Emerging Trends in Early Lung Cancer Detection via LDCT: A Critical Review

Gagan Thakral
J. C. Bose University of Science and Technology,
YMCA Faridabad, India

Umesh Kumar
J. C. Bose University of Science and Technology,
YMCA Faridabad, India

Sapna Gambhir
George Mason University,
Fairfax Virginia, United States

ABSTRACT
Lung cancer remains one of the leading sources of cancer-related mortality throughout the world. The early detection of lung cancer plays a very crucial role in improving a patient's survival rate. Low-Dose Computed Tomography (LDCT) has emerged as a powerful tool for early lung cancer screening. LDCT scan uses reduced x-ray dose as compared to normal CT scan. LDCT maintains a balance between reduced radiation exposure and high sensitivity. This systematic review follows the PRISMA framework to explore the latest advancements in LDCT-based lung cancer detection. This systematic review also includes artificial intelligence-driven detection support and integration of LDCT scan with multi-modal biomarkers. The review discusses the impact of large-scale screening programs and highlights key challenges such as false positives. The review also provides publicly available datasets. The manuscript provides a systematic review in tabular form with complete details like results, shortcomings, and future scope of the selected papers. The authors also discussed over diagnosis, accessibility, and examined potential solutions to enhance screening efficacy. Additionally, we analyze emerging trends in AI-powered image analysis and the role of deep learning in improving detection accuracy. The study also provides insights about the explain ability in lung cancer detection. LDCT scan has suggestively improved early detection rates and efficiency of the system. Further research is required to improve screening procedures and address its limitations. This systematic review provides understandings of the current state of LDCT for lung cancer detection and future directions for advancing early diagnosis.

Keywords
Lung Cancer, Artificial Intelligence, early detection, LDCT scan, screening

1. INTRODUCTION
Studies have shown that when lung cancer is diagnosed at stage I (initial stage), then the five-year survival rate can exceed up to 60-70%. The five-year survival rate is always less for late-stage identification [3]. Low-Dose Computed Tomography (LDCT) screening has proven that it is more effective in identifying lung cancer at initial stages. According to National Lung Screening Trial (NLST), it also reduces the mortality rate up to 25% as compared to normal chest X-rays. Early detection enhances the patient outcomes and quality of life while reducing the overall burden of the disease [3]. Early detection also provides more treatment options like minimally invasive surgery and targeted therapies.

Lung cancer remains one of the prominent sources of cancer-related deaths worldwide. This is due to high incidence and mortality rates. Another reason is specifically due to late-stage diagnosis of the tumour [1]. Fig. 1 describes the structure of the tumour. Lung cancer is responsible for about 19.4% of cancer-related fatalities and 12.3% of all new cancer cases each year, according to global cancer statistics. The disease is more dominant among smokers, but sometimes a significant proportion of cases also occurs in non-smokers. Generally, this will happen due to genetic and environmental factors [2]. Despite the advancements in treatment with time, the five-year survival rate remains low, especially for late-stage lung cancer. So, it is the requirement of the system to highlight the urgent need for early detection strategies to improve patient outcomes.

Early detection of lung cancer significantly improves the survival rates of the patients. The main reason behind it is timely intercession before the disease progresses to advanced stages.

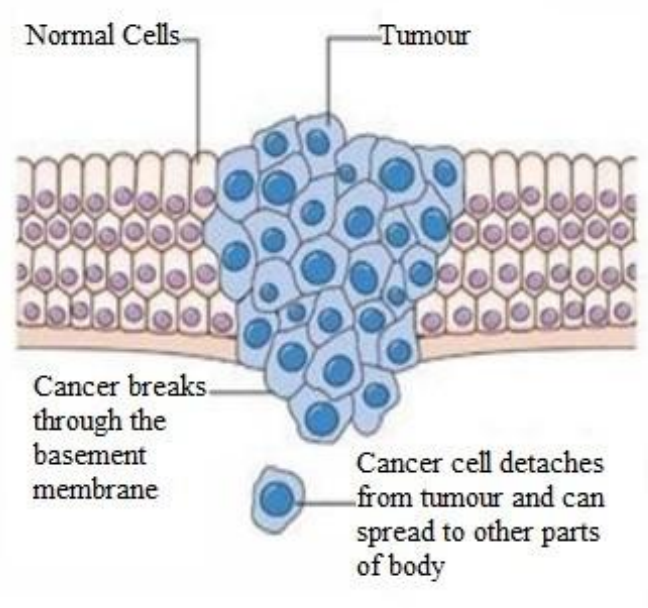

Fig. 1: The Tumor's Structure [source:Thakral et al. [4]]

LDCT scan is a non-offensive imaging technique that has been developed as a critical screening tool for early lung cancer detection. It is also important for high-risk individuals such as long-term smokers. Unlike conventional chest X-rays, LDCT provides cross-sectional, high-resolution images of the lungs. Which allow for the detection of abnormalities or small nodules at an initial stage when treatment is most effective [5]. The NLST demonstrated that LDCT screening reduces lung cancer mortality as compared to normal CT scans and standard X-rays. That's why LDCT becomes the preferred method for early identification of lung cancer. Due to its ability to identify lung cancer before symptoms appear, LDCT plays a vital role in improving survival rates and helping timely medical involvement [6].

In this article, the progress and difficulties of early lung cancer detection using LDCT are systematically reviewed. Within the document, there are several sections. The procedure for the Lung Cancer Detection System using LDCT images is explained in Section

In Section III, the manuscript's technique is explained. In Section IV, several datasets for the identification of lung cancer are offered. A review of the manuscript's literature is given in Section V. In Section VI, explainability is offered. Section VII discusses the difficulties in diagnosing lung cancer using LDCT scan images. Lastly, Section VIII serves as the paper's conclusion.

2. LUNG CANCER DETECTION SYSTEM PROCESS

Fig 2. illustrates the Lung Cancer Detection System Process using Low-Dose Computed Tomography (LDCT) scan images in four key steps:

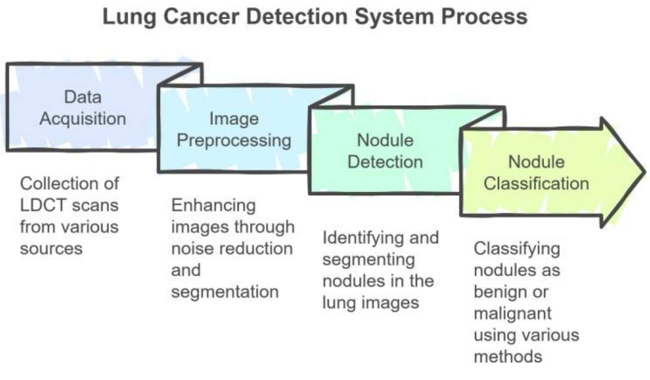

Fig. 2: Lung Cancer Detection System Process with LDCT images

- **Data Acquisition:** This is the first step where LDCT scans of the lungs are collected from different sources, such as hospitals and medical imaging centers. These days, there are some publicly available data sources, which are detailed in Section IV. These scans serve as the input for the detection system [7].
- **Image Preprocessing:** Once the LDCT images are collected, they go through preprocessing to enhance their quality. This involves reducing noise (unwanted distortions) and segmenting important regions in the image to make the lung structures clearer.
- **Nodule Detection:** In this step, the system identifies and segments nodules (small abnormal growths)

present in the lung images. These nodules can be potential indicators of lung cancer.

