Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 20 No. 10 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v20i10.49585

PAPER

Lung Sound Classification for Respiratory Disease Identification Using Deep Learning: A Survey

Thinira Wanasinghe[1], Sakuni Bandara[1], Supun Madusanka[1], Dulani Meedeniya[1](✉), Meelan Bandara[1], Isabel De La Torre Díez[2]

[1]Department of Computer Science & Engineering, University of Moratuwa, Moratuwa, Sri Lanka

[2]Department of Signal Theory and Communications, and Telematics Engineering, University of Valladolid, Valladolid, Spain

dulanim@cse.mrt.ac.lk

ABSTRACT

Integrating artificial intelligence (AI) into lung sound classification has markedly improved respiratory disease diagnosis by analysing intricate patterns within audio data. This study is driven by the widespread issue of lung diseases, which affect around 500 million people globally. Early detection of respiratory diseases is crucial for delivering timely and effective treatment. Our study consists of a comprehensive survey of lung sound classification methodologies, exploring the advancements made in leveraging AI to identify and classify respiratory diseases. This survey thoroughly investigates lung sound classification models, along with data augmentation, feature extraction, explainable techniques and support tools to improve systems for diagnosing respiratory conditions. Our goal is to provide meaningful insights for healthcare professionals, researchers and technologists who are dedicated to developing methodologies for the early detection of pulmonary diseases. The paper provides a summary of the current status of lung sound classification research, highlighting both advancements and challenges in the use of AI for more accurate and efficient diagnostic methods in respiratory healthcare.

KEYWORDS

artificial intelligence, classification, explainability, respiratory diseases, sound processing

1 INTRODUCTION

Lung diseases are one of the common types of diseases that have invaded about 500 million people across the globe [1]. It is vital to identify different lung conditions and symptoms for effective diagnosis and treatment. Generally, the diagnosis of respiratory diseases is done by experienced physicians by listening to lung sounds via a stethoscope or analysing the scanned images of the respiratory system. Although this process requires expert knowledge and time, the World Health Organisation (WHO) [2] states that 45% of the WHO member states have less than 1 physician per 1000 population [3]. Therefore, the task of identifying lung conditions has become less efficient due

Wanasinghe, T., Bandara, S., Madusanka, S., Meedeniya, D., Bandara, M., De La Torre Díez, I. (2024). Lung Sound Classification for Respiratory Disease Identification Using Deep Learning: A Survey. *International Journal of Online and Biomedical Engineering (iJOE)*, 20(10), pp. 115–129. https://doi.org/10.3991/ijoe.v20i10.49585

Article submitted 2024-04-10. Revision uploaded 2024-05-06. Final acceptance 2024-05-06.

to the lack of skilled physicians in developing countries and the time taken to examine and analyse individuals. Moreover, the results of manual auscultation depend on the expertise of medical practitioners and may be prone to human error. Thus, considering the limited availability of healthcare resources, the accurate diagnosis of patients may be subjected to variability and inconsistencies, highlighting the need for a standardised approach to automate lung sound classification as a support tool for physicians.

The inclusion of AI in healthcare, such as lung condition detectors, cancer classifiers, and eye disease identifiers, has significantly improved efficiency and performance in medical informatics [4, 5]. Recently, deep learning (DL) has evolved in the medical domain by providing techniques to analyse complex data types, improve diagnostic accuracy, and facilitate medical research and decision-making processes [6], [7]. Lung sound classification is important for the effective and efficient identification of respiratory diseases. Different DL classifiers can be trained to learn the features in lung sound signals to distinguish different lung conditions such as pneumonia, asthma, COVID, and chronic obstructive pulmonary disease (COPD) [8–13]. Accordingly, employing DL-based classifiers and developing support tools utilising these models to provide a diagnosis for lung sounds would be beneficial to the healthcare sector [14]. This kind of system would act as an assisting tool for doctors as well as a learning tool for medical students. Since respiratory diseases are a global health issue, having a simple and effective classification system could improve patient care, make healthcare processes smoother and help manage diseases better.

This study provides a survey of related literature and offers a comprehensive analysis of existing research findings. These serve as valuable resources, consolidating diverse studies to provide a more subtle understanding of the subject matter and aiding researchers and practitioners in making informed decisions based on the collective evidence in the field. Although there are systematic reviews of other respiratory disease diagnoses using images such as chest X-rays [4], there is a lack of surveys conducted on lung sound classification and analysis. The existing lung sound classification surveys and reviews [15–19] are limited in their coverage, particularly regarding explainability techniques and support tools integrated into the classification process. Understanding the intricacies of lung sound classification is crucial for improving diagnostic accuracy and enhancing patient care in respiratory medicine. By addressing this gap, our survey paper aims to extend the current understanding of lung sound classification methodologies, thereby providing researchers and clinicians with a more robust foundation for developing effective diagnostic tools and treatment strategies.

Additionally, this paper contributes to advancing the field by identifying areas for further research and innovation, ultimately facilitating better healthcare outcomes for patients with respiratory conditions. In this paper, we have addressed these requirements with a comprehensive analysis of literature from 2018–2024 (Q1), where Q1 represents the first quarter of the year, covering the main areas of data preprocessing, feature extraction, classification models, explainability, and support tools in the lung sound domain. We also discuss existing challenges and state possible future research directions in the considered domain of study. The goal of this survey is to explore the following research questions by analysing the utilisation of DL techniques within the lung sound identification domain.

RQ1: Classification performance: How accurately can the DL models differentiate between lung sound classifications with different diseases?
RQ2: Impact of data augmentation: Can adding noise or altering pitch improve model accuracy in lung sound classification?
RQ3: Audio inherent features: Will different types of features relevant to audio data impact the performance of classification models differently?

RQ4: Optimal architectures: Which DL models classify with high accuracy?
RQ5: Explainability: How to make the model's decisions more understandable by highlighting relevant regions in the audio waveform?

