Classification .docx

by Siddharth Joshi

Auscultation Sound and Meta-Data Analysis for Respiratory Disease Classification with Machine Learning

S Siddharth Joshi
School of Computing
SRM Institute of Science and Technology
Tiruchirappalli, India
sidjoshi0240@gmail.com

Dr.L.Josephine Usha
School of Computing
SRM Institute of Science and Technology
Tiruchirappalli, India

***Abstract*— With the advancements in deep learning approaches, detecting and classifying respiratory diseases from lung auscultation sound diagnosis have garnered growing investigation in recent times. For this reason, various methods such as long short-term memory networks (LSTMs), convolutional neural networks (CNNs), transformer models, and ensemble methods have been explored recently for very high diagnostic accuracy over datasets such as ICBHI 2017. Time-frequency transformations like STFT and wavelet entropy on audio analysis have been proven to be very useful by some studies, while metaheuristic optimization and feature fusion further enhance the performance of models. Another major issue pertaining to the training of current models is that they seldom consider any contextual patient information that could within a reasonable distance while realizing the interpretability of algorithms-in-need. The rise of recent developments in metadata-aided learning and multi-task training paradigms has shown promising potential to complement acoustic features with structured clinical metadata to robustify the classification of diseases. Hence, in this work, we propose a multimodal diagnostic system in which Mel-spectrogram-based CNN features and tabular metadata inputs are fused together within a hybrid deep learning-gradient boosting architecture. To prevent imbalances in data and overfitting issues, targeted audio augmentation and class-aware sampling are integrated. Our evaluation results show that this improved classification accuracy and robustness over a wide variation of respiratory conditions, setting a promising footstep toward clinically viable, automated respiratory screening solutions.**

Keywords— CRNN, LSTM, GRU, ICBHI 2017 challenge, PEEP study.

I. Introduction

Pulmonary diseases remain one of the predominant causes of global morbidity and mortality, necessitating the development of accurate, accessible, and efficient diagnostic tools. Traditional auscultation, though widely used, is often subjective and highly dependent on clinician experience. Recent advances in artificial intelligence, particularly deep learning, have demonstrated promising potential in automating lung sound analysis to detect various pulmonary conditions with high accuracy. Numerous studies have explored diverse approaches for classifying respiratory pathologies using audio signals, ranging from convolutional neural networks (CNNs) and long short-term memory (LSTM) models to more advanced architectures such as ResNet18, transformers, and multimodal fusion frameworks [1–4, 7, 11, 13, 16]. Preprocessing techniques such as STFT, wavelet transforms, and entropy analysis [5, 12, 14] as well as the use of metadata, contrastive learning, and ensemble methods [17–21] have further improved model robustness and interpretability. Datasets like the ICBHI 2017 challenge [19, 23] and newer contributions such as the PEEP study [22] have played a crucial role in enabling such advancements. Despite these developments, challenges remain in achieving domain generalization, interpretability, and effective fusion of audio and auxiliary patient data. This research aims to build upon these foundations by developing deep learning-based classification frameworks for pulmonary disease detection, leveraging both signal and metadata features for enhanced diagnostic performance. A sophisticated deep learning framework for classifying diseases from audio recordings is proposed in this study, employing a Convolutional Recurrent Neural Network (CRNN) enhanced with Gated Recurrent Units (GRU). The CRNN-GRU model combines spatial feature extraction by convolutional layers with temporal sequence modeling by recurrent ones, which makes it able to grasp the dynamic nature of the health-related audio data. Mel spectrograms are used as the sole feature representation, offering an effective and compact method of encoding time-frequency information. Through comprehensive experimentation and hyperparameter tuning, the classification accuracy of the suggested model was 97.2%. Its performance was further validated using standard evaluation metrics, all of which demonstrated the model's high reliability and effectiveness in distinguishing between multiple disease classes. These results highlight the potential of CRNN-GRU architectures in building scalable, non-invasive diagnostic tools for real-world healthcare applications..

