International Journal of Computing and Digital Systems
2025, VOL. 17, NO. 1, 1–12

http://dx.doi.org/10.12785/ijcds/1571032811

A Systematic Review on Effectiveness and Contributions of Machine Learning and Deep Learning Methods in Lung Cancer Diagnosis and Classifications

Jayapradha Jayaram[1,2], Su-Cheng Haw[2,*], Naveen Palanichamy[2], Elham Anaam[2] and Senthil Kumar Thillaigovindhan[1]

[1]*Department of Computing Technologies, School of Computing,SRM Institute of Science and Technology,Kattankulathur, Tamil Nadu 603203, India*

[2]*Faculty of Computing and Informatics, Multimedia University, Jalan Multimedia, 63100, Cyberjaya, Selangor.*

Received 24 May 2024, Revised 26 October 2024, Accepted 28 November 2024

Abstract: In the present scenario of people's health, lung tumour is the major reason of cancer-related losses, and its death rates are steadily rising. Several researchers have developed automated techniques for quickly and accurately predicting the development of cancer cells using medical imaging and machine learning techniques. As advances in computerized systems have been made, deep learning techniques have been thoroughly investigated to aid in understanding the results of computer-aided diagnosis (CADx) and computer-aided detection (CADe) in computed tomography (CT), magnetic resonance imaging (MRI), and X-ray for the identification of lung cancer. To provide a thorough review of the deep learning methods created for lung cancer diagnosis and detection, this study is being done. The present study offers an articulation of deep learning and machine learning methods and aims for applications in the diagnosis of lung cancer as well as an analysis of the advancements made in the techniques explored. In order to detect and screen for lung cancer, two main deep learning techniques are used in this study: analysis and categorization across the internal and external organizational environment and its markets. The benefits and drawbacks of the deep learning models that are currently in use will also be covered. Deep Learning technologies can deliver accurate and efficient computer-assisted lung tumor detection and diagnosis, as shown by the subsequent analysis of the scan data. This study ends with a description of potential future studies that might enhance the use of deep learning to the creation of computer-assisted lung cancer detection systems.

Keywords: Lung Cancer, Diagnosis, Classification, Medical Imaging, Machine Learning, Deep Learning and Analysis.

1. INTRODUCTION

Yearly, more than 2 million human beings are affected by lung tumour [1], and nearly 74% of those patients died because of illness within a short span of time [2]. Because of high intra-tumor heterogeneity (ITH), treatment for cancer becomes more difficult and intelligent cancer cells that develop drug defiance [3]. In the last several decades, significant technical improvements in cancer research have made feasible many large-scale joint endeavours. Various medical, therapeutic imagery, and sequencing archives have been developed by these systems [4]. These databases make it easier for researchers to look at the patterns of lung cancer, including how the disease is diagnosed, treated, and responds to clinical results [5]. Specifically, recent research on analysis, including transcriptomics, proteomics, metabolomics, and genomes, has increased the research tools and capacities. Clinical tasks involving diverse and high-dimensional types of information still require a great deal of time and expertise, even with the help of dimension reduction techniques like matrix and tensor factorizations, and researchers face a great deal of difficulty in analyzing the rapidly growing databases associated with cancer [6].

The earlier cancer diagnosis with higher accuracy can direct to appropriate and effective treatment, which may also increase human life [7]. Irrespective of the medical elements, medical specialists are needed to evaluate the medical data about the disease diagnosis. It is more common for medical experts make have dispersed opinions because of the intricacy of the medical data. Hence, there is a need for an intellectual and automated diagnosis model in the field of health or clinical diagnosis. In recent times, the emergence of Artificial Intelligence (AI) algorithms has helped in evaluating and interpreting medical images to

E-mail address: jayapraj@srmist.edu.in, sucheng@mmu.edu.my, p.naveen@mmu.edu.my, anaamelham@gmail.com, senthilt2@srmist.edu.in

accurately diagnose diseases. The emergence of computer-assisted diagnosis models in the field of disease analysis is exceedingly challenging, and more research is needed to process many kinds of complicated disease diagnoses. The recent growth of DL technology has used computer-aided models for automated analysis of the visual features of diseases, which causes the effective employment of several medical image processing models. The two main types of lung cancer are given as follows,

i Small Cell Lung Cancer (SCLC)

ii Non-Small Cell Lung Cancer (NSCLC)

This work intends to comprehensively analyse the AI methodologies employed to treat lung cancer diagnosis and classification. Furthermore, the paper compares the effectiveness of many diagnosis models using dataset images. This paper also emphasizes the requirement for reliable and effective models for CT-oriented automated diagnostic models. An interesting solution for enhancing the accuracy of lung cancer diagnosis in an automated and precise manner is also discussed in this work. As mentioned, the benefits and limitations of several existing models are presented in this paper, including recommendations for further research.

