iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** I **Month of publication:** January 2026

DOI: https://doi.org/10.22214/ijraset.2026.76714

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue I Jan 2026- Available at www.ijraset.com

A Comparative Review of Machine Learning and Deep Learning Models for Lung Disease Classification Across Medical Imaging Modalities

Mrs. Ashwini Bhavsar, Atharv Borhade, Aniket Gawali, Ashish Saini
Dept. of Computer Science Engineering, Pimpri Chinchwad College of Engineering & Research, Pune, India

Abstract: Purpose – The increasing integration of artificial intelligence into medical imaging has significantly advanced lung disease detection using chest X-rays and related modalities. This survey conducts a comprehensive comparative analysis of 12 researchworkspublishedbetween 2021 and 2025, with emphasis on five key studies employing machine learning, deep learning, and hybrid algorithms for image-based lung disease classification.

Design/methodology/approach–Thereviewexaminesmethodologiesinvolving segmentation-classification pipelines, attention-enhanced CNNs, hybrid CNN–ELM architectures, multi-branch deep networks, and fusion models combining handcrafted and deep features. Datasets used across these studies include NIH ChestX-ray14, JSRT, MC, LC25000, SIIM-ACR, LIDC-IDRI,andCOVID-19imagerepositories.Eachalgorithmisanalyzedinterms of preprocessing strategies, model design, evaluation metrics, computational efficiency, and applicability to different imaging types such as chest X-rays, CT scans, and histopathologyslides.

Findings – The analysis shows thatnosinglearchitectureperformsbestacrossalltasks;instead, each model demonstrates unique advantages depending on data modality and clinical objective. Attention-based CNNs excel in feature refinement, hybrid CNN-ELM models offer lightweight computation, and dual-branch frameworks improve multi-label chest disease detection..

Originality/value – By consolidating state-of-the-art image-based lung disease classification research, this survey highlights methodological strengths, limitations, and emerging trends. Although key research gaps are identified for future inclusion, the present review provides a structured foundation for developing more scalable and reliable diagnostic systems.

Keywords: Lung disease classification, Chest X-rays, Deep learning, CNN, Hybrid models, Medical imaging Paper type Review article

I. INTRODUCTION

Lung diseases such as COVID-19, pneumonia, pulmonary fibrosis, emphysema, edema, consolidation, pleural thickening, atelectasis, cardiomegaly, lung nodules, masses, infiltration, hernia, effusion, and lung cancer continue to impose a substantial health burden worldwide. Around the world, respiratory illnesses remain among the leading causes of morbidity and mortality, with chest radiography and computed tomography (CT) serving as the most widely utilized imaging modalities for early diagnosis and clinical assessment. However, the interpretation ofmedicalimagesremainschallenging.Theselimitationsoftenresultindiagnostic variability and reduced accuracy, particularly in high-volume clinical settings.

The emergence of machine learning (ML) and deep learning (DL) has significantlytransformed the field of medicalimageanalysisbyenablingautomated,data-driveninterpretationofcomplex imaging patterns. Techniquessuchasconvolutionalneuralnetworks(CNNs),attention-enhanced architectures,hybridlearningframeworks,segmentation–classificationpipelines, transformer-based models, and object-detection systems have demonstrated substantial improvements in extracting discriminative features and detecting subtle disease signatures from CXR, CT, histopathology, and multimodal scans. Theseadvancementshavecontributedtofaster diagnosis, improved sensitivity, and greater clinical reliability. Nevertheless, the performanceof these models remains heavily dependent on dataset size, imagequality,andarchitecturaldesign, creating the need for a structuredunderstandingofwhichalgorithmsperformbestunderspecific conditions.

Despite the rapid expansion of ML/DL research for lung disease detection, existing survey studies exhibit several notable limitations. Most prior reviews focus primarily on lung cancer, limiting their applicability across the broader spectrum of thoracic diseases encountered in clinical practice. Other reviews are constrained to a single imaging modality, typically either CXR or CT, overlooking the value of multi-modality insights.

Critically, the literature lacks a unifiedperformancecomparisonthatsystematicallyevaluatesdifferentarchitectural families—ranging from classical ML models to CNNs, attention networks, hybrid CNN-ELM frameworks, detection models, and transformer-based pipelines—across multiple lung diseases and imaging types. With a significant increase in research output between 2021 and 2025, an updated and consolidated review is required to contextualize these developments and provide clarity regarding the comparative strengths and limitations of state-of-the-art approaches.

Contributions of This Review

Toaddressthesegaps,thisreviewmakesthefollowingkeycontributions:

1) A comprehensive comparative analysis of 12 recent ML/DL studies (2021–2025) covering a broad spectrum of lung diseases across CXR, CT, histopathology, and multimodal imaging.
2) A unified summary of all models,detailingpreprocessingstrategies,architecturaldesign, segmentation components, classification methods, performance metrics, and identified limitations.
3) A structured performance comparison acrossdiversealgorithmfamilies,offeringinsights into how different architectures behave under varying imaging modalities and disease characteristics.

II. REVIEW METHODOLOGY

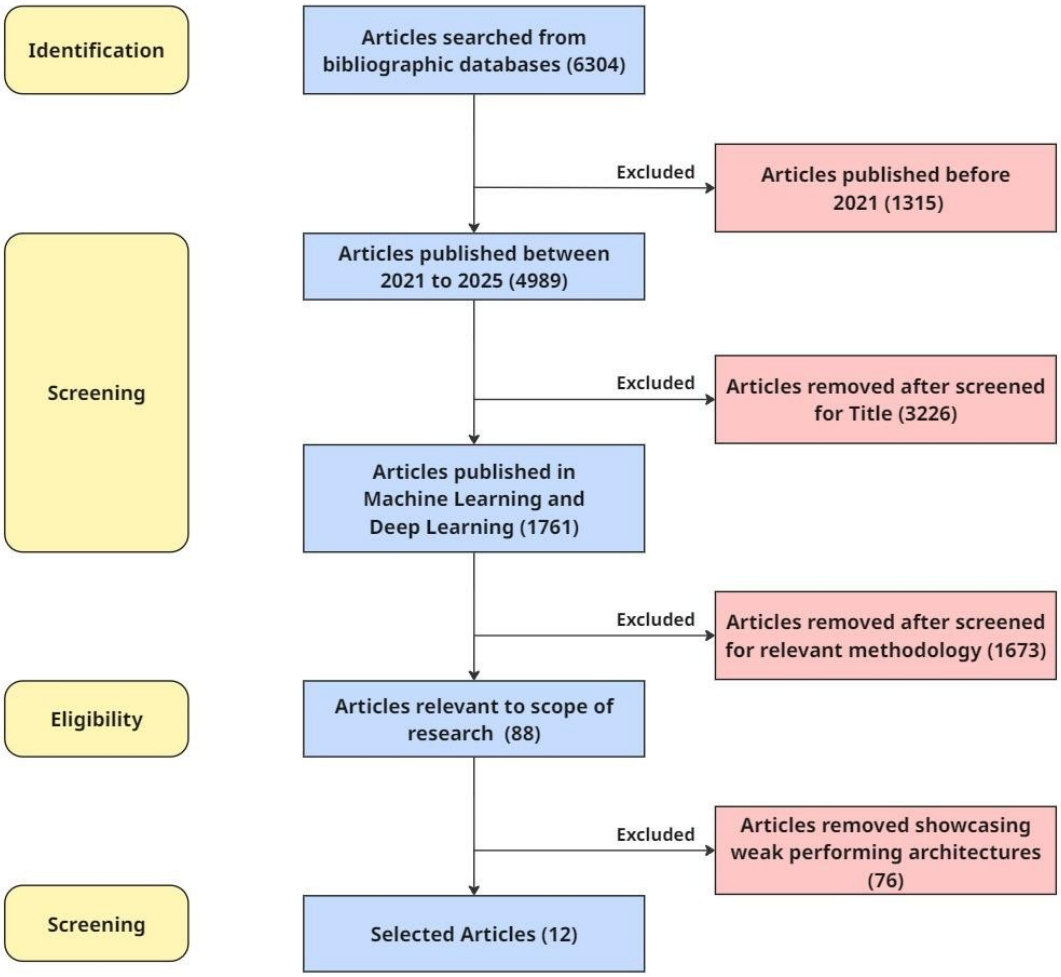


Figure 1: Prisma Flow Diagram of the Review

This review follows a structured methodology inspired by principles of systematic and scoping reviews to ensure transparent, reproducible, and unbiased selection of relevant studies. The methodology includes database selection,keyword-drivensearchstrategies,predefinedinclusion and exclusion criteria, a PRISMA-style screening workflow, and a standardized data extraction process applied to all selected papers.

A. Literature Search Databases

The literature search was conducted primarily across IEEE Xplore, which served as the most significant source of relevant studies due to its strong emphasis on computer science, medical image processing, and artificial intelligence research. To broaden the search and minimize publication bias,supplementaryquerieswereperformedusingGoogleScholarandSpringerLink, enabling access to interdisciplinary and medical imaging–focused research. These three sources collectively ensured coverage of a wide variety of journals, conference papers, and emerging research preprints relevant to medical image classification.

Table1:Summary of Literature Survey

Sr. No.	ArticleTitle-Author	Modality	Architecture	Advantages	Disadvantages
1	MI2A:AMultimodal Information Interaction Architecture for AutomatedDiagnosis of Lung Nodules Using PET/CT Imaging - Kai Li, Tongtong Li, Lei Zhang, Junfeng Mao, Xuerong Shi, Zhijun Yao,LeiFang,BinHu (2025)	PET/CT	MI2A, a multimodal informationinteraction architecture using U-Net–basedROIextraction, a PET–CT imaging encoder, cross-attention multimodal fusion, and alignment lossfor lung noduleclassification.	The cross-attention mechanismexplicitly models interactionsbetween anatomical and metabolic features, improving multimodal representation learning.	High metabolic activity in certain benign lesions introduces ambiguity,limiting classification performance despitemultimodal fusion.
2	Attention Enhanced InceptionNeXt - Based Hybrid Deep Learning Model for Lung Cancer Detection - BurhanettinOzdemir, Emrahaslan ,AndishakPacal(2025)	CT	An attention-enhanced hybrid deep learning model integrating InceptionNeXt blocks with grid and block attention mechanisms, combiningCNNsandVision Transformersforlungcancer detection and subtype classification.	The hybrid design effectivelycaptures both fine-grained and global contextual features with high accuracy whilemaintaininga lightweight parameter count.	The model is evaluated only on CTdataandlimited public datasets, which may constrain generalization across populations and imaging protocols.
3	Classification based deeplearningmodels for lung cancer and diseaseusingmedical images - Ahmad Chaddad, Jihao Peng, Yihang Wu (2025)	CT + Histopathology + CXR	A ResNet-based classification framework (ResNet+) enhanced with ResNet-D downsampling and convolutional attention modulesforlungcancerand disease classification across multiple imaging modalities.	The modified residual design preserves feature informationduring downsamplingand improvesfocuson diagnostically relevant regions across diverse medical images.	Themodeldepends on extensive data augmentation and multi-dataset training,whichmay introduce modality-specific bias and increase trainingcomplexity.