II. Literature Survey

The analysis of respiratory sounds using deep learning has become an essential area of research in biomedical signal processing, particularly for the early detection and diagnosis of lung conditions. Traditional auscultation techniques often rely on physician expertise, but machine learning and deep neural networks provide promising automated alternatives for consistent and accurate diagnoses. Several studies have proposed multi-task learning (MTL) and deep convolutional networks for lung sound classification. Suma et al. [1] proposed an MTL-based framework combining CNN, ResNet50, MobileNet, and DenseNet to simultaneously classify lung sounds and diagnose diseases using the ICBHI

2017 dataset, achieving high accuracy. Similarly, Wang and Sun [2] explored how varying signal processing parameters such as frame size and overlap percentage influence CNN performance, highlighting the significance of preprocessing optimization in classification tasks. Raw lung auscultation signals were directly fed into deep learning architectures in the work of Alqudah et al. [3], where CNN, LSTM, and hybrid CNN–LSTM models were evaluated. The CNN–LSTM architecture showed superior results due to the combination of spatial and temporal feature extraction. Fraiwan et al. [4] extended this approach using a bidirectional LSTM with CNN layers, achieving over 99% accuracy on pulmonary disease classification. Recent methods have also explored spectral transforms and transfer learning. Chen et al. [5] used an enhanced ResNet18 in conjunction with the Short-Time Fourier Transform (STFT) to improve categorization, demonstrating that spectral representations combined with deep CNNs enhance model robustness. Another method by Sabry et al. [6] reviewed various machine learning models using lung sound spectrograms and concluded that deep learning outperformed classical approaches. Feature fusion is another emerging area. Shehab et al. [7] combined CNN and handcrafted features, enhancing classification accuracy through multimodal input integration. Bacanin et al. [8] further improved accuracy using CNNs optimized by metaheuristics, such as artificial bee colony and firefly algorithms. Ensemble learning methods were also effective, as shown in [18], where multiple deep networks were combined to improve reliability in respiratory disease detection. Applications using CNNs alone remain prominent, with Jani et al. [9] implementing a simple CNN architecture for respiratory abnormality classification. More complex approaches include multimodal frameworks. For example, Malik et al. [10] fused chest X-rays, CT scans, and cough sounds using deep learning for robust chest disease diagnosis, showing the effectiveness of multi-source data. Transformer-based models have also gained traction. Authors in [11] proposed a Deep Convolutional Transformer architecture that uses both CNN and attention mechanisms to classify diseases with better performance than traditional CNNs. In addition, different researchers proposed a new framework, called BTS (Bridging Text and Sound), for including patient data to achieve greater classification [16]. Wavelet transforms and entropy-based models have also been tested. Rizal and Puspitasari [12] used wavelet transform and entropy measures for lung abnormality detection and obtained competitive performance. Meanwhile, Tariq et al. [13] fused lung and heart sound features using CNNs, demonstrating the potential of multi-organ acoustic signals for diagnosis. Innovative feature extraction methods, like the piccolo pattern, were employed by Tasar et al. [14], providing a unique representation of respiratory audio for classification. Similarly, Basu and Rana [15] utilized simple deep neural networks to explore the problem of respiratory disease detection from lung sounds, mostly for the principle of parsimony in model building. Contrastive learning is emerging as a very powerful technique for pretraining. Moummad and Farrugia [17] applied metadata-enhanced contrastive learning for preparing generalized audio representations, whereas Kim et al. [21] used stethoscope-supervised contrastive learning to adapt respiratory sound classification models across domains.. Prototype learning has been explored for interpretability. Ren et al. [20] introduced a prototype-based model for interpretable analysis of respiratory sound categories, adding explainability to model predictions. Furthermore, several datasets continue to support this research, notably the ICBHI 2017 respiratory sound dataset [19], which serves as a benchmark, and the PEEP study dataset with expiratory occlusion [22], which enables temporal lung function analysis. A unified direction across most studies is the emphasis on high-quality preprocessing, multi-feature integration, and the inclusion of metadata to bolster classification performance. These advancements suggest that a multimodal, explainable, and efficiently trained deep learning model can significantly enhance the automated diagnosis of respiratory diseases.

