GKU Journal of Multidisciplinary Research (GKUJMR)
ISSN: 3049 – 298X, Volume – II, Issue – I, April – 2026

A Comprehensive Review of Lung Disease Detection Using Machine Learning and Deep Learning Techniques

[1] Mr. Inzamam Kausar, [2] Mr. Gurpreet Singh, [3] Mr. Rohit Kumar, [4] Dr. Rajinder Kumar
[1,2,3] Department of Computer Science & Engineering, Guru Kashi University, Talwandi Sabo, Bathinda.
[4] Associate Professor, Department of Computer Application, Guru Kashi University, Talwandi Sabo, Bathinda.
[4] ORCID ID: 0009-0001-4129-0388, E-mail: [4] drrajinder1983@gmail.com

Accepted: 10.04.2026 Published: 30.04.2026 Page No. 111 – 120 DOI: 10.5281/zenodo.20020870

Abstract

Respiratory diseases, such as pneumonia, tuberculosis (TB), chronic obstructive pulmonary disease (COPD) and COVID-19, are among the world's major causes of illness and death. Timely and precise diagnosis plays a pivotal role in patient outcomes; nevertheless, conventional diagnostic processes are limited by subjectivity, resource availability, and the requirement for skilled professionals. The emergence of integrating machine learning (ML) and deep learning (DL) in the diagnosis of lung diseases has introduced a new paradigm that may improve diagnostic accuracy, shorten the diagnostic process, and broaden access to health care, especially to those in remote settings. This review paper provides an overview of the research using ML and DL techniques for the diagnosis of lung diseases, focusing on chest X-rays and computed tomography (CT) scans. This review examines popular datasets, feature extraction techniques, classification models, convolutional neural network (CNN) models, and transfer learning methods. It also explores image preprocessing techniques, data augmentation methods, and performance metrics. The research highlights the challenges in this field - lack of labeled data, class-imbalance, and interpretability of the models - and suggests future research opportunities. The work is primarily an analysis of the literature, and it does not include any original experiments or clinical studies.

Keywords: lung disease diagnosis, machine learning, deep learning, convolutional neural network, transfer learning, chest X-ray, CT scan.

1. Introduction

1.1 Background of Lung Diseases

Lung diseases are a major global health issue in the 21st century. The World Health Organization (WHO) lists lower respiratory tract infections as one of the world's top four causes of death, resulting in more than four million deaths per year (WHO, 2022). Tuberculosis alone accounted for the deaths of around 1.6 million people in 2021, and COVID-19 has led to a global health burden not previously experienced in modern history since its onset in late 2019 (Dong et al., 2020). COPD, a progressive lung disease marked by airflow limitation, affects over 300 million individuals worldwide and often goes unnoticed until later in the disease course (Global Initiative for Chronic Obstructive Lung Disease [GOLD], 2022). These figures highlight the need for effective, scalable, and early detection techniques.

1.2 Need for Early Detection

Timely diagnosis of lung-related diseases enhances patient outcomes and decreases the financial burden on health care. Research shows that the five-year survival rate for early-stage lung cancer, detected before metastasis, is greater than 55%, whereas late-stage lung cancer survival rates are below 5% (Siegel et al., 2021). Likewise, early diagnosis of pneumonia and TB allows for early antibiotic treatment, thereby decreasing disease spread and death rates. But the lack of radiologists - especially in developing and middle-income countries - and the subjectivity of visual interpretation present significant challenges to early diagnosis (Patel et al., 2019). The need for computer systems that can help clinicians in making a diagnosis is therefore important to public health.

1.3 Machine Learning and Deep Learning

Artificial intelligence (AI), including traditional machine learning and the more recent deep learning technique, has shown extraordinary promise in medical imaging applications. Machine learning techniques such as Support Vector Machines (SVM), k-Nearest Neighbors (kNN), and Random Forests can learn to classify radiomic features with moderate accuracy on relatively small datasets. Deep learning models, particularly CNNs, can automatically identify hierarchical features from pixel data, removing the need for feature engineering and producing state-of-the-art results on large-scale medical imaging datasets (LeCun et al., 2015; Rajpurkar et al., 2017). Classical models such as VGG16, ResNet50, InceptionV3, and DenseNet121 have been effectively fine-tuned for lung disease classification with transfer learning methods, achieving diagnostic performance that in some cases equals or exceeds board-certified radiologists.