The significant paper impacts are stated as results.

i The paper presents a clear analysis of how ML and DL are used to diagnose lung cancer using medical data.

ii Briefing the methodologies for lung tumor diagnosis and classification

iii Comparative analysis of the various learning model's results and effectiveness.

iv Summarizing the limitations of the current models and suggesting some potential ideas for enhancing the detection accuracy.

The article's structure is given as follows: Section 2 describes and compares the various ML and DL models for lung cancer diagnosis, processing the different medical data inputs. Section 3 explains the available datasets for disease diagnosis. The shortcomings of the existing disease diagnosis models are discussed in Section 4. The assumption and ideas for upcoming enhancements are set in Section 5

2. MACHINE LEARNING CONTRIBUTIONS ON EARLIER DISEASE DIAGNOSIS USING MEDICAL DATA:

Earlier detection of cancer disease is a significant process for dipping the demise rate of people due to cancer. For evaluations, low dose computed tomography (CT) images are used. Moreover, CAD-based diagnosis models are modelled to improve the results of diagnosis that help medical specialists in medical data processing and analysis. The model is reflected as an important tool for medical practitioners to discuss and provide second opinions that may change the patient's life.

The traditional computed aided model based on the extracted image features is given in three levels,

i Nodule Segmentations

ii Feature Extraction and Selections

iii Classification

A method that uses the textural features of some specific nodules from the image inputs, combined with the patients' medical history, is used for training the ML model. The classifier can be a linear discriminant analysis (LDA) and logistic regression (LR), which are used for detecting the malignancy probability. The nodule Size, nodule type, position, count, and border are some of the nodule measures provided by CT scans; however, they are sometimes subjective and may not provide a thorough enough image to identify malignant nodules. Researchers are investigating methods to use DL techniques, like convolutional neural networks (CNNs), to decrease the frequency of false positives and the running time of CAD systems, while simultaneously increasing the accuracy of lung tumour identification. These models usually have three stages: detecting and segmenting nodules, extracting features from nodules, and inferring clinical judgment [8]. Different from attributes-based CAD systems, DL-based CAD systems may properly describe the 3D structure of a suspicious nodule by autonomously recovering and extracting intrinsic properties [9]. To extract three 2D-view feature vectors of the nodule from CT scans. The authors in [10] used OverFeat to build a model [11]. All nodules detected by CT scans may now be thoroughly investigated with the help of the newly integrated CNN models. The authors [12] developed a complementary CNN framework in which nodule segmentation is employed using a spherical harmonic process [13] and feature extraction along with texture differentiation is done with a Deep Convolution Neural Network (DCNN) [14]. Further, the classification operations are processed with downstream classifications established on the shape and appearance of the medical image. Similarly, the work in [15] utilized ensemble-based classification for disease diagnosis. The models 2D-ResNet50 [16] and 3D-Inception-V1 [17] are involved in the process of classification that performs feature extraction of pulmonary nodules and integrates those features to provide input for the classifications. Considerably, some research works utilized CNN models to make final clinical decisions, which is crucial for the treatment process. An end-to-end disease diagnosis model has been discussed in [18], which performs localization and risk classification of lung cancer. The model combined the following Convolutional Neural Networks models,

i Mask-RCNN [19] for lung nodule segmentation.

ii Modified RetinaNet [20] for RoI detection iii 3D inflated Inception- V1 [21], defines the risk rate based on malignancy.

On the other hand, like CT images, the NN models are commonly utilized for lung cancer diagnosis using histological images as inputs. The histological images provide more data about the cellular levels of cancer tissues than the CT images. Concerning that, the author in [22] employed Micro-Net [23] for detecting the cancer tissue contours and the classifications are done with SC-CNN [24].

Another study applied the Inception-V3 model to segmenting whole-slide H&E-stained histopathological images into normal and malignant tissues of LUAD and LUSC [25]. A good thing for this paper is the model can detect somatic mutations across many tissues in lung tumour driver genes such as TP53, SETBP1, FAT1, EGFR, KRAS, andSTK11. It is significant to note that several researchers have used transfer learning to improve the efficacy and robustness of their training of new models due to the dataset's complex structure and substantial resource requirements. Although ML techniques are already widely used in computer-aided design (CAD), the primary barrier is the limited number of annotated images. Overfitting may happen when a complex CNN model is trained using a few training datasets.