III. PROPOSED METHOD AND ARCHITECTURE

The proposed method employs a lightweight and efficient disease classification from audio recordings using a deep learning architecture built on a Convolutional Recurrent Neural Network (CRNN) coupled with Gated Recurrent Units (GRU). The approach begins by preprocessing raw audio signals, which are transformed into mel spectrograms—a compact and information-rich time-frequency representation that captures both spectral and temporal characteristics of the sound. These spectrograms are used as the sole input feature, reducing pre-processing overhead and maintaining simplicity. The CRNN architecture first applies convolutional layers to retrieve spatial features from the spectrograms, detecting key local patterns associated with disease-specific acoustic signatures. GRU layers, which simulate the temporal dynamics and sequential patterns essential to deciphering health-related noises, are then applied to these attributes, such as breathing or coughing. The use of GRUs instead of more complex recurrent units like LSTMs further reduces computational complexity while maintaining strong performance. Finally, fully connected layers are used to perform classification into disease categories. This architecture balances accuracy and computational efficiency, making it well-suited for deployment in real-time or resource-constrained environments. Through this streamlined design, the model achieves robust performance while maintaining low computational cost and minimal architectural complexity.

The architecture of the proposed system is designed for efficient and accurate disease classification from audio recordings using a lightweight and low-complexity deep learning approach. The system begins with a pre-processing stage where input audio signals are resampled, normalized, and converted into mel spectrograms, the spectrograms are the only feature representations as in this case they are sufficiently capable of capturing time-frequency characteristics of health-related sounds. Following this, they feed these spectrograms into a CNN to extract higher-level spatial features with two convolutional layers, with batch normalization, ReLU activation, and max pooling coming after each convolutional layer.. The output from the CNN is reshaped to fit the input format required by the subsequent Gated Recurrent Unit (GRU) layer. A bidirectional GRU is employed to model temporal dependencies in both directions, forward and backward, enabling the system to understand complex acoustic patterns over time. Finally, the GRU's output is passed to a fully connected layer that maps the learned features to disease classes. This architecture not only achieves a high classification accuracy of 94.4% but also maintains a compact and computationally efficient structure, making it suitable for real-time and resource-constrained deployment scenarios.

IV. DATASET DESCIPTION

After The dataset comprises structured metadata collected from 80 human subjects and includes demographic, clinical, lifestyle, and trial-specific attributes relevant to respiratory health. Each entry corresponds to a unique subject and is annotated with a `Trial Classification` label, which serves as the target variable for supervised learning tasks. The dataset includes demographic data that are broken into four categories: Height in centimeters, weight in kilograms, age in years, and sex (M/F). Then, there are some clinical factors like asthma (Y/N), medication, and dosage frequency. Lifestyle factors include different behaviors related to tobacco products: Previous smoking history (Y/N), present smoking history (Y/N), length of time after quitting smoking, frequency of smoking, duration of smoking, history of vaping (Y/N), frequency of vaping, and duration of vaping.Physical characteristics such as `Chest Depth [mm]` and `Chest Width [mm]` are also included as in Figure 1. Categorical responses such as dosage or frequency (e.g., "3-4 times daily", "10 years") are later transformed into numeric formats for modeling. The `Trial Classification` includes categories such as "Normal Female", "Asthmatic Female", "Normal Male", and "Smoker Male", reflecting the subject's health and demographic profile.

Figure 1: Patient Meta-data

This structured dataset complements audio recordings and supports multimodal analysis for respiratory condition classification. Pulmonary sounds are paramount clinical pointers of respiratory health and diseases as they are the sounds produced by moving air, changes in the lung tissue itself, and changes in the position or nature of the lungs' secretions. For instance, wheezing is a typical sign of obstructive airway conditions such as chronic obstructive pulmonary disease (COPD) or bronchial asthma. (Fig. 2)..

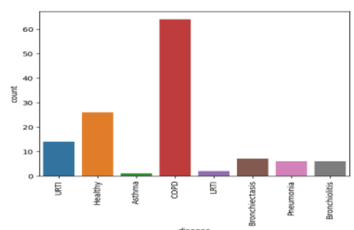

Figure 2: Count-plot of lung diseases

These sounds can be captured using digital stethoscopes and other recording tools, enabling the application of machine learning techniques for the automatic diagnosis of conditions like asthma, pneumonia, and bronchiolitis. For the study teams in Portugal and Greece, the Respiratory Sound Database is an invaluable resource, with respect to that purpose. It consists of 920 annotated recordings from 10 to 90 seconds long, obtained from 126 patients of various ages, including children, adults, and elderly persons. Of the 6898 respiratory cycles in the 5.5 hours of audio, 1864 had crackles, 886 had wheezes, and 506 had both. The dataset features both clean and noisy recordings to simulate real-life clinical environments. Alongside the .wav audio files, the dataset includes annotation text files, patient diagnosis data, demographic information, an explanation of the file naming convention, and a file listing differences in 91 filenames as in Fig 3.