4	CNN-O-ELMNet: Optimized Lightweight and Generalized Model for Lung Disease Classification and Severity Assessment -SaurabhAgarwal, Yogesh Kumar Meena,K.V.Arya (2024)	CXR	CNN-O-ELMNet, a lightweight hybrid architecture combining CNN-based feature extractionwithanoptimized Extreme Learning Machine classifier tuned using the Imperialistic Competitive Algorithm for lung disease classification and severity assessment.	Themodelachieves high accuracy with extremely low parameter count, enabling fast inference and suitability for resource-constraine d clinical environments.	Theseparationof featureextraction andclassification may limit end-to-end feature optimization and reduceadaptability to highly complex disease patterns.
5	AMulti-ModelDeep Learning Framework and Algorithmsfor Survival Rate PredictionofLung Cancer Subtypes With Region of Interest Using Histopathology Imagery - MattakoyyaAharonu, Lokeshkumar Ramasamy (2024)	Histopathology	Amulti-modeldeeplearning framework consisting of LCSCNet for lung cancer subtype classification and LCSANet for survival rate prediction, enhanced with ROI-based processing and VGG16-based feature learning.	The framework jointly supports cancersubtyping and survival analysis while ROI-basedlearning improves focus on diagnostically relevant tissue regions.	Theapproachrelies on high-quality histopathology annotations and is computationally intensive, limiting scalabilityforlarge whole-slidedatasets.
6	Lung-RetinaNet: Lung Cancer DetectionUsinga RetinaNet With Multi-ScaleFeature Fusion and Context Module - Praveena K.,S.VenakataSatya Krishna, R. Dinesh Kumar, P. Keerthika (2024)	CT	Lung-RetinaNet, a single-stageRetinaNet-based detector augmented with multi-scale feature fusion andadilatedcontextmodule for early lung tumor localization and classification.	The architecture effectivelyenhances small nodule detection by integrating contextual and multi-scale features while maintaining high detection accuracy.	The detection-focused design increases architectural complexity and relies on precise anchorgeneration, which may affect robustness on heterogeneousCT datasets.

7	SynergisticAnalysis of Lung Cancer's Impact on Cardiovascular Disease using ML-Based Techniques - Gunasekaran Raja; Balakumar Ramkumar;Bhargavi Rajendiran; Sahaya Beni Prathiba; Thamodharan Arumugam; Kalimuthu Karuppanan; Lewis Nkenyereye; Kapal Dev (2024)	CT	LCDP system combining AdaDenseNetwithtransfer learning and Prox-SMOTE for lung cancer detection, followed by VGG16-based featureextractionandSVM classification for cardiovascular disease prediction.	The dual-module designenablesjoint analysis of lung cancer and cardiovascular risk while mitigating class imbalance in CT datasets.	The framework relies on multiple sequential models, increasing system complexity and limitingend-to-end optimization.
8	TransferLearning BasedMulti-Class Lung Disease Prediction Using Textural Features DerivedFromFusion Data - Kurupati Sai Charan , Omtri Vijaya Krishna, A. K. Ilavarasi , Palla Venkata Sai (2024)	CXR	A transfer-learning-based multi-class lung disease classification framework using fused CXR datasets and textural feature enhancement via Local Binary Patterns integrated with CNN backbones such asEfficientNetandResNet.	The use of texture-guided learningimproves discrimination between visually similar lung diseases in multi-class settings.	Performance is sensitivetodataset fusionstrategyand handcraftedtexture extraction parameters.
9	ADual-Branch Network for DiagnosisofThorax DiseasesfromChest X-Rays - G. Jignesh Chowdhary, Vivek Kanhangad (2022)	CXR	Adual-branchdeeplearning framework combining lung-region segmentation using R-I UNet with global and local feature extraction viatwopre-trainedAlexNet models followed by GRU-based feature fusionformulti-labelthoraxdisease classification.	The architecture captures both localized lung pathology and global contextual cues, enabling effective multi-label diagnosis and improved representation of diseases extending beyond lung boundaries.	The pipeline introduces additional computational overhead due to explicit lung segmentation and recurrent processing, increasingtraining and inference complexity.
10	A Weakly-Supervised Framework for COVID-19 Classification and LesionLocalization From Chest CT - Xinggang Wang, Xianbo Deng, Qing Fu, Jiapei Feng, Hui Ma, Wenyu Liu , Chuansheng Zheng (2022)	CT	A weakly supervised 3D deep learning framework (DeCoVNet)thatperforms lung segmentation using a pre-trained U-Net and applies a lightweight 3D CNN for COVID-19 classificationandlesion localization using patient-levellabelsonly.	Themodelreduces annotation burden by avoiding voxel-level lesion labeling while still enabling lesion localizationandfast patient-level inference.	Weaksupervision may limit localization precision,andthe framework is task-specific, restricting adaptability to broader thoracic diseasedetection.

11	STBi-YOLO: A Real-Time Object DetectionMethodfor Lung Nodule Recognition-Kehong Liu (2022)	CT	STBi-YOLO, a real-time single-stage lung nodule detector derived from YOLO-v5s, enhanced with stochastic-pooling–based spatial pyramid pooling, bidirectionalfeaturepyramid fusion, and EIoU loss optimization.	Thearchitecture improves small-nodule localizationwhile maintaining real-timeinference and reduced memory footprint.	The model is optimizedprimarily for detection tasks and does not directly address disease classification or clinical staging.
12	Detection and classificationoflung diseases for pneumonia and Covid-19 using machine and deep learningtechniques-Shimpy Goyal, Rajiv Singh (2021)	CXR	A fusion and normalization–based framework(F-RNN-LSTM) thatintegratesimagequality enhancement, adaptive ROI extraction, handcrafted lung-specific feature extraction (HOG, texture, intensity, geometric moments), followed by classification using soft computingmethodsandan RNN with LSTM for pneumoniaandCOVID-19 detection.	The framework reduces computational overhead while maintaininghigh accuracy by focusing on lung-specific featuresandenables disease severity analysis through ROI-based feature normalization.	The reliance on handcrafted feature engineering and multiple preprocessingstages increases pipelinecomplexity and limits full end-to-endlearning scalability.

B. Search Keywords

Thesearchstrategyemployedcombinationsofdisease-related,imaging-related,and technique-specific keywords to ensure comprehensive retrieval of ML/DL studies focused on lung-disease prediction from medical images.

Disease terms such as “lung disease,” “pulmonary disease,” “lung cancer,” “COVID-19,” “pneumonia,” “coronavirus,” and “tuberculosis (TB)” were paired with imaging descriptors including “chest X-ray,” “CT image,” “medical image,” “radiograph*,”and“CXR.”Thesewere further combined with computational terminology such as “machine learning,” “deep learning,” “neural network,” “CNN,” “convolutional neural network,” “U-Net,” “transfer learning,” “segmentation,” “classification,” and “ensemble.” Boolean operators (AND/OR) were used to refineandstructurethequery,ensuringthatonlystudiesapplying ML/DLmethodsto lung-image analysis were retrieved.

This expanded keyword selection was broad enough to capture general-purpose classification pipelines as well as highly specialized architectures designed for specific lung conditions and imaging modalities.

C. Inclusion Criteria

Studies were included if they met four essential conditions: they were published within the 2021–2025 timeframe, utilized image-based datasets such as CXR, CT, histopathology, or multimodal scans, implemented machine learning or deep learning models, and reported quantitative performance metrics that enabled objective comparison. These criteria ensured that the selected studies were methodologically robust and relevant to modern AI-driven clinical workflows.

D. Exclusion Criteria

To maintain thefocusoncomputationalimagingapproaches,studieswereexcludediftheyrelied solely on clinical or epidemiological data, lacked an ML/DL component, or used datasets too small to provide meaningfulgeneralization. Papersthatutilizedaudio signals (e.g.,lungsounds), textual records, or non-imaging modalities were also removed from consideration.Thisfiltering ensured that the final set of studies aligned strictly with the objectives of evaluating ML/DL algorithms for image-based lung disease prediction.

E. Data Extraction Process

For each selected study, a structureddataextractionprotocolwasfollowedtoensureconsistency and enable meaningful comparison between models. Information was collected on the imaging modality employed, the specific datasets used, and allpreprocessingorsegmentationtechniques applied. Particular attention was given to architectural details of theproposedmodels,including feature extraction strategies and classification mechanisms. Quantitative performance metrics such as accuracy, AUC, precision, and recall were recordedalongsidethereportedstrengthsand limitations of each approach. Finally, the clinical application of each model was noted to contextualize whether it addressed single-disease, multi-disease, or severity-based classification scenarios. This systematic extraction ensured that the comparative evaluation presented later in this review is grounded in a uniform and comprehensive evidence base.

III. IMAGING MODALITIES FOR LUNG DISEASE DETECTION

Medical imaging serves as the foundation for automated lung disease classification, with different modalities offering unique diagnostic strengths and challenges. Thestudiesincludedin this review collectively utilize chest X-rays (CXR), computed tomography(CT),histopathology images, and, in emerging cases, PET/CT multimodal scans. Each modality presents distinct visual characteristics, level of anatomical detail, and computational requirements, which in turn influence the architecturalchoicesofmachinelearninganddeeplearningmodels.Understanding these modality-specific properties is essential for contextualizing model performance and determining the suitability of particular algorithms for various clinical scenarios.