1.4 Review Purpose

This article is a systematic review of peer-reviewed articles indexed in databases such as IEEE Xplore, PubMed, Scopus, and Google Scholar published from 2015 to 2024. The authors have not undertaken any experiments, datasets, or clinical studies. The paper aims to provide an overview of the current state of the art, assess the methodological strategies, identify the gaps and future scope of applying ML and DL techniques for lung disease detection. The rest of the paper is organised as follows: Section 2 offers a clinical overview of major lung diseases and traditional diagnostic techniques; Section 3 reviews of the benchmark datasets and the challenges faced; Sections 4 and 5 describe classical ML and deep DL approaches respectively; Section 6 covers image preprocessing; Section 7 compares model performance; and Section 8 concludes with a discussion of limitations and future work.

2. Background of Lung Diseases and Diagnosis

2.1 Classification and Characteristics of Major Lung Diseases

Acute inflammation of the lung parenchyma, most commonly due to bacterial, viral or fungi infection. The most frequent pathogens are Streptococcus pneumoniae and Influenza A. Lobar consolidation, interstitial infiltrates, and ground-glass opacities (GGOs) are typical CXR findings. The variability of its radiological features makes its classification challenging (Wang et al., 2017).

Tuberculosis is one of the world's oldest and most persistent infectious diseases, with a primary focus on the lungs. CXR findings in TB include upper lobe consolidation, cavitation, and hilar lymphadenopathy in the presence of Mycobacterium tuberculosis infection. The radiological appearance of TB lesions is like other lung diseases and requires laboratory confirmation for diagnosis; however, automated diagnostic tools have been proven effective in resource-poor environments (Jaeger et al., 2014).

COVID-19 pneumonia, due to SARS-CoV-2 virus, can be seen with bilateral peripheral GGOs, crazy-paving pattern, and consolidation on CT. The global pandemic of COVID-19 prompted the need for automated tools to diagnose SARS-CoV-2 pneumonia and differentiate it from other viral and bacterial infections (Ai et al., 2020).

COPD is an inflammatory lung disease with irreversible airflow limitation due to emphysema and chronic bronchitis. CT findings include hyperinflation, bullae and thickening of the airway walls. Lung cancer, including adenocarcinoma, squamous cell carcinoma and small cell types, requires early detection and differentiation of lung nodules to facilitate surgical resection.

2.2 Diagnostic Imaging Modalities

Chest X-ray (CXR) is the most widely used imaging technique in clinical practice for its affordability, ubiquity, and speed. But CXR has poor three-dimensional resolution and a high degree of superimposition, making it difficult to identify subtle abnormalities. CT offers better anatomical resolution and cross-sectional images, allowing detection of small nodules, interstitial and vascular abnormalities, but with increased radiation and scan time. Positron emission tomography (PET)-CT and magnetic resonance imaging (MRI) are used for specific clinical purposes, such as mediastinal staging and pleural disease, respectively (Litjens et al., 2017).

The clinical use of these imaging techniques is time-consuming, requires expert interpretation and is prone to variability. These challenges have sparked interest in the development of AI-based diagnostic systems that offer reliable, objective, and efficient image analysis.

3. Datasets and Challenges

3.1 Benchmark Datasets

The development of large-scale, annotated medical image datasets has played a key role in driving improvements in AI-based medical diagnosis of lung disease. The National Institutes of Health (NIH) ChestX-ray14 dataset (Wang et al., 2017) comprises 112,120 anteroposterior CXR images with 14 different thoracic diseases, and is the most popular benchmark dataset. The Shenzhen and Montgomery County TB datasets (Jaeger et al., 2014) are CXR images annotated for tuberculosis screening. The Kaggle RSNA Pneumonia Detection dataset, sourced from the Society for Thoracic Radiology, includes more than 30,000 CXR images, and has been widely used for pneumonia classification and localization. The COVID-19 Radiography Database (Chowdhury et al., 2020) is a multi-class CXR dataset that includes COVID-19, viral pneumonia, and normal images, while the LIDC-IDRI dataset offers CT images for the detection of pulmonary nodules.