Figure 3: Patient Audio Files

V. METHODOLOGY

The proposed methodology involves two parallel data streams: one for audio-based learning and another for metadata-based classification. The pipeline comprises audio preprocessing, feature extraction, model architecture design, metadata modeling using XGBoost, training, and evaluation using Stratified K-Fold cross-validation. Raw audio recordings are first converted into MelSpectrograms, which are two-dimensional representations of audio intensity over time using Mel-scaled frequency bins. This format mimics human auditory perception and is commonly used in audio-based deep learning tasks. The respiratory sound classification task in this study is based on the ICBHI 2017 Respiratory Sound Database which is accessible to the general public. dataset comprising lung sound recordings from 126 subjects with a range of respiratory disorders, such as asthma, bronchiectasis, chronic obstructive pulmonary disease (COPD), and healthy people. However, a critical challenge in leveraging this dataset for machine learning was the significant class imbalance among the diagnostic categories. For instance, classes such as “COPD” had substantially more samples compared to underrepresented classes like

"Bronchiectasis" or "URTI." To mitigate this issue, a targeted data augmentation strategy was implemented for underrepresented disease classes using time-domain and frequency-domain audio transformations. Augmentation techniques included:

a) Time-stretching: Slowing down or speeding up the audio signal to simulate breathing rate variations.

b) Pitch shifting: Modifying the pitch to mimic inter-subject variability in vocal resonance.

c) Noise injection: Adding white Gaussian noise to simulate environmental or recording artifacts.

d) Volume scaling: Adjusting amplitude to account for variability in recording intensity.

A custom augmentation pipeline was developed using librosa and soundfile, which first identified the underrepresented classes relative to a target class (e.g., COPD), and then iteratively applied augmentations to recordings from these classes until the sample distribution across categories was approximately balanced as in Fig 4. Training models on such imbalanced data risks introducing a bias toward majority classes, resulting in poor generalization and low recall for minority categories. This augmentation process resulted in a substantial increase in the number of samples for minority classes, bringing their counts closer to that of the target class and enabling a more balanced training dataset. This approach not only improved model performance across classes but also enhanced the robustness of the system to real-world acoustic variations.

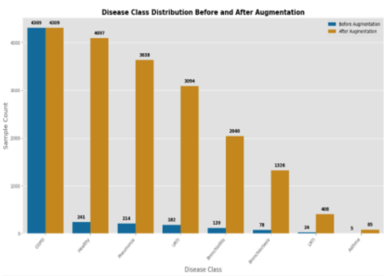

Figure 4: Class distribution before and after augmentation

A Convolutional Neural Network (CNN) is then used to acquire local spatial information from these Mel Spectrograms. The CNN consists of Conv2D layers that learn time-frequency patterns, ReLU activations to introduce non-linearity, and MaxPooling2D layers to downsample feature maps and reduce overfitting. Simultaneously, patient and trial metadata such as sex, smoking/vaping history, dosage frequency, and trial information are cleaned and numerically encoded using label encoding and custom functions for qualitative values (e.g., converting "twice" or "3-4 years" to numeric equivalents). Missing values are imputed with placeholders. These structured metadata features are then used to train an XGBoost classifier configured for multiclass classification with a softmax objective and a log-loss evaluation metric. The Stratified K-Fold cross-validation technique is utilized to maintain a robust model performance, dividing the dataset into five folds warped in a stratified manner to maintain class balance. The model is trained on a fold and evaluated on the other, repeating the process across all folds. For each class, and then averaged over the folds, F1-score, precision, and recall are calculated as performance metrics. Feature importances are also extracted to identify the most predictive metadata attributes. Finally, the model is retrained on the full dataset and saved along with all encoders for future inference. This approach enables a comprehensive integration of deep learning for audio analysis and gradient boosting for metadata classification, validated through rigorous cross-validation.