- **Nodule Classification:** Finally, the detected nodules are analyzed and classified as either benign (non-cancerous) or malignant (cancerous) using various machine learning or deep learning techniques. This classification helps doctors make informed decisions about further diagnosis and treatment.

3. METHODOLOGY

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) framework provides a structured approach for conducting and reporting systematic reviews[2]. Below is a PRISMA-based methodology for this research paper:

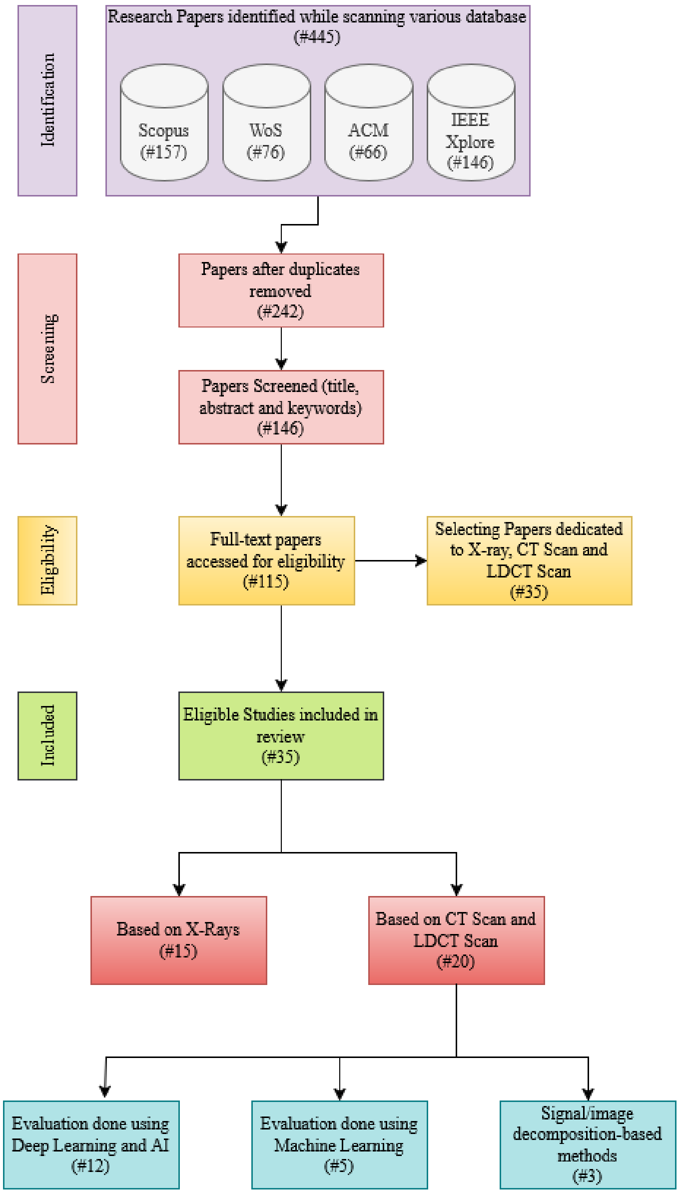

Fig. 3: An illustration of the search process with inclusion and exclusion criteria

- **Information Sources and Search Strategy:** A systematic literature search was conducted across multiple databases, like Scopus, PubMed, IEEE Xplore, Google Scholar, and Web of Science. These databases are used to identify relevant studieson early lung cancer detection using LDCT Scans. The search was restricted to peer-reviewed journal articles, conference proceedings, and systematic reviews published between the years 2023 and 2018.

- **Search Keywords:** Main keywords used for searching are: "LDCT, Lung Cancer Detection, Lung Cancer Detection with LDCT, Lung Nodule Detection, Early Lung Cancer Detection." Alternate keywords used for searching are: "Artificial Intelligence in LDCT, Machine Learning in LDCT, False Positives in LDCT, Deep Learning in LDCT, Lung

- Cancer Survival Rate, Challenges in Lung Cancer Screening." Boolean operations used for searching are: AND, OR, NOT (e.g., "LDCT" AND "Lung Cancer Detection" NOT "X-ray").

- **Eligibility Criteria:** Studies were included or excluded based on the following criteria: The Inclusion Criteria focuses on LDCT-based early lung cancer detection. Mainly selected publications have quantitative performance metrics (e.g., sensitivity, specificity, accuracy). Also, those publications are also included which discusses technological advancements, challenges, and screening guidelines. Those publications are also selected which have meta-analyses relevant to the topic. The Exclusion Criteria for Studies is the papers discussing lung cancer detection without LDCT involvement (e.g., PET, MRI, or X-ray-based methods). Also, the papers which have lacking full-text availability are excluded. The Papers Studies which focusing on non-human trials or animal models are also excluded. The manuscripts which are not available in English are also excluded. Duplicate publications are also excluded.

- **Study Selection Process:** The study selection followed a three-stage screening process. The first step is title and abstract screening. In this step two independent reviewers screened the study titles and abstracts to remove irrelevant articles. The second step is Full-Text review. In this step articles that passed the abstract screening were reviewed in full for relevance. The third and final step is Final Inclusion. In this step any discrepancies were resolved through discussion among reviewers or by consulting a third reviewer. A PRISMA flow diagram (fig. 3) summarizes the study selection process, showing the number of articles identified, screened, excluded, and included in the final review.

4. DATASETS

The below listed lung cancer datasets provide valuable resources for research in medical imaging, diagnostics, and AI-driven analysis. The National Lung Screening Trial (NLST) and the PLCO Cancer Screening Trial focus on lung cancer screening mortality data and outcomes [8]. The Cancer Imaging Archive (TCIA) provides a large collection of de-identified CT and MRI scans for research. The Cancer Genome Atlas (TCGA) offers genomic data. which includes mutations and gene expression.

