

AI-Based Multi-Disease Detection Application Using Multimodal Deep Learning

Medha B S¹, Amrutha A B¹, Nishchitha K M¹, Swathi M¹, Naveena C²

¹Students, Department of Artificial Intelligence & Machine Learning, Vivekananda College of Engineering & Technology, Puttur, Karnataka, India

²Assistant Professor, Department of Artificial Intelligence & Machine Learning, Vivekananda College of Engineering & Technology, Puttur, Karnataka, India

Email: medhabs286@gmail.com, amruthaab84@gmail.com, nishchithabindhu90@gmail.com, swathiaiml40@gmail.com

Abstract - Respiratory diseases remain a leading cause of illness and death worldwide, yet diagnosing them usually means reading a chest X-ray, listening to breath sounds and reviewing clinical notes as three separate exercises. This paper presents a multimodal deep learning application that brings these sources together to screen for asthma, chronic obstructive pulmonary disease, pulmonary fibrosis, pneumonia and tuberculosis. Chest X-rays are analysed with a ResNet50 network, breathing recordings are converted into Mel-frequency cepstral coefficients and classified by a dense neural network, and clinical notes are vectorised with term frequency-inverse document frequency and classified by a second dense network. The three calibrated confidence scores are combined by late weighted fusion, with weights of 45%, 30% and 25% for image, audio and text, and the fused score is mapped to a disease decision and a risk level. Gradient-weighted class activation maps show which lung regions influenced the image prediction. The models sit inside a secure web application that stores patient records and produces downloadable reports. The system is built on a balanced dataset of 1,500 patient records, and its end-to-end behaviour is demonstrated on disease-present and normal cases. It is meant to support clinicians, not replace them.

Key Words: Multimodal Deep Learning, Respiratory Disease Detection, ResNet50, Late Fusion, Grad-CAM, Chest X-ray.

1. INTRODUCTION

Breathing is something most of us never think about until it becomes difficult. Asthma, chronic obstructive pulmonary disease (COPD), pulmonary fibrosis, pneumonia and tuberculosis (TB) are among the respiratory conditions that place a heavy burden on patients and health systems, and all of them are easier to manage when they are found early. The difficulty is that the person who first sees a patient is often not a radiologist or a pulmonologist, and specialist help is not always close by.

In routine practice, diagnosis leans on three separate sources of evidence: a chest X-ray, the sound of the patient's breathing through a stethoscope, and the clinical history. Each one is useful, but each is usually read on its own, and the final judgement depends heavily on the experience of the physician. Electronic health records store the results, yet they seldom combine them into a single computed assessment [1], [2]. The practical outcome is a process that is slow, hard to standardise and difficult to use for routine screening.

Deep learning has already improved how each source can be analysed by itself. Convolutional neural networks (CNNs)

classify chest radiographs with good accuracy [3], [4], [5], and models built on Mel-frequency cepstral coefficients (MFCCs) or spectrograms can separate normal from abnormal lung sounds [6], [7], [8]. Reviews of the field agree that models which fuse several kinds of data tend to outperform those that rely on one [9], [2]. Even so, many published systems focus on a single modality or a single disease, and few are packaged as usable tools that report confidence, explain themselves and produce a record a clinician can file.

This paper describes an application that tries to close that gap. A user logs in, chooses one of five respiratory diseases and uploads a chest X-ray, a breathing recording and a clinical note for the same patient. Three dedicated networks analyse the inputs separately: a ResNet50 CNN for the image, a dense neural network (DNN) on MFCC features for the audio, and a DNN on term frequency-inverse document frequency (TF-IDF) features for the text. Their calibrated outputs are merged by late weighted fusion into one confidence score, which is turned into a decision and a risk level. Gradient-weighted class activation mapping (Grad-CAM) highlights the regions of the X-ray that drove the image prediction.

The main contributions of this work are as follows. First, we build a patient-level multimodal pipeline that handles image, audio and text for five respiratory diseases in one workflow. Second, we combine the three branches with a calibrated late-fusion scheme that produces a single confidence score and a four-level risk category. Third, we add visual explanation through Grad-CAM. Finally, we wrap everything in a secure web application with authentication, record storage and downloadable reports, so that the models can be used rather than only benchmarked.

The rest of the paper is organised as follows. Section 2 reviews related work. Section 3 explains the proposed methodology. Section 4 describes the dataset and preprocessing. Section 5 presents the implementation, evaluation approach and results, Section 6 discusses explainability, and Section 7 concludes with limitations and future scope.

