Ashmi Jomon, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Multimodal Deep Learning for Respiratory Disease Prediction Using Lung Sounds and Chest Images

Ashmi Jomon
Department of Computer Science Engineering Rajagiri School of Engineering and Technology Kochi, India

Abstract- Pneumonia remains one of the leading causes of mortality worldwide, necessitating accurate and timely diagnostic tools. Conventional diagnostic approaches often rely on a single modality such as chest X-rays or CT scans, each providing valuable but distinct clinical information. This paper presents a multimodal deep learning framework that integrates three com-plementary diagnostic modalities—Chest X-Ray images, Chest CT Scan images, and Lung Sound audio recordings—for robust and flexible pneumonia detection. Three independent deep learning models are developed: a DenseNet121 architecture for Chest X-Ray classification, a ResNet50 architecture for CT Scan analysis, and a custom Convolutional Neural Network (CNN) for Lung Sound classi-fication, where raw audio recordings are converted into Mel spectrogram images prior to inference. An attention-based Late Fusion mechanism dynamically combines the probability outputs of the individual models by assigning learned trust weights to each modality through an attention network and producing a final consensus prediction via a dedicated consensus network. The complete system is deployed as a Flask-based web applica-tion supporting both single-modality and comprehensive multi-modal prediction modes, enabling adaptability across different clinical scenarios. Experimental evaluation demonstrates that the proposed system effectively supports reliable predictions across individual modalities while also enabling enhanced inference when multiple modalities are available, evaluated using standard metrics including Accuracy, Precision, Recall, F1-Score, and ROC-AUC. The system demonstrates significant potential as an accessible and clinically meaningful decision support tool for early pneu-monia detection.

Keywords: Pneumonia detection, multimodal deep learning, chest X-ray, CT scan, lung sound classification, mel spectrogram, DenseNet121, ResNet50, late fusion, attention mechanism, con-volutional neural network, respiratory disease.

I. INTRODUCTION

Pneumonia is a severe respiratory infection that inflames the air sacs in one or both lungs and represents one of the leading causes of morbidity and mortality globally, particularly among children under five years of age and elderly popula-tions. According to the World Health Organization, pneumonia accounts for approximately 14% of all deaths of children under five years old worldwide, making early and accurate detection a critical clinical priority [1]. Despite advances in medical imaging and diagnostic technology, pneumonia diagnosis re-mains challenging due to overlapping visual and acoustic characteristics shared with other respiratory conditions such as bronchitis, COPD and pleural effusion.

Traditional computer-aided diagnosis systems for pneumo-nia often rely on a single modality, most commonly chest X-rays [2]. While chest X-rays are widely available and cost-effective, they primarily capture structural information and may vary in interpretation across observers [3]. CT scans provide higher structural resolution and detailed visualization of lung abnormalities, albeit with increased cost and radia-tion exposure. Lung sound auscultation, performed using a stethoscope, captures important acoustic characteristics—such as crackles and wheezes—that provide complementary clinical insights and are increasingly being explored in automated diagnostic systems [4]. Each modality therefore offers distinct and valuable information about the disease.

Recent advances in deep learning have demonstrated sig-nificant success in medical image and signal analysis tasks [5]. Convolutional Neural Networks (CNNs), including archi-tectures such as ResNet and DenseNet, have achieved strong

performance in chest X-ray and CT scan classification tasks [6]. In parallel, deep learning approaches applied to audio signals using spectrogram-based representations have shown promising results in respiratory sound classification [7].

While many existing approaches focus on individual modal-ities, integrating multiple complementary data sources can further enhance the robustness and flexibility of diagnostic sys-tems. Multimodal learning, which combines information from heterogeneous sources, has emerged as an effective paradigm in medical diagnosis [8]. Fusion strategies—including early fusion, feature-level fusion and late fusion—offer different trade-offs in terms of flexibility, computational complexity and performance [9]. Among these, late fusion is particularly suitable for combining independently trained models operating on fundamentally different data types.