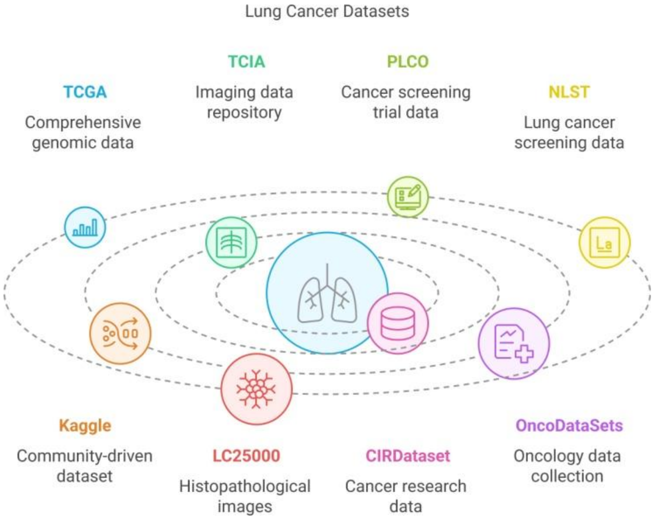

Fig.4:DatasetsofLungCancer

The LC25000 dataset consists of 25,000 histopathological images for training AI models in lung cancer detection. The Kaggle Lung Cancer Dataset contains LDCT and CT scan images for lung cancer detection and generating predictive models. CIR Dataset provides lung nodule radiomics and malignancy predictions, whereas Onco DataSets is a collection of datasets covering cancer survival, biomarkers, and genetic studies.

Table I: Lung Cancer Datasets

DatasetName	Description	AccessLink
The Cancer GenomeAtlas (TCGA)	Genomic data for lung cancer, including mutations, gene expression, and clinical data.	TCGA Dataset
CIR Dataset	Clinicallyinterpretable lung nodule radiomics and malignancy predictions.	CIR Dataset
Prostate,Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial	Extensive data on lung cancer screening, incidence, and mortality from a large-scale trial.	PLCODataset
National Lung Screening Trial (NLST)	Dataset focused on lung cancer screening outcomes and imaging results.	NLSTDataset
Lung Cancer Dataset on Kaggle	Demographic and medical data for predictive modelingin lung cancer detection.	KaggleDataset
The Cancer Imaging Archive (TCIA)	Large repository of de-identified medical images such as CT and MRI scans.	TCIADataset

Lung and Colon Cancer Histopathological Image Dataset (LC25000)	25,000 histopathological imagesfortraining image-based diagnostic models.	LC25000 Dataset
ICCRLungCancer Datasets	Standardizeddatasets for lung cancer reporting and analysis.	ICCRDataset
OncoDataSets	Collection of datasetscovering cancer survival, genetic studies, and biomarkers.	OncoDataSets
Lung Cancer CT Scan Dataset on Hugging Face	CT scan images categorized into adenocarcinoma,large cellcarcinoma, squamous cell carcinoma, and normal lung tissue.	HuggingFaceDataset

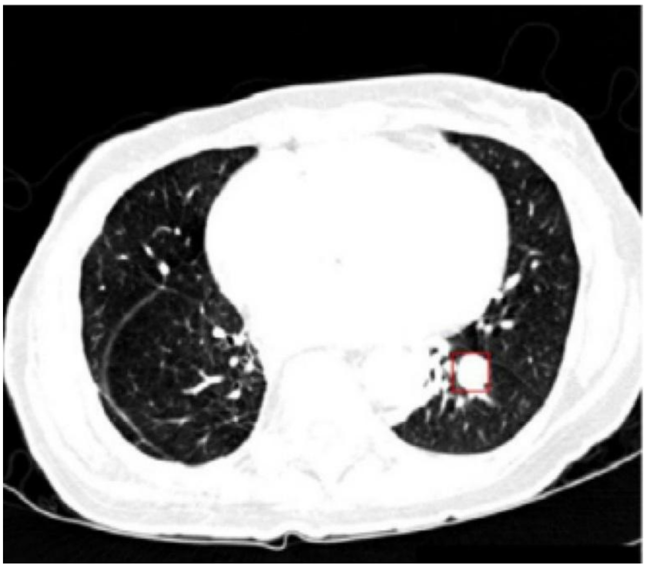

Fig.5: Image fromTCIA Dataset

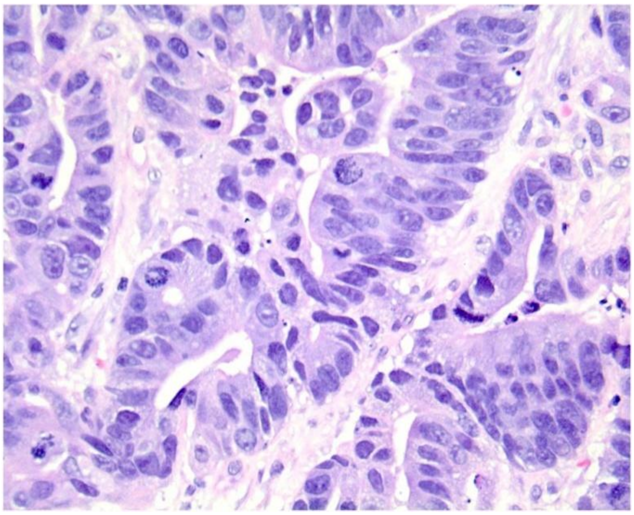

Fig. 6: An example of a histological image from the LC25,000 dataset

The Hugging Face Lung Cancer CT Scan Dataset classifies CT images into different types of lung cancer, and the ICCR Lung Cancer Datasets standardize lung cancer reporting for clinical research. These datasets collectively support advancements in AI, deep learning, and early detection methods for lung cancer diagnosis. Fig. 4 is used to describe various datasets. Table 1 represents various datasets with description and Access link. Fig. 5 represents the sample image from TCIA dataset. Fig. 6 represents an example of a histological image from the LC25,000 dataset.

5. LITERATURE REVIEW

This overview of the literature shelters light on Early detection of Lung Cancer with LDCT Scan images. It also focuses on emphasizing significant developments, difficulties, and prospects for further study. The purpose of this review is to find the responses to the following simple research questions:

- **RQ1:** Which deep learning algorithms are most frequently employed by researchers?
- **RQ2:**Whichdatasetsareopenlyaccessibleandavailabl eto researchers?
- **RQ3:** Are strategies for explainability used to make it easier to apply them in practical settings?
- **RQ4:** What kinds of difficulties did the researchers have while they were trying to identify lung cancer?

Table 1 displays the analysis of all 20 of the chosen publications that were approved for review.

Alsheikhy et al. [9] proposes a fully automated Computer-Aided Diagnosis (CAD) system. The main aim of the system is to detect and classify lung cancer in early stages. The hybrid model is trained and tested on several lung cancer datasets from LUNA16 grand challenge and Kaggle. The system integrates Long Short-Term Memory networks (LSTMs) and Deep Convolutional Neural Network (DCNN), specifically VGG-19. The proposed system achieved an average accuracy of 99.42%, which was the main purpose of the system. The recall, precision, and F-score values of the system are 99.76%, 99.88%, and 99.82%, respectively. These results suggest that the hybrid deep learning approach can effectively assist doctors in accurately identifying lung cancer.

Siddiqui et al. [10] proposes an enhanced approach for detecting and classifying lung cancer in CT scan images. The authors used Gabor filters as pre-processing techniques. To improve the accuracy of lung cancer detection system they used a Deep Belief Network (DBN) to extract meaningful features from the images. This method powers the texture analysis capabilities of Gabor filters. The deep learning strengths of DBNs to successfully distinguish between malignant and benign lung nodules. The study demonstrates that this hybrid approach achieves higher classification accuracy compared to traditional methods. Also, the authors suggested its potential as a valuable tool in computer-aided diagnosis systems for early lung cancer detection.