2. RELATED WORK

2.1 Chest X-ray classification

Transfer learning has become the standard way of classifying chest radiographs when labelled data is limited. Irtaza et al. [4] compared ten pretrained CNNs on the NIH ChestX-ray14 collection, which has 112,120 images across 14 pathologies. MobileNet performed best with 93.4% binary accuracy, but recall was only 57%, a reminder of how badly

class imbalance can hurt. Adding synthetic images from a generative adversarial network (GAN) raised the F1-score by 5%. Constantinou et al. [3] fine-tuned five pretrained networks on 33,920 images from the COVID-QU database, and ResNet101 reached 96% accuracy, precision, recall and F1-score. Khan et al. [5] reported 96.13% accuracy with EfficientNetB1 on a four-class problem, and Brunese et al. [10] paired X-ray classification with visual explanations for pulmonary disease and COVID-19 detection.

Work beyond plain radiographs points in the same direction. Thillai et al. [11] trained a 3D U-Net to segment lung, airway, vascular and fibrotic regions in computed tomography (CT) scans of patients with idiopathic pulmonary fibrosis and segmented 97.8% of scans successfully, and Huang et al. [12] reviewed 79 studies on self-supervised pretraining, which lowers the dependence on labelled images.

2.2 Respiratory sound analysis

Lung-sound classification has been studied for longer, and MFCCs appear in a large share of the work. Sreejith et al. [6] combined wavelet denoising, MFCCs, a CNN and a temporal convolutional network on the ICBHI 2017 respiratory sound database (from the International Conference on Biomedical and Health Informatics challenge) and reached 95.49% accuracy. Ayappan et al. [13] used the same database with a denoising autoencoder and an optimised bidirectional LSTM (long short-term memory) network and reported 96.84% accuracy over five classes. Engin et al. [14] obtained 97.3% with gammatonegram images and DenseNet201 on 400 respiratory cycles. Simpler feature-based models also do well on small sets: a multilayer perceptron (MLP) on energy, zero-crossing rate, spectral centroid and MFCC features reached 94% test accuracy [7], and a quadratic support vector machine (SVM) separated asthma from healthy subjects with 99.86% accuracy [8]. Haider and Behera [15] separated asthma, COPD and controls with a decision tree at 99.3% accuracy.

The headline numbers need careful reading. Hassan et al. [16] reported perfect scores inside one database, but accuracy fell to between 86.10% and 89.97% when the model was tested on a different one. Petmezas et al. [17] used a CNN-LSTM with focal loss and reached 73.69% on the four-class ICBHI task, and Kim et al. [18] reached 86.5% on 1,918 recordings from 871 patients. Reviews of the area repeatedly point to small single-centre datasets, noisy recordings and weak external validation [19].

2.3 Multimodal fusion and clinical data

The review by Estay Zamorano et al. [9] covers chest radiographs, CT, respiratory sounds and health records and concludes that multimodal models are generally more accurate than unimodal ones, while also listing scarce labels, class imbalance, demographic bias and poor transparency as open problems. Sunil and Joshi [2] describe early, late and hybrid fusion for chronic disease prediction, and Dimakatso et al. [1] found multimodal systems reaching between 85.03% and 92.80% accuracy across different diseases. A concrete example of late fusion is the study by Xu et al. [20], where ResNet features from CT scans were combined with clinical and laboratory data, giving up to 97.7% accuracy and an improvement of 15% to 42% over single-modality models. Graph-based fusion has also been explored [21], and Sahu et al. [22] ran a multi-disease CNN and tree-ensemble pipeline on an NVIDIA Jetson Nano edge device with accuracies between 90% and 99%.

On the clinical side, Kocks et al. [23] showed how much structured data can help: a classifier using only 12 clinical parameters separated asthma, COPD and their overlap with 73% accuracy, ahead of primary care physicians (50%) and pulmonologists (61%). Feng et al. [24] survey the wider use of machine learning in asthma and COPD care.

2.4 Research gap

Taken together, the literature shows strong results for individual modalities and encouraging evidence for fusion. What is still hard to find is a working system that takes an X-ray, a breathing recording and a clinical note for the same patient, covers several respiratory diseases, reports a calibrated confidence and a risk level, explains its image reasoning and generates a record for follow-up. The limitations that keep appearing, namely small datasets, class imbalance, limited external validation and low interpretability [9], [19], shaped several of our design choices: a class-balanced dataset, a visual explanation for the image branch, and a deliberate framing of the tool as decision support.