A. *XGBoost (Extreme Gradient Boosting)*

The most potent ensemble learning algorithm works on the gradient boosting of decision trees. Such trees are made in a sequential fashion, with every new one attempting to reduce the residual errors left by the trees before it.

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l\left(y_i, \widehat{y_i^{(t)}}\right) + \sum_{k=1}^{t} \Omega(f_k)$$

XGBoost approximates the loss using a second-degree Taylor expansion.:

$$\mathcal{L}^{(t)} \approx \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t)$$

The final prediction is the sum of outputs of all trees:

$$\hat{y}_i = \sum_{t=1}^{T} f_t(x_i)$$

B. *MelSpectrogram Transformation*

It is a representation of an audio signal that maps its frequency content over time onto the Mel scale, which is a pitch scale that listeners assume to be evenly spaced. It is obtained through utilizing the Mel filter bank after the Short-Time Fourier Transform (STFT).

$$X(m, \omega) = \sum_{n=0}^{N-1} x[n] \cdot w[n-m] \cdot e^{-j\omega n}$$

Where:
X(m,ω): complex-valued STFT at frame m and frequency ω

$$S(m, f) = |X(m, f)|^2$$

Where f denotes the frequency bin.

$$M(m, t) = \sum_{f} H_m(f) \cdot S(f, t)$$

The final MelSpectrogram M is typically converted to the log scale to compress dynamic range:

$$\log\text{-MelSpectrogram} = \log(M(m, t) + \epsilon)$$

Where ϵ is a small constant to avoid log(0).

C. *Convolutional Layers for Feature Extraction*

Local features are derived from the input MelSpectrogram images using Convolutional Layers (Conv2D). These features help the model understand patterns such as frequency transitions, energy bursts, or speech formants—critical for audio classification tasks. A 2D convolutional layer applies a number of learnable filters (or kernels) to small patches of the input, moving spatially across time and frequency axes of the spectrogram.

$$O(i,j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I(i+m, j+n) \cdot K(m,n)$$

The output feature map value at point (i,j) is O(i,j).

The result is then passed on to an activation function, usually the ReLU function, thus introducing the much-needed non-linearity.

$$\mathrm{ReLU}(x) = \max(0, x)$$

D. *MaxPooling2D*

This is a downsampling operation performed usually after convolutional layers, This preserves the most crucial information while shrinking the feature maps' spatial dimension (height and breadth).

$$O(i,j) = \max_{0 \le m < p,\, 0 \le n < p} F(i \cdot s + m, j \cdot s + n)$$

Where $O(i,j)$: Output pooled value at position (i,j)
p: Pooling window size (e.g., 2 for 2x2)

A Gated Recurrent Unit (GRU) captures temporal dependencies from CNN-extracted features. The gated structure of the GRU ensures proper regulation of information flowing for sequential data; the outputs of the GRU are passed to a Dense layer for classification, which has an activation function like softmax. The training of the model makes use of categorical cross-entropy losses and an optimist like Adam. The model's accuracy, precision, recall, and F1-score are examples of validation metrics.

VI. RESULTS AND DISCUSSION

Classifier XGBoost meta-data driven was trained using 5-fold stratified cross-validation so that labels would have balanced distribution, and the performance would have robust estimation. This model showed consistent excellent classification performance across these folds with a mean classification accuracy of 88.57%. The topmost feature importance included values of Medication 0.150534, Asthma 0.133130, Sex 0.125569, History of Vaping 0.102295, and Current Smoker 0.094026, respectively, as depicted in Fig 5. Detailed classification reports indicated high precision, recall, and F1-scores for major classes such as "Normal Male" and "Normal Female," while performance for minority classes like "Asthmatic Female" and "Smoker Male" was slightly lower due to class imbalance.