Shahetal.[11] presents an ensemble methodology utilizing two-dimensional(2D) CNN. The main motive of the research is to provide a better system with more accuracy. The researchers utilized the LUNA16 dataset for CT scan images. They train, test and validate their models to enhance the accuracy of the system. The investigators integrate three distinct CNN designs with varying kernels, layers, and pooling techniques. The proposed ensemble model achieved a combined accuracy of 95%, which is better than baseline methods. The authors suggested its potential as a valuable tool in computer-aided diagnosis systems for early lung cancer detection.

Ali et al. [12] presents a novel approach to lung cancer identification using deep learning techniques. The authors developed an end-to-end 2D CNN model. The main purpose of this model is to accurately distinguish between lung nodules and non-nodules. Another purpose of the system is to reduce the false positive rate. The proposed CNN architecture was trained, tested, and evaluated using the LUNA16 dataset. 80% images were utilized for training, 10% images for testing, and the remaining 10% for validation. The model was able to achieve a sensitivity of 68.66% and a specificity of 98.42%. The study employed data augmentation techniques to address the issue of data imbalance between nodule and non-nodule images. Notably, the model maintained an average false positive rate of just 1.5%, which indicates its prospective effectiveness in clinical settings for early lung cancer detection.

Lanjewar et al. [13] introduces an advanced methodology for classifying lung cancer subtypes using CT scan images. They utilized the Kaggle CT scan dataset. The authors enhanced the DenseNet201 architecture to improve feature extraction capabilities by adding multiple layers. To address the high dimensionality of features, the authors used feature selection techniques. They utilized Minimum Redundancy Maximum Relevance (MRMR) and Extra Trees Classifier (ETC) as feature selection techniques. The refined features were then classified using machine learning algorithms. They utilized Support Vector Machine (SVM), Random Forest (RF), Gaussian Naive Bayes (GNB), Decision Tree (DT), Logistic Regression (LR), and K-Nearest Neighbors (KNN). The proposed approach achieved a classification accuracy of 100% with SVM and RF classifiers. The authors demonstrated its efficacy in accurately detecting lung cancer subtypes.

Manickavasagam et al. [14] proposes a novel approach to enhance the classification accuracy of pulmonary nodules in CT scan images. CNN-5CL architecture was developed by the authors. 11 layers are proposed in the model, of which 5 are convolutional layers. The main purpose of this model was to design automatic feature extraction and classification. The proposed model utilized the LIDC-IDRI dataset. The proposed method achieved very good sensitivity of 99.62%, an accuracy of 98.88%, and specificity of 93.73%. The area under the ROC curve (AUC) of the proposed model was about 0.928. These results indicate that the CNN-5CL approach outperforms traditional methods such as K-nearest neighbor, Naïve Bayes, support vector machine, and adaptive neuro-fuzzy inference systems. This approach outperforms other deep learning-based techniques in detecting and classifying lung nodules.

Guo et al. [15] explores the application of deep neural networks to enhance the detection of small lung nodules. They utilized low-dose fluorodeoxyglucose positron emission tomography (FDG PET) imaging. The main purpose of the model is to eliminate false positives and enhance the accuracy of the system. A deep neural network-based method was developed by the authors for low-dose FDG PET images to improve nodule detection. Also, the investigators focus on reduced radiation exposure. The FDG PET images are very crucial for early lung cancer analysis. Small lung nodules were detected more specifically and sensitively by the proposed approach compared to traditional methods. Furthermore, they suggested that integrating DNNs with low-dose FDG PET imaging could be a promising strategy for effective lung cancer screening. New future directions for the detection of lung cancer in early stages are provided by this study.

Jain et al. [16] proposes an advanced method for detecting lung tumors with the help of histopathological image analysis. Authors also described machine configuration and other important information to implement deep learning models. Deep learning architecture was utilized for their purpose system. First, the authors performed preprocessing on the datasets to remove noise, artifacts, and enhance the quality of images. After that, feature extraction is performed using a combination of CNN integrated with Kernel Principal Component Analysis (KPCA). KPCA is employed to capture complex patterns in images within the feature extraction layer of the CNN. Subsequently, to differentiate between tumorous and non-tumorous lung cells, a Fast Deep Belief Neural Network (FDBNN) is used. Various histopathological image datasets were analyzed in the experimental study. This method achieves improved performance metrics like accuracy, recall, precision, and F-score as compared to existing methodologies. The confusion matrix in the study illustrates the actual versus predicted classifications of tumors in input images. The efficacy in lung tumor detection is highlighted by the proposed methodology.

A hybrid approach to improve the detection of lung nodules with the help of chest X-ray images was developed by Choudhry and Qureshi [17]. Chest X-ray images from the National Institutes of Health (NIH) dataset were utilized by the authors. This dataset containsinformationaboutmorethan30,000uniquepatients.In the purposed modal, authors utilized pre-trained VGG and Inception models for features extraction. For classification authors utilized Support Vector Machines (SVM). The proposed method achieved an Area Under the Curve (AUC) of 0.92 and an accuracy of 96.87% in detecting lung nodules. The specificity of the model is 99%. The proposed method achieved a sensitivity of 97.24%. These results suggest that combining manually extracted features with transfer learning can increase the performance of computer-aided diagnostic systems for lung nodule detection.

Xu et al. [18] introduces a novel deep learning framework to classify non-small cell lung cancer(NSCLC).The main purpose of the model was to increase the accuracy of the system and reduce the false positives. The researchers focused on main regions (segmentation) within the CT scan images. The proposed design combines attention mechanisms with CNN. The model's ability to differentiate between benign and malign ant tissues is improved by the attention mechanism. Both local and global features of NSCLC can be effectively captured by this model. Multiple datasets were cross-validated to validate the model in terms of classification accuracy and detection precision.

Wang and Charkborty [19] present a deep learning-based system for early lung cancer detection and risk assessment. The proposed framework consists of two main modules named hierarchical Recurrent Neural Networks (RNNs) and a

Nodule Detection Module. 3D-CNN models are used by the Cancer Risk Evaluation Module to evaluate nodule malignancy based on morphological characteristics. NoduleDetectionModuleconsistsofanensembleof3D-CNNs.This module is mainly used to improve the recall of pulmonary nodule identification. Multiple deep learning models are highlighted in the study as capable of improving reliability in lung cancer detection. The combination of these heterogeneous deep learning architecture enhances the accuracy and clinical interpretability of lung cancer prediction.

Rafael-Palou et al. [20] introduces a novel method for tracking pulmonary nodules across the successive CT scans without the need for image registration. The authors developed a two-stage automated system. In the first stage, the system detects nodules in individual scans. After that system generates a 3D Siamese Neural Network (SNN) to match these nodules between scans of the same patient. This approach allows for the judgement of nodule growth over time. Thesystemshowedanoduledetectionaccuracyof88.8%,and94.7 % sensitivity. These results suggest that the proposed method effectively facilitates the monitoring of pulmonary nodules.