2 METHODS

This survey explores literature published between 2018 and 2024 (Q1) in the domain of lung sound classification using DL techniques. We followed the preferred reporting for systematic reviews and meta-analysis extension for scoping reviews (PRISMA-ScR) guidelines proposed by Tricco et al. [20] to identify, screen, select eligibility, and filter the included articles for this study, as shown in Figure 1.

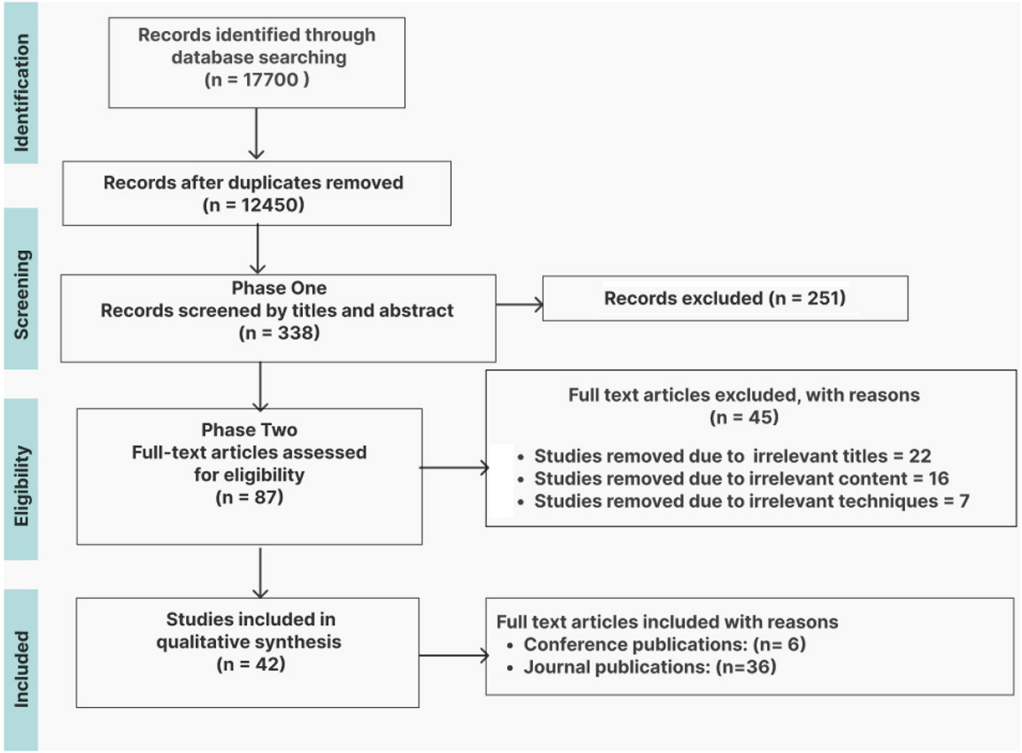

Fig. 1. PRISMA process for article selection

Identification criteria. We used Google Scholar search to identify all the possible studies in the domain of 'lung sound classification using DL.' We filtered the studies based on the available filtering options, such as custom year range, study areas and language, for which we chose English to narrow down the search results.

Screening criteria. The search strategy of the screening process was based on articles from publishers such as IEEE, Elsevier, Springer, ACM and MDPI. The main reason behind selecting these platforms was the high chance of finding many relevant studies. We removed the duplicates and screened the identified articles from Google Scholar search for their eligibility by their title and abstract. This process is mainly done by the first three authors. The contradictory opinions were resolved by analysing those articles by the fourth author.

Eligibility criteria. In this process, we excluded articles based on irrelevant titles, irrelevant content, and unrelated techniques. The first three authors have divided the articles equally and performed the data extraction process. The accuracy and consistency were checked by the rest of the authors.

Inclusion criteria. Accordingly, for this survey, we have included 42 articles, including 6 conference publications and 36 journal publications on lung sound classification.

3 TAXONOMY

The taxonomy for the related studies in the lung sound classification domain is shown in Figure 2. Different techniques have been used for predictions based on the purpose of the application and the available dataset. Data preprocessing approaches including normalisation [21], augmentation [22–24], feature extraction [23], [25], [26], [19], [22], classification [27], [26], [28], [10], [29], and explainability [30], [31], [32], [33], [34] were considered. As technology evolves, the taxonomy changes through the inclusion of new techniques, thereby contributing to the advancement of DL applications in lung sound analysis.