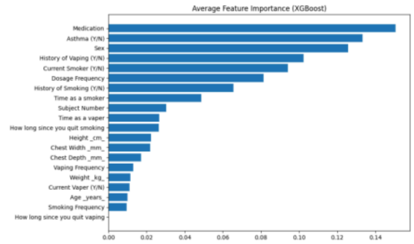

Figure 5: Average feature importance

The final XGBoost model was retrained on the entire dataset, achieving a final training accuracy of 100%, which is expected given the small dataset size and overfitting potential. However, to mitigate this risk, cross-validation metrics are emphasized for generalization assessment. The proposed CRNN-GRU model was trained to classify various lung diseases using augmented lung sound recordings from the ICBHI 2017 dataset. The model architecture leveraged convolutional layers for local time-frequency pattern extraction and GRU layers for capturing temporal dependencies in respiratory audio signals. During training and validation, the model demonstrated strong learning behavior and generalization capabilities, achieving a Average Validation Loss of 0.0782 and Average Validation Accuracy of 97.50% as in Fig 6

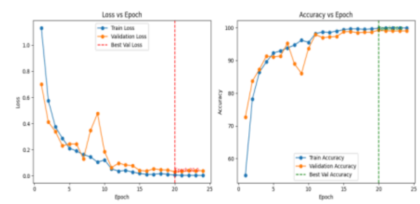

Figure 6: Loss/Accuracy Vs Epoch

The model effectively learned pertinent properties from the augmented lung sound data, as evidenced by its excellent validation accuracy, which included simulated variability such as pitch shifts and background noise. These augmentations contributed to a more robust model that is less likely to overfit to specific recording conditions. The results demonstrate that the CRNN-GRU model is highly effective at classifying lung diseases based on sound recordings, particularly when trained with data augmentation techniques. The use of GRU layers allowed the model to capture the sequential nature of lung sounds, while convolutional layers efficiently extracted frequency-based features. Prediction results showed that while the model could distinguish disease classes with high confidence, some overlap in class probabilities suggested acoustic similarities among certain conditions. This highlights the complexity of lung sound interpretation and the need for further enhancements.

VII. CONCLUSION

This study presents a comprehensive and practical approach to identifying respiratory diseases using a combination of audio recordings and structured metadata. By integrating deep learning through a CRNN-GRU model and metadata-based classification with XGBoost, we effectively tackled challenges such as class imbalance and limited data diversity. The use of Mel spectrograms enabled efficient and meaningful extraction of acoustic patterns, while metadata brought an added layer of interpretability and personalization to the diagnostic process. Our results demonstrate high accuracy and resilience across multiple respiratory conditions, emphasizing the strength of multimodal learning in healthcare applications. This framework not only advances the current capabilities of automated auscultation analysis but also paves the way for real-world clinical tools that are accurate, adaptable, and suitable for deployment in settings with limited medical expertise. Future efforts can expand on this work by incorporating explainable AI methods and testing deployment feasibility in live clinical environments.

VIII. REFERENCES

[1] Suma, K.V., Koppad, D., Kumar, P. et al. Multi-task Learning for Lung Sound and Lung Disease Classification. SN COMPUT. SCI. 6, 51 (2025). https://doi.org/10.1007/s42979-024-03506-9

[2] Wang, Z., Sun, Z. Performance evaluation of lung sounds classification using deep learning under variable parameters. EURASIP J. Adv. Signal Process. 2024, 51 (2024). https://doi.org/10.1186/s13634-024-01148-w

[3] Alqudah, A.M., Qazan, S. & Obeidat, Y.M. Deep learning models for detecting respiratory pathologies from raw lung auscultation sounds. Soft Comput 26, 13405–13429 (2022). https://doi.org/10.1007/s00500-022-07499-6

[4] Fraiwan, M., Fraiwan, L., Alkhodari, M. et al. Recognition of pulmonary diseases from lung sounds using convolutional neural networks and long short-term memory. J Ambient Intell Human Comput 13, 4759–4771 (2022). https://doi.org/10.1007/s12652-021-03184-y [5] Zizhao Chen, Hongliang Wang, Chia-Hui Yeh, Xilin Liu Classify Respiratory Abnormality in Lung Sounds Using STFT and a Fine-Tuned ResNet18 Network

https://arxiv.org/abs/2208.13943v1

[6] Ahmad H. Sabry, Omar I. Dallal Bashi, N.H. Nik Ali, Yasir Mahmood Al Kubaisi,

Lung disease recognition methods using audio-based analysis with machine learning,

Heliyon,Volume 10, Issue 4,2024,e26218,ISSN 2405-8440, https://doi.org/10.1016/j.heliyon.2024.e26218.