3. PROPOSED METHODOLOGY

3.1 System architecture

The application follows a client-server architecture, shown in Fig. 1. A React.js frontend handles user authentication, disease selection, patient-data upload, result visualisation, access to analysis history and PDF-report download. A Flask backend manages authentication with JSON Web Tokens (JWT), file validation, preprocessing, model inference, multimodal fusion, Grad-CAM generation, database storage and report generation.

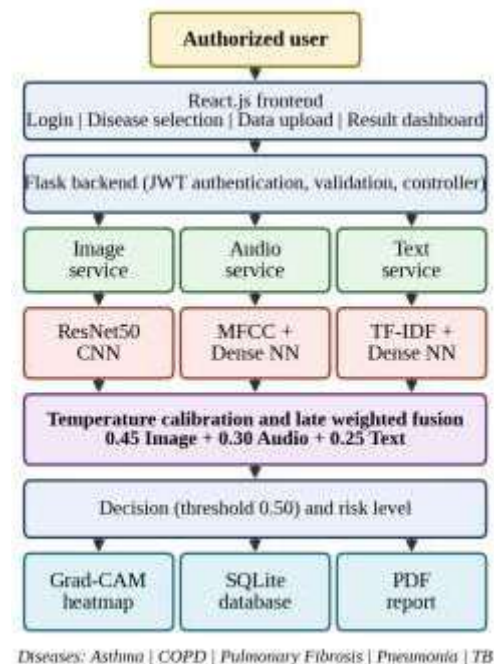


Fig -1: Proposed multimodal system architecture

The chest X-ray, breathing audio and clinical note are processed through separate prediction branches because they have different formats and need different feature-extraction techniques. Each branch has its own preprocessing service and trained model. The outputs are then calibrated and combined by a late weighted fusion mechanism. The final output includes the disease decision, overall confidence score, modality-wise confidence scores, risk level, Grad-CAM

heatmap and a downloadable PDF report. User and prediction records are stored in a SQLite database.

In a typical session the user logs in, selects Asthma, COPD, Pulmonary Fibrosis, Pneumonia or Tuberculosis, and uploads the patient's chest X-ray (PNG, JPG or JPEG), breathing recording (WAV) and clinical note (TXT). The backend validates the inputs, runs the three branches, fuses the scores, assigns a risk level, generates the Grad-CAM heatmap when an X-ray is available, stores the record and displays the result dashboard.

3.2 Chest X-ray processing

The chest X-ray is resized to 224×224 pixels and its pixel values are normalised to the 0–1 range by dividing by 255. Analysis uses a ResNet50-based convolutional neural network with transfer learning. Global average pooling, batch normalisation, dense and dropout layers are added above the pretrained base model. The image models are built with the Keras Functional API so that their internal convolutional feature maps can be accessed for Grad-CAM generation, and callbacks such as EarlyStopping and ModelCheckpoint are used to save the best-performing model. A disease-specific image model is maintained for each disease (Fig. 2).

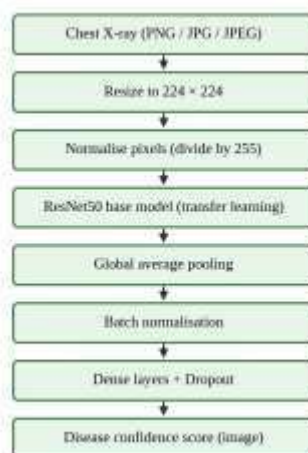


Fig -2: Chest X-ray processing using ResNet50

3.3 Breathing-audio processing

Breathing recordings are loaded at a sampling frequency of 22,050 Hz using Librosa. Each recording is set to a duration of 10 seconds: longer recordings are trimmed and shorter ones are zero-padded. Forty Mel-frequency cepstral coefficients (MFCCs) are extracted from the processed audio, and mean pooling over time gives a feature vector of length 40. MFCCs capture the characteristics of breathing sounds that are useful for identifying patterns such as wheezing and crackling. The features are standardised and passed to a dense neural network that produces the audio confidence score (Fig. 3).