In this paper, we present a flexible multimodal deep learning system that integrates Chest X-Ray images, Chest CT Scan images and Lung Sound audio recordings for pneumonia detection. The proposed system is designed to operate effec-tively with individual modalities while also enabling enhanced inference when multiple modalities are available.

The key contributions of this work are as follows:

- Development of three independent deep learning pipelines—DenseNet121 for Chest X-Ray, ResNet50 for CT Scan and a custom CNN for Lung Sound—each with modality-specific preprocessing, data balancing and optimized training strategies for pneumonia detection.
- A novel audio-to-image conversion pipeline that trans-forms raw lung sound recordings into 128-band Mel spectrogram images using librosa, enabling CNN-based classification of acoustic respiratory patterns.
- An attention-based Late Fusion mechanism that dy-namically assigns learned trust weights to each modal-ity's prediction and produces a final consensus diag-nosis, while also providing interpretability through per-modality attention weights.
- A complete end-to-end Flask-based web application supporting both single-modality and multi-modal pre-diction, enabling practical deployment as an accessible clinical decision support tool.

II. RELATED WORK

A. Deep Learning for Chest X-Ray Analysis

Chest X-ray pneumonia detection has been extensively stud-ied using deep convolutional neural networks. Rajpurkar et al.
[3] introduced CheXNet, a DenseNet121 model trained on the CheXpert dataset that achieved radiologist-level performance across 14 pathologies including pneumonia. Wang et al. [10] proposed the ChestX-ray14 benchmark and demonstrated the effectiveness of multi-label classification using weakly super-vised localization. More recent work by Irvin et al. [11] ex-tended this to the CheXpert dataset with uncertainty handling for ambiguous labels. While these works demonstrate strong performance on X-ray classification, they operate exclusively on a single modality and do not incorporate complementary diagnostic information from CT scans or acoustic data.

B. Deep Learning for CT Scan Analysis

CT scan-based pneumonia detection has gained significant attention particularly during the COVID-19 pandemic. Wang et al. [12] proposed a noise-robust learning framework for CT scan classification achieving high sensitivity for COVID-19 pneumonia detection. Ardila et al. [13] demonstrated that deep learning models trained on CT scans could match or exceed radiologist performance for lung cancer and pneumonia screening. ResNet-based architectures have been consistently favored for CT scan analysis due to their residual connections that enable effective training of very deep networks on high-resolution medical images [14]. These approaches, however, remain confined to CT scan data alone.

C. Lung Sound Classification

Respiratory sound analysis using deep learning has emerged as a promising non-invasive diagnostic approach. Pham et al.
[7] demonstrated that CNN models applied to Mel spectro-gram representations of lung sounds could effectively clas-sify respiratory conditions including

pneumonia, COPD and asthma. Demir et al. [15] applied deep learning to the ICBHI Respiratory Sound Database — the same dataset used in this work — achieving competitive classification performance using spectrogram-based CNN architectures. The conversion of audio signals to Mel spectrogram images has become the dominant preprocessing strategy in this domain due to its ability to capture both temporal and frequency characteristics of respiratory sounds simultaneously.

D. Multimodal Fusion in Medical Diagnosis

Multimodal fusion for medical diagnosis has been explored across several disease domains. Baltrusaitis et al. [9] provided a comprehensive survey of multimodal machine learning, categorizing fusion strategies into early, late and hybrid ap-proaches. In the context of respiratory disease, Kocheturov et al. [16] demonstrated that combining CT and clinical data improved pneumonia classification over single-modality base-lines. Attention-based fusion mechanisms have shown partic-ular promise, as they allow models to dynamically weight the contribution of each modality based on input confidence rather than relying on fixed predefined weights [17]. The present work extends this paradigm by integrating three fundamentally different modalities — imaging and audio — through a learned attention-based late fusion architecture specifically designed for pneumonia detection.