GAN (Generative adversarial network) systems are used recently to generate fake images to improve discriminative classifier performance [26]. Artificial lung nodule CT images were originally generated using a deep convolutional GAN (DCGAN) framework [27]. GAN models have been combined with other CNN models in recent research to tackle the challenge of overfitting in the category of lung cancer. The authors [28] used a two-step method: an AlexNet was used to identify lung cancer using both actual and fake datasets, and a DCGAN was utilized to create artificial images of the disease. In the work [29] comparable investigations were carried out. DCGAN was also used for data augmentation. Further, the authors have created the VGG-DF transfer learning model, which is regularization-enhanced, to improve performance to tackle the overfitting problem that occurs when using pre-trained models.

A. Lung Cancer Diagnosis with Medical Imaging:

CT scans use the basic principle of X-ray absorption to image internal organs. A computer's analysis of the X-ray attenuation data results in cross-sectional body images, which may be further processed to create three-dimensional views. When assessed to further medical imagery methods, CT imaging offers several benefits. One of them is its non-invasiveness since it doesn't involve any invasive procedures or incisions. CT scanning is a quick procedure as it requires minimal time for processing. Additionally, a CT scan provides healthcare professionals with highly detailed images from within the body, uncovering abnormalities that may go unnoticed with traditional imaging techniques. CT imaging does have some drawbacks, however, such the use of ionizing emission, which may ultimately raise the risk of cancer. As a result, it needs sufficient radiation shielding and ought to be used sparingly. Furthermore, anomalies seen in a condition known as a syndrome may not be abnormal, and good results from CT scans may still occur. This might result in pointless, costly, and perhaps dangerous medical tests and procedures. DL techniques have enhanced the detection of lung cancer through CT imaging. DL utilizes multi-layered networks to uncover information from vast quantities of data.

Established ML algorithms could not understand the connections and the patterns in the data that these networks can identify. However, the DL algorithms can be trained to automatically identify the nodules and injuries from the CT scan images. Dl algorithms also help to categorize the images into various regions such as liver, lung and brain. Such categorization helps the medical practitioners to find the irregularities and defects in various parts of the body. As DL algorithms provides several advantages of studying CT scan images, it may be used to train huge volume of data, able to cope up with various input data sources, such as CT scan images with several contrast and brightness and noise. Furthermore, DL algorithms can infer the images accurately which improves the effectiveness of the diagnosis process. It also improves the correctness and consistency of the image analysis, thus the probability of erroneous results and the false positive and negative are reduced.

Lot of challenges are faced while processing the CT scan images using DL algorithms. Due to non-availability of vast, high quality CT scan images dataset, training the DL algorithms become a major challenge. To train the DL algorithms, the datasets need to be properly labelled such that the algorithms can recognize the characteristics of the dataset properly.

Comprehending DL algorithms can be difficult due of their complicated nature. To address the challenges associated with using DL techniques in clinical settings and to fully evaluate the effectiveness and safety of these approaches for CT image interpretation, a thorough evaluation is necessary.

Automation techniques in computer-aided diagnosis (CAD) may be used to diagnose a variety of ailments. With this method, symptoms will be divided, predicted, identified, and classified using software. The existence and severity of illnesses will then be inferred. The paper's main aim is to examine CAD techniques used to reveal cancer lumps in lung CT images. Lung cancer nodules are usually easier to see using a CT scan, especially if they are larger and indicate a more advanced stage of the disease. Finding the nodules as soon as possible is crucial, however, since they are often little until a patient gets a lung tumour the extent of a golf ball. Figure 1 illustrates the medical imaging methods used to identify lung cancer.