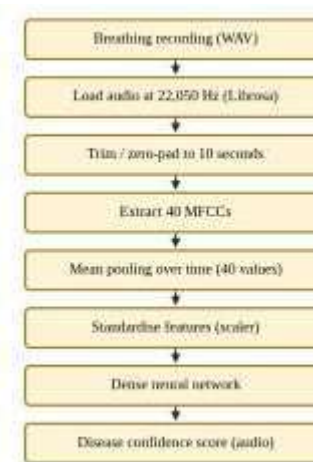


Fig -3: Breathing-audio processing using MFCC

3.4 Clinical-text processing

Clinical notes are cleaned and transformed into a numerical representation using term frequency-inverse document frequency (TF-IDF). The text is represented by 500 features, which lets the system pick out important medical terms while reducing the influence of frequently used words that carry little diagnostic value. The resulting vector is passed to a second dense neural network that produces the text confidence score (Fig. 4).

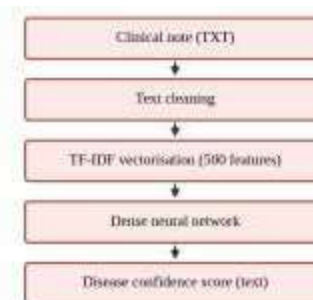


Fig -4: Clinical-text processing using TF-IDF

3.5 Multimodal fusion

The three model outputs are passed to a calibration and late-fusion stage, where temperature calibration is applied to the modality scores. The calibrated scores of the image, audio and text branches are then combined using the configured late-fusion weights: the chest X-ray contributes 45%, the breathing audio 30% and the clinical text 25%. These values are implementation-configured weights used to reflect the intended relative contribution of the modalities; they are not presented as experimentally optimized weights in the absence of a dedicated weight-search study:

$$S = 0.45 \times p_{\text{image}} + 0.30 \times p_{\text{audio}} + 0.25 \times p_{\text{text}} \quad (1)$$

This lets the system use complementary information from all three modalities while preserving the individual strengths of each model. Figure 5 summarises the fusion and decision step. Because the branches are modular, one of them, for example the audio model, can be updated without affecting the image or text modules.

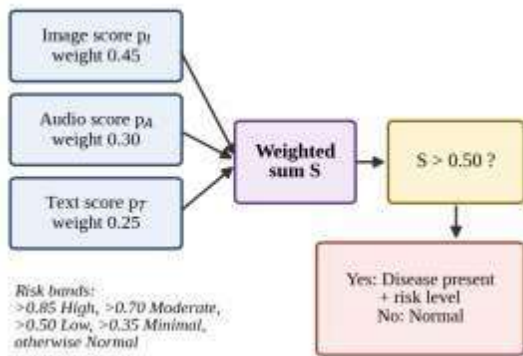


Fig -5: Weighted fusion architecture and decision logic

3.6 Disease prediction and risk level

The fused score S is compared with the decision threshold of 0.50. A score greater than 0.50 gives the output “Disease present”, whereas a score of 0.50 or below gives “Normal”. The fused score is also mapped to the implemented risk categories: above 0.85 is High, above 0.70 and up to 0.85 is Moderate, above 0.50 and up to 0.70 is Low, above 0.35 and up to 0.50 is Minimal, and 0.35 or below is Normal.

4. DATASET AND PREPROCESSING

4.1 Dataset

The study uses a multimodal dataset assembled from existing publicly available Kaggle datasets containing three forms of patient information: chest X-ray images, breathing-sound recordings and clinical notes. The system targets five respiratory diseases: asthma, COPD, pulmonary fibrosis, pneumonia and tuberculosis. Each disease has 300 patient records, consisting of 150 disease-present and 150 normal cases, giving 1,500 records in total (Table 1). The source data were downloaded from Kaggle and reorganised into patient-level records for this project. The dataset was used for academic research after removing or excluding personally identifiable information. Records were included when the required modality files were readable, correctly labelled and associated with the same patient case. Corrupted or duplicate files, records with missing required modalities, inconsistent labels and cases lacking sufficient information for preprocessing were excluded. The exact Kaggle dataset names, contributor details and URLs should be cited in the final submission using the specific source pages used by the project team. Sample images are shown in Fig. 6.