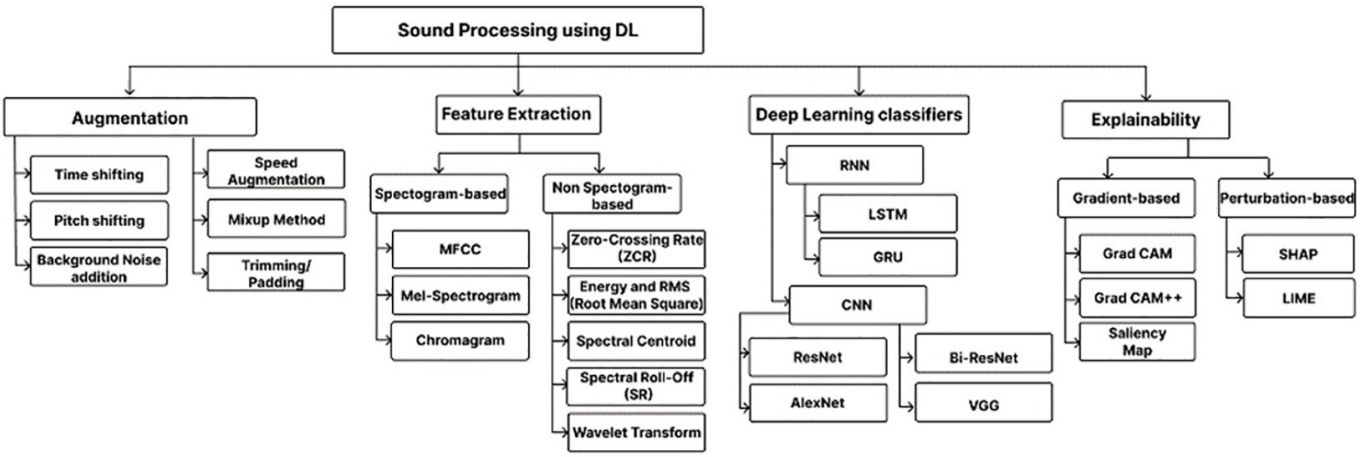

Fig. 2. Taxonomy of techniques

Considering the data pre-processing, normalisation methods such as min-max normalisation are used to adjust the scale of features, ensuring that audio characteristics remain consistent even when recorded using different devices [21]. Min-max normalisation re-scales numerical features to fit within a certain range, usually between 0 and 1, which helps maintain consistency across all features [35]. Data augmentation increases the training dataset and thereby improves the performance of models by reducing overfitting, improving generalisation, and making the model more robust. Data augmentation applies different transformations or adjustments to existing data samples, creating new samples that still belong to the same category as the original data [36]. For augmenting audio data, noise injection, time shifting, pitch shifting, mix-up methods, and speed changes can be applied. Numpy offers a straightforward approach for handling noise injection and time shifting, whereas Librosa is a library that allows for pitch shifting and speed changes with minimal code [37]. Pitch shifting is an augmentation technique that is commonly used in the lung sound classification domain. In this process, the frequency or pitch of the audio is modified by one or more semitones while preserving its duration [36]. On the other hand, time shifting shifts the audio either to the left or right by a random number of seconds [37]. In time shifting, the time series is stretched by a fixed rate. Moreover, the noise addition technique adds random values to the data.

This can improve the model's ability to withstand environmental noise or variations in recording conditions [36].

Various feature extraction methods incorporate time-domain, frequency-domain, and time-frequency domain features. Prominent methods for feature extraction in the frequency domain include Mel-frequency cepstral coefficients (MFCCs), Mel-spectrograms and chromagrams in the domain of lung sound classification. A spectrogram visually represents the spectrum of frequencies in a signal, representing the changes over time. A typical format of a spectrogram consists of a two-dimensional graph where one axis denotes time, another represents frequency, and the intensity or colour of each point in the image indicates the amplitude of a specific frequency at a given time [10]. Apart from frequency domain features, we can extract features such as root mean square (RMS), spectral centroid (SC) and zero crossing rate (ZCR). For a continuous-time waveform, the RMS value is calculated by squaring the function defined in the continuous waveform. The SC, a metric in digital signal processing, characterises a spectrum by indicating the location of its centre of mass [10]. ZCR measures the rate at which the signal transitions between positive and negative values.

Among the classifiers, the most preferred model in the lung sound classification domain is convolutional neural networks (CNNs), as it automatically learns to extract low-level features of the input audio, such as edges, and high-level features, such as complex patterns. CNNs are specialised deep-learning models designed for grid-like data, such as images or spectrograms. By utilising convolutional layers with small filters, CNNs automatically learn hierarchical features through the convolution operation. On the other hand, recurrent neural networks (RNNs) are rarely applied in lung sound classification, offering unique advantages in handling sequential data. RNNs have connections that form a directed cycle, allowing them to capture the temporal dependencies in lung sound recordings. This helps in understanding the sequential nature of respiratory data, where patterns are useful for accurate classification. By employing long short-term memory (LSTM) networks or gated recurrent units (GRUs), challenges such as vanishing gradients can be resolved, and the performance of the classifier can be improved.

Although DL-based models interface with the medical domain as an assistive application, it is important to show the transparency of the prediction process. Explainability techniques support showing the regions of interest of the input that cause the predictions by the classifier. Explainable artificial intelligence (XAI) is an evolving approach and helps to increase the trustworthiness of the DL model [38]. Some of these techniques include gradient-weighted class activation mapping (Grad-CAM), saliency maps, layer-wise relevance propagation (LRP) and local interpretable model-agnostic explanations (LIME). Grad-CAM is a commonly used XAI technique in image classification [39]. This provides a way to interpret the most relevant and crucial regions utilised in the prediction given by the classification model. The saliency map, on the other hand, is an XAI method with more control than Grad-CAM. The saliency map gives a value from 0 to 1 for each pixel in an image to indicate the level of contribution of the pixel to the final prediction [40].

4 RELATED STUDIES

Several studies are available for lung sound classification, and the performance of these approaches highly relies on the specific techniques employed for data preprocessing, feature extraction and classification.

Table 1. Comparison of used techniques

Technique	Advantages	Limitations	Studies
Spectrograms: MFCC Mel-Spectrogram Chromagram	Spectrograms are less sensitive to noise and variations. Represent time-frequency domain.	Loss of temporal information. Sensitivity to windowing parameters.	[10], [11], [12], [23], [24], [26], [28], [29], [41], [42], [43]
CNNs: ResNet VGG Alexnet	Learn spatial relationships in the data. Learn frequency patterns in sound.	Need a large set of labelled data to adjust weights and biases for accurate predictions.	[10], [13], [21], [27], [29], [42], [44], [45]
RNNs: LSTM GRU	Used for sequential data analysis.	Do not perform well in handling spatial information.	[24], [29], [46]
SVM	Relatively insensitive to noise. Perform well in small datasets.	Can not handle spatial data, which leads to low accuracy.	[43], [46], [47], [48]