III. METHODOLOGY

A. System Architecture Overview

The proposed system consists of four interconnected layers: the Frontend Interface Layer, the Backend Processing Layer, the Model Inference Layer and the Fusion Decision Layer. Three independently trained deep learning models handle their respective input modalities, and an attention-based Late Fusion layer combines their outputs into a single final diagnosis. The complete system is deployed as a Flask web application accessible through a standard browser interface.

B. Dataset and Preprocessing

1. **Chest X-Ray Dataset:** The Chest X-Ray Pneumonia dataset sourced from Kaggle (paultimothymooney/chest-xray-pneumonia) [18] was used for training the X-Ray model. The dataset contains chest X-ray images organized into NORMAL and PNEUMONIA classes. The dataset was split into 70% training, 15% validation and 15% test sets. All images were resized to 224×224 pixels and normalized using ImageNet mean [0.485, 0.456, 0.406] and standard deviation [0.229, 0.224, 0.225] values.
2. **CT Scan Dataset:** CT Scan Dataset: The CT Scan Pneu-monia dataset (harshkumarcn005/pneumonia-ct-scan-image-dataset) [20] was downloaded from Kaggle us-ing the kagglehub library. The dataset was recursively searched for NORMAL and PNEUMONIA class folders and split into 70% training, 15% validation and 15% test sets using a stratified splitting strategy. Images were resized to 224×224 pixels and normalized using the same ImageNet parameters.
3. **Lung Sound Dataset:** The ICBHI Respiratory Sound Database (vbookshelf/respiratory-sound-database) [19] was used for lung sound classification. The dataset contains 920 audio recordings from 126 patients with associated diagnosis labels. A binary mapping was ap-plied: recordings from patients diagnosed with Pneu-monia were assigned label 1, and all other diagnoses were assigned label 0 (not pneumonia). A patient-level stratified split was performed to prevent data leakage, ensuring all recordings from a single patient appeared exclusively in one split. This resulted in a 70/15/15 train/validation/test split.
4. **Audio Preprocessing — Mel Spectrogram Conver-sion:** Raw audio recordings were converted to Mel spec-trogram images through the following pipeline: audio files were loaded and resampled to 22050 Hz using librosa; recordings were padded or trimmed to a fixed duration of 5 seconds; a 128-band Mel spectrogram was computed using an FFT window of 1024, hop length of 512, minimum frequency of 50 Hz and maximum frequency of 8000 Hz; the power spectrogram was converted to decibel scale using librosa.power to db; values were normalized to the range [0, 255] and stacked into a 3-channel RGB image of size

224×224. All spectrograms were cached to disk prior to training for computational efficiency.

5. **Data Augmentation:** For CT and X-Ray training sets, augmentation included random cropping, horizontal flip-ping, random rotation, color jitter and Gaussian blur to improve model generalization. For lung sound spec-trograms, SpecAugment-style augmentation was applied using two RandomErasing operations - one simulating time masking and one simulating frequency masking - in addition to horizontal flipping, random affine transforms and color jitter.
6. **Class Imbalance Handling:** The datasets exhibited significant class imbalance with more PNEUMONIA samples than NORMAL samples. Two complementary strategies were employed simultaneously across all three models. First, a WeightedRandomSampler assigned higher sampling probability to minority class samples ensuring approximately balanced training batches. Sec-ond, a Weighted CrossEntropyLoss was applied with per-class weights computed as the inverse of class counts, causing wrong predictions on minority class samples to incur proportionally larger gradient updates during backpropagation.