Sorietal.[21] introduces a novel approach to improve lung cancer detection by addressing the complex morphology of pulmonary nodules. They worked on image noise as well. The authors propose a “denoising first” two-path convolutional neural network (DFD-Net). First, the model preprocesses and improves the quality of CT scan images. The primary goal of the preprocessing stage is to eliminate noise and offer a more precise examination of nodule attributes like size and shape. It then detects nodules using a residual learning denoising model (DR-Net). The two-path CNN then processes the denoised images to detect cancerous nodules. The study highlights the importance of careful analysis of suspected nodules. It is also suggested that combining information from multiple nodules can improve detection accuracy of the system.

Lizetal.[22]discussed that the Pneumonia in children under five is discussed as a crucial problem, accounting for about 15% of all pediatric deaths globally. The main purpose of the authors was to generate a model which will reduce mortality. The authors developed ensemble models of CNN to analyze chest X-ray images for pneumonia detection. The study introduces a novel called Explainable Artificial Intelligence (XAI) procedure to enhance the interpretability of the system. This technique combines individual heat maps from each model in the ensemble. The model was trained, tested and validated on a dataset (950 pediatric X-rays). This approach highlights the most relevant are as of the images contributing to the classification decision. The ensemble models outperformed as compared to other models and achieved highly competitive results. The robustness of the proposed method was further cross validated on an additional dataset. Elnakib et al. [23] proposes a CAD system aimed at early detection of lung nodules from LDCT images. The system enhances the quality of image with preprocessing techniques. After that extracts features using various deep learning architectures, like VGG16, VGG19, and AlexNet.

Table II: Literature Review

References	DataSet	Specificity (%)	Sensitivity (%)	Modality used	Results	Shortcomings	Future Scope
Alsheikhy etal.[9] 2023	Kaggle, LUNA16	98.42	99.76	DeepLearning, Hybrid CNN	Accuracy: 99.42%	High computational cost due to deep learning models, requires large dataset for training	Integration with cloud-based medical systems for scalable lung cancer detection
Siddiquiet al. [10] 2023	CT Scans	N/A	N/A	DBN+Gabor Filters	Improved classification accuracy	Feature selection complexity, requires extensive preprocessing for optimal classification	Integrating explainability models to improve confidence in lung cancer classification
Shah etal.[11] 2023	LUNA16	N/A	N/A	Ensemble 2D CNN	Improved lung cancer detection	Higher complexityof ensemble CNNs, trade-offbetween accuracy and computational efficiency	Extending to 3D models for improved spatial feature extraction in lung nodule detection
Alietal. [12]2023	LUNA16	98.42	68.66	2DCNN	Improved false-positive reduction	Lowersensitivityfor smaller nodules, difficultyindetecting early-stage lung cancer	Improving classification robustness, refining CNN layers for enhanced sensitivity
Lanjewar et al. [13] 2023	Kaggle CTScan dataset	100	100	DenseNet + MLClassifiers	100%accuracy with SVM, RF	Potential datasetbias, feature selection methodology may limit generalizability	Generalizing to additional lung cancer datasets, refining feature selection techniques

Manickavas et al. [14] 2022	gLaImDC/IDR	93.73	99.62	CNN-5CL Model	Accuracy: 98.88%	Computationally expensive,model complexitylimitsits application inreal-time scenarios	Deploying on embeddedAIsystems for real-time lung cancer screening applications
Guoetal. [15]2022	FDG PET dataset	N/A	N/A	Deep Neural Networks	Enhancedsmall lesion detection	Limited generalizability, lacks validation on diverse datasets, high dependence on FDG PET imaging	Adapting the model to different imaging modalities, enhancing diagnostic applicability
Jainetal. [16]2022	Histopathol dataset	ogNi/cAal	N/A	KPCA+CNN +FDBNN	Higher classification accuracy	Computational overhead due to kernel PCA and deep belief networks, training time is high	Optimizing computational efficiency, making deeplearningmodels more practical in clinics
Choudhry and Qureshi [17]2022	NIH Chest X-ray dataset	99	97.24	Transfer Learning + SVM	96.87% accuracy,AUC: 0.92	High dependency on manually extracted features, limited adaptability to automated pipelines	Integrating with CAD systems to streamline automated lungcancerscreening in hospitals
Xuetal. [18]2022	Non-Small Cell Lung Cancer dataset	N/A	N/A	CNN + Attention Mechanism	Higher classification accuracy	Need for morediverse datasets, sensitivity to variations in tumor morphology	Hybridizing with biomarker-based diagnostictechniques forimprovedlung cancer detection
Wang and Charkborty [19]2021	LDCT Scans	N/A	N/A	3DCNN,RNN	Enhanced detection and classification	Lack of real-time adaptability,sensitivity to hyperparameter tuning,hightraining cost	Enhancing real-time detectioncapabilities, reducing dependency on manual hyperparameter tuning
Rafael-Palou et al. [20] 2021	CT images dataset	88.8	92.0	3D Siamese Neural Networks	Noduledetection sensitivity 92%	Noisy data affects accuracy, struggles with small nodule detection in early stages	Improving noise handling and small noduledetection through adaptive feature extraction
Sorietal. [21]2021	CT Scans	N/A	N/A	Denoising + CNN	Improved false-positive reduction	Sensitive to input noise variations, performancefluctuates with denoising effectiveness	Optimizingdenoising networks, exploring hybrid techniques to enhance robustness
Lizetal. [22]2021	Pediatric X-ray dataset	N/A	N/A	Ensemble CNN	Improved pediatric pneumonia diagnosis	Requires large datasetsforrobust performance,model interpretability remains a challenge	Enhancing clinical applicability, training on larger datasetswith diverse patient samples
Elnakib etal.[23] 2020	I-ELCAP	95	97.5	CNN,Genetic Algorithm	Accuracy: 96.25%	Limited dataset size, potential overfitting, lacks generalizationtodiver sepopulations	Expanding dataset diversity, incorporating multi-center studies toimproverobustness

Zhangand Kong [24]2020	Multi-scene dataset	N/A	N/A	Multi-scene DeepLearning	High detection rate	Requires multiple scene inputs, challengingintegration into real-time medical systems	Applying the approach to real-world hospital systems, improving real-timedeployment
Sajja etal.[25] 2019	LIDC Dataset	N/A	N/A	Transfer Learning	Higher accuracy than baseline CNNs	Potential overfitting due to transfer learning, struggles with feature extraction from CT scans	Expanding to other datasets to validate modelgeneralization, reducing overfitting risks
Xiaoet al. [26]2019	LUNA16	91.7	91.7	Multi-scale3D CNN	CPM: 0.874, Sensitivity 91.7%	Limited interpretability, multi-scale CNN increases computationalburden and memory usage	Applying the framework to other medicalimaging fields,improving general diagnostic capabilities
Eunetal. [27]2018	LUNA16	99	97.24	Single-view2D CNN	AUC:0.922	Lower sensitivity in differentiating non-nodules,increased false positive reduction efforts	Improving interpretability using explainable AI techniques, reducing false positives
Jinetal. [28]2018	LUNA16	N/A	N/A	3D Residual CNN	High detection performance in LUNA16	High computational requirement,increased processing timefor3Dresidual CNN training	Enhancing generalizability through transfer learning on larger multi-sourcedatasets

The most pertinent attributes are then chosen via a genetic algorithm to maximize these traits, and they are subsequently categorized using various classifiers. Researchers used 320 LDCT images from I-ELCAP Dataset to validate the model. The model achievedthehighestdetectionaccuracyof96.25%,withaspecificity of 95% and sensitivity of 97.5% by using VGG19 architecture and a SVM classifier. These results suggest that the proposed system holds promise for improving early lung cancer detection.