Table 2. Summary of related studies

Study/Year	Dataset [#classes]	Techniques	Performance
[11]/2024	ICBHI 2017 [10]	FE: Mel, MFCC, Chromagram CL: CNN XAI: Grad-CAM, Saliency map	Acc: 91.04%
[32]/2023	Clinical dataset [6]	FE: Mel spectrogram CL: CNN, VGG16 XAI: Grad-CAM	Acc: 92.56%
[10]/2022	ICBHI 2017 [6]	FE: MFCC, Mel-Spectrogram, Chromagram CL: CNN	Acc: 99%
[22]/2022	ICBHI 2017 [2,7]	FE: Chromagram, RMS, SC, Bandwidth, MFCC, ZCR, Poly CL: kNN, SVM, CNN, Logistic Regression	F1: 0.983
[23]/2021	ICBHI 2017 [2]	FE: Mel-Spectrogram, MFCC, Chromagram, Q-Chromagram CL: CNN	Sen: 92% Spe: 92%
[42]/2021	Clinical data [4]	FE: Mel-spectrogram CL: CNN, VGG	Acc: 85.7%
[49]/2021	ICBHI 2017 [6]	FE: Entropy features CL: Boosted DT	Acc: 98%
[27]/2020	ICBHI 2017 [2,6]	CL: Lightweight CNN FE: CWT and EMD scalogram	Acc: 98.92%
[26]/2020	ICBHI 2017 [6]	Augmented with random noise FE: FCC CL: CNN	Acc: 95.67%
[44]/2020	RALE, Think Labs [4]	FE: Scalograms CL: CNN	Acc: 93.78%
[50]/2019	ICBHI 2017 [4]	FE: Spectrogram CL: CNN, SVM	Acc: 65%
[3]/2018	ICBHI 2017 [4]	FE: Spectral features CL: Boosted DT	Acc: 85%

Notes: FE: feature extraction, CL: classification, XAI: explainable AI, Acc: accuracy, Sen: sensitivity, Spe: specificity, F1: F1 score.

Table 1 shows a comparison of used techniques in the literature, and a summary of related studies is depicted in Table 2. Here, the stated abbreviations are defined as follows: RMS-root mean square, SC-spectral centroid, ZCR-zero crossing rate, CWT-continuous wavelet transform and EMD-empirical mode decomposition.

Among them, utilising CNN models with feature-based fusion to classify lung and heart sounds done by Tariq et al. [10] has shown promising results with the ICBHI 2017 respiratory dataset. They applied augmentation techniques such as noise addition, time stretching, and pitch shifting to increase the dataset. The sliding window technique was used to improve accuracy by analysing consistent data in three-second clips. They extracted audio-inherent features such as MFCC, Mel-spectrogram, and chromagram from lung sound audio. CV2 and NumPy libraries were used to visually represent the feature vectors extracted from these three types of spectrograms and were transformed into JPG images with dimensions [128 × 128]. They trained three classification models namely, FDC-1, FDC-2, and FDC-3, to extract features concurrently and independently by feeding the spectrograms to each classification model. These individual classifiers were then transferred into the FDC-FS fusion model to deliver the final prediction. The fusion model was built using the output features of three models: FDC-1, FDC-2, and FDC-3. This study showed a classification accuracy of 97% for six classes in the original dataset and 99% accuracy for the augmented dataset.

Another study by Brunese et al. [22] proposed a two-step method for classification. The initial phase distinguished between healthy and abnormal lung sounds. The second classifier identified the respiratory disease type, including asthma, bronchiectasis, bronchiolitis, COPD, pneumonia, and lower or upper respiratory tract infections. They combined nine features into a single feature vector to feed into machine learning (ML) models. The feature vector included chromagram, RMS, SC, bandwidth, Tonnetz, MFCC, ZCR and poly features, which were extracted from audio files. They utilised supervised ML algorithms, namely, K-nearest neighbours (KNN), SVM, CNN, and logistic regression, for disease prediction. The highest F-measure was achieved by the CNN, which features one convolutional layer with a filter size of 5 × 5 and a pool size of 2 × 2, each with 100 feature maps. The F-measure for disease detection and disease categorisation was 0.983 and 0.923, respectively. This study indicated the feasibility of utilising a two-step classifier to identify lung disease rather than solely distinguishing between healthy individuals and those affected by lung diseases.

A lightweight CNN is proposed by Shuvo et al. [27], utilising scalogram images of lung sounds to classify respiratory diseases. The authors used the ICBHI 2017 challenge dataset for a three-class and a six-class classification. The authors have utilised six-s audio clips and extracted the features to form scalograms employing empirical mode decomposition (EMD) and continuous wavelet transform (CWT). They developed a CNN model comprising four convolutional layers and five dense layers, followed by the output layer. The same architecture was used for both three-class and six-class classifications. The suggested approach has yielded weighted accuracy of 98.92% and 98.70% for three-class chronic classification and six-class pathological classification, respectively.

Moreover, Srivastava et al. [23] have proposed a CNN classifier to identify COPD. They used the lung sounds from the ICBHI 2017 dataset that were subjected to data pre-processing, where they trimmed the audio samples into 20-s windows using a Python library named Librosa. They utilised five feature extraction techniques,

including MFCC, Mel-spectrogram, chromagram, constant Q-chromagram, and chroma energy normalised variant. They assigned 40 as the 'n' value for each extracted feature to maintain consistency across features. The authors claimed that the classifier is resistant to noisy lung sound inputs and computationally efficient with less memory storage. The best performance was observed by the features extracted from augmented data using MFCC, which achieved 92% for both sensitivity and specificity.

Another study by Wanasinghe et al. [11] proposed a multi-class classifier to classify the input lung sound audio data into 10 classes, including healthy conditions and nine abnormal conditions. This model has achieved the highest accuracy of 91.04% for 10 classes. Additionally, they have applied XAI techniques to interpret the given prediction. They used the Grad-CAM technique to compute the weighted sum of the gradients of the output in the final convolutional layer. Thus, the target class visualises the relevance of the input feature for the final prediction of the classifier [51]. Unlike pure image classifications, the interpreted areas correlate to specific frequency bands; thus, issues such as the way that the frequency band is connected to the model's prediction emerged. Next, the saliency approach was applied to backtrack and analyse the model's interpretations to examine the behaviour of the audio waveform. The authors used two different saliency techniques, both of which used back-propagation to evaluate the relative importance of each input layer feature. The regular ReLU activation function was used in the first method, while in the second approach, to mitigate the contribution loss that occurred from nodes that resulted from the usage of pooling layers, they employed a guided ReLU function [40]. Both approaches returned values ranging from 0 to 1 for each input feature to represent the contribution levels for a particular prediction. Further, they applied a threshold value by considering the values in the saliency map to hide the low-contribution features. Three distinct thresholds—low, average, and high—were calculated to show the contribution and improve the visualisation. For each input feature, both methods provide values between 0 and 1 that represent the contribution levels to a given prediction.