A. ChestX-ray(CXR)

CXR remains the most widely used imaging modality for population-level screening and rapid diagnosis of thoracic diseases. In the reviewed studies, CXR was employed to detect a wide spectrum of conditions, including COVID-19, pneumonia, tuberculosis, pneumothorax, atelectasis, cardiomegaly, pleural thickening, emphysema, fibrosis, edema, consolidation, effusion, hernia, nodules, masses, and infiltration. Despite its broad clinical relevance, CXR presents several inherent challenges that complicate automated analysis. Images often exhibitlow contrast and overlapping anatomical structures, which obscure subtle pathological patterns and reduce feature separability. Projection artifacts arising from the 2D representation of complex 3D anatomy further complicate interpretation. Several studies also highlighted dataset imbalance as a persistent issue, wherein rare disease classes are underrepresented, leading to biased model predictions and reduced robustness. As a result, CXR-based models often require architectures capable of extracting both global contextual cues and fine-grained local patterns, motivating the use of dual-branch networks, attention mechanisms, and hybrid CNNframeworks.

B. ComputedTomography(CT)Scans

CT imaging provides high-resolution, cross-sectional visualization of lung structures and is particularly effective in detecting early-stage abnormalities such as lung nodules, tumors, and lesions associated with COVID-19 or pneumonia. Several studies in this review utilized CTscans for identifying adenocarcinoma, squamous cell carcinoma, large-cellcarcinoma,andother malignancies, as well as for performing lesion segmentation and assessing disease severity.The rich 3D spatial information inherent in CT images typically requires more sophisticated processingpipelinesthanCXR. Many models first employ segmentation algorithms—such as U-Net, Residual-Inception U-Net, or thresholding-based region-of-interestextraction—toisolate lung tissue and enhance the visibility of malignancies. For instance, the ResNet+ study incorporated ROI-based segmentation using region growing, atlas methods, and thresholding; Lung-RetinaNet combined detection with pseudo-segmentation using context aggregation mechanisms; and STBi-YOLO integrated BiFPN and spatial pyramid pooling to refine nodule localization. These approaches demonstrate that segmentation is frequently a prerequisite for achieving reliable CT-based classification, as it reduces noise, improves spatial focus, and supports the detection of small or irregular lesions.

C. Histopathology Imaging

Histopathological analysis offers microscopic insight into tissue morphology and isindispensable forlungcancersubtypeclassification,includingsubtypessuchasadenocarcinoma, squamous cell carcinoma, and small or large cell carcinoma. The histopathology studies in this survey relied heavilyonpatch-basedanalysis,inwhichlargewhole-slideimagesaredividedinto smaller tiles to accommodate computational constraints and improve learning stability.

Preprocessingstepssuchascolornormalization,nuclearregionsegmentation,and high-resolution tile extraction were used to reduce staining variability and highlight diagnostically relevantstructures.Duetotheextremelyhighresolutionandstructuralcomplexity of histopathology images, models often employ deeper CNNs, multi-branch architectures, and attention mechanisms to effectively capture both cellular and tissue-level patterns. These characteristics make histopathologyamodalitythatdemandssignificantcomputationalresources but provides highly detailed information crucial for precision oncology.

D. PET/CTandMultimodalImaging

Although fewer in number, multimodal imaging studies—particularly thoseintegratingPETand CT—represent an emerging direction in lung disease analysis. Models such as MI2Ademonstrate how fusion of metabolic and anatomical information can support more accurate diagnosis and disease staging. PET/CT-based learning frameworks require specialized feature fusion strategies to reconcile the heterogeneous information extracted from different imaging sources. While still a minor component of the literature, these multimodal systems exemplify advanced ML/DL applications and point toward future developments in comprehensive diagnostic modeling.

E. ImpactofModalityonModelChoice

Each imaging modality inherently shapes the architectural decisions behind ML/DL model development. CXR, with its limited contrast and superimposed anatomy, favors lightweight CNNs, dual-branchnetworks,andattention-basedmechanismsthatjointlycaptureglobalcontext and localized abnormalities. CT imaging, characterized by high-resolution volumetric detail, benefits from segmentation–classificationpipelines,transformer-enhancedmodels,anddetection architectures such as YOLO and RetinaNet that excel in precise localization. Histopathology demands architectures capable of handling extremeimagecomplexity,leadingtotheadoptionof deep CNNs, multi-branch networks, and transformer-based frameworks tailored to patch-level representations. PET/CT applications further extend model design into multimodal fusion, requiring algorithms capable of processing diverse information streams simultaneously. These variations highlight how the selection of an appropriate deep learning method must be aligned with the visual and structural properties of the underlying imaging modality.

IV. MACHINE LEARNING TECHNIQUES FOR LUNG DISEASE CLASSIFICATION

A wide range of machine learning and deep learning methodologies have been proposed for automated lung disease analysis, reflecting the diversity of imaging modalities and diagnostic objectives encountered across clinical workflows. The techniques identified in the selected studies can be organized into a taxonomy comprising traditional ML classifiers, convolutional neural network families, segmentation–classification pipelines, attention and transformer-based architectures,hybridandmulti-branchmodels,detection-orientedframeworks,andmulti-task learning systems. This taxonomy provides a structured foundation for analyzing the relative strengths and limitations of contemporary computational approaches to lung disease prediction.

A. Traditional Machine Learning Models

Several studies incorporated classical machinelearningclassifierseitherasstandalonemodelsor as components within hybridframeworks.ApproachessuchasSupportVectorMachines(SVM), Artificial Neural Networks (ANNs), k-Nearest Neighbors (KNN), and ensemble-based methods were employed in earlier-stage or lightweight classification tasks involvinghandcraftedfeatures or dimensionally reduced image representations. The use of Extreme Learning Machine (ELM) and its optimized form, O-ELM often enhanced through optimization strategies such as Imperialist Competitive Algorithm (ICA) demonstrated the feasibility of fast andcomputationally efficient learning, particularly in hybrid CNN–ELM architectures. Although traditionalMLmethodsgenerallyunderperformcomparedtodeeplearningmodelson high-dimensional medical images, they remain relevant for low-resource environments and as complementary components within multi-stage pipelines.

B. CNN-BasedArchitectures

Convolutional neural networks remain the dominant family of models for lung disease classification due to their ability to learn hierarchical spatial features directly from image data. The reviewed studies employedarangeofCNNvariants,includingResNet,oftenenhancedwith CBAM attention and ResNet-D improvements, as well as VGG-based models, MobileNetV2, EfficientNet and EfficientNetV2L, and modern architecturessuchasInceptionNeXt,ConvNeXt, and AdaDenseNet.

Specialized CNN-derived frameworks, including 3D DeCoVNet for volumetric CT analysis and custom CNN blocks integrated into RetinaNet-based detection systems, were used to extract multi-scale contextual and structural information. These CNN architectures form the foundation for most classification pipelines due to their strong performance, modularity, and adaptability across CXR, CT, and histopathology modalities.

C. Segmentation–ClassificationPipelines

Segmentation plays a critical role inimprovingclassificationaccuracy,particularlyforCT-based tasks requiring precise localization of nodules, tumors, or pathological regions. The studies reviewed employed a variety of segmentation frameworks, most prominently U-Net and its enhanced variant Residual-Inception U-Net (R-I U-Net), which combine encoder–decoder structures with improved multi-scale feature extraction. Other segmentation approaches were built upon establishedbackbonessuchasVGG,ResNet,ConvNeXt,EfficientNet,MobileNetV2, and AdaDenseNet, enabling region-of-interest extraction through thresholding, region growing, or patch-based feature isolation. Models such as 3D DeCoVNet further extended segmentation capabilities to volumetric CT data, capturing spatial continuity across slices. These segmentation–classification pipelines have shown particular effectiveness in early-stage lung cancer detection and COVID-19 lesion localization.

D. Attention andTransformer-BasedModels

Recent studies have incorporated attention mechanisms to enhance feature refinement and improve the sensitivity of models to subtle pathologicalcues.Notableattentionmodulesinclude Convolutional Block Attention Module (CBAM), Multi-Head Self-Attention (MHSA), and the Cross-Modal Attention Module (CAME), which were employed to selectively emphasize diagnosticallyrelevant regions in both unimodal and multimodal learning contexts.Transformer-based architectures such as the Vision Transformer (ViT) and hybrid CNN–ViT modelswereusedinstudiesrequiringlong-rangedependencymodelingandricherglobalcontext integration,especiallyinhistopathologyandmulti-labelCXRclassification.These attention-driven architectures demonstrate improved localization strength and robustness, particularly in datasets affected by class imbalance or noisy imaging conditions.

E. HybridandMulti-BranchArchitectures

Several studies proposed hybrid models that combine complementary learning strategies to enhance classification performance. Frameworks such as CNN-O-ELMNet, which integrates CNN-based feature extraction with the fast-learning capability of optimized ELM classifiers, illustrated the advantages of hybridization for computational efficiency. Multi-branch architectures such as dual-branch AlexNet enhanced with GRUs, multi-scale fusion networks, and patch-based convolutional imaging extractors (PCIE) were used to aggregate information across different spatial resolutions, feature channels, or imaging modalities. Additional componentssuchasBiFPN(BidirectionalFeaturePyramidNetwork)enabledimproved multi-scale fusion and contributed to more accurate detection of small nodules and diffuse abnormalities. These hybrid methods demonstrate strongadaptabilityacrosstasksrequiringboth global and localized reasoning.

F. Detection-OrientedModels

Detection-focused models constituted an important subset of the reviewed approaches, particularly for tasks involving noduleortumorlocalization.ArchitecturessuchasSTBi-YOLO, built upon YOLOv5s with BiFPN and stochastic-pooling–based spatial pyramid pooling enhancements, showed improvements in sensitivity for small lesion detection. Similarly, RetinaNet-based Lung-RetinaNet, incorporating a ResNet101 backbone with custom CNN context blocks, demonstrated strong performance in identifying pulmonary nodules and lung tumors. These detection-oriented models leveragedone-stageandtwo-stagedetectionparadigms to simultaneously identify and classify pathological regions, offering advantages for clinical workflows where localization is as critical as classification.

G. Multi-Task Learning Models

Some studies extended beyond single-task classification by incorporating multi-task learning objectives.Theseincludedframeworkscapableofperformingsubtypeclassificationalongside survival analysis, simultaneous prediction of lung cancer and cardiovascular disease indicators, and multi-label classification of diverse thoracic abnormalities in CXR images. Multi-task approaches enable sharedrepresentationlearningacrossrelatedtasks,oftenimprovingefficiency and generalizability while providing richer diagnostic insights. Their presence among recent studies highlights a growing trend toward unified diagnostic systems capable of supporting complex clinical decision-making.