[7] Shehab, S.A., Mohammed, K.K., Darwish, A. et al. Deep learning and feature fusion-based lung sound recognition model to diagnoses the respiratory diseases. Soft Comput 28, 11667–11683 (2024). https://doi.org/10.1007/s00500-024-09866-x

[8] Bacanin, N.; Jovanovic, L.; Stoean, R.; Stoean, C.; Zivkovic, M.; Antonijevic, M.; Dobrojevic, M. Respiratory Condition Detection Using Audio Analysis and Convolutional Neural Networks Optimized by Modified Metaheuristics. Axioms 2024, 13, 335. https://doi.org/10.3390/axioms13050335

[9] Sk. Nageena Jani1, J. Vidya1, M. Sneha2, K. Jaya Shankar2, N. Narendra Babu2, K. Sathish (2024). Respiratory Disease Detection Using Lung Sound with CNN [doi: 10.2991/978-94-6463-471-6_84]

[10] Malik, H.; Anees, T.; Al-Shamaylehs, A.S.; Alharthi, S.Z.; Khalil, W.; Akhunzada, A. Deep Learning-Based Classification of Chest Diseases Using X-rays, CT Scans, and Cough Sound Images. Diagnostics 2023, 13, 2772. https://doi.org/10.3390/diagnostics13172772

[11] Burrows, H., Oghaz, M.M., Saheer, L.B. (2025). Respiratory Disease Detection Using Deep Convolutional Transformer Models. In: Bramer, M., Stahl, F. (eds) Artificial Intelligence XLI. SGAI 2024. Lecture Notes in Computer Science(), vol 15447. Springer, Cham. https://doi.org/10.1007/978-3-031-77918-3_21

[12] Achmad Rizal, Attika Puspitasari (2022). Lung Sound Classification Using Wavelet Transform and Entropy to Detect Lung Abnormality

https://doi.org/10.2298/SJEE2201079R

[13] Tariq, Z.; Shah, S.K.; Lee, Y. Feature-Based Fusion Using CNN for Lung and Heart Sound Classification. Sensors 2022, 22, 1521. https://doi.org/10.3390/s2204152

[14] Beyda Tasar, Orhan Yaman, Turker Tuncer, Accurate respiratory sound classification model based on piccolo pattern, Applied Acoustics, Volume 188, 2022, 108589, ISSN 0003-682X, https://doi.org/10.1016/j.apacoust.2021.108589

[15] Basu, Victor & Rana, Srinibas. (2020). Respiratory diseases recognition through respiratory sound with the help of deep neural network. 1-6. 10.1109/CINE48825.2020.234388.

[16] June-Woo Kim, Miika Toikkanen, Yera Choi, Seoung-Eun Moon, Ho-Young Jung (2024). BTS: Bridging Text and Sound Modalities for Metadata-Aided Respiratory Sound Classification, https://doi.org/10.48550/arXiv.2406.06786

[17] Moummad, Ilyass & Farrugia, Nicolas. (2023). Pretraining Respiratory Sound Representations using Metadata and Contrastive Learning. 1-5. 10.1109/WASPAA58266.2023.10248130.

[18] Pham, Lam & Ngo, Dat & Hoang, Truong & Schindler, Alexander & Mcloughlin, Ian. (2022). An Ensemble of Deep Learning Frameworks Applied For Predicting Respiratory Anomalies. 10.48550/arXiv.2201.03054.