Subsequently, the study mentioned above has extended to a binary classification model followed by a multi-class classifier [14]. As the initial steps of both approaches, they applied pitch shifting by one semitone as a data augmentation technique, followed by normalisation. Secondly, they used MFCC, Mel-spectrogram and chromagram, which are audio-inherent features. Third, they stacked the extracted spectrogram types on top of each other to form a 3D feature representation for every audio sample in the dataset. Fourth, they fed the created 3D features to the CNNs according to the specific approach. In the initial binary classification phase, lung sounds are identified as healthy or abnormal, enabling early exits for normal cases. The second phase categorises abnormal sounds into one of nine specific diseases. The highest accuracy achieved for disease detection using binary classification and disease prediction using multi-class classification were 94.09% and 92.59%, respectively. Additionally, the proposed binary classification model, followed by a multi-class classifier, has been deployed as a user-friendly mobile application for seamless integration with the classification models. The mobile application allows users to upload lung sound files and predicts the top three potential diagnoses. The proposed two-phase strategy has optimised computational resources, providing high accuracy and efficiency. The mobile application extended the practical application of diagnostic models, with a system usability score above 73.2% showing good usability.

A comparison of related studies is stated in Table 3. Accordingly, the study by Tariq et al. [10] has shown the highest accuracy of 99%. However, they have employed augmentation techniques to increase the dataset size by about ten times

the original dataset and labelled it into six classes. Moreover, Chen et al. [52] utilised ResNet-50 with various features, consisting of over 23 million training parameters, and achieved an accuracy of 98.79%. However, their classification was limited to only three classes. In contrast, Wanasinghe et al. [11] developed a model with fewer trainable parameters, allowing for the execution of a 10-class classification model without a significant decrease in accuracy in a resource-constrained mobile environment. Furthermore, Choi et al. [32] employed a CNN with an attention module to classify lung sounds into six classes that achieved an accuracy of 92.5%. Additionally, they utilised Grad-CAM as an explainable AI technique to interpret the crucial regions of the input features, which mostly contribute to the model's prediction process.

Table 3. Comparison of used techniques

Study/Year	Accuracy/F1-Score	Classes	Explainability	Real-Time Application
[11]/2024	91.04%	10	✓	X
[14]/2024	92.59%	10	✓	✓
[32]/2023	92.56%	6	✓	X
[10]/2022	99%	6	X	X
[22]/2022	0.923%	7	X	X
[23]/2021	93%	1	X	X
[42]/2021	85.7%	3	X	X
[49]/2021	98%	6	X	X
[27]/2020	98.70%	6	X	X
[44]/2020	83.78%	4	X	X
[13]/2019	98.79%	3	X	X
[50]/2019	65%	4	X	X
[3]/2018	85%	4	X	X
[53]/2018	93.3%	2	X	X

5 DISCUSSION

5.1 Lessons learned

This paper concludes the following aspects in the domain of lung sound classification using DL techniques, reflecting the research questions considered:

Classification performance (RQ1). Different models proposed by recent studies have achieved varied accuracy in identifying respiratory diseases based on lung sound data. For instance, according to Table 2, the fusion model developed by Tariq et al. [10] achieved a high accuracy of 99% in classifying six lung diseases, while the CNN developed by Brunese et al. [22] showed 0.92 as the F-measure for disease categorization of six lung conditions.

Impact of data augmentation (RQ2). Augmentation techniques support increasing the dataset size and addressing the class imbalance issue. For example, in [10], the model accuracy increased by 6%, indicating the importance of augmentation. Moreover, the approach presented by Wanasinghe et al. [11] employed pitch shifting by one semitone to augment samples of two classes. The classification

performance slightly increased, but an answer could not be drawn since they only augmented two classes, and the increase in accuracy could also be due to the nature of the classifier.

Audio inherent features (RQ3). According to Table 2, many researchers have utilised individual features such as MFCC and Mel spectrograms to feed CNN models. However, the authors of [11] experimentally confirmed the impact of different types of features. When utilising 2D convolutional layers, a 3D input can be given as input to the CNN. A single type of feature extraction function with different parameters can be employed to generate a 3D input. After experimenting, different results were achieved by the authors for different types of inherent audio features. Combining various feature types and feeding the CNN yielded higher results, thus indicating the significance of using different audio inherent features combined.

Optimal architectures (RQ4). Researchers have explored both traditional ML models and deep neural networks for classification purposes [22], [10], [27], [23], and [11]. According to the results of classifications corresponding to performance metric values in Table 2, CNNs have performed well in classifying lung sounds into different classes.

Transparency of the decisions made by the classifiers (RQ5). It is important to show the reasons behind the automated prediction models so that users can use the DL-based applications with confidence. According to Table 3, only a few studies have explored explainability. Wansinghe et al. [11] used saliency maps to visualise the most crucial regions in the original lung sound waveform to interpret the prediction with different relevance levels. These regions were evaluated by masking the most relevant features from each stacked 3D representation of spectrograms created for the test dataset to generate a new masked dataset and assess the original model. The regions depict the time intervals so that a medical practitioner can carefully listen to that specific part of the audio to make decisions. Unlike highlighting spectrograms, this method allows doctors and medical students to analyse the lung sound audio to recognise the interpretation of the model's results.