C. Model Architectures

1) DenseNet121 for Chest X-Ray: A DenseNet121 archi-tecture pretrained on ImageNet was adopted for chest X-ray classification. DenseNet's dense connectivity — where each layer receives feature maps from all pre-ceding layers — enables maximum feature reuse and is particularly effective at capturing the subtle texture differences present in chest X-ray images. The original classifier layer was replaced with a custom head con-
sisting of Linear(1024→512), ReLU, Dropout(0.4) and Linear(512→2). Training followed a two-phase progres- sive fine-tuning strategy: in Phase 1 (Epochs 1–4), the first 80% of DenseNet parameters were frozen and only the classification head was trained with a learning rate of 1e-4; in Phase 2 (Epoch 5 onwards), all parameters were unfrozen and fine-tuned with a reduced learning rate of 1e-5.

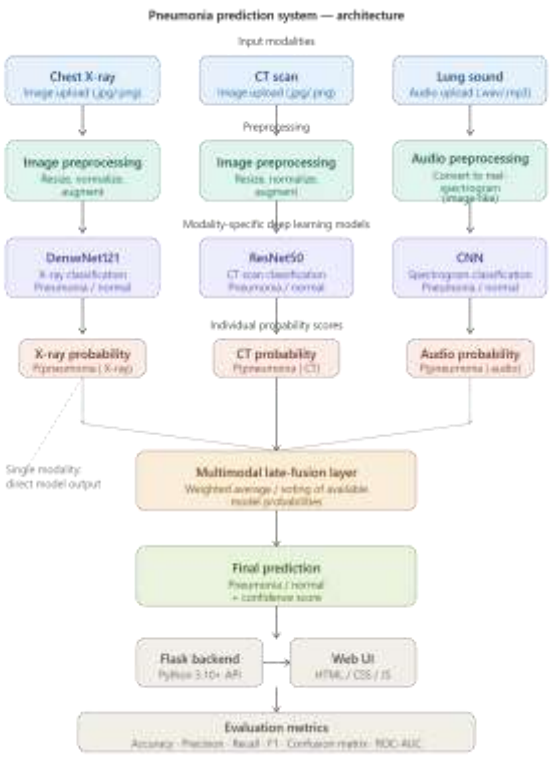

Fig. 1. Prediction system with input modalities and preprocessing steps.

2) ResNet50 for CT Scan: A ResNet50 architecture pretrained on ImageNet V2 weights was adopted for CT scan classification. The final fully connected layer was replaced with a custom head consist-ing of Linear(2048→512), BatchNorm1d(512), ReLU, Dropout(0.5) and Linear(512→2). Training followed a three-phase progressive fine-tuning strategy: Phase 1
(Epochs 1–5) trained only the FC head with the entire backbone frozen at learning rate 1e-4; Phase 2 (Epochs 6–11) unfroze layer3 and layer4 of ResNet50 at a reduced learning rate of 2e-5; Phase 3 (Epochs 12–25) unfroze the entire network at a further reduced learning rate of 5e-6. Gradient clipping with maximum norm 1.0 was applied throughout training to prevent exploding gradients.

3) Custom CNN for Lung Sound:

A custom LungSoundCNN architecture was designed from scratch for Mel spectrogram classification. The model consists of four progressive ConvBlocks, each

containing two Conv2d layers with BatchNorm2d and ReLU activations followed by MaxPool2d, with channel dimensions growing from 3→32→64→128→256. An AdaptiveAvgPool2d layer reduces the spatial dimensions to 4×4, followed by a Linear(4096→512) → Batch-Norm1d → ReLU→Dropout(0.5) → Linear (512→128) → ReLU → Dropout (0.3) → Linear (128→2). Conv layers were initialized using Kaiming Normal initialization and Linear layers using Xavier Uniform initialization. The entire network was trained from scratch using Adam optimizer with learning rate 3e-4, weight decay 1e-4, gradient clipping and ReduceLROnPlateau scheduling with early stopping of patience 7.