Table 1: Summary of Benchmark Lung Disease Datasets

Dataset	**Disease Focus**	**# Images**	**Format**	**Access**
NIH ChestX-ray14	14 Pathologies	112,120	PNG	Public
Shenzhen TB Set	Tuberculosis	662	PNG	Public
Montgomery TB Set	Tuberculosis	138	PNG	Public
Kaggle RSNA	Pneumonia	30,000+	DICOM	Public
COVID-19 Radiography	COVID-19	21,165	PNG	Public
LIDC-IDRI	Lung Nodules (Cancer)	1,018 CT scans (~100k+ slices)	DICOM	Public (with agreement)
JSRT Dataset	Lung Nodules	247	IMG (grayscale)	Research / Restricted

Table 1. Comparative overview of publicly available benchmark datasets utilized in lung disease detection research.

3.2 Dataset Challenges

Despite the availability of these databases, there are common challenges in research. Class imbalance is a common issue in medical imaging, where the number of pathological cases is significantly lower than that of non-pathological cases, making models trained on such distributions prone to overestimate their accuracy (Johnson & Khoshgoftaar, 2019). Limited sample sizes are an issue in CT studies due to the expense of image acquisition and privacy concerns. Moreover, labeling inconsistencies due to inter-radiologist variability result in label noise, which may hinder model training and testing. Domain shift, where models trained with data obtained from one CT scanner or institution do not generalise to other acquisition protocols, is a major limiting factor to translation (Raghu et al., 2019). Finally, CT data is three-dimensional, which introduces computational challenges not present in CXRs.

4. Machine Learning Techniques

4.1 Feature Extraction Strategies

Traditional machine learning approaches for medical image classification involve the extraction of hand-designed feature representations before classification. These can be broadly construed as textural, morphological, and statistical. Textural features, such as those obtained from the Gray-Level Co-occurrence Matrix (GLCM), Local Binary Patterns (LBP) and Gabor filters, describe spatial intensity distributions of abnormal tissue. Morphological features capture shape information of the segmented regions (lesions or lung regions), whereas statistical features capture global or local statistics of the pixel intensity distribution of the image. Other features that have been used for lung image analysis include the Histogram of Oriented Gradients (HOG) and the Scale-Invariant Feature Transform (SIFT), which can be used to describe nodule shapes (El-Baz et al., 2013). Dimensionality reduction techniques, such as Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), are often used before classification.

4.2 Support Vector Machines

Support Vector Machines (SVMs) are a family of supervised classifiers that find the hyperplane that best separates two or more classes in a high-dimensional feature space, while providing the greatest margin between the classes. SVMs are especially well-adapted to binary classification problems in high-dimensional low-sample sizes. The kernel trick allows SVMs to fit a nonlinear hyperplane using radial basis function (RBF), polynomial or sigmoid kernels. Several recent studies have reported high classification accuracy for TB screening with GLCM (84-92%) and pneumonia detection with HOG descriptors using SVMs. But SVM classifiers are not suitable for multi-class problems, or when the features have different distributions (Cortes & Vapnik, 1995).

4.3 k-Nearest Neighbors and Decision Trees

The k-Nearest Neighbors (kNN) classifies a new instance by computing the majority label of its k nearest neighbors in the training data using a distance function like Euclidean or Minkowski distance. Although intuitive and straightforward, kNN is inefficient at inference, dependant on the choice of k, and prone to the curse of dimensionality in high-dimensional spaces. Decision trees recursively split the feature space based on information gain or Gini impurity, and produce easily interpretable rules for classification. Ensemble variants such as Random Forests and Gradient Boosting Machines significantly improve the model's performance by reducing variance (bagging) and sequentially correcting the errors (boosting) respectively, and have demonstrated high accuracy in radiological classification problems (Breiman, 2001).

5. Deep Learning Techniques

5.1 Convolutional Neural Network Architecture

Convolutional Neural Networks (CNNs) are a type of feed-forward artificial neural network selectively designed for spatial processing. Convolution, activation, pooling and fully connected layers are the building blocks of a CNN. Convolutional layers employ learned sets of filters to the input feature map, extracting local spatial features via weight sharing and translational equivariance. The Rectified Linear Unit (ReLU) activation function adds nonlinearity without suffering from the vanishing gradient issues of sigmoid and tanh units. Max-pooling layers successively downsample spatial dimensions while retaining feature responses. Fully connected layers combine global feature responses and generate class probability scores using a Softmax function (LeCun et al., 2015).