Table -1: Dataset distribution

Disease	Positive	Normal	Total
Asthma	150	150	300
COPD	150	150	300
Pulmonary Fibrosis	150	150	300
Pneumonia	150	150	300
Tuberculosis	150	150	300
Total	750	750	1,500

The available samples were divided into training, validation, and testing subsets at the patient level using an 80:10:10 split. For each disease, 240 patient records were allocated to training, 30 to validation, and 30 to testing. Across the five diseases, this resulted in 1,200 training records, 150 validation records, and 150 testing records, giving a total of 1,500 patient records. Patient-level splitting

ensured that records from the same patient did not occur in more than one subset.

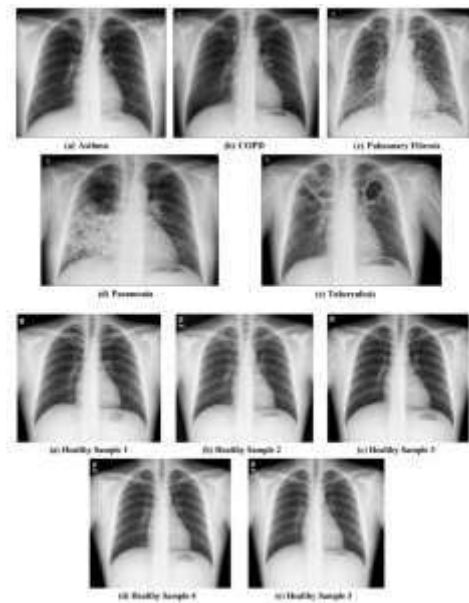


Fig -6: Sample chest X-rays from the dataset: disease-present cases (top) and normal cases (bottom)

4.2 Preprocessing

Each modality is preprocessed separately. Chest X-rays are resized to 224×224 pixels and normalised to the 0–1 range. Breathing sounds are loaded at 22,050 Hz, fixed to 10 seconds by trimming or zero-padding, converted to 40 MFCCs and mean-pooled over time. Clinical notes are converted into 500 TF-IDF features. The processed image, audio and text are then grouped by patient record, so that the three modalities of the same patient are analysed and fused together. The fitted audio scalars, the TF-IDF vectorizer and the calibration values are saved with the models, so the backend applies the same transformations during prediction as during training.

5. IMPLEMENTATION, RESULTS AND DISCUSSION

5.1 Implementation

The system is implemented in Python. TensorFlow and Keras are used to build, train and save the ResNet50-based image model and the dense neural networks for audio and text. OpenCV handles image resizing and the overlay of Grad-CAM heatmaps, Librosa handles audio loading and MFCC extraction, scikit-learn provides the TF-IDF vectorization and evaluation metrics, and NumPy and Matplotlib support numerical work and plots. The backend uses Flask, the frontend uses React.js, records are stored in SQLite, and ReportLab generates the PDF reports. The model manager loads the models for a disease only when it is needed, which reduces memory consumption.

For normal use of the web interface, the minimum hardware is an Intel Core i3 (or equivalent) processor at 1.9 GHz or higher, 4 GB of RAM (8 GB recommended), 100 GB of storage, a 64-bit Windows 10/11, macOS or Linux system and a stable internet connection. A CUDA-enabled NVIDIA GPU can significantly reduce model training and inference time.

5.2 Evaluation approach

The image, audio and text models are first examined independently on the reserved testing samples using their ground-truth labels, and the fused result is then evaluated separately. The evaluation uses accuracy, precision, recall, specificity, F1-score and the confusion matrix. Accuracy is the proportion of correctly classified samples, precision indicates the correctness of positive predictions, recall measures the ability to identify positive disease cases, specificity measures the correct identification of normal cases, and the F1-score combines precision and recall. The calibration stage is checked for consistent confidence generation, the fusion stage is verified against the configured modality weights, and the risk classification is checked against the implemented confidence ranges.

5.3 Representative Case Studies

Figure 7 shows the result pages of the application for two asthma-screening cases, and Table 2 lists the scores. The following two cases are representative qualitative demonstrations of the end-to-end system and are not intended to represent aggregate test-set performance. In the first case, the chest X-ray, breathing audio and clinical text returned 17.6%, 18.0% and 17.5%, and the overall confidence of 17.7% led to the output Normal with a Normal risk level. In the second case, the three modalities returned 68.2%, 68.0% and 68.5%, and the overall confidence of 68.2% led to the output Disease present with a Low risk level. Each result page also shows the Grad-CAM heatmap, and the analysis can be downloaded as a PDF report or a new analysis can be started.