D. Late Fusion Mechanism

The Late Fusion layer combines the softmax probability outputs of all three models into a single final diagnosis. The fusion layer consists of two networks operating on the concatenated 6-dimensional probability vector (3 models × 2 class probabilities). The Attention Network takes the 6-dimensional input and produces 3 attention weights through Linear(6→64)→ReLU→Dropout(0.2)→Linear(64→3)→Softmax. These weights, which always sum to 1.0, represent the dynamic trust assigned to each modality's prediction for a given input. The Consensus Network takes the same 6-dimensional input and produces the final prediction through Linear(6 → 64) → ReLU → BatchNorm1d→Dropout(0.3) → Linear (64→32) → ReLU→Dropout(0.2) → Linear(32→2).

The final output is passed through Softmax to produce

NORMAL and PNEUMONIA confidence percentages. All linear layers in both networks were initialized using Xavier Uniform initialization. When fewer than three modalities are provided, the system gracefully falls back to partial fusion or individual prediction without requiring retraining.

E. Web Application Deployment

The complete system was deployed as a Flask web application. All three models are loaded into memory at application startup. The application provides four prediction endpoints: /predict/ct, /predict/xray, /predict/lung and /predict/comprehensive. Files are validated for type and size (maximum 16 MB), preprocessed through their respective pipelines and deleted automatically after prediction. Prediction results including individual model outputs, attention weights and final fused diagnosis are returned as JSON responses to the frontend dashboard.

IV. EVALUATION AND RESULTS

A. Experimental Setup

All three models were trained on Google Colab using an NVIDIA Tesla T4 GPU with 16 GB VRAM. PyTorch 2.5.1 with CUDA support was used as the deep learning framework. The Adam optimizer was used across all three models with ReduceLROnPlateau learning rate scheduling. Early stopping was applied to prevent overfitting. Model performance was evaluated on held-out test sets using five standard classification metrics: Accuracy, Precision, Recall, F1-Score and ROC-AUC.

B. Individual Model Performance

Each of the three models was independently trained and evaluated on its respective held-out test set. Table I presents the evaluation results.

Table I
Individual Model Performance On Test Sets

Metric	DenseNet121 (X-Ray)	ResNet50 (CT Scan)	Custom CNN (Lung Sound)
Accuracy	0.9423	0.9924	0.8741
Precision	0.9447	0.9949	0.3333
Recall	0.9641	0.9898	0.3846
F1-Score	0.9543	0.9924	0.3571
Roc-Auc	0.9812	0.9998	0.8101

C. Multi-Modal Prediction Capability

Beyond individual model evaluation, the system supports flexible multi-modal prediction where users can submit any combination of the three modalities simultaneously. When multiple inputs are provided, the attention-based Late Fusion mechanism combines the predictions of all available models dynamically.

The fusion layer operates as follows:

- When one modality is provided - the corresponding individual model prediction is returned directly without fusion
- When all three modalities are provided - the full late fu-sion pipeline activates, combining CT, X-Ray and Lung Sound predictions through the attention and consensus networks for maximum diagnostic accuracy

D. Discussion

The individual model results in Table I demonstrate strong performance of the X-Ray and CT Scan models. The DenseNet121 model achieves an accuracy of 94.23% and ROC-AUC of 0.9812, confirming the effectiveness of dense feature reuse for capturing subtle lung opacity patterns in chest X-ray images. The ResNet50 model achieves the highest performance with an accuracy of 99.24% and near-perfect ROC-AUC of 0.9998, demonstrating the strong suitability of ResNet50 with three-phase progressive fine-tuning for CT scan pneumonia classification. The Custom CNN for Lung Sound achieves an overall accuracy of 87.41% and ROC-AUC of 0.8101, indicating reasonable discriminative capability at the audio signal level. However, the lower Precision (0.3333) and F1-Score (0.3571) for the pneumonia class reflect the inherent challenge of the ICBHI dataset, which contains a highly imbalanced distribution of pneumonia recordings relative to other respiratory conditions. The small number of confirmed pneumonia recordings in the dataset limits the model's ability to learn sufficiently discriminative pneumonia-specific acoustic features despite the application of WeightedRandomSampler and Weighted CrossEntropy Loss. These results are consistent with findings reported by Demir et al. [15] on the same dataset, where class imbalance was identified as a primary limiting factor.