Figure 1: Convolutional Neural Network Architecture for Lung Disease Classification

Input 224×224×3	**Conv +relu** Feature maps	**Pooling** Downsampling	**Conv +relu** Deep features	**Flatten** Feature vector	**Fc+softmax** Class output

Figure 1. General CNN architecture pipeline from raw image input to disease classification output. Arrows indicate forward propagation through successive processing stages.

5.2 Landmark Deep Learning Models

The VGGNet of Simonyan and Zisserman (2014) is a deep network with 16-19 weight layers using 3×3 convolutional filters that has been widely fine-tuned for pneumonia and TB classification and has reported 93-96% accuracy on benchmark tasks. He et al. (2016) proposed ResNet50, which uses residual skip connections to overcome the problem of degradation in very deep networks, allowing training of networks with 50-152 layers and superior performance in multi-class thoracic disease classification. InceptionV3 (Szegedy et al., 2016) uses factorized convolutions and inception modules to efficiently build feature maps in parallel without compromising representational power, and has shown strong performance when applied to TB classification. DenseNet121 (Huang et al., 2017) uses dense connectivity, where each layer receives feature maps from all preceding layers and has achieved state-of-the-art performance on the NIH ChestX-ray14 benchmark with an area under the receiver operating characteristic curve (AUC) of more than 0.80 for 14 thoracic diseases.

5.3 Transfer Learning

Transfer learning involves transfer of representational knowledge learned by a deep neural network through pretraining on a large-scale dataset (typically ImageNet (Deng et al., 2009), a collection of 1.2 million natural images) and fine-tuning the network for a particular medical imaging task. Such an approach is especially valuable in medical imaging, where the available labeled training data is limited in comparison to the number of parameters of deep neural networks. The main transfer learning approaches are featuring extraction, where the pretrained convolutional layers are frozen and only the classification layer is retrained; and fine-tuning, where all or a subset of layers are retrained with a lower learning rate. Research has shown that transfer learning consistently yields significant improvements in accuracy over training from scratch on small medical imaging datasets, with 5-15% improvement in accuracy being reported (Tajbakhsh et al., 2016).

5.4 Hybrid and Ensemble Models

Hybrid models that combine CNN feature representation with recurrent neural networks (RNNs) or attention-based mechanisms have also been applied to capture temporal information in longitudinal imaging studies, and to enforce spatial attention on salient regions of interest. The use of squeeze-and-excitation and transformer-based models (including Vision Transformers or ViTs) has also proved beneficial in modelling global spatial dependencies in high-resolution CT scans. Ensemble learning, which combines predictions from multiple individually-trained models, further reduces variance and enhances model calibration, and has been used in several successful lung disease detection challenges to deliver state-of-the-art performance (Rajpurkar et al., 2017).

Table 2: Comparative Performance of Deep Learning Models for Lung Disease Detection

Model	Architecture	Accuracy (%)	Parameters	Dataset	Disease
VGG16	16 Layers CNN	94.5	138M	NIH CXR	Pneumonia
ResNet 50	Residual Net	96.2	25M	COVID-19 DB	COVID-19
InceptionV3	Inception Mod.	95.8	23M	Shenzhen	TB
DenseNet121	Dense Connect.	97.1	8M	NIH CXR14	Multi-class
MobileNetV2	Lightweight	93.7	3.4M	Kaggle	Pneumonia
EfficientNetB4	Compound Scal.	97.5	19M	RSNA	Nodule Det.

Table 2. Summary of reported accuracy values for major deep learning architectures across diverse lung disease datasets. Results are sourced from literature and are approximate.

6. Image Processing and Preprocessing

6.1 Image Enhancement

Medical images are often degraded by noise, poor contrast, and other scanner-specific artifacts, which require systematic pre-processing before feeding them to the model. Contrast-Limited Adaptive Histogram Equalization (CLAHE) is commonly used for local contrast enhancement of CXR images to boost detection of subtle lung infiltrates and nodular opacities while avoiding global image saturation (Pisano et al., 1998). Gaussian and median filtering are used for noise reduction, while histogram normalization ensures consistent image intensity distributions across different scanners and acquisition techniques to reduce domain shift. Bone suppression (removing bony structures from CXR images) has been shown to enhance the detection of pulmonary infiltrates obstructed by rib and clavicular bones (Suzuki et al., 2006).