Convolutional neural networks, or CNNs, are a specific kind of neural network that does very well in picture classification. A CNN filter emulates the function of a neuron

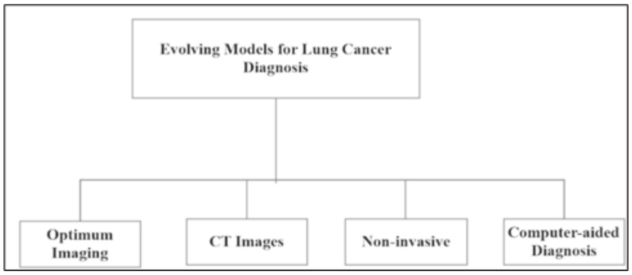

Figure 1. Inputs for Lung Cancer Diagnosis

by using a collection of receptive fields to examine certain areas of an image. Most of the neurons are grouped in hierarchical layers, where the lower layers can identify more complicated patterns because their receptive fields are bigger than those of the upper levels. CNNs may alternatively be seen as many layers as possible of moveable windows that move over an image, like tiny neural networks. CNNs' capacity to recognize patterns independent of their location is one of its advantages. The term "invariance of location" describes this. Sliding windows are used to facilitate the filter's ability to detect patterns in pictures.

Because of its hierarchical nature, CNNs also have the benefit of automatically identifying more complicated patterns. The lower layers can recognize shapes and limits, the middle layers are able to discern details like these, and the upper layers are able to recognize the general forms of objects. Instead of dealing with individual slices of CT scan 2D images, CNNs can be configured to work with 3D images. Instead of a sliding pane, a sliding cube can be used to build 3D CNNs that capture features at each iteration while traversing three dimensions. A compilation of the uses, advantages, and disadvantages of several lung imaging methodology is depicted in Table I.

CT scans are frequently utilized for lung cancer detection because they provide in-depth lung images. DL models can detect the affected nodules in CT scans by examining the sizes, shapes, surface characteristics, and brightness levels. 3D CT scans present an additional widespread view of lung capacity associated to 2D CT images by providing a complete visualization of the lungs.

DL is a kind of ML that makes use of complex multi-layered networks to extract important characteristics from large datasets. These networks may uncover intricate relationships and patterns from data that AI based ML and DL approaches cannot comprehend. Because of its effective competence to study the complicated features from various scan images such as CT scan, MRI, 2D, 3D, DL and ML algorithms are progressively used in the identification of lung cancer as shown in Figure 2. As the CT scan images always stipulate the precise images of the lungs, they are often employed to diagnosis the lung cancer. DL algorithms may recognize the affected parts of the lung using CT scan image evaluation to examine the size of the nodules, figures

TABLE I. Uses, Pros and Cons of Medical Imaging Techniques

Medical Imaging	Applications	Pros	Cons
x-Ray	Used for Lung Cancer initial screening, Detecting Rib Fractures and Pneumonia screening	Fast, Cost-effective, and easily accessible	Minimal rates of accuracy, and chances of missing earlier diagnosis
CT Imaging	Diagnosis and screening of lung cancer pre and post-surgery and diagnosis of pulmonary embolism	Aiding the analysis of lung nodules and earlier disease diagnosis	Higher cost and Higher dose of radiation
Magnetic Resonance Imaging (MRI)	Measuring lung functions and lung cancer invasions	Minimal radiation exposure and result accuracy	Longer time, higher cost and less accessible
Ultrasound Imaging	Evaluating diaphragm performance, detecting pleural effusions	No radiations and Non-invasive	Limitations in scanning lungs and operator-dependent
Position Emission Tomography-Computed Tomography (PET-CT)	Evaluating lung cancer stages, monitoring the post-treatment effects and detecting cancer reoccurrence	Higher rate of sensitivity and accuracy	This may cause higher rate of false positives because of inflammations, Longer time, higher cost, higher radiations and less accessible

and grains. As comparable to 2D scan images, the 3D scan mages provide a more comprehensive examination of lung ability as it could provide broad imagery. Nowadays, several research have implemented DL algorithms to precisely identify the size of the nodules and other abnormalities from 3D scan images. In order to reduce the various emission exposures during lung cancer examining, various techniques such as image augmentation, denoising, and low dose technologies can be adopted. Even the blood flow and tissue intensity can be monitored and exposed in MRI when compared to CT scan images.

DL algorithms can study the MRI images to find out the lumps and irregularities based on their distinctive appearances. To enhance image characteristic and minimize noise, filtering methods like Gaussian and Median filtering are commonly used for pre-processing incoming images. In cases of lung cancer, candidate detection involves identifying suspicious areas in the image to identify potential nodules or masses. Prospective candidates may be found using a variety of DL techniques, such as sliding window approaches and region proposal networks.