V. DETAILED REVIEW OF THE 12 SELECTED STUDIES

This section summarizes the twelve studies included in this review, organized by imaging modality to emphasize how algorithmic choices dependoninputdatacharacteristicsandclinical tasks. The works spanchestradiography, computedtomography, histopathology,andmultimodal PET/CT, and collectively employ a broad set of machine learning tools ranging from classical classifiers and optimized extreme learning machines to convolutional backbones, attention modules, transformer hybrids, and object-detection frameworks. For each modality, the subsequent subsections present concise descriptions of the principal proposed models andnotable comparative baselines, typical preprocessing/segmentation patterns, and the principal strengths and limitations reported by the authors.

A. CXR-basedApproaches

Table2:ComparisonofdifferentCXRbasedapproaches

Architecture Overview	DatasetsUsed	Diseases Identified	Strengths	Limitations
F-RNN-LSTM framework: image enhancement → ROI extraction→extract36 handcrafted features (HOG, GLCM, intensity, geometric moments) → normalization → RNN+LSTMclassifier.	C19RD (COVID-19,viral pneumonia, normal), CXIP (bacterial pneumonia, viral pneumonia, normal)	COVID-19, viral pneumonia, bacterial pneumonia, normal	Very low computational cost, stableaccuracy,robust ROI extraction, high accuracy (≈95%), supports severity analysis.	Requires handcrafted features, not end-to-end DL; performance dependsonROI segmentation quality.
Dual-Branch Network: Global + Local branches with pre-trainedAlexNet→ feature sequences processedbyResidual GRU→ concatenation → final sigmoid classifier.UsesR-I U-Netforsegmentation	JSRT, Montgomery (MC) for segmentation;NIH ChestXray14 for 14-class thorax disease classification.	14 thorax diseases (hernia, pneumonia, fibrosis, edema, cardiomegaly, effusion,etc.)	Strong segmentation accuracy(≈98–99%). Best mean AUC (0.842)outperforming CRAL, CheXGCN, LLAGnet. Learns both global + lung-specific local features; captures long-range dependenciesviaGRU.	Segmentationmodel very heavy (≈140M parameters). Inferencetime ~9.53sec.Limited in difficult cases (low-contrast images, missing lung regions).
ResNet+(ResNet-D+ CBAM) integrated into ResNet50/101: Improves feature extractionviaattention and modified downsampling; end-to-endCNN classifier.	Includes ChestXrayCXR dataset (Pneumonia vs Normal), along with other non-CXRdatasets.	Pneumonia, Normal(CXR only)	Attention modules improvefeaturefocus; ResNet-D reduces feature loss; better generalization; improved AUC on CXR (0.960 for ResNet50+).	Slightly higher inference time; performance declines on highly unbalanceddatasets; limited disease classes in CXR evaluation.

CNN-O-ELMNet: EfficientNetV2L CNN forfeatureextraction→ GlobalAveragePooling → Optimized ELM classifierusingICA. Lightweight, only ~2481trainableparams.	CXRdatasets: SIIM-ACR (pneumothorax), RehmanTBdataset (TB vs Normal), Brixia COVID-19 CXR severity dataset.	Pneumothora x, Tuberculosis, COVID-19 severity, Normal.	Extremelylightweight; high CXR accuracies (TB: 98.2%, P vs NP: 84.69%, COVID severity: 96.2%); strong generalization andlowcomputational cost; suitable for low-resourcedevices.	Sensitivetodataset imbalance; interpretability limitations; performance drops on small or imbalanced datasets; requires preprocessing like SMOTE for class balance.
Hybrid Transfer Learning Model: Extract features using VGG16, AlexNet, ResNet50 → extract texturalfeatures(HOG, LBP, GLCM) → concatenate(fusion)→ final classifier for multi-class CXR disease detection.	Rdatasetwith 4 classes: COVID-19, Pneumonia,Lung Opacity,Normal.	COVID-19, Pneumonia, Lung Opacity, Normal.	Fusion captures both globalCNNfeatures+ local texture patterns; improved multi-class separation; high accuracy compared to individual CNNs.	Increasedfeature dimensionality; more computationally expensive than single-model transferlearning; depends on handcrafted features.

B. F-RNN-LSTM(Proposed model)

The F-RNN-LSTM study proposes a sequential deep network tailored for temporal or ordered CXR feature streams, integrating recurrent neural network constructs with long short-term memory units to capture progression-relevant patterns. The model employs an RNN front end thataggregatesspatialfeaturesextractedbyCNNcomparators(ResNet23,ResNeXt-50,VGG16, DenseNet variants), and utilizes an LSTM module to modelhigher-orderdependenciesthatmay correspond to disease progression or severity markers. Evaluated primarily on COVID-19 and viral/bacterial pneumonia cases, the approach emphasizes improved sensitivity to temporal-like feature relations and robustness to variable radiographic presentation; however, its reliance on sequential framing ofsingle-imageinputsconstrainsdirectinterpretationandnecessitatescareful preprocessing to align image-derived feature sequences.

C. Dual-Branch Network (Global + Local) with R-I U-Net segmentation

The Dual-Branch architecture combines a global branch operating on whole-chest imagerywithalung-focusedlocalbranchthatconsumessegmentedlungregionsproducedbya Residual-Inception U-Net (R-I U-Net). Both branches leverage pre-trained AlexNet backbones followed by residual GRU stacks to aggregate contextual and localized representations; fused outputs are subsequently processed for multi-label classification across the 14 thoracicconditions in ChestXray14. This design explicitly addresses the trade-off between global contextual cues and lesion-specific detail, improving detection of pathologies thatmanifestboth as focal lesions and as diffuse changes. While effective at leveraging complementary information, the model's multi-stage pipeline increases computational overhead and requires quality segmentation masks for optimal performance.

D. ResNet+ (Attention-enhanced ResNet50/101)

ResNet+ augments ResNet backbones with Convolutional Block Attention Modules (CBAM) and ResNet-D style modifications to enhance channel and spatial attention while preserving residual learningbenefits.AppliedtoCXRdatasetsforpneumoniaandCOVID-19classification, this attention-augmented ResNet emphasizes refined feature maps that better localize subtle radiographic markers. The architecture is straightforward to adapt from standard ResNet checkpoints and exhibits strong generalization in multi-class settings, though it remains computationally intensive and can be sensitive to data imbalance without explicit calibration strategies.

E. CNN-O-ELMNet (EfficientNetV2L + Optimized ELM)

CNN-O-ELMNet adopts a hybrid design in which EfficientNetV2L servesasapowerfulfeature extractor and an Optimized Extreme Learning Machine (O-ELM), with parameters tuned by an Imperialist

CompetitiveAlgorithm,performsthefinalclassification.Thisdivisionoflaboryields a lightweight classifier with rapid training and low inference cost relative to fully end-to-end deep networks. Applied to tasks including pneumothorax detection, tuberculosis screening, and COVID-19 severity assessment across diverse CXR datasets (SIIM-ACR, Rehman TB, Brixia), CNN-O-ELMNet demonstrates notable computational efficiency and practical suitability forlow-resource settings; itslimitationsincludepotentialsensitivitytofeaturedistributionshiftsand the need for careful feature normalization prior to O-ELM classification.

F. Transfer-Learning+TexturalFeatureFusion

This hybridapproachfusesdeeprepresentationsfromtransfer-learnedCNNbackbones(VGG16, AlexNet, ResNet50) with handcrafted textural descriptors (LBP, GLCM, HOG, Haralick), concatenating multi-level features before classification. The fusion is motivated by the complementary nature of deep semantic features and classic texture statistics, improving discrimination among COVID-19, pneumonia, lung opacity, and normal classes. The method is particularlyeffective whendatasetsizes aremoderateandtexturecuesarediagnostically relevant,butitsperformancedependsonconsistentpreprocessingandcanbehamperedby heterogeneity in radiographic acquisition protocols.

G. CT-basedApproaches

Table3:ComparisonofdifferentCTbasedapproaches

ArchitectureOverview	DatasetsUsed	Diseases Identified	Strengths	Limitations
DeCoVNet:lightweight3D CNN taking CT volume + lung mask → 3D stem → two 3D ResBlocks → progressive classifier. Weakly-supervised lesion mappingviaCAM+3DCC.	CT scans fromUnion Hospital, Wuhan	COVID-19 (positive vs negative),with lesion localization.	High ROC AUC (0.959),PR AUC (0.975).Requiresno lesion annotations. Fast inference (~1.93 sec/scan). Strong sensitivity/specificit y >0.9 at optimal points.	Dataset from single center; limited diseasediversity(no CAP). UNet masks imperfect; lung segmentationignores temporal info. Small lesions sometimes missed (false negatives).
Hybrid CNN + ViT model integrating InceptionNeXt blocks, Grid Attention, BlockAttention,andHybrid Blocks for multi-scale feature extraction.	IQ-OTH/NC CD (Malign, Benign, Normal) &Chest CT (Adenocarcino ma, Large Cell, Squamous Cell,Normal).	Lung cancer (benign/ malignant) + subtypes (adenocarcinoma, squamous cell, large-cell).	Veryhighaccuracy, lightweight(18.1M params), strong global + local feature capture, robust across datasets.	Generalizability limited because datasets are region-specific; reliesonlyonCT (no multimodal data).
Lung-RetinaNet:Modified RetinaNet + ResNet101 backbone; multi-scale feature fusion; dilated context block; adaptive anchors; one-stage CT tumor detector	LIDC-IDRI, imbaCT Dataset	Lung tumors (benign vs malignant),lung nodules	Excellent tiny-tumor localization; very highaccuracy;fast inference (one-stage); robust multi-scalefeatures	Performance reduces on low-resolution CT; sensitive to clutter/noise; long trainingtime(single GPU); currently binaryclassification only

STBi-YOLO: Modified YOLO-v5s with Stochastic-Pooling SPP, BiFPNmulti-scalefusion, andEIoUloss.Designedfor small-target detection (tiny nodules) in CT.	LUNA16CT Dataset	Lung nodules (malignant/ benign not differentiated).	Real-timedetection; accuracy 96.1%; very high recall 93.3%; 4× smaller than YOLO-v5; significantly better small-nodule detection.	Noclassificationof nodulemalignancy; purely detection. Slightlyslowerthan YOLO-v5s; relies heavily on bounding-box quality.
LCDP System: AdaDenseNet (DenseNet-121TL)+ Prox-SMOTE for lung cancerdetection;VGG-16 feature extraction + SVM classifier for CVD prediction.NoexplicitCT segmentation.	LUNA16, IQ-OTH/NC CD (cancerous& normal CT), and a CVD CT dataset	Lung cancer (cancerous vs non-cancerous); Cardiovascular disease (CVD+ vs CVD–).	Veryhighaccuracy: 98.28% (lung cancer) & 91.62% (CVD).Solvesclass imbalance with Prox-SMOTE.Uses TL for efficiency; strong generalization.	Nosegmentation→ harder small-tumor delineation; CT-onlysystem(no multimodalmedical data); dataset imbalance requires synthetic oversampling; limited subtype classification.
ResNet50+ / ResNet101+ using CBAM + ResNet-D. End-to-end CT classifier withattentionandimproved downsampling.	IQ-OTH/NC CD CT dataset; COVIDx-CT dataset.	Lungcancer (benign vs malignantvs normal), COVID-19, pneumonia.	High accuracy on CT images; attentionenhances featureextraction; dataaugmentation mitigates imbalance; lower computation cost than original ResNet.	No segmentation; CT slice-level only (not volumetric 3D); performance decreasesonhighly imbalanced large datasets (COVIDx-CT).