Zhang and Kong [24] proposes an efficient identification system for lung nodules using a Multi-Scene Deep Learning Framework (MSDLF). For pre-processing system uses a vesselness filter. The researchers focused on main regions (segmentation) within the CT scan images. The purposed model combines two image scenes into a four-channel CNN model. The purposed model utilized the LIDC/IDRI dataset and cross validated on another dataset also. This approach aims to improve the accuracy of detecting four-stage nodules by leveraging the radiologist's knowledge through the integration of multiple image scenes.

Sajjaetal.[25] proposes a method for classifying lung tissues as malignant or benign using a deep neural network based on GoogleNet. It is a pre-trained model based on CNN. The authors sparsified the densely connected architecture by deploying 60% of all neurons on dropout layers. The main purpose of this is to reduce computing costs and prevent overfitting. The performance of this modified network was evaluated using the LIDC dataset. Also, the model was cross validated with other pre-trained CNN Models like ResNet50, AlexNet, and GoogleNet. The results demonstrated that the proposed network achieved better classification accuracy than the other compared networks.

Xiao et al. [26] introduces a novel approach to enhance the accuracy of pulmonary nodule detection in CT scan images. The main purpose of the proposed model is to reduce false positives. The authors used a Multi-Scale Heterogeneous (MSH) CNN (Three-Dimensional) model. The suggested model used two distinct CNN branches (3-D) to extract different feature representations. In other words, authors claim that the model is able to capture heterogeneous features. Additionally, a weighted feature fusion strategy is used to increase the model's judgement ability. The model was applied on the LUNA16 dataset. Two false positives per scan resulted in a sensitivity of 91.7% and an average CPM score of 0.874. The proposed system significantly reduced false positives compared to traditional models. For future direction, it is recommended that MSH-CNN has strong potential for improving the reliability of CAD systems in lung cancer detection.

Eun et al. [27] proposes an innovative Deep Learning based framework to detect pulmonary nodule with the help of CT scan images. The system was enhanced by the researchers to improve its accuracy and efficiency. They tried to address challenges such as the imbalance between non-nodules and nodules. An ensemble of single view2DCNNmodelswasusedtocoverthediverseappearances of non-nodules. Computational power is improved by this approach ascomparedtotraditional3DCNNmethods.ThepurposedModel is also memory efficient. Non-nodules are automatically classified using k-means clustering after characteristics are retrieved by an auto encoder. To improve the model's capacity to differentiate between different types of non-nodules and nodules, each 2D CNN is trained on the same nodule data but distinct non-nodule categories. The proposed framework evaluated on the LUNA16 dataset and achieved a competition performance metric score of 0.922. The model was also able to reduce false positives in pulmonary nodule detection.

Jin et al. [28] proposes an advanced method to enhance the accuracy of pulmonary nodule detection in CT scan images. The authors developed a deep 3D residual CNN model. The main purpose of the model was to reduce false positives and improve efficiency. The purposed framework incorporates a cropping (SPC) layer and spatial pooling layer to extract multilevel contextual information. Additionally, to enhance the network's performance on difficult samples, like nodules with irregular forms, an online hard sample selection technique was used during training. The proposed framework was evaluated on LUNA16 Challenge dataset (888 CT scans). The proposed method demonstrated high detection performance and reduced false positives. The designed model indicatesitsrobustnessandpotentialapplicabilitytoother3Dobjec t detection tasks in medical image processing.

6. EXPLAINABILITY IN LUNG CANCER DETECTION

When combining AI-driven investigative tools into clinical processes, explain ability in lung cancer detection is very essential. Inthesedays,3DneuralnetworksandCNNarefrequentlyused for lung nodule detection. But due to the black-box nature of deep learning models, gaining clinician trust is very difficult. To address these kinds of issues explainable AI(XAI) techniques are used. SHAP (Shapley Additive Explanations) and Grad-CAM (Gradient-weighted Class Activation Mapping) are some examples of XAI techniques. Generally, these techniques can highlight key regions in CT scans and X-rays that influence model decisions. Furthermore, attention mechanisms in deep learning models are used to increase transparency. Generally, Nodule classification techniques are promoted by feature visualization techniques. Examples of these techniques are Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE). These techniques have ability to translate high-dimensional medical images into understandable representations. Rule-based models and decision trees are combined with deep learning to provide a methodical justification for classifications. This will improve significantly interpretability of the system.

However, there are many challenges like data bias due to imbalanced datasets, interpretability trade-offs, complexity and lack of standardization in XAI frameworks for medical imaging. Some other challenges are like computational overhead. Some explainability techniques may increase processing time, which makes real-time deployment very difficult. Future research in this area focuses on hybrid explainable models. These models may combine deep learning with domain knowledge for decision making. Further, AI decisions should be validated before final diagnosis. The main goal of XAI is to make AI reliable, transparent, and clinically acceptable in lung cancer detection. Finally, the automated systems can support radiologists in early and accurate lung cancer diagnosis.

7. DISCUSSION AND CHALLENGES

LDCTscanhasmodernizedearlylungcancerdetection.Itwillimpr ovethesurvivalratesofthepatientsbysupporting the identification of small pulmonary nodules at an early stage. LDCT-based screening has been shown to reduce lung cancer mortality among high-risk persons by 20–26% in several clinical trials, such as the NLST and NELSON trial. The combination of AI and deep learning models into LDCT imaging has further enhanced diagnostic accuracy. This will further reduce false positives and also improve nodule classification. The combination of image-based analysis with clinical and genetic biomarkers is also an emerging area. It will further provide a multimodal approach for lung cancer detection. AI-powered CAD systems, powered by CNN, radiomics-based AI models, and deep neural networks, can differentiate between malignant and benign lung nodules with higher precision than traditional radiologists.

However, even with these improvements, several challenges remain that limit the extensive clinical acceptance of LDCT for lung cancer screening. These challenges mostly relate to radiation exposure, false positives, cost, accessibility, and standardization of AI-based diagnostic models. Fig. 7 insights the several challenges, which are generally faced by researchers in early detection of lung cancer.