B. *Computer-Aided Lung Cancer Diagnosis from CT Medical Images*

Extensive research has been conducted on DL algorithms for diagnosing lung tumours using CT scans. CT images frequently show varying image attenuation patterns for scans of healthy and sick individuals. By using Gray-level thresholding, shape-based approaches and numerical methods, segmentation of lungs had been carried out to separate them from surrounding tissues in a straightforward manner. Some authors introduced a method for segmenting the chest region using automated, knowledge-based technology in their study published in [30]. The required inputs for this method include the estimated size, form, position, and X-ray absorption of organs. Brown et al. developed an automated segmentation technique with expertise in [31] to obtain important data from the CT image dataset. Unintended rateable evaluations of individual lung performance that can't be obtained through traditional pulmonary function testing can be generated automatically. Additionally, the research in [32] created a fully automated technique for segmenting lungs from three-dimensional pulmonary X-ray data. When comparing the root mean square accuracy between the machine and the human evaluation of the recommended technique using 3-Dimensional CT image from 8 healthy participants, a disparity of 0.8 pixels was found.

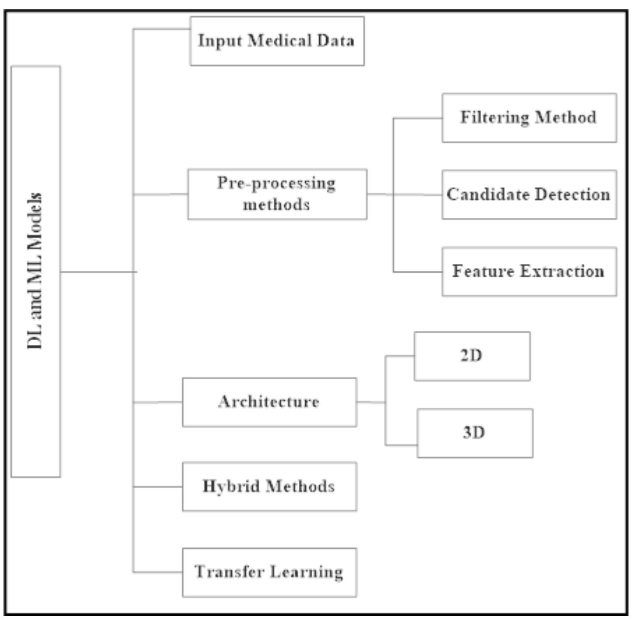

Figure 2. Structure of Procedures in Lung Cancer Diagnosis

By using two sets of classification requirements about position data, size and circularity together with a pixel-value cut-off applied on a slice-by-slice basis, the study in reference [33] presented a completely automated approach for segmenting lungs.The authors had achieved nearly 94% segmentation accuracy with the help of 2950 thick slice pictures approximately and 97% segmentation accuracy using 1100 slim piece images approximately in a test with 101 CT patients. In addition, a filtering anisotropic and an interpolation wavelet-based method were proposed for the segmentation and visualization of lung volumes in [34]. The effectiveness and reliability of the suggested strategy were shown using percentage increases in volume overlap and volume difference using single-detector CT images.

In the publication [35], a level-set formulation-based segmentation technique was used with a traditional segmentation approach to estimate the active dense displacement field. The results of the investigation demonstrated that the proposed technique performed more accurately than the separate procedures of segmentation and registration. Based on a common shape model for lung nodules, the research in [36] presented a novel technique for segmenting lung nodules on CT images. One advantage of the recommended approach is that it is not dependent on the location or kind of nodules. A parameter-free segmentation technique was presented in [37] to focus on juxta pleural nodules, using a related idea to improve the precision of lung nodule identification. With the help of nodules (403 juxta pleural) present in the Lung Imaging Database Consortium (LIDC), the research showed a nearly 92% reintegration rate. Additionally, an automatic lung segmentation method and a hybrid geometric active contour model were created [38]. The use of the global region and edge information enhances the method's performance where there are a narrow bands or regions with somewhat indistinct separation. As forces are applied on the boundaries of the target lung, the segmentation approach described in [39] expands from an initial sphere inscribed in the lung. The results in the 40 CT scan indicated the success rate of the algorithm to be 22% for both.

Scientists have spent the past decade examining the robustness of CNNs for computer vision tasks. Several methods have been proposed using CNN for analysing medical images and processing natural images. Numerous methods have been proposed for lung cancer diagnosis, using artificial intelligence and CT scans. In [40], a three-dimensional CNN with three modules was created to detect and classify lung nodes as an illustration.