5.2 Existing challenges and future research directions

One prominent limitation of the literature is the tendency of studies to restrict their datasets to a single source. Notably, Table 2 shows that the ICBHI dataset is the most widely used. The number of maximum classes utilised for classification is six, since classes such as asthma and URTI are not used due to insufficient sample sizes. The number of conditions can be increased by combining with other datasets [11] to address the different natures of diseases. Another limitation that has not been significantly addressed by previous research is the combination of different audio-inherent features. Most of the studies have relied on individual feature representation techniques regarding the extraction of inherent audio features from lung sound audio samples. However, this method may constrain models such as CNNs to grasp broader spatial patterns present within the input data. Moreover, few studies have explored XAI to interpret the classification predictions, primarily relying on interpretations based on frequency. Thus, in real-world applications, it would be crucial to classify lung sounds into various categories and interpret the prediction within the input waveform.

There hasn't been much research exploring the transition from building classification models to creating practical support tools for medical practitioners and interns. It is possible to develop a tool like a mobile application consisting of developed

models, as done by the authors in [14]. Different methods could be explored to integrate classification models into usable support tools.

While researchers may gradually address the aforementioned challenges, issues such as data imbalances are likely to persist. This is primarily due to the difficulty in obtaining a dataset that equally represents various lung conditions, as certain diseases are rare while others occur more frequently. In the future, advancements in lung sound classification can be driven by several key extensions.

Increase the generalisability of the lung sound dataset. There is a need to create diverse datasets while adhering to ethical guidelines. This could be beneficial to improve classification performance with a greater number of lung sound audio samples.

Employ explainability to interpret the prediction of the classification model. The exploration of explainability techniques requires the attention of researchers in this domain to interpret the model's predictions. This will foster trust among medical practitioners in integrating AI into the healthcare sector.

Develop support tools to assist medical practitioners and medical students. The development of user-friendly support tools that integrate classification models and preprocessing techniques has the potential to improve respiratory care and provide clinicians with accessible insights during patient examinations.

6 CONCLUSION

This paper has explored the literature on lung sound classification using DL techniques. The analysis of various studies highlights the dynamic evolution within the field, reflecting the continuous efforts to enhance the accuracy and efficiency of classification models. As evidenced by the related applications, DL has proven to be a promising and adaptable tool in the lung sound classification domain to diagnose respiratory diseases. The ongoing advancements in techniques such as feature extraction, data augmentation, and explainability contribute to the robustness and versatility of these models. Despite the progress made in related studies, future research should aim to address challenges such as the implementation of real-world applications, model interpretability, and the incorporation of real-world clinical data to further validate the practical utility of DL in the domain of lung sound analysis.

7 REFERENCES

[1] R. S. Gupta, A. Koteci, A. Morgan, P. M. George, and J. K. Quint, "Incidence and prevalence of interstitial lung diseases worldwide: A systematic literature review," *BMJ Open Respiratory Research*, vol. 10, no. 1, p. e001291, 2023. https://doi.org/10.1136/bmjresp-2022-001291

[2] F. C. Schmitt, *et al.*, "The world health organization reporting system for lung cytopathology," *Acta Cytologica*, vol. 67, no. 1, pp. 80–91, 2023. https://doi.org/10.1159/000527580

[3] G. Chambres, P. Hanna, and M. Desainte-Catherine, "Automatic detection of patient with respiratory diseases using lung sound analysis," in *2018 International Conference on Content-Based Multimedia Indexing (CBMI)*, 2018, pp. 1–6. https://doi.org/10.1109/CBMI.2018.8516489

[4] D. Meedeniya, H. Kumarasinghe, S. Kolonne, C. Fernando, I. De La Torre Díez, and G. Marques, "Chest X-ray analysis empowered with deep learning: A systematic review," *Applied Soft Computing*, vol. 126, p. 109319, 2022. https://doi.org/10.1016/j.asoc.2022.109319

[5] N. Wijethilake, D. Meedeniya, C. Chitraranjan, I. Perera, M. Islam, and H. Ren, "Glioma survival analysis empowered with data engineering—a survey," *IEEE Access*, vol. 9, pp. 43168–43191, 2021. https://doi.org/10.1109/ACCESS.2021.3065965

[6] T. Shyamalee, D. Meedeniya, G. Lim, and M. Karunarathne, "Automated tool support for glaucoma identification with explainability using fundus images," *IEEE Access*, vol. 12, pp. 17290–17307, 2024. https://doi.org/10.1109/ACCESS.2024.3359698

[7] L. Gamage, U. Isuranga, D. Meedeniya, S. De Silva, and P. Yogarajah, "Melanoma skin cancer identification with explainability utilizing mask guided technique," *Electronics*, vol. 13, no. 4, p. 680, 2024. https://doi.org/10.3390/electronics13040680

[8] N. S. Haider, B. K. Singh, R. Periyasamy, and A. K. Behera, "Respiratory sound based classification of chronic obstructive pulmonary disease: A risk stratification approach in machine learning paradigm," *J. Med. Syst.*, vol. 43, no. 8, 2019. https://doi.org/10.1007/s10916-019-1388-0

[9] M. Fraiwan, L. Fraiwan, B. Khassawneh, and A. Ibnian, "A dataset of lung sounds recorded from the chest wall using an electronic stethoscope," *Data in Brief*, vol. 35, p. 106913, 2021. https://doi.org/10.1016/j.dib.2021.106913

[10] Z. Tariq, S. K. Shah, and Y. Lee, "Feature-based fusion using CNN for lung and heart sound classification," *Sensors*, vol. 22, no. 4, p. 1521, 2022. https://doi.org/10.3390/s22041521

[11] T. Wanasinghe, S. Bandara, S. Madusanka, D. Meedeniya, M. Badara, and I. De La Torre Díez, "Lung sound classification with multi-feature integration utilizing lightweight CNN model," *IEEE Access*, vol. 12, pp. 21262–21276, 2024. https://doi.org/10.1109/ACCESS.2024.3361943

[12] B. M. Rocha, D. Pessoa, A. Marques, P. Carvalho, and R. P. Paiva, "Automatic classification of adventitious respiratory sounds: A (un)solved problem?" *Sensors (Basel)*, vol. 21, no. 1, p. 57, 2021. https://doi.org/10.3390/s21010057

[13] H. Chen, X. Yuan, Z. Pei, M. Li, and J. Li, "Triple-classification of respiratory sounds using optimized s-transform and deep residual networks," *IEEE Access*, vol. 7, pp. 32845–32852, 2019. https://doi.org/10.1109/ACCESS.2019.2903859

[14] T. Wanasinghe, S. Bandara, S. Madusanka, D. Meedeniya, and M. Badara, "CNN-based optimization for lung sound classification with mobile accessibility," in *International Conference on Smart Computing and Systems Engineering (SCSE) 2024*, 2024.