Despite the significant advancements in LDCT for early lung cancer detection there are various challenges remain that hinder its wide spread clinical adoption. The high false-positive rate, in which many lung nodules that are discovered are benign, is one of the most urgent problems. Studies indicate that more than 20% of nodules detected in LDCT screenings require further testing but do not develop into cancer. To address this concern, AI-driven false-positive reduction techniques are developed. These techniques perform well for lung nodule classification. Another major challenge is radiation exposure which is already handled by LDCT Scans. The radiation exposure from an LDCT scan is substantially lower than that of a standard CT scan. Researchers are exploring low-dose AI modelsand image reconstruction algorithms to maintain image quality while reducing radiation exposure further.

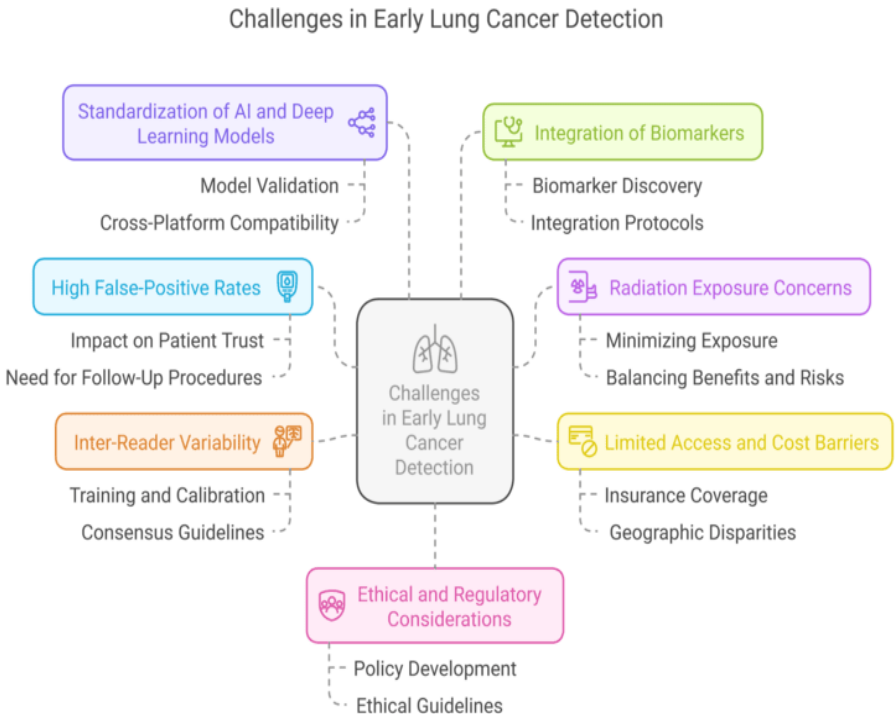

Fig.7:Challenges in Early Lung Cancer Detection

High prices and restricted access are other obstacles to the adoption of LDCT, especially in low-resource environments where many hospitals lack LDCT equipment and AI-based diagnostic tools. Additionally, insurance coverage and screening guidelines vary across different regions. This will impact the affordability of lung cancer screening programs. Subjective differences in classifying nodules can lead to inconsistent diagnoses. AI-assisted decision support systems offer potential solutions by providing more consistent assessments. But due to the lack of standardization, AI models remain hampered in their clinical acceptance. AI models trained on heterogeneous, multi-institutional datasets could help generalize performance and improve reliability across hospitals.

Furthermore, the regularization of deep learning and AI models remains an issue. Most of the AI tools are trained on differentdatasetswithvaryingimagingprotocols.Thismaycausei nconsistent results. To solve this issue, federated learning techniques are being investigated. In which AI models are taught across several hospitals while protecting patient privacy. Additionally, the recent studies suggest that combining AI-driven imaging analysis with biomarker-based screening techniques will improve diagnostic accuracy. Examples of biomarker-based screening are genomic profiling and liquid biopsy. However, due to variations in data formats,processingpipelines,andregulatoryissues,integratingth ese multimodal techniques continues to be difficult.

Finally, regulatory and ethical considerations play a significant role in the deployment of AI-driven LDCT screening. Regulatory agencies (FDA and EMA) require AI-based medical tools to demonstrate transparency, clinical reliability, and explainability before approval. In ethical concerns, we consider bias in AI models, data privacy, and liability issues. In this regard, explainable AI techniques like SHAP-based and Grad-CAM feature visualization are being developed. The purpose of these techniques is to improve trust and transparency in AI-driven lung cancer detection. Addressing these challenges through standardized AI protocols, federated learning, multimodal integration, and regulatory compliance will be crucial.

8. CONCLUSION

Despite the significant progress in LDCT scan-based lung cancer detection. But there are many technical, ethical, and clinical challenges that need to be addressed to ensure broader clinical acceptance. The efforts in this field should focus on reducing false positives, improving accessibility and optimizing AI-driven diagnostics. Furthermore, combining biomarkers with LDCT Scan images to produce better results. Researchers should focus on compliance of AI based detection system. Future research should focus on AI based model generalization which is accepted by all radiologists. Furthermore, radiologists are able to validate AI based generated detection to enhance reliability and trust. By overcoming these obstacles, AI-assisted screening and LDCT will be able to establish themselves as the gold standard for early lung cancer diagnosis.

9. REFERENCES

[1] S. Xun, D. Li, H. Zhu, M. Chen, J. Wang, J. Li, M. Chen, B. Wu, H. Zhang, X. Chai et al.,"Generative adversarial networks in medical image segmentation: A review," Computers in Biology and Medicine, vol. 140, p. 105063, 2022.

[2] G. Thakral and S. Gambhir,"Early detection of lung cancer with low-dose CT scan using artificial intelligence: A comprehensive survey," SN Computer Science, vol. 5, no. 5, p. 441, 2024.

[3] N. Borodinov, W.-Y. Tsai, V. V. Korolkov, N. Balke, S. V. Kalinin, and O. S. Ovchinnikova,"Machine learning-based multidomain processing for texture-based image segmentation and analysis," Applied Physics Letters, vol. 116, no. 4, 2020.

[4] G. Thakral, S. Gambhir, and N. Aneja,"Proposed methodology for early detection of lung cancer with low-dose CT scan using machine learning," in 2022

International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), vol. 1. IEEE, 2022, pp. 662–666.

[5] R. M. Devarapalli, H. K. Kalluri, and V. Dondeti,“Lung cancer detection of CT lung images,” International Journal of Recent Technology and Engineering, vol. 7, no. 5S4, pp. 413–416, 2019.

[6] Thakral, Gagan, Umesh Kumar, and Sapna Gambhir,“Robust pre-processing strategies for early lung cancer diagnosis with low-dose CT scans,” in 2025 2nd International Conference on Computational Intelligence, Communication Technology and Networking (CICTN), pp. 305–311. IEEE, 2025.

[7] G. Thakral, U. Kumar, and S. Gambhir,“Implementation of deep learning-based segmentation technique on LDCT scan images for detection of lung cancer in early stages,” in 2024 International Conference on Computing, Sciences and Communications (ICCSC). IEEE, 2024, pp. 1–6.