Figure 2: Image Preprocessing Pipeline for Lung Disease Detection Systems

Raw image Dicom/png	**Resize & normalize** 224×224 px	**Enhance** Clahe/he	**Segment** Lung region	**Augment** Flip/rotate	**Model input** Ready data

Figure 2. Schematic representation of the sequential image preprocessing pipeline, from raw DICOM/PNG acquisition through normalization, enhancement, segmentation, and augmentation.

6.2 Lung Segmentation

Lung field segmentation highlights the areas of interest in the lungs by removing extraneous thoracic structures, such as the cardiac silhouette, mediastinum, and diaphragm, thus enhancing the signal-to-noise ratio of subsequent feature extraction or deep learning processing. Traditional methods include thresholding, region growing and graph-cut, while deep learning models - such as U-Net and its extensions - have set new standards for automated lung segmentation (Ronneberger et al., 2015). U-Net's contracting and expanding architecture with skip connections allows for accurate detection of boundaries even with limited training data, making it well-adapted to medical applications. Volumetric segmentation of CT scans with 3D-UNet allows for volumetric measurements of lung nodules and extraction of quantitative radiomic features.

6.3 Data Augmentation

Data augmentation is a crucial strategy for enlarging the apparent size of small training sets and avoiding overfitting in deep learning models.

Common geometric augmentations of medical images involve horizontal and vertical flipping, random rotation (usually ±15 degrees), scaling, translation, and elastic deformation. Random variations in image intensity (brightness and contrast jitter, Gaussian noise) also contribute to intra-class variations. A more advanced approach to augmentation is through generative adversarial networks (GANs), where synthetic pathological images are generated to complement real image data; several studies have reported accuracy gains of 2-5% with GAN augmentation for rare disease classes (Frid-Adar et al., 2018). Mixup and CutMix policies, originally designed for natural image classification, have also been applied to lung image classification with promising results.

Figure 3: End-to-End System Flow for Automated Lung Disease Detection

1. Patient CXR / CT Scan Acquisition
▼
2. DICOM/Image File Storage & Retrieval
▼
3. Image Preprocessing (CLAHE, Normalization, Denoising)
▼
4. Lung Segmentation (U-Net / Classical Methods)
▼
5. Data Augmentation (Geometric / GAN-based)
▼
6. Feature Extraction / CNN Forward Pass
▼
7. Classification (ML Classifier / Softmax Layer)
▼
8. Disease Prediction & Confidence Score Output
▼
9. Radiologist Review & Clinical Decision Support

Figure 3. End-to-end system architecture for automated lung disease detection, from image acquisition through preprocessing, feature learning, classification, and clinical decision support.

7. Performance Evaluation and Comparison

7.1 Evaluation Metrics

Performance analysis of lung disease classifiers requires the consideration of several performance metrics, as no single statistic is suitable to fully describe the classifier's performance in all aspects relevant to clinical practice. Accuracy - the fraction of correctly predicted cases among all cases - is an intuitive measure but can be misleading in the presence of class imbalance, where a simple majority-class predictor might yield very high accuracy. Precision (positive predictive value) represents the fraction of positive predictions that are true positives (i.e., true pathological cases), while Recall (sensitivity) represents the fraction of true positive instances that are detected. The F1-score, defined as the harmonic mean of precision and recall, is a balanced metric that is appropriate for imbalanced data. The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) measures the discriminative power of a classifier over all possible decision thresholds, and is considered the preferred metric for binary medical classification problems. Other metrics such as Specificity, Negative Predictive Value (NPV) and the Dice Similarity Coefficient (for segmentation problems) are also used in the literature we review.

7.2 Comparative Analysis of Approaches

Table 3: ML vs. Deep Learning Performance Comparison

Criterion	SVM	KNN	Decision Tree	CNN (DL)
Accuracy	Moderate–High	Moderate	Moderate	Very High
Data Requirement	Low–Medium	Low	Low	High
Feature Engineering	Required	Required	Required	Automatic
Interpretability	Moderate	High	High	Low
Training Time	Fast	Very Fast	Fast	Slow
Scalability	Limited	Limited	Moderate	High

Best Use Case	Small datasets	Simple tasks	Structured data	Image data

Table 3. Comparative analysis of classical machine learning algorithms and deep learning frameworks across key performance dimensions for lung disease detection.