[15] R. Palaniappan, K. Sundaraj, N. Ahamed, A. Arjunan, and S. Sundaraj, "Computer-based respiratory sound analysis: A systematic review," *IETE Tech. Rev.*, vol. 30, no. 3, p. 248, 2013. https://www.tandfonline.com/doi/abs/10.4103/0256-4602.113524

[16] R. X. A. Pramono, S. Bowyer, and E. Rodriguez-Villegas, "Automatic adventitious respiratory sound analysis: A systematic review," *PLoS One*, vol. 12, no. 5, p. e0177926, 2017. https://doi.org/10.1371/journal.pone.0177926

[17] H. Sfayyih, N. Sulaiman, and A. H. Sabry, "A review on lung disease recognition by acoustic signal analysis with deep learning networks," *J. Big Data*, vol. 10, no. 1, 2023. https://doi.org/10.1186/s40537-023-00762-z

[18] T. Nguyen and F. Pernkopf, "Chapter 9-computational lung sound classification: A review," in *State of the Art in Neural Networks and Their Applications*, A. S. El-Baz and J. S. Suri, Eds. Academic Press, 2023, pp. 193–215. https://doi.org/10.1016/B978-0-12-819872-8.00016-1

[19] R. Dubey and R. M Bodade, "A review of classification techniques based on neural networks for pulmonary obstructive diseases," in *Proceedings of Recent Advances in Interdisciplinary Trends in Engineering & Applications (RAITEA)*, 2019. https://doi.org/10.2139/ssrn.3363485

[20] C. Tricco, *et al.*, "Prisma extension for scoping reviews (prisma-scr): Checklist and explanation," *Annals of Internal Medicine*, vol. 169, no. 7, pp. 467–473, 2018. https://doi.org/10.7326/M18-0850

[21] 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*, Beijing, China, 2020, pp. 2902–2906. https://doi.org/10.21437/Interspeech.2020-2487

[22] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "A neural network-based method for respiratory sound analysis and lung disease detection," *Applied Sciences*, vol. 12, no. 8, p. 3877, 2022. https://doi.org/10.3390/app12083877

[23] Srivastava, S. Jain, R. Miranda, S. Patil, S. Pandya, and K. Kotecha, "Deep learning based respiratory sound analysis for detection of chronic obstructive pulmonary disease," *PeerJ Computer Science*, vol. 7, p. e369, 2021. https://doi.org/10.7717/peerj-cs.369

[24] J. Acharya and A. Basu, "Deep neural network for respiratory sound classification in wearable devices enabled by patient specific model tuning," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 14, no. 3, pp. 535–544, 2020. https://doi.org/10.1109/TBCAS.2020.2981172

[25] G. Serbes, S. Ulukaya, and Y. P. Kahya, "An automated lung sound preprocessing and classification system based on spectral analysis methods," in *Precision Medicine Powered by pHealth and Connected Health: ICBHI 2017*, Thessaloniki, Greece, Greece: Springer, 2018, pp. 45–49. https://doi.org/10.1007/978-981-10-7419-6_8

[26] V. Basu and S. Rana, "Respiratory diseases recognition through respiratory sound with the help of deep neural network," in *4th International Conference on Computational Intelligence and Networks (CINE)*, Kolkata, India: IEEE, 2020, pp. 1–6. https://doi.org/10.1109/CINE48825.2020.234388

[27] S. B. Shuvo, S. N. Ali, S. I. Swapnil, T. Hasan, and M. I. H. Bhuiyan, "A lightweight CNN model for detecting respiratory diseases from lung auscultation sounds using EMD-CWT-based hybrid scalogram," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 7, pp. 2595–2603, 2021. https://doi.org/10.1109/JBHI.2020.3048006

[28] D. Perna and A. Tagarelli, "Deep auscultation: Predicting respiratory anomalies and diseases via recurrent neural networks," in *IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*, Cordoba, Spain: IEEE, 2019, pp. 50–55. https://doi.org/10.1109/CBMS.2019.00020

[29] G. Petmezas, G. A. Cheimariotis, L. Stefanopoulos, B. Rocha, R. P. Paiva, A. K. Katsaggelos, and N. Maglaveras, "Automated lung sound classification using a hybrid CNN-LSTM network and focal loss function," *Sensors (Basel)*, vol. 22, no. 3, p. 1232, 2022. https://doi.org/10.3390/s22031232

[30] M. Ancona, E. Ceolini, C. Öztireli, and M. Gross, "Towards better understanding of gradient-based attribution methods for deep neural networks," *arXiv preprint arXiv*:1711.06104, 2017. https://doi.org/10.48550/arXiv.1711.06104

[31] M. Du, N. Liu, and X. Hu, "Techniques for interpretable machine learning," Communications of the ACM, vol. 63, no. 1, pp. 68–77, 2019. https://doi.org/10.1145/3359786

[32] Y. Choi and H. Lee, "Interpretation of lung disease classification with light attention connected module," *Biomedical Signal Processing and Control*, vol. 84, p. 104695, 2023. https://doi.org/10.1016/j.bspc.2023.104695

[33] I. Topaloglu, P. D. Barua, A. M. Yildiz, T. Keles, S. Dogan, M. Baygin, H. F. Gul, T. Tuncer, R.-S. Tan, and U. R. Acharya, "Explainable attention resnet18-based model for asthma detection using stethoscope lung sounds," *Engineering Applications of Artificial Intelligence*, vol. 126, p. 106887, 2023. https://doi.org/10.1016/j.engappai.2023.106887