Future work involving larger and more balanced lung sound datasets is expected to substantially improve classifica-tion performance for this modality. The attention-based Late Fusion capability further extends the system by enabling multi-modal consensus diagnosis when more than one input is avail-able. The per-modality attention weights returned alongside each multi-modal prediction provide clinical explainability by indicating which modality contributed most to the final diagnosis decision.

V. CONCLUSION AND FUTURE SCOPE

A. Conclusion

This paper presented the design, implementation and evalua-tion of a flexible multimodal deep learning system for pneumo-nia detection that integrates three complementary diagnostic modalities — Chest X-Ray images, Chest CT Scan images and Lung Sound audio recordings. Three independent deep learning models were developed with modality-specific pre-processing pipelines, data balancing strategies and optimized training schedules: a DenseNet121 architecture for Chest X-Ray classification with a two-phase progressive fine-tuning strategy, a ResNet50 architecture for CT Scan classification with a three-phase progressive fine-tuning strategy and a custom LungSoundCNN architecture trained from scratch that converts raw audio recordings into 128-band Mel spectrogram images prior to classification. A key strength of the proposed system is its flexible prediction architecture.

Each model is fully capable of producing accurate standalone predictions when used independently, making the system immediately applicable in clinical scenarios where only one diagnostic modality is available. When all three modalities are provided simultaneously, the attention-based Late Fusion mechanism activates and dynamically assigns learned trust weights to each available modality, producing a final consensus diagnosis that draws upon complementary information from multiple diagnostic sources. This dual capability — strong individual model performance combined with advanced multi-modal fu-sion — represents a meaningful advancement in accessible and flexible clinical decision support for pneumonia detection. A key strength of the proposed system is its flexible pre-diction architecture. When a single modality is provided, the corresponding individual model produces a direct prediction with confidence score. When multiple modalities are provided simultaneously, an attention-based Late Fusion

mechanism dynamically assigns learned trust weights to each available modality through an attention network and produces a final consensus diagnosis through a dedicated consensus network. This design ensures the system remains fully functional re-gardless of which diagnostic inputs are available in a given clinical scenario. The ResNet50 CT Scan model achieved near-perfect performance with 99.24% accuracy and ROC-AUC of 0.9998. The DenseNet121 X-Ray model achieved 94.23% accuracy and ROC-AUC of 0.9812. The Custom CNN Lung Sound model demonstrated reasonable discriminative capability with 87.41% accuracy and ROC-AUC of 0.8101, with class-level performance reflecting the inherent challenges of the ICBHI dataset. The complete system is deployed as a Flask-based web application requiring no specialized hardware, making it practically accessible across a wide range of healthcare settings.

B. Future Scope

Several important directions remain for future development. The most critical next step is clinical validation with real patient data collected from hospital settings. While the current system demonstrates strong performance on publicly available benchmark datasets, evaluation with diverse real-world clinical data is necessary before meaningful claims about therapeutic or diagnostic value can be made. Improving the lung sound classification module represents an important technical prior-ity. Training on larger and more balanced lung sound datasets with confirmed pneumonia recordings is expected to sub-stantially improve class-level performance for this modality. Transformer-based architectures such as Audio Spectrogram Transformers (AST) for lung sound classification and Vision Transformers (ViT) for image modalities represent promising directions that may capture longer-range dependencies beyond the capability of current CNN-based models. Feature-level fusion strategies that combine intermediate layer representa-tions across modalities — rather than final output probabilities — could be explored to capture cross-modal interactions at a deeper representational level. Expanding disease coverage beyond binary pneumonia detection to include conditions such as COPD, asthma and bronchitis would substantially increase clinical applicability. Explainability enhancements such as Grad-CAM visualization for CT and X-Ray models integrated into the web interface would further improve clinical trust in system outputs. Finally, integration with hospital Electronic Health Record systems and compatibility with portable digital stethoscope devices would facilitate real-world point-of-care deployment.