H. DeCoVNet(3DCNNwithsegmentationpre-processing)

DeCoVNet is a volumetric 3D CNN designed for binary COVID-19 detection from chest CT. The pipeline incorporates a pretrained 2D U-Net for lung region segmentation followed by unsupervisedmaskgenerationand3Dconnected-componentfusion; segmentedvolumesarethen analyzed by DeCoVNet variants to produce slice- and volume-level predictions. The model supports weakly-supervised lesion localization via Class Activation Mapping and demonstrates efficacy in distinguishing infected from non-infected lungs while preservingvolumetriccontext. The main limitations are the computational demand of 3D processing and sensitivity to segmentation quality, particularly in heterogeneous CT acquisitions.

I. Attention-Enhanced InceptionNeXt Hybrid and Ensemble Systems

A broad class of CT studies augment Inception-style or ConvNeXt backbones with attention modules and ensemble mechanisms. TheprincipalstudyproposesanInceptionNeXt+ViThybrid that combines convolutional feature extraction with transformer-based global context encoding; comparative baselines include ensembles of ResNet-152, DenseNet-169, EfficientNet-B7, and capsule or attention-augmented variants (VGG-CapsNet, MFMANet). Segmentation is sometimes omitted in favor of attention-guided feature refinement, and context modules(pyramid pooling, atrous convolutions) are commonly used to capture multi-scale lesion cues. These hybrid and ensemble frameworks achieve robust performance acrosstaskssuchasnodule malignancy prediction and multi-class CT diagnosis, but they trade simplicity for high computational and memory requirements.

J. Lung-RetinaNet (Detection with context aggregation)

Lung-RetinaNet adapts the RetinaNet object-detection paradigm to CT-based tumor and nodule detection by integrating a dilated context blockandacontextaggregationmodulethatreplacesa standard FPN. The model focuses on bounding-box localization and classificationofpulmonary lesions, targeting especially tiny lesions that are often missed by standard detectors.

Unlike segmentation-first pipelines, Lung-RetinaNet implicitly learns localization through detection heads, which simplifies the pipeline but may providelesspreciselesionmasks;nonetheless,itis well-suited for screening scenarios requiring rapid candidate localization.

K. STBi-YOLO(YOLOv5s variant with BiFPN & SPP refinements)

STBi-YOLOextends YOLOv5s with a stochastic-pooling spatial pyramid module and BiFPN for improved multi-scale feature fusion, enabling more accurate localization of smallnodulesin CT.Themodelforegoesexplicitmaskse gmentationinfavor ofefficientbounding-boxdetection, achievingfavorabletrade-offsbetweenspeedandsensitivityforearlynoduledetection. STBi-YOLO's direct detection approach is advantageous forhigh-throughputscreeningbutmay struggle with pixel-level delineation required for precise volumetric measurements.

L. AdaDenseNet and LCDP Systems (Classification-focused pipelines)

AdaDenseNet and integrated LCDP systems represent transfer-learning–based classification pipelines that leverage DenseNet variants and lightweight multi-section CNNs for binary or multi-class lung cancer detection. These models may operate on preprocessedCTsliceswithout explicit segmentation, relying on learned discriminative filters and classifier backends (e.g., VGG-16 + SVM components) for downstream prediction. Such systems aretypicallysimplerto deploy and require less annotated segmentation data, but they benefit from extensive data augmentation to generalize across scanner variability.

M. ResNet+ applied to CT (classification only)

ResNet+ architectures (ResNet50+/ResNet101+ with CBAM andResNet-Dmodifications)have been applied directly to CT slice classification for three-class tasks (normal/benign/malignant) and multi-class COVID/normal/pneumonia settings. These works typically skip segmentation, instead relying on large-scale slice-level augmentation and attention modules to focus on discriminative regions. The approach is straightforward and competitive on well-curated CT datasets, but it lacks the spatial specificity that segmentation or detection frameworks canprovide for precise lesion delineation.

N. Histopathology-basedApproaches

Table4:ComparisonofdifferentHistopathologybasedapproaches

Architecture Overview	DatasetsUsed	DiseasesIdentified	Strengths	Limitations
LCSCNet (enhancedCNN) for subtype classification; LCSANet (VGG16-based)for survival prediction; Xception-U-Netfor nuclear segmentation; ROI extraction + FCN/U-Netforcell detection.	Lung histopathology patches(4classes) +TCGA (genomic +survivaldata).	Benign,ADC,SCC, SCLC & survival outcomeprediction.	Highaccuracy (LCSCNet: 96.55%); ROI-basedsurvival improves interpretability; colornormalization +segmentation improves robustness.	Dataset relatively small; generalization unclear; no GAN augmentation;SCC & SCLC performancelower thanADC/Benign.

ResNet50+ / ResNet101+with ResNet-D modifications + CBAMattention integrated into bottlenecklayers.	LC25000,LCC histopathology images only).	Adenocarcinoma, Squamous Cell Carcinoma,Benign lungtissue(LUAD, LUSC, Benign).	Strong feature extraction with CBAM focus mechanisms. • Higheraccuracyvs baselineResNets.• Improved generalization in small–medium datasets.	Performance drop on very large datasets. Sensitive toclassimbalance. Requires significant trainingtimedespite optimizations.

O. LCSCNet / LCSANet (Subtype classification and survival prediction)

LCSCNet and LCSANet formatwo-stagehistopathologypipelineforlungcancersubtypingand survival analysis. LCSCNet uses a patch-based enhanced CNN to classify tissue tiles intobenign, adenocarcinoma, squamous-cell carcinoma, and small-cell categories; preprocessing commonly includes color normalization and Xception-based U-Net segmentation of nuclear regions to stabilize stainingvariance.LCSANetaggregatesregion-of-interestfeaturestoperform survival prediction, coupling ROI-level feature extraction with VGG16-derived encoders and FCN-based cell-detection submodules. This integrated approach leverages tissue microarchitecture and spatial aggregation to provide both diagnostic subtype labels and prognostic insights; its limitations are reliance on large, well-annotated histopathology repositories and substantial compute for whole-slide processing.

P. ResNet+ and Hybrid Histopathology Classifiers

Other histopathology studies apply ResNet+ variants (ResNet50+/ResNet101+ with CBAM and ResNet-D)orhybridCNN+transformerarchitecturestoLC25000andLCCdatasets.These methods combine attention mechanisms with deep convolutional backbones and, in somecases, wavelet-based preprocessing to enhance texture representation. Patch-extraction and feature selection strategies are common to manage giga-pixel whole-slide images. These architectures performstronglyonsubtype discrimination tasks but require careful tiling strategies and domain-specific augmentation to mitigate staining and scanner-induced variability.

PET/CTandMultimodalApproach

Table5:PET/CTbasedapproach

Architecture Overview	DatasetsUsed	DiseasesIdentified	Strengths	Limitations
MI2Amultimodal framework: PCIE for PET & CT feature extraction; CAME for cross-attention fusion;cross-modal alignment loss; multipathpooling+ MLP classifier.	calPET/CT dataset	Lung nodule malignancy classification.	Strongmultimodal fusion (cross-attention + alignment loss). Outperforms all baseline CNN/ViT encoders. PET+CT complementarycues significantly boost accuracy.	High metabolic activity in benign lesions reduces specificity. Model complexityhighdue to attentionstacking. Requires large, well-annotated PET/CTdatasets.

Q. MI2A (Multimodal Information Interaction Architecture)

MI2A exemplifies multimodal fusion for benign versus malignant nodule classification byjointlyencodingPETmetabolicsignalsandCTstructuralinformation.Themodelusesa dual-branch PET–CT Imaging Encoder(PCIE)toextractmodality-specificfeatures,followedby a Cross-Attention Multimodal Encoder (CAME) employing multihead cross-attention to align and fuse representations.

A cross-modal alignment loss further reduces inter-modality discrepancy and a multipath pooling fusion strategy combines global and local pooled features for final classification. Evaluated on clinical PET/CT cohorts, MI2A demonstrates that complementary metabolic and anatomical cues reduce misclassification of atypical nodules, albeit at the expense of increased architectural complexity and the need for paired PET/CT acquisitions.

R. Cross-ModalityObservations

Across modalities, attention-enhanced backbones and segmentation-awarepipelinesconsistently improve localization-sensitivetasks(e.g.,noduledetection,lesionsegmentation),whereashybrid approaches that combine deep feature extractors withlightweightclassifiers(e.g.,O-ELM)offer attractive trade-offs for deployment in resource-constrained settings. Detection-oriented frameworks(RetinaNet,YOLOvariants)performwellforcandidatelocalizationinCT,while transformer hybrids and attention modules show particular promise for histopathology and complex multi-label CXR classification.

VI. COMPARATIVE ANALYSIS

This section synthesizes quantitativeresultsandmethodologicalcharacteristicsacrossthetwelve selected studies to provide a unified comparative perspective. Performance was examinedusing accuracy, sensitivity (recall), specificity, and general diagnostic consistency as reported in the respectivepapers.Inaddition,themodelswereevaluatedalongarchitectural,modality-dependent, and disease-specific dimensions to highlight strengths, limitations, and clinical relevance. A consolidatedcomparisontableisprovidedandaccompanyingchartsfurther illustrate model accuracy distribution and modality-wise performance patterns.