[8] S.Cui, S.Ming, Y.Lin, F.Chen, Q.Shen, H.Li,G. Chen,X. Gong, and H. Wang, “Development and clinical application of deep learning model for lung nodules screening on ctimages,” Scientific reports, vol. 10, no. 1, pp. 1–10, 2020.

[9] A. A. Alsheikhy, Y. Said, T. Shawly, A. K. Alzahrani, and H. Lahza,“A CAD system for lung cancer detection using hybrid deep learning techniques,” Diagnostics, vol. 13, no. 6, p. 1174, 2023. [Online]. Available: https://doi.org/10.3390/diagnostics13061174

[10] E. A. Siddiqui, V. Chaurasia, and M. Shandilya,“Detection and classification of lung cancer computed tomography images using a novel improved deep belief network with Gabor filters,” Chemometrics and Intelligent Laboratory Systems, vol. 235, p. 104763, 2023. [Online]. Available: https://doi.org/10.1016/j.chemolab.2023.104763

[11] A. A. Shah, H. A. M. Malik, A. Muhammad, A. Alourani, and Z. A. Butt,“Deep learning ensemble 2D CNN approach towards the detection of lung cancer,” Scientific Reports, vol. 13, no. 1, p. 2987, 2023. [Online]. Available: https://doi.org/10.1038/s41598-023-29656-z

[12] S. Ali, S. Asad, Z. Asghar, A. Ali, and D. Kim,“End-to-end 2D convolutional neural network architecture for lung nodule identification and abnormal detection in cloud,” Computers, Materials & Continua, vol. 75, no. 1, pp. 461–475, 2023. [Online]. Available: https://doi.org/10.32604/cmc.2023.035672

[13] M. G. Lanjewar, K. G. Panchbhai, and P. Charanarur,“Lung cancer detection from CT scans using modified DenseNet with feature selection methods and ML classifiers,” Expert Systems with Applications, vol. 224, p. 119961, 2023. [Online]. Available: https://doi.org/10.1016/j.eswa.2023.119961

[14] R. Manickavasagam, S. Selvan, and M. Selvan,“CAD system for lung nodule detection using deep learning with CNN,” Medical & Biological Engineering & Computing, vol. 60, no. 1, pp. 221–228, 2022. [Online]. Available: https://doi.org/10.1007/s11517-021-02462-3

[15] H. Guo, J. Wu, Z. Xie, I. W. K. Tham, L. Zhou, and J. Yan,“Investigation of small lung lesion detection for lung cancer screening in low-dose FDG PET imaging by deep neural networks,” Frontiers in Public Health, vol. 10, p. 1047714, 2022. [Online]. Available: https://doi.org/10.3389/fpubh.2022.1047714

[16] D. K. Jain, K. M. Lakshmi, K. P. Varma, M. Ramachandran, and S. Bharati,“Lung cancer detection based on kernel PCA-convolution neural network feature extraction and classification by fast deep belief neural network in disease management using multimedia data sources,” Computational Intelligence and Neuroscience, vol. 2022, p. 3149406, 2022. [Online]. Available: https://doi.org/10.1155/2022/3149406

[17] I. A. Choudhry and A. N. Qureshi,“Detection of lung nodules on X-ray using transfer learning and manual features,” Computers, Materials & Continua, vol. 72, no. 1, pp. 1445–1463, 2022. [Online]. Available: https://doi.org/10.32604/cmc.2022.025208

[18] Z. Xu, H. Ren, W. Zhou, and Z. Liu,“ISANet: Non-small cell lung cancer classification and detection based on CNN and attention mechanism,” Biomedical Signal Processing and Control, vol. 77, p. 103773, 2022. [Online]. Available: https://doi.org/10.1016/j.bspc.2022.103773

[19] W. Wang and G. Charkborty,“Automatic prognosis of lung cancer using heterogeneous deep learning models for nodule detection and eliciting its morphological features,” Applied Intelligence, vol. 51, no. 4, pp. 2471–2484, 2021. [Online]. Available: https://doi.org/10.1007/s10489-020-01990-z

[20] X. Rafael-Palou, A. Aubanell, I. Bonavita, M. Ceresa, G. Piella, V. Ribas et al.,“Re-identification and growth detection of pulmonary nodules without image registration using 3D Siamese neural networks,” Medical Image Analysis, vol. 67, p. 101823, 2021. [Online]. Available: https://doi.org/10.1016/j.media.2020.101823

[21] W. J. Sori, J. Feng, A. W. Godana, S. Liu, and D. J. Gelmecha,“DFD-Net: Lung cancer detection from denoised CT scan image using deep learning,” Frontiers of Computer Science, vol. 15, no. 2, p. 152701, 2021. [Online]. Available: https://doi.org/10.1007/s11704-020-9050-z

[22] H. Liz, M. Sánchez-Montañés, A. Tagarro, S. Domínguez-Rodríguez, R. Dagan, and D. Camacho,“Ensembles of convolutional neural network models for pediatric pneumonia diagnosis,” Future Generation Computer Systems, vol. 122, pp. 220–233, 2021. [Online]. Available: https://doi.org/10.1016/j.future.2021.03.001

[23] A. Elnakib, H. M. Amer, and F. E. Z. Abou-Chadi,“Early lung cancer detection using deep learning optimization,” International Journal of Online and Biomedical Engineering (iJOE), vol. 16, no. 6, pp. 82–96, 2020. [Online]. Available: https://doi.org/10.3991/ijoe.v16i06.13657

[24] Q. Zhang and X. Kong,“Design of automatic lung nodule detection system based on multi-scene deep learning framework,” IEEE Access, vol. 8, pp. 90380–90389, 2020. [Online]. Available: https://doi.org/10.1109/ACCESS.2020.2993872

[25] T. Sajja, R. Devarapalli, and H. Kalluri,“Lung cancer detection based on CT scan images by using deep

transfer learning," Traitement du Signal, vol. 36, no. 4, pp. 339–344, 2019. [Online]. Available: https://doi.org/10.18280/ts.360406

[26] Z. Xiao, N. Du, L. Geng, F. Zhang, J. Wu, and Y. Liu,"Multi-scale heterogeneous 3D CNN for false-positive reduction in pulmonary nodule detection, based on chest CT images," Applied Sciences, vol. 9, no. 16, p. 3261, 2019. [Online]. Available: https://doi.org/10.3390/app9163261

[27] H. Eun, D. Kim, C. Jung, and C. Kim,"Single-view 2D CNNs with fully automatic non-nodule categorization for false-positive reduction in pulmonary nodule detection," Computer Methods and Programs in Biomedicine, vol. 165, pp. 215–224, 2018. [Online]. Available: https://doi.org/10.1016/j.cmpb.2018.08.012.

[28] H. Jin, Z. Li, R. Tong, and L. Lin,"A deep 3D residual CNN for false-positive reduction in pulmonary nodule detection," Medical Physics, vol. 45, no. 5, pp. 2097–2107, 2018. [Online]. Available: https://doi.org/10.1002/mp.12846