REFERENCES

1. World Health Organization, "Pneumonia," WHO Fact Sheet, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/pneumonia
2. P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, Bagul, C. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," arXiv preprint arXiv:1711.05225, 2017.
3. S. Raoof, D. Feigin, A. Sung, S. Raoof, L. Irugulpati, and E. C. Rosenow III, "Interpretation of plain chest roentgenogram," Chest, vol. 141, no. 2, pp. 545–558, 2012.
4. A. R. A. Sovijarvi, "Characteristics of breath sounds and adventitious respiratory sounds," Eur. Respir. Rev., vol. 10, pp. 591–596, 2000.
5. G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sa´nchez, "A survey on deep learning in medical image analysis," Med. Image Anal., vol. 42, pp. 60–88, 2017.
6. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE CVPR, pp. 770–778, 2016.
7. L. Pham, I. McLoughlin, H. Phan, M. Tran, T. Nguyen, and R. Palaniappan, "Robust deep learning framework for predicting respiratory anomalies and diseases," in Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), pp. 164–167, 2020.
8. D. Ramachandram and G. W. Taylor, "Deep multimodal learning: A survey on recent advances and trends," IEEE Signal Process. Mag., vol. 34, no. 6, pp. 96–108, 2017.

9. T. Baltrusaitis, C. Ahuja, and L. P. Morency, "Multimodal machine learning: A survey and taxonomy," IEEE Trans. Pattern Anal. Mach. Intell., vol. 41, no. 2, pp. 423–443, 2018.
10. X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks," in Proc. IEEE CVPR, pp. 2097–2106, 2017.
11. J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya, et al., "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in Proc. AAAI, vol. 33, pp. 590–597, 2019.
12. S. Wang, Y. Zha, W. Li, Q. Wu, X. Li, M. Niu, M. Wang, X. Qiu, H.
Li, H. Yu, and W. Gong, et al., "A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis," Eur. Respir. J., vol. 56, no. 2, 2020.
13. D. Ardila, A. P. Kiraly, S. Bharadwaj, B. Choi, J. J. Reicher, L. Peng,
Tse, M. Etemadi, W. Ye, G. Corrado, and D. P. Naidich, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," Nat. Med., vol. 25, no. 6, pp. 954–961, 2019.
14. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in Proc. IEEE CVPR, pp. 4700–4708, 2017.
15. F. Demir, A. Sengur, and V. Bajaj, "Convolutional neural networks based efficient approach for classification of lung diseases," Health Inf. Sci. Syst., vol. 8, no. 1, p. 4, 2019.
16. A. Kocheturov, P. M. Pardalos, and A. Karakitsiou, "Massive datasets and machine learning for computational biomedicine: trends and chal-lenges," Ann. Oper. Res., vol. 276, no. 1, pp. 5–34, 2019.
17. J. Yao, X. Zhu, F. Zhu, and J. Huang, "Deep correlational learning for survival prediction from multi-modality data," in Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Interv. (MICCAI), pp. 406–414, 2017.
18. P. Mooney, "Chest X-Ray images (pneumonia)," Kaggle Dataset, 2018. [Online]. Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia
19. B. M. Rocha, D. Filos, L. Mendes, I. Vogiatzis, E. Perantoni, E. Kaimakamis, P. Natsiavas, A. Oliveira, C. Ja´come, A. Marques, and R. P. Paiva, "A respiratory sound database for the development of automated classification," in Proc. Int. Conf. Biomed. Health Inform. (ICBHI), Springer, pp. 33–37, 2017.
20. H. Kumar, "Pneumonia CT scan image dataset," Kaggle Dataset, 2021. [Online]. Available: https://www.kaggle.com/datasets/harshkumarcn005/pneumonia-ct-scan-image-dataset