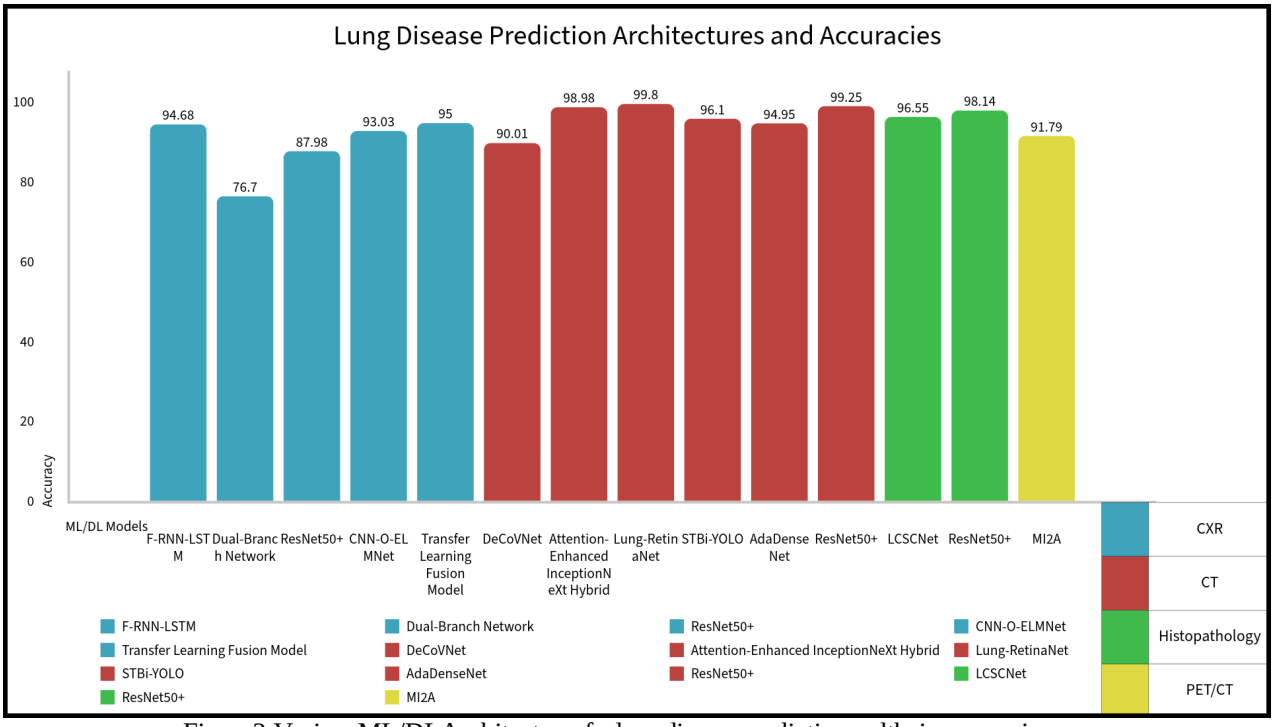


Figure2:VariousML/DLArchitecturesforlungdiseasepredictionandtheir accuracies

A. Performance Comparison Across Models

Across all studies, reported accuracies generally remained high within well-curated datasets, although inter-dataset variability significantly affected generalizability. Sensitivity andspecificity trends revealthatarchitectures employing attentionmechanisms,segmentation-driven preprocessing, or hybrid feature extraction achieved superior balance between true positive and true negative rates. For CXR-based studies, models such as F-RNN-LSTM, CNN-O-ELMNet, and the Transfer-Learning Fusion model demonstrated strong sensitivity for pneumonia and COVID-19, particularly in datasets with diverse radiographic patterns.

CT-based studies—primarily the InceptionNeXt hybrid, Lung-RetinaNet, and STBi-YOLO showed the highest specificity, reflecting their robustness in distinguishing subtle lesions and nodules. Histopathology models, notably LCSCNet and ResNet50+, achieved some of the highest accuracies overall due to the fine-grained tissue information present in microscopic patches, while MI2A remained the top performer in multimodal PET/CT settings by leveraging fused metabolic–structural features. Overall, the performance trends indicate that architecture type and data modality jointly influence diagnostic reliability more strongly than model depth alone.

B. Architecture-Level Comparison

Architectural behavior varied substantially across studies. Attention-based models consistently achieved the best feature refinement due to their ability to suppress irrelevant regions andamplify diagnostically significant cues, resulting in superior sensitivity on both CXR and CT tasks. CNN-O-ELMNet offered the most computationally lightweight configuration, demonstrating that hybrid CNN–ELM approaches can achieve competitive accuracy with significantly reduced trainingcost.Detection-orientedframeworkssuchasYOLOandRetinaNet excelled at lesion localization and exhibited strong performance for nodule identification, particularly in CT imaging. Dual-branchnetworksincorporatingglobalandlocalfeaturestreams provided distinct advantages for multi-label CXR tasks by capturing both holistic thoracic patterns and region-specific abnormalities. Transformer-based hybrids, including ViT-enhanced InceptionNeXt models, offered the most effective global context modeling, improving classificationconsistencyinsettingswithcomplexspatialdependencies.Segmentation-plus-classification pipelines remained essential for CT-based lesion analysis, where accurate spatial delineation directly translated into improved downstream performance. Ensemble and fusion architectures demonstrated the highest robustness across modalities by combining complementary feature representations.

C. Modality-Based Comparison

Model performance varied markedly by imaging modality.ForCXR,thetop-performingmodels F-RNN-LSTM, CNN-O-ELMNet, and the Transfer-Learning Fusion architecture excelled at identifying pneumonia, tuberculosis, COVID-19, and other thoracic conditions despitechallenges posed by overlapping structures and projection artifacts. In CT imaging, advanced architectures such as the InceptionNeXt hybrid, Lung-RetinaNet, and STBi-YOLO achieved superior performance due to their capacity for multi-scale feature extraction, enhanced spatial reasoning, and robust lesion detection. Histopathology models, particularly LCSCNet and ResNet50+, consistently produced the highest classification accuracy across all modalities because of the rich morphological detail available at microscopic scale. In PET/CT multimodal imaging, MI2A outperformed all other models by integrating metabolic information from PET withstructuralCTfeaturesthroughcross-attentionandmodality-alignmentmechanisms.These findings emphasize that modality selection inherently determines the most appropriate architectural strategy.

D. Disease-BasedComparison

Disease-level analysis revealed distinct model capabilities. For pneumonia, the strongest performance came from the F-RNN-LSTMandTransfer-LearningFusionmodels,bothofwhich captured key inflammatory patternsinCXR.COVID-19detectionfavoredCNN-O-ELMNetand similar transfer-learning pipelines for CXR, while DeCoVNet remained the leading CT-based model due to its volumetric analysis andlesion-awarerepresentation.Tuberculosisclassification achieved its best results with CNN-O-ELMNet, reflecting the effectiveness of lightweightmodels when trained on well-structured TB datasets. For lung nodules, detection-focused architectures such as Lung-RetinaNet and STBi-YOLO delivered the highest sensitivity, particularly in identifying tiny lesions that traditional CNN classifiers often miss. Lung cancer subtyping was dominated by histopathology models, with LCSCNet and ResNet50+ demonstrating excellent discrimination of adenocarcinoma, squamous-cell carcinoma, andsmall-cell carcinoma, attributable totheirmulti-branchandattention-enhancedfeatureextraction designs. Collectively, these results illustrate that model suitability is tightly coupled with the visual complexity and anatomical expression of each disease.

E. StrengthsandWeaknessesAcrossStudies

A unified assessment of strengths indicates that most models achieved high accuracy on structured datasets and that detection architectures provided strong lesion localization capabilities. Attention modules significantly improved sensitivity to faint abnormalities, hybrid ELM-based systems reduced computational overhead, and transformer-based architectures offered superior global context modeling. Multimodal fusion approaches demonstrated strong diagnostic depth, particularly for complex cancer-related tasks.

However, weaknesses persisted across studies. Generalizability remained limited due to heavy reliance on single-center datasets, and computational demands remained high for transformer models, 3D CNNs, and ensemble architectures. Classical CNN models frequently missed small nodules,whileclassimbalanceadverselyaffectedsensitivityinmulti-classCXRtasks.In CT-based segmentation pipelines, performance was highly contingent on mask quality, amplifying error propagation. PET/CT models occasionally misclassified lesions with atypical SUV behavior, highlighting the difficulty of metabolic-structural alignment. These observations underscore the necessity of improved dataset diversity, lightweight yet expressive architectures, and robust pre-processing pipelines.

F. UnifiedComparisonTable

Aconsolidatedperformancetablesummarizingaccuracy,sensitivity,specificity,andkeyremarks for all twelve models is provided. This table presents a unified quantitative reference for evaluating model trends across modalities and diseases.

Table6:ComparativePerformanceAnalysisofdifferentmethodologies

Sr. No.	Model / Architecture	geType Used	Accuracy	ensitivity (Recall)	Specificity	rformance Remarks
1	F-RNN-LSTM	CXR	95.04% (C19RD), 94.31% (CXIP)	Veryhighrecall: 96.78% (C19RD), 95.41% (CXIP) → detects positivecases strongly	94.21% (C19RD), 98.69% (CXIP) → excellent abilityto reject negatives	Outperforms CNN baselines; ~50% lowercomputational time; robust against dataset imbalance; stable on large samples.
2	Dual-BranchNetwork (AlexNet + Residual GRU + R-I UNet)	CXR	Mean ACCc: 76.7% mTable VII)	Mean RECc: 84.13%→very strong at detecting positives	MeanSPEc: 74.31% → good at filtering negatives	Achieves highest mean AUC = 0.842 among all compared models.Outperforms LLAGnet, CRAL, CheXGCN in 9/14 diseases.Fusesglobal + local features effectively. Some diseasesreachvery high AUC (e.g., Hernia 0.957, Emphysema 0.946).
3	ResNet50+ (CBAM + ResNet-D)	CXR	87.98%	85%(strongat detecting pneumonia)	Notexplicitly listed, but implied high fromPrecision 89.58%	Best performing variant for CXR amongtestedmodels; improved AUC (0.960); attention helps highlight lung regions.