Table 4: Reported Performance Ranges by Methodological Approach

Approach	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
SVM (Hand-crafted)	82–88	80–86	79–85	79–86
KNN	75–83	73–82	72–81	73–82
Random Forest	84–90	83–89	82–88	82–89
Basic CNN	88–93	87–92	86–91	87–92
Transfer Learning	93–97	92–96	91–96	92–96
Hybrid DL Models	95–98	94–97	94–97	94–97

Table 4. Summary of accuracy, precision, recall, and F1-score ranges reported across the reviewed literature, categorized by algorithmic approach.

The benchmarking results in Tables 3 and 4 demonstrate a clear hierarchy in performance between deep learning methods, especially transfer learning and hybrid approaches, and classical ML classifiers across all benchmarking metrics, when applied to large imaging datasets. Transfer-learned CNNs typically attain classification accuracies in the 93-97% range, whereas SVM and Random Forest classifiers typically achieve 82-90% accuracy. But this performance gain comes with the assumption of adequate training data; in low-data regimes (less than 500 labeled images), traditional ML approaches with strong feature engineering can match or outperform undertrained CNNs. Rajpurkar et al. (2017) showed that a DenseNet-based model outperformed the average radiologist on the National Institutes of Health (NIH) CXR14 data set measured by the area under the curve (AUC) metric, suggesting deep learning has the potential to be a clinical decision support system, rather than just a benchmark system.

8. Challenges, Future Directions, And Conclusion

8.1 Persistent Challenges

Despite the significant advances highlighted in this review, several challenges remain that hinder the transition of AI-based lung disease detection systems from research to clinical settings. The lack of large, diverse, and accurately annotated medical imaging data sets - and regulatory limitations on cross-institutional data sharing - limits model generalisability and external validity. Class imbalances, particularly for rare lung diseases, are a common problem that requires the application of sampling techniques, cost-sensitive learning, or data augmentation approaches. The lack of interpretability of deep neural networks - the "black-box" issue - presents a major hurdle to clinician acceptance and regulatory approval; while gradient-weighted class activation mapping (Grad-CAM) and integrated gradients can provide post-hoc saliency maps, they don't explain the decision-making of these models (Selvaraju et al., 2017). Domain shift due to variability in imaging equipment, protocols, and demographics presents a challenge to model generalization and requires domain adaptation or federated learning strategies. In addition, ethical issues related to algorithmic bias, privacy, and liability in AI-assisted diagnoses need to be addressed (Obermeyer & Emanuel, 2016).

8.2 Future Research Directions

The review of literature suggests several valuable research opportunities. Federated learning (distributed model training across multiple sites in a privacy-preserving manner) presents a scalable approach to address data scarcity and domain shift while preserving privacy, and has been applied to lung nodule detection in initial studies (Roth et al., 2020). Unsupervised and semi-supervised learning, leveraging large quantities of unlabeled data to learn generic representations, is an attractive alternative to supervised learning approaches that may suffer from limited training data in clinical practice. The incorporation of multi-modal data - for instance, combining imaging features with clinical metadata, electronic health record data and genomic data - has the potential to significantly enhance diagnostic performance and prognostic value compared to imaging-based models alone. Vision Transformers, which employ global self-attention, can be used as a complementary approach to CNNs to learn long-range spatial relationships in high-resolution CT images. Lastly, the need for uniform assessment and validation of diagnostic systems, prospective clinical trials, and regulatory procedures is crucial for the safe and equitable use of AI-based diagnostic solutions in clinical practice.

8.3 Conclusion

This paper has provided a systematic review of machine learning and deep learning algorithms for detection and classification of major lung diseases, such as pneumonia, tuberculosis, COVID-19, COPD, and lung cancer. This is a literate review based solely on analysis of published peer-reviewed studies, and has distilled evidence across benchmark datasets, feature extraction methods, classification algorithms, CNN architectures, transfer learning techniques, and image preprocessing approaches. The results of this review confirm that deep learning (especially transfer-learned CNN models like ResNet50, DenseNet121 and EfficientNet) consistently outperforms classical machine learning algorithms in diagnostic accuracy across large-scale image data, while also confirming that classical ML approaches hold many advantages in small data and interpretability-constrained settings. Critical issues of data availability, class imbalance, domain shift, and model interpretability must be comprehensively resolved with innovative approaches to federated learning, self-supervised learning, and explainable AI for safe and equitable clinical translation of AI-powered pulmonary diagnosis. It is hoped that this review will provide a useful guide to researchers and clinicians tackling the interdisciplinary challenge of leveraging AI to benefit global pulmonary health.