Table -2: Results for two asthma-screening cases (confidence in %)

No.	Image	Audio	Text	Fused	Risk	Outcome
A	17.6	18.0	17.5	17.7	Normal	Normal
B	68.2	68.0	68.5	68.2	Low	Disease present



Fig -7: Result dashboard for a normal case (left) and a disease-present case (right)

The generated PDF report contains the patient information, the overall result, confidence score and risk level, the modality-wise breakdown, the original chest X-ray with its Grad-CAM heatmap, and an interpretation section that can be used for documentation and further clinical review.

5.4 System testing

The application was tested at several levels. Unit testing covered functions such as user authentication, file validation, image preprocessing, MFCC extraction, TF-IDF vectorization, model loading, fusion-score calculation, risk classification, Grad-CAM generation, database insertion and PDF report creation. Integration testing verified the communication between the React frontend and the Flask backend, the transfer of the three modality scores to the fusion module, and the link between the prediction service, Grad-

CAM component, SQLite database, report service and result dashboard. Functional and system testing followed the whole workflow from login to the final result. Usability, security and error-handling tests covered invalid credentials, protected routes, unsupported files, missing inputs and failed processing, where the system is expected to give suitable feedback instead of terminating unexpectedly. Repeated end-to-end runs were used to check reliability.

5.5 Discussion and limitations

By combining complementary patient information, the framework reduces dependence on a single diagnostic source and supports more consistent and explainable respiratory disease screening. The system is a supportive screening tool and does not replace professional medical diagnosis. Further clinical validation using larger and more diverse datasets is required before deployment in real-world healthcare settings.

6. EXPLAINABILITY USING GRAD-CAM

To make the image prediction easier to understand, the system uses Gradient-weighted Class Activation Mapping (Grad-CAM) [25]. Grad-CAM highlights the regions of the chest X-ray that influenced the prediction of the image model. Because the image models are built with the Functional API, their internal convolutional feature maps can be accessed. OpenCV is used to combine the generated heatmap with the original X-ray, with red and yellow regions showing where the CNN focused most. The heatmap is displayed on the result page and included in the PDF report next to the original X-ray. Figure 8 shows examples from the generated reports for a normal case and a disease-present case; in the disease-present case the highlighted areas fall on the lung regions.



Fig -8: Original chest X-ray and Grad-CAM heatmap from the generated PDF reports: normal case (top) and disease-present case (bottom)

Since Grad-CAM is applied to the image model, the heatmap explains only the chest X-ray prediction and not the audio or text models. It helps users and clinicians verify that the model is looking at clinically relevant lung regions.

7. CONCLUSION AND FUTURE SCOPE

This paper presented an integrated approach to respiratory disease screening using chest X-ray images, breathing audio and clinical notes. ResNet50, MFCC with a dense neural network, and TF-IDF with a dense neural network analyse the three modalities, and their predictions are combined by late weighted fusion with weights of 45% for image, 30% for audio and 25% for text. The system supports the detection of Asthma, COPD, Pulmonary Fibrosis, Pneumonia and

Tuberculosis while providing confidence scores, risk classification, Grad-CAM visual explanations and downloadable reports. With authentication, data storage, validation and an interactive interface, the application offers a practical framework for explainable respiratory disease screening. However, it is intended to support medical professionals and should not replace clinical diagnosis.

In future, the system can be trained with larger and more diverse clinical datasets to improve reliability across patient populations. Additional respiratory information such as spirometry and oxygen-saturation measurements may be incorporated to strengthen the multimodal analysis. A separate React Native/Expo mobile interface could be developed using the existing Flask backend and trained models. Improvements to Grad-CAM and other explainable AI methods may give clearer interpretations of chest X-ray predictions, and integration with hospital information systems and electronic health records could support secure clinical use, subject to validation, privacy protection and regulatory approval. Continuous performance monitoring, periodic model updating, personalised risk assessment and multilingual clinical-text processing are further enhancements.

DECLARATIONS

Ethical approval and consent: The study used anonymised, non-identifiable project data and did not involve direct recruitment, intervention or interaction with human participants. No personally identifiable patient information was retained in the application. Institutional ethical requirements and data-use permissions should be confirmed by the authors' institution before any clinical deployment or external publication.

Data and code availability: The source data were obtained from publicly available Kaggle datasets and reorganised for this project. The exact Kaggle dataset names, URLs and version details should be included in the final reference list by the authors. The project source code and trained model files are maintained as project materials and are not publicly deposited at present; they may be made available by the corresponding author on reasonable request, subject to institutional permission and applicable data-use conditions.