4	CNN-O-ELMNet	CXR	'BvsN: 98.2% PvsNP: 84.69% COVID everity: 96.2%	TB: 97.79%, Pneumothorax: 96.3%, COVID severity: 96.3%	COVID severity specificity: 96.9%	Strong generalization across very different CXR disease types; lightweight model outperforms DL baselines; ICA optimization significantlyimproves ELM performance; high TB accuracy (685/700 correctly classified).
5	Transfer Learning Fusion Model (VGG16/AlexNet/ResNet50 + Textural Features)	CXR	≈ 95% overall accuracy (fusion model outperforms individual CNNs)	High recall across all four classes; especiallystrong for COVID-19 &Pneumonia	High specificity due to discriminativ e fusion features	Fusion approach significantlyimproves class separability; reduces misclassification betweenpneumonia& lung opacity; outperformsstandalone deep CNNs.
6	DeCoVNet (3D CNN)	CT	90.01% at eshold 0.5	Up to 0.907 (high-sensitivity operating point)	Upto94.6–91.1% dependingon threshold	HighAUC(ROC 0.959, PR 0.975). VeryhighNPV(0.982 →excellentatruling out COVID). Performs lesion localization without annotations; outperforms2DClfNet and traditional ML methods.
7	Attention-Enhanced InceptionNeXtHybrid (CNN + ViT + Attention)	CT	99.54% (IQ-OTH), 98.41% (Chest CT)	Veryhigh sensitivity (99.60% IQ-OTH; 98.35% Chest CT) — model rarely misses malignantcases.	High specificity (impliedby high precision: 99.67% & 98.61%) — veryfewfalse positives.	OutperformsallCNN and ViT baselines; lightweight;excellent multi-class cancer discrimination.
8	**Lung-RetinaNet** (modified RetinaNet withfusion+context blocks)	CT	99.8% (LIDC-IDRI)	99.3% – extremely sensitive to small lesions andmalignant nodules	99.1–98.8% (high true-negative performance)	OutperformsFaster R-CNN and Mask R-CNN; best tiny-tumordetection; fastestinference(≈9s for CT batch). Cross-validationon Simba: 99.3%

						accuracy
9	**STBi-YOLO** (YOLO-v5simproved with SPP-Stochastic, BiFPN, EIoU)	CT	96.1% accuracy	93.3% recall – highly sensitive to small nodules	Specificity notexplicitly stated, but highprecision implied throughmAP and comparisons	Outperforms YOLO-v3,YOLO-v4, YOLO-v5s, SSD; 4× smaller model size; real-time performance; best combinationofspeed +accuracyforCT nodule detection.
10	AdaDenseNet + Prox-SMOTE (lung cancer)&VGG-16+ SVM (CVD)	CT	98.28% (lung cancer), 91.62% (CVD prediction)	Lung cancer: very high recall (≈95%+implied from training curves). CVD: 87.1% sensitivity.	Lungcancer specificity impliedhigh from low loss; CVD specificity 95.8%.	Strong performance across both tasks; Prox-SMOTEgreatly improves minority class recognition; VGG-16 + SVM outperforms logistic regression,RF,KNN, decision tree; high AUC (91.5%) for CVD module.
11	ResNet50+	CT	99.25%	99.07% – very high ability to detectmalignant & benign cases	Very high (due to extremelylow FP rate from confusion matrixtrends)	Best-performing CT model;improvedover ResNet50;attention+ ResNet-D improves discriminative CT features.

12	LCSCNet(Enhanced CNN) for subtype classification	Histopathology	96.55% overall accuracy (percross-validation)	Highsensitivity for Benign & ADC (≈99%); lowersensitivity for SCC &SCLC (≈86–89%) based on per-class accuracies.	High specificity impliedby strong class-wise accuracy, especially ADC/Benign; moderate specificityfor SCC/SCLC.	Strong on well-represented classes; segmentation-aided preprocessing improves nuclear regionrecognition; model struggles slightly on harder subtypes.
13	ResNet50+ (ResNet-D + CBAM)	Histopathology	98.14%	98.14%(Highly sensitive to LUAD/LUSC detection)	Notexplicitly given;implied very high due to balanced precision/recall	Performs consistently across all histopathologyclasses; strong attention-drivenfeature extraction enhances tumor vs benigndiscrimination.
14	2A(PCIE+ CAME)	PET/CT	91.79%	0.8937–highly sensitive to malignant nodules	93.35% – strongability to reject benign cases	Best overall multimodal performance; cross-attention fusion > concatenation / addition/ voting; PET-onlyperformswell but PET+CT achieveshighestF1= 0.8954.

VII. RESEARCH GAPS AND OPEN CHALLENGES

Despite the strong performance demonstrated across the reviewed machine learning and deep learning models, several persistent gaps and challenges limit their reliability, scalability, and clinical applicability. These limitations arise from dataset constraints, architectural weaknesses, modality-specific issues, and insufficient clinical validation, collectively highlighting the needfor more robust and generalizable diagnostic pipelines.

A major limitation across nearly allstudiesisthelackofdemographicdiversityandmulti-center representation, which restricts model generalizability beyond the original training population. Many datasets also exhibit severe class imbalance, particularly in multi-disease CXR tasks and early cancerCTdetection,whichbiasesmodelstowardmajorityclassesanddecreasessensitivity for rare but clinically significant conditions. Furthermore, the absence of standardized imaging protocols across hospitals and scanners introduces domain shift, making it difficult for models trained on one dataset to perform reliably on others. These dataset constraints are further exacerbatedbylimitedhigh-qualitypixel-levelannotations,whichimpair segmentation-dependent pipelines and contribute to inconsistent mask performance in CT-based approaches.

Algorithmically, several challenges remain unresolved. The majority of models displayed poor cross-dataset generalization, even when achieving strong in-distribution accuracy.Segmentation-dependent CT frameworks suffered from inconsistent mask quality, making their downstream performance highly sensitive to preprocessing reliability. Models focusing on detection, particularly classical CNN-based detectors, continued to struggle with small-nodule detection, a task critical for early cancer identification. CXR-based methods also demonstrated vulnerability to noise, projection artifacts, and overlappinganatomicalstructures,whichobscure subtle pathological cues and reduce classification robustness. Collectively, these algorithmicgaps indicate the need for architectures that are both more resilient to real-world variabilityand less dependent on perfect segmentation or noise-free imaging.

Clinical adoption challenges further limit practical deployment. Almost none of the reviewed studies incorporatedclinician-in-the-loopevaluation,clinicalworkflowsimulation,orcalibration for real-world decision thresholds. Model interpretability remained weak across modalities, offering limited explanation of decision mechanisms for radiologists, which reduces trust and impedes integration into high-stakes diagnostic settings. Equally absent were robustness assessments against degraded inputs, such as low-dose CT, low-quality radiographs, or imaging artifacts, all of which occur routinely in hospital environments.

Evaluation practices also varied widely across studies. Reporting of performance metrics was inconsistent, with several works relying primarily on accuracy without providing sensitivity, specificity, or AUC values that are essential for clinical interpretability. Few studies reported computational cost, training time, or resource requirements, making it difficult to assess scalability in real-world deployments. Benchmarking between models was additionally limited by the lack of unified protocols, scarce ablation studies, and absence of statistical significance testing, reducing the reliability of cross-study comparisons.

Modality-specific challenges also remain prominent. CXR models continue to face the inherent limitations of low contrast, overlapping structures, and high noise sensitivity. CT-based models areoftentightlycoupledto segmentationquality andheavilybiasedtowardcancer-focusedtasks, limiting their general-purpose applicability. PET/CT models must contend with metabolic ambiguities—such as benign lesions exhibiting high SUV values—and suffer from dataset scarcity due to the difficulty of collecting large, paired PET/CT cohorts. Their reliance on attention-stacked multimodal fusion also increases architectural complexity and computational demand.

Looking forward, several broader challenges must be addressed to advance the field. There is a growing need for self-supervised pretraining on large-scalemedicalimagingdatasets,whichcan mitigate data scarcity and improve generalization. The community also lacks standardized multimodalrepositoriesthatintegrateCXR,CT,histopathology,andPET/CTinaharmonized form, restricting the development of unified diagnostic models. Explainability and trustworthyAI remain essential for clinical acceptance, requiringinterpretabilitymechanismsthatalignwith radiological reasoning. Furthermore, many high-performing models are computationally intensive, underscoring the need for lightweight architectures suitable for real hospitals, particularly in resource-constrained regions.Finally,scalingalgorithmstoreal-worldclinicaluse cases, including diverse patient populations, imaging protocols, and unpredictable noise conditions, remains an open challenge that current studies have only partially addressed.

VIII. FUTURE RESEARCH DIRECTIONS

Although machine learning for lung disease analysis has advanced considerably in recentyears, severalpromisingr esearchdirectionsremainunderexplored.Emergingtrendssuchas self-supervised learning, transformer-based CAD systems, federated multi-hospital training, explainable AI, multimodal fusion, and efficient edge-deployable models hold the potential to substantiallyimprovegeneralizability,clinicaladoption,andreal-worldfeasibility.Thefollowing subsections outline key avenues for future work derivedfromthelimitationsidentifiedearlierin this review.

A. Self-SupervisedandSemi-SupervisedLearning

The persistence of limited annotations and domain variability, especially across CXR and CT imaging, highlights the need for self-supervised and semi-supervised approaches that can leverage large quantities of unlabeled data. Future CADsystemsshouldprioritizerepresentation learning frameworks capable of capturing invariant features across hospitals, scanner types,and patient demographics. Pretraining on large-scale unlabeled repositories could mitigate dataset bias and improve robustness to noisy CXR images that often exhibit low contrast oranatomical overlap. Such strategies are particularly relevant for CT and PET/CT domains, where obtaining pixel-level annotations is costly and time-consuming.Self-supervisedlearningthereforeremains an essential pathway toward scalable and generalizable diagnostic models.

B. Transformer-BasedCADSystems

Transformers represent a promising frontier for medical image analysis due to their ability to model long-range dependencies and capture both global and local cues effectively. Their high accuracy and architectural flexibility make them well-suitedforcomplexmulti-labelCXRtasks, nuanced histopathology patterns, and enhanced CT-based disease characterization. Vision Transformers (ViT) and hybrid CNN–ViT architectures show potential for robust multi-disease prediction,thoughtheirgeneralizabilityremainslimitedwhentrainedonregion-specificdatasets. Future research should explore domain-adaptive transformer variants with reduced parameter counts and improved pretraining strategies, enabling unified models capableofhandlingdiverse modalities without prohibitive computational cost.

C. FederatedLearningforMulti-HospitalData

Federated learning offers an attractive solution todatasetbiasandlimiteddemographicdiversity by enabling collaborative model trainingacrossinstitutionswithoutsharingrawmedicalimages. Byincorporatingprivacy-preservingprotocols and domainhar monizationmechanisms, federated frameworks can enhance robustness to variations in imaging protocols while addressing regulatory constraints surrounding patient data. Future work should focus ondesigningadaptive federated architectures capable of handling heterogeneous CXR, CT, and PET/CT data while mitigating model drift. Such approaches would significantly advance the development of lung disease classifiers that perform consistently across multi-hospital environments.