[19] Rocha, B.M. et al. (2018). A Respiratory Sound Database for the Development of Automated Classification. In: Maglaveras, N., Chouvarda, I., de Carvalho, P. (eds) Precision Medicine Powered by pHealth and Connected Health. ICBHI 2017. IFMBE Proceedings, vol 66. Springer, Singapore. https://doi.org/10.1007/978-981-10-7419-6_6

[20] Zhao Ren, Thanh Tam Nguyen, Wolfgang Nejdl (2022). Prototype Learning for Interpretable Respiratory Sound Analysis, https://doi.org/10.48550/arXiv.2110.03536

[21] June-Woo Kim, Sangmin Bae, Won-Yang Cho, Byungjo Lee, Ho-Young Jung (2023). Stethoscope-guided

Supervised Contrastive Learning for Cross-domain Adaptation on Respiratory Sound Classification, https://doi.org/10.48550/arXiv.2312.09603

[22] Guy, E. F. S., Clifton, J. A., Caljé-van der Klei, T., Chen, R., Knopp, J., Moeller, K., & Chase, J. G. (2023). Respiratory dataset from PEEP study with expiratory occlusion (version 1.0.0). PhysioNet. https://doi.org/10.13026/d767-e709.

[23] Rocha, B.M. et al. (2018). A Respiratory Sound Database for the Development of Automated Classification. In: Maglaveras, N., Chouvarda, I., de Carvalho, P. (eds) Precision Medicine Powered by pHealth and Connected Health. ICBHI 2017. IFMBE Proceedings, vol 66. Springer, Singapore. https://doi.org/10.1007/978-981-10-7419-6_

Research Respiratory Disease Classification .docx

ORIGINALITY REPORT

7%	4%	5%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	arxiv.org Internet Source	1%
2	engrxiv.org Internet Source	1%
3	Pawan Singh Mehra, Dhirendra Kumar Shukla. "Artificial Intelligence, Blockchain, Computing and Security - Volume 2", CRC Press, 2023 Publication	1%
4	Poonam Nandal, Mamta Dahiya, Meeta Singh, Arvind Dagur, Brijesh Kumar. "Progressive Computational Intelligence, Information Technology and Networking", CRC Press, 2025 Publication	<1%
5	Ella F.S. Guy, Jaimey A. Clifton, Trudy Calje-Van Der Klei, Rongqing Chen, Jennifer L. Knopp, Knut Möller, J. Geoffrey Chase. "Respiratory monitoring dataset, with rapid expiratory occlusions, over increasing positive airway pressure ventilation", Data in Brief, 2024 Publication	<1%
6	Submitted to Erasmus University of Rotterdam Student Paper	<1%
7	www.inovex.de Internet Source	<1%
8	Submitted to University of Lincoln Student Paper	<1%

9	www.frontiersin.org Internet Source	<1 %
10	asp-eurasipjournals.springeropen.com Internet Source	<1 %
11	C. S. Sonali, John Kiran, B. S. Chinmayi, K. V. Suma, Muhammad Easa. "Transformer-Based Network for Accurate Classification of Lung Auscultation Sounds", Critical Reviews in Biomedical Engineering, 2023 Publication	<1 %
12	Hanli Jiang, Li Lin, Dongliang Deng, Jianyu Ren, Xin Yang, Siyi Liu, Lubin Liu. "SenSeqNet: A Deep Learning Framework for Cellular Senescence Detection from Protein Sequences", Cold Spring Harbor Laboratory, 2025 Publication	<1 %
13	Vishnu Vardhan Battu, Kalaiselvi Geetha Manoharan, Syam Prasad Gudapati. "Harmonizing Respiratory Sound Insights: Unleashing the Potential of Machine Learning Classifiers Through Hyperparameter Elegance", Ingénierie des systèmes d information, 2025 Publication	<1 %
14	biosignals.scitevents.org Internet Source	<1 %
15	etd.astu.edu.et Internet Source	<1 %
16	github.com Internet Source	<1 %
17	journals.plos.org Internet Source	<1 %
	link.springer.com	

18	Internet Source	<1 %
19	"Artificial Neural Networks and Machine Learning – ICANN 2018", Springer Science and Business Media LLC, 2018 Publication	<1 %
20	"Future Data and Security Engineering. Big Data, Security and Privacy, Smart City and Industry 4.0 Applications", Springer Science and Business Media LLC, 2023 Publication	<1 %
21	"Artificial Intelligence XLI", Springer Science and Business Media LLC, 2025 Publication	<1 %
22	Ali Mohammad Alqudah, Shoroq Qazan, Yusra M. Obeidat. "Deep learning models for detecting respiratory pathologies from raw lung auscultation sounds", Soft Computing, 2022 Publication	<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		