References

[1] Ai, T., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., & Xia, L. (2020). Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China. Radiology, 296(2), E32–E40. https://doi.org/10.1148/radiol.2020200642

[2] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[3] Chowdhury, M. E. H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. B., & Islam, M. T. (2020). Can AI help in screening viral and COVID-19 pneumonia? IEEE Access, 8, 132665–132676.

[4] Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

[5] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). ImageNet: A large-scale hierarchical image database. Proceedings of the IEEE CVPR, 248–255.

[6] Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. The Lancet Infectious Diseases, 20(5), 533–534.

[7] El-Baz, A., Beache, G. M., Gimel'farb, G., Suzuki, K., Okada, K., Elnakib, A., & Abdollahi, B. (2013). Computer-aided diagnosis systems for lung cancer: Challenges and methodologies. International Journal of Biomedical Imaging, 2013, 1–46.

[8] Frid-Adar, M., Diamant, I., Klang, E., Amitai, M., Goldberger, J., & Greenspan, H. (2018). GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing, 321, 321–331.

[9] Global Initiative for Chronic Obstructive Lung Disease (GOLD). (2022). Global strategy for

the diagnosis, management, and prevention of COPD: 2022 Report. https://goldcopd.org

[10] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE CVPR, 770–778.

[11] Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. Proceedings of the IEEE CVPR, 4700–4708.

[12] Jaeger, S., Candemir, S., Antani, S., Wang, Y. X. J., Lu, P. X., & Thoma, G. (2014). Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. Quantitative Imaging in Medicine and Surgery, 4(6), 475–477.

[13] Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. Journal of Big Data, 6(1), 1–54.

[14] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.

[15] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., & van Ginneken, B. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.

[16] Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—Big data, machine learning, and clinical medicine. New England Journal of Medicine, 375(13), 1216–1219.

[17] Patel, A., Jernigan, D. B., & 2019-nCoV CDC Response Team. (2019). Initial public health response and interim clinical guidance for the 2019 novel coronavirus outbreak. MMWR Morbidity and Mortality Weekly Report, 69(5), 140–146.

[18] Pisano, E. D., Zong, S., Hemminger, B. M., DeLuca, M., Johnston, R. E., Muller, K., & Pizer, S. M. (1998). Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. Journal of Digital Imaging, 11(4), 193–200.

[19] Raghu, M., Zhang, C., Kleinberg, J., & Bengio, S. (2019). Transfusion: Understanding transfer learning for medical imaging. Advances in Neural Information Processing Systems, 32.

[20] Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., & Lungren, M. P. (2017). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv preprint arXiv:1711.05225.

[21] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. MICCAI, Lecture Notes in Computer Science, 9351, 234–241.

[22] Roth, H. R., Chang, K., Singh, P., Neumark, N., Li, W., Gupta, V., & Xu, D. (2020). Federated learning for breast density classification: A real-world implementation. Domain Adaptation and Representation Transfer, Distributed and Collaborative Learning, 181–191.

[23] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE ICCV, 618–626.

[24] Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2021). Cancer statistics, 2021. CA: A Cancer Journal for Clinicians, 71(1), 7–33.

[25] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[26] Suzuki, K., Abe, H., MacMahon, H., & Doi, K. (2006). Image-processing technique for suppressing ribs in chest radiographs by means of massive training artificial neural network (MTANN). IEEE Transactions on Medical Imaging, 25(4), 406–416.

[27] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. Proceedings of the IEEE CVPR, 2818–2826.

[28] Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B., & Liang, J. (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Transactions on Medical Imaging, 35(5), 1299–1312.

[29] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). ChestX-ray8: Hospital-scale chest X-ray database and benchmarks. Proceedings of the IEEE CVPR, 2097–2106.

[30] World Health Organization (WHO). (2022). Global tuberculosis report 2022. https://www.who.int/publications/i/item/9789240061729