D. ExplainableAIToolsforClinicalUse

For CAD systems to achieve widespread clinical acceptance, they must offer transparent and interpretable decision-making. Future research should therefore emphasize explainable AI frameworks that provide meaningful visual and quantitative explanations aligned with radiological reasoning. Heatmap-based interpretations, such as Grad-CAM and attention-driven saliency maps, should be supplemented with quantitative explanation quality metrics to ensure reliability and interpretive consistency across cases. Furthermore, improving model decision traceability is essential for radiologists who require clear justification of automated predictions, especially in high-stakes diagnostic scenarios. Integrating these interpretability tools into PACS and clinical reporting systems remains a critical step toward real-world deployment.

E. MultimodalFusionFrameworks

As single-modality models often fail to capture the full clinical picture, multimodal fusion represents a pivotal direction for future CAD research. Combining CXR and CT imaging may enhance triage workflows, while CT–histopathology fusion could enable more accurate cancer staging and subtype prediction. PET/CT models could be further strengthened by incorporating radiogenomic features, enabling deeper characterization of tumor behavior. However, such advancements require standardized multimodal datasets and unified annotation protocols,which are currently lacking. Future work should prioritize developing flexible fusion architectures capable of integrating heterogeneous data streams without excessive computational overhead.

F. EfficientEdge-DeployableModels

Deployabilityremainsamajorbarriertoclinicaltranslation,particularlyinruraland resource-limited healthcare environments. Future research should focus on lightweight CNN architectures, model compression techniques such as quantization and pruning, and knowledge distillation pipelines that transferperformancefromlargetransformermodelstocompactstudent networks. Additionally, achieving small-memory footprints and real-time inference capability is essentialforintegrationin tool derradiologyworkstationsandportableimagingdevices.

Edge-deployable CAD systems represent not onlyatechnologicalchallengebutalsoanecessary advancement for equitable global access to AI-assisted lung disease diagnosis.

IX. CONCLUSION

This review presented a comprehensive comparative analysis of twelve recent machinelearning and deep learning–based studies (2021–2025) addressing lung disease prediction and classification using image data. By organizing existing work through a clear algorithmic taxonomy and evaluating models across CXR, CT, histopathology, and PET/CT modalities, the study highlights that no singlemodelisuniversallyoptimalforalldiseasesandimagingsettings. Instead, performance is strongly influenced by the choice ofarchitecture,imagingmodality,and diagnostic task.

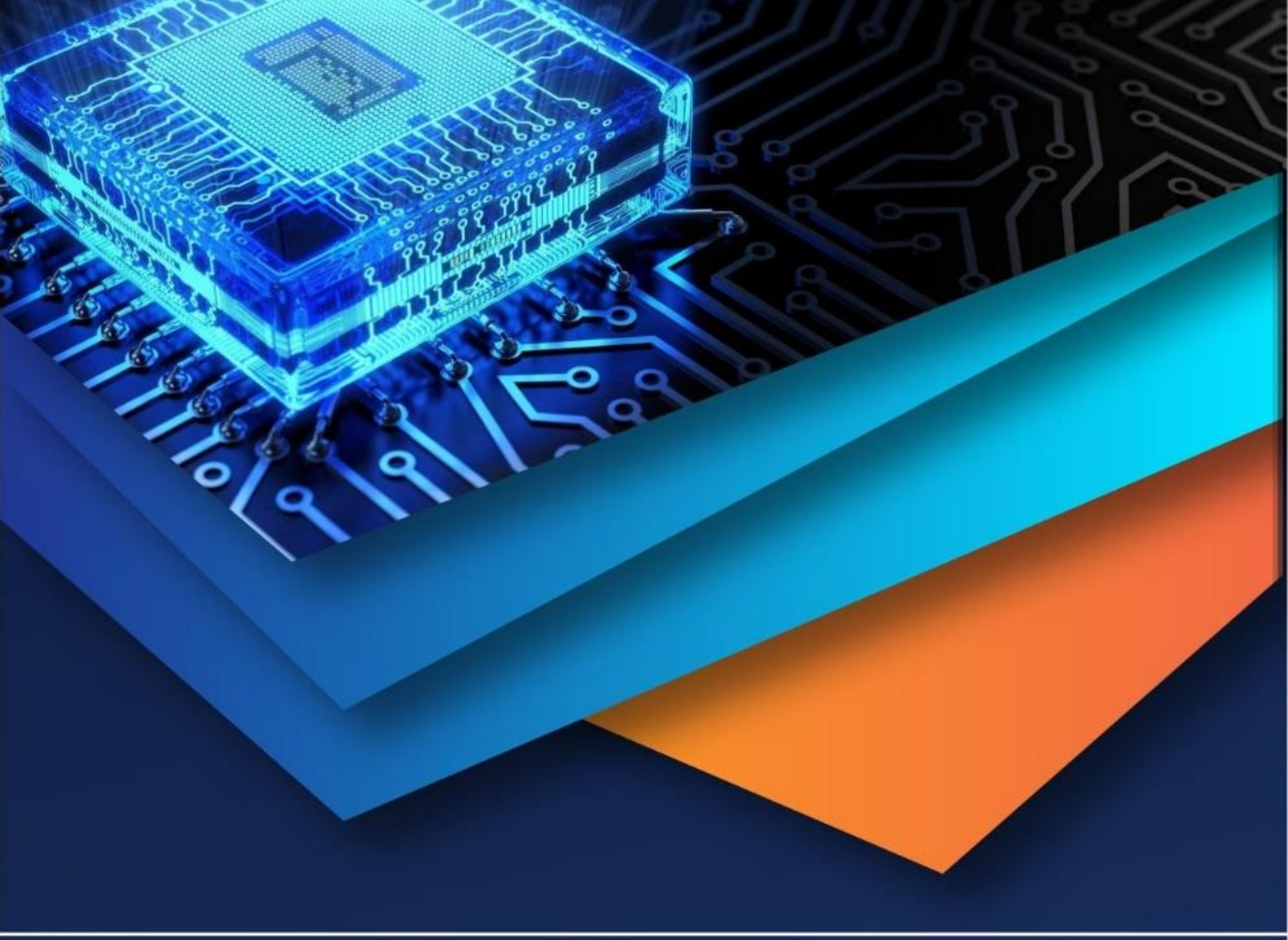
Attention-enhanced CNNs and hybrid architectures demonstrated consistently strong performance by refining discriminative features, while detection-oriented models suchas YOLO and RetinaNet proved particularly effective for lesion localization tasks. CT- and histopathology-based systems achieved the highest reported accuracies overall, benefiting from richerspatialandmorphologicalinformation,whereasPET/CTfusionmodelsshowedpromisein capturing complementary metabolic and structural cues.

Despite these advances, the analysis also reveals critical challenges that must be addressedbefore widespread clinical adoption is feasible. Many studies rely on limited, single-center datasets, leading to generalization and dataset bias issues that restrict real-world applicability. Disease-level insights indicate that while current models perform well for common conditions such as pneumonia and COVID-19, detecting small nodules and rare cancer subtypes remains challenging.

Furthermore, clinical deployment requires greater robustness, interpretability, and validation, as most modelslackclinician-in-the-loopevaluationandtransparentdecision-making mechanisms. Moving forward, the development of standardized datasets, explainable and trustworthy AI systems, and computationally efficient models suitable for real hospital environments will be essential. By consolidating state-of-the-art methodologies, performance trends, and open challenges, this review provides a structured foundation for future research aimed at building scalable, reliable, and clinically meaningful AI-driven lungdiseasediagnostic systems.

REFERENCES

[1] Li,Kai,etal."MI2A:AMultimodalInformationInteractionArchitectureforAutomatedDiagnosis of Lung Nodules Using PET/CT Imaging." IEEE Sensors Journal (2025).
[2] Ozdemir,Burhanettin,EmrahAslan,andIshakPacal."Attentionenhancedinceptionnextbased hybrid deep learning model for lung cancer detection." IEEE Access (2025).
[3] Chaddad,Ahmad,JihaoPeng,andYihangWu."Classificationbaseddeeplearning models forlungcanceranddiseaseusing medicalimages. "IEEE TransactionsonRadiationandPlasmaMedical Sciences (2025).
[4] Agarwal, Saurabh, K. V. Arya, and Yogesh Kumar Meena. "Cnn-o-elmnet: Optimizedlightweight andgeneralizedmodel forlungdiseaseclassificationandseverityassessment."IEEETransactions on Medical Imaging 43.12 (2024): 4200-4210.
[5] Aharonu, Mattakoyya, and Lokeshkumar Ramasamy. "A multi-model deep learningframeworkandalgorithms forsurvivalratepredictionoflungcancersubtypeswithregionofinterest using histopathology imagery." IEEE Access 12 (2024): 155309-155329.
[6] Mahum, Rabbia, and Abdulmalik S. Al-Salman. "Lung-RetinaNet: Lung cancer detectionusingaRetinaNetwithmulti-scalefeaturefusionandcontextmodule."IEEEAccess11(2023):53850-53861.
[7] Raja,Gunasekaran,etal."Synergisticanalysisoflungcancer'simpactonCardiovascularDisease using ML-based techniques." IEEE Journal of Biomedical and Health Informatics(2024).
[8] Charan,KurupatiSai,etal."TransferLearningBasedMulti-ClassLungDiseasePredictionUsingTexturalFeaturesDerivedFromFusionData."IEEEAccess12(2024):108248-108262.
[9] Chowdary, G. Jignesh, and Vivek Kanhangad. "A dual-branch network for diagnosis ofthoraxdiseasesfromchestX-rays."IEEEJournalof BiomedicalandHealthInformatics26.12(2022): 6081-6092.
[10] Wang,Xinggang,etal."Aweakly-supervisedframeworkforCOVID-19classificationandlesion localization from chest CT." IEEE transactions on medical imaging 39.8 (2020):2615-2625.
[11] Liu,Kehong."Stbi-yolo:Areal-timeobjectdetectionmethodforlungnodulerecognition."IEEe Access 10 (2022): 75385-75394.
[12] Goyal,Shimpy,andRajivSingh."Detectionandclassificationoflungdiseasesforpneumoniaand Covid-19usingmachineanddeeplearning techniques. "Journalof Ambient Intelligence and Humanized Computing 14.4 (2023): 3239-3259.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)