[34] E. Livieris, E. Pintelas, N. Kiriakidou, and P. Pintelas, "Explainable image similarity: Integrating siamese networks and grad-cam," *Journal of Imaging*, vol. 9, no. 10, p. 224, 2023. https://doi.org/10.3390/jimaging9100224

[35] D. Meedeniya, *Deep Learning: A Beginners' Guide*, CRC Press, 2023. https://doi.org/10.1201/9781003390824

[36] J. M. Herrera, "Audio data augmentation: Techniques and methods," https://blog.pangeanic.com/audio-data-augmentation-techniques-and-methods, accessed: 2024-1-25.
[37] E. Ma, "Data augmentation for audio," *Medium*, 2019. https://medium.com/@makcedward/data-augmentation-for-audio-76912b01fdf6, Accessed: 2024-1-25.
[38] Singh, S. Sengupta, and V. Lakshminarayanan, "Explainable deep learning models in medical image analysis," *Journal of Imaging*, vol. 6, no. 6, p. 52, 2020. https://doi.org/10.3390/jimaging6060052
[39] F. Nunnari, M. A. Kadir, and D. Sonntag, "On the overlap between grad-cam saliency maps and explainable visual features in skin cancer images," in *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*, Springer, 2021, pp. 241–253. https://doi.org/10.1007/978-3-030-84060-0_16
[40] K. Chung, "Explainable AI: How to implement saliency maps," Coders Kitchen, 2021. https://www.coderskitchen.com/explainable-ai-how-to-implement-saliency-maps/ Accessed: 2024-2-2.
[41] M. Pahar, M. Klopper, R. Warren, and T. Niesler, "Covid-19 cough classification using machine learning and global smartphone recordings," *Computers in Biology and Medicine*, vol. 135, p. 104572, 2021. https://doi.org/10.1016/j.compbiomed.2021.104572
[42] 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, 2021. https://doi.org/10.1038/s41598-021-96724-7
[43] F. Demir, A. M. Ismael, and A. Sengur, "Classification of lung sounds with CNN model using parallel pooling structure," *IEEE Access*, vol. 8, pp. 105376–105383, 2020. https://doi.org/10.1109/ACCESS.2020.3000111
[44] S. Jayalakshmy and G. F. Sudha, "Scalogram-based prediction model for respiratory disorders using optimized convolutional neural networks," *Artif. Intell. Med.*, vol. 103, no. 101809, p. 101809, 2020. https://doi.org/10.1016/j.artmed.2020.101809
[45] K. K. Lella and A. Pja, "Automatic diagnosis of covid-19 disease using deep convolutional neural network with multi-feature channel from respiratory sound data: Cough, voice, and breath," *Alexandria Engineering Journal*, vol. 61, no. 2, pp. 1319–1334, 2022. https://doi.org/10.1016/j.aej.2021.06.024
[46] L. Pham, H. Phan, R. Palaniappan, A. Mertins, and I. McLoughlin, "CNN-moe based framework for classification of respiratory anomalies and lung disease detection," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 8, pp. 2938–2947, 2021. https://doi.org/10.1109/JBHI.2021.3064237
[47] Monaco, N. Amoroso, L. Bellantuono, E. Pantaleo, S. Tangaro, and R. Bellotti, "Multi-time-scale features for accurate respiratory sound classification," *Applied Sciences*, vol. 10, no. 23, p. 8606, 2020. https://doi.org/10.3390/app10238606
[48] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "A neural network-based method for respiratory sound analysis and lung disease detection," *Appl. Sci. (Basel)*, vol. 12, no. 8, p. 3877, 2022. https://doi.org/10.3390/app12083877
[49] L. Fraiwan, O. Hassanin, M. Fraiwan, B. Khassawneh, A. M. Ibnian, and M. Alkhodari, "Automatic identification of respiratory diseases from stethoscopic lung sound signals using ensemble classifiers," *Biocybernetics and Biomedical Engineering*, vol. 41, no. 1, pp. 1–14, 2021. https://doi.org/10.1016/j.bbe.2020.11.003
[50] F. Demir, A. Sengur, and V. Bajaj, "Convolutional neural networks based efficient approach for classification of lung diseases," *Health Information Science and Systems*, vol. 8, no. 1, p. 4, 2019. https://doi.org/10.1007/s13755-019-0091-3
[51] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on computer vision*, Cambridge, MA, USA, 2017, pp. 618–626. https://doi.org/10.1109/ICCV.2017.74

[52] H. Chen, X. Yuan, Z. Pei, M. Li, and J. Li, "Triple-classification of respiratory sounds using optimized s-transform and deep residual networks," *IEEE Access*, vol. 7, pp. 32845–32852, 2019. https://doi.org/10.1109/ACCESS.2019.2903859

[53] M. A. Islam, I. Bandyopadhyaya, P. Bhattacharyya, and G. Saha, "Multichannel lung sound analysis for asthma detection," *Computer Methods and Programs in Biomedicine*, vol. 159, pp. 111–123, 2018. https://doi.org/10.1016/j.cmpb.2018.03.002

8 AUTHORS

Thinira Wanasinghe is a B.Sc. Eng. undergraduate at the University of Moratuwa, Sri Lanka.

Sakuni Bandara is a B.Sc. Eng. undergraduate at the University of Moratuwa, Sri Lanka.

Supun Madusanka is a B.Sc. Eng. undergraduate at the University of Moratuwa, Sri Lanka.

Dulani Meedeniya is a professor at the University of Moratuwa, Sri Lanka (E-mail: dulanim@cse.mrt.ac.lk).

Meelan Bandara is a B.Sc. Eng. graduate at the University of Moratuwa, Sri Lanka.

Isabel De La Torre Díez is a Professor in the Department of Signal Theory and Communications at the University of Valladolid, 47011 Valladolid, Spain.