Funding: This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Conflict of interest: The authors declare that they have no conflict of interest related to this study.

ACKNOWLEDGEMENT

The authors thank the Principal, Dr. Mahesh Prasanna K., and the Head of the Department of Artificial Intelligence & Machine Learning, Dr. Radhika Shetty D S, for their support and encouragement. We are also grateful to the management, and to the teaching and non-teaching staff of the department, for providing the facilities and help needed to carry out this work.

REFERENCES

1. T. Dimakatso, V. Kuthadi, R. Selvaraj, and O. Dinakenyane, "Pragmatic review on progressions in multimodal disease prediction with combination of machine learning, deep learning and electronic health records," in Proc. 4th IEEE Int. Conf. ICT in Business Industry & Government (ICTBIG), 2024, doi: 10.1109/ICTBIG64922.2024.10911239.

2. S. D. M. Sunil and P. V. Joshi, "Early prediction of chronic diseases through multimodal data fusion in machine learning: A survey," in Proc. 7th Int. Conf. Intelligent Sustainable Systems (ICISS), 2025, pp. 1108–1115, doi: 10.1109/ICISS63372.2025.11076438.
3. M. Constantinou, T. Exarchos, A. G. Vrahatis, and P. Vlamos, "COVID-19 classification on chest X-ray images using deep learning methods," Int. J. Environ. Res. Public Health, vol. 20, no. 3, Art. no. 2035, 2023, doi: 10.3390/ijerph20032035.
4. M. Irtaza, A. Ali, M. Gulzar, and A. Wali, "Multi-label classification of lung diseases using deep learning," IEEE Access, vol. 12, pp. 124062–124080, 2024, doi: 10.1109/ACCESS.2024.3454537.
5. E. Khan, M. Z. U. Rehman, F. Ahmed, F. A. Alfouzan, N. M. Alzahrani, and J. Ahmad, "Chest X-ray classification for the detection of COVID-19 using deep learning techniques," Sensors, vol. 22, no. 3, Art. no. 1211, 2022, doi: 10.3390/s22031211.
6. R. Sreejith, R. K. Ramasamy, W.-N. Mohd-Isa, J. Abdullah, S. M. Jamal, F. A. Hamzah, and S. Thanjappan, "RespireNet: Enhancing lung sound classification using CNN-TCN hybrid approach," in Proc. 5th Int. Conf. Artificial Intelligence and Application Technologies (AIAT), Kyoto, Japan, Dec. 2025, pp. 1–10, doi: 10.1145/3787120.3787132.
7. W. Zhang, X. Li, Q. Liu, X. Zheng, Y. Ge, X. Pan, and Y. Zhou, "High-accuracy lung sound classification for healthy versus unhealthy diagnosis using artificial neural network," Frontiers in Bioengineering and Biotechnology, vol. 13, Art. no. 1583416, 2025, doi: 10.3389/fbioe.2025.1583416.
8. I. Topaloglu, G. Ozduygu, C. Atasoy, G. Batihan, D. Serce, G. Inanc, M. O. Güçsav, A. M. Yildiz, T. Tuncer, S. Dogan, and P. D. Barua, "Machine learning-driven lung sound analysis: Novel methodology for asthma diagnosis," Advances in Respiratory Medicine, vol. 93, Art. no. 32, 2025, doi: 10.3390/arm93050032.
9. B. Estay Zamorano, A. Dehghan Firoozabadi, P. Adasme, W. Montiel Piña, M. Chávez Muñoz, D. Zabala-Blanco, P. Palacios Játiva, and C. A. Azurdia-Meza, "Multimodal deep learning approaches for lung disease detection: A review," Medicina, vol. 62, Art. no. 1223, 2026.
10. L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays," Comput. Methods Programs Biomed., vol. 196, Art. no. 105608, 2020, doi: 10.1016/j.cmpb.2020.105608.
11. M. Thillai, J. M. Oldham, A. Ruggiero, F. Kanavati, T. McLellan, G. Saini, S. R. Johnson, F.-X. Ble, A. Azim, K. Ostridge, A. Platt, M. Belvisi, T. M. Maher, and P. L. Molyneaux, "Deep learning-based segmentation of computed tomography scans predicts disease progression and mortality in idiopathic pulmonary fibrosis," Am. J. Respir. Crit. Care Med., vol. 210, no. 4, pp. 465–472, Aug. 2024, doi: 10.1164/rccm.202311-2185OC.
12. S.-C. Huang, A. Pareek, M. Jensen, M. P. Lungren, S. Yeung, and A. S. Chaudhari, "Self-supervised learning for medical image classification: A systematic review and implementation guidelines," npj Digital Medicine, vol. 6, Art. no. 74, 2023, doi: 10.1038/s41746-023-00811-0.
13. G. Ayappan, S. Sumathi, V. Mani, and P. Esther Jebarani, "Novel deep learning-based optimization framework for the classification of respiratory diseases using lung sound analysis," Scientific Reports, vol. 16, Art. no. 19018, 2026, doi: 10.1038/s41598-026-44936-0.
14. M. A. Engin, R. Uzun Arslan, İ. S. Yapici, S. Aras, and A. Gangal, "Deep learning-based classification of common lung sounds via auto-detected respiratory cycles," Bioengineering, vol. 13, no. 2, Art. no. 170, 2026.
15. N. S. Haider and A. K. Behera, "Computerized lung sound based classification of asthma and chronic obstructive pulmonary disease (COPD)," Biocybernetics and Biomedical Engineering, vol. 42, pp. 42–59, 2022, doi: 10.1016/j.bbe.2021.12.004.
16. U. Hassan, A. Singhal, and G. Gupta, "Neural network based AI model for lung health assessment," Scientific Reports, vol. 15, Art. no. 25177, 2025, doi: 10.1038/s41598-025-09524-8.
17. 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, vol. 22, no. 3, Art. no. 1232, 2022, doi: 10.3390/s22031232.

18. Y. Kim, Y. Hyon, S. S. Jung, S. Lee, G. Yoo, C. Chung, and T. Ha, "Respiratory sound classification for crackles, wheezes, and rhonchi in the clinical field using deep learning," *Scientific Reports*, vol. 11, Art. no. 17186, 2021, doi: 10.1038/s41598-021-96724-7.
19. X. Xu and R. Sankar, "Classification and recognition of lung sounds using artificial intelligence and machine learning: A literature review," *Big Data and Cognitive Computing*, vol. 8, no. 10, Art. no. 127, Oct. 2024, doi: 10.3390/bdcc8100127.
20. M. Xu, L. Ouyang, L. Han, K. Sun, T. Yu, Q. Li, H. Tian, L. Safarnejad, H. Zhang, Y. Gao, F. S. Bao, Y. Chen, P. Robinson, Y. Ge, B. Zhu, J. Liu, and S. Chen, "Accurately differentiating between patients with COVID-19, patients with other viral infections, and healthy individuals: Multimodal late fusion learning approach," *J. Med. Internet Res.*, vol. 23, no. 1, Art. no. e25535, 2021, doi: 10.2196/25535.
21. S. Zheng, Z. Zhu, Z. Liu, Z. Guo, Y. Liu, Y. Yang, and Y. Zhao, "Multi-modal graph learning for disease prediction," *IEEE Trans. Med. Imaging*, vol. 41, no. 9, pp. 2207–2216, Sep. 2022, doi: 10.1109/TMI.2022.3159264.
22. O. Sahu, A. Singh, and D. Prakash, "Multi-disease detection using deep learning," in *Proc. 13th Int. Conf. Intelligent Systems and Embedded Design (ISED)*, 2025, doi: 10.1109/ISED67359.2025.11405003.
23. J. W. H. Kocks, H. Cao, B. Holzhauer, A. Kaplan, J. M. FitzGerald, K. Kostikas, D. Price, H. K. Reddel, I. Tsiligianni, C. F. Vogelmeier, S. Bostel, and P. Mastoridis, "Diagnostic performance of a machine learning algorithm (Asthma/Chronic Obstructive Pulmonary Disease [COPD] Differentiation Classification) tool versus primary care physicians and pulmonologists in asthma, COPD, and asthma/COPD overlap," *J. Allergy Clin. Immunol. Pract.*, vol. 11, no. 5, pp. 1463–1474, 2023.
24. Y. Feng, Y. Wang, C. Zeng, and H. Mao, "Artificial intelligence and machine learning in chronic airway diseases: Focus on asthma and chronic obstructive pulmonary disease," *Int. J. Med. Sci.*, vol. 18, no. 13, pp. 2871–2889, 2021, doi: 10.7150/ijms.58191.
25. 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 *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 618–626.