A similar artificial neural network (ANN) was used in the work aimed at detecting Lung Cancer with accuracy of 96. 67% accuracy rate [41]. In addition, there was a improvement in lung image quality and an increase in the diagnostic accuracy of lung cancer in up to 98 percent with a combination of the ITNN (Instantaneously Trained Neural Networks) method and the IPCT(Improved Profuse Clustering Technique) method of the study. 42%.According to the research cited in [42], lung nodule detection and classification may be accomplished with accuracies of 0.909 and 0.872, respectively, by using a conventional CDNN and a double convolutional deep neural network (DCDNN). A CAD system developed by the authors [43] was able to identify nodules with excellent accuracy and minimal false positive and negative detection rates. This strategy is distinct. Using lung pictures rather than random initialization, the inception-v3 transfer learning strategy produced a sensitivity rate of 95.41% for the deep model in reference [44]. A unique patch-based learning system was described in [45] to diagnose lung cancer which has a sensitivity rate of 80.06% and 94%, respectively, with 4.7 false positives per scan and 15.1 false positives per scan.

The study in [46] presented DenseBTNet, a parameter-efficient dense convolutional binary-tree network for multi-scale feature extraction. Li et al. underscored how important early identification in lowering lung cancer death rates in their research study [47]. They introduced the DL-CAD system, which classifies lung nodules less than 3 mm and predicts whether they will become malignant. This method uses DL. The sensitivity of the system was assessed using the NLST and LIDC-IDRI datasets, producing an accuracy rate of 86.2%. [48] Reported the publication of a novel 3D residual CNN designed to minimize false positives in CT scans for automated lung nodule identification. The network is comparable to earlier techniques. Their 27-layer network achieved a 98.3% sensitivity rate using the LUNA-16 dataset. Contextual data was extracted at many levels using a SPC (Spatial Pooling and Cropping) layer.

To improve the probability of automating the classification of lung cancers, a new DCNN with sectors of pooling, fully connected and convolutions were introduced in [49]. They also trained the DCNN with only 76 cancer cases to achieve 71% classification accuracy and pointed out several limitations to the use of other datasets such as SURBOOST. Another network developed by the research in [50] for assisting clinicians in diagnosing lung nodules is a 3D convolutional neural network for volumetric CT data. The LUNA16 was applied to test this model, comprising of multiple groups of three dimensional convolutional layers, coupled with fully connected layers, appropriate max pooling as well as softmax layers. Their findings suggest that the use of 3D CNNs might improve detection accuracy (94.4% SNR). Using DL algorithms and CT scans, two studies [51] estimated the survival rate of lung adenocarcinoma, the existence of EGFR mutation, and its subtype classification. A comprehensive evaluation of several studies focused on the use of DL methods for the segmentation and classification of lung nodules on CT scans was conducted [52].

Rather of using a high-dose method, the research team in [53] built a whole lung cancer detection system from the ground up using low-dose chest CT scans and a three-dimensional deep-learning model. Furthermore, the researchers [54] used DL algorithms in combination with mobile low-dose CT scans to detect lung cancer in resource-poor settings. The research [55] looks closely at the application of DL methods for finding and identifying lung nodules in CT scans. Likewise, the model described in [56] is used to non-small-cell lung cancer to predict the EGFR mutation and the expression status of PD-L1. This is done by CT scans.

In the research [57], lung CT image segmentation was achieved by the application of deep neural networks and the classification technique. According to the author, a DL model trained on CT scan data achieved 96.3% accuracy in lung cancer diagnosis. The effort has produced [58] qualitative research that examines the diagnostic efficacy and precision of DL models for chest radiography in clinical settings, and it also examines the use of CT scans in the diagnosis of lung cancer. It was demonstrated in the article [59] that the method of DL applied for the lung cancer diagnosis has the sensitivity of 93%. 34% and, respectively, a specificity of 91%. 5%. In creating the model in [60] the focus was a case with lower nodules to assess the use of immune checkpoint drugs and DL based on CT scans of NSCLC to estimate PD-L1. of 0, they achieved an F1 score of 0, due to false negative cases that could not be sorted into any category. In the work [61], the authors demonstrated how to utilize DL for the info extraction: stage of lung cancer from CT data in 848. The authors of [62] provided a viable example in how the use of a CAD system powered by DL could help in pinpointing nodules on 1-mm-thick CT images and therefore support their idea of using machine learning in the identification of benign, pre-invasive, and invasive lung nodules in scans. In the same way, a DL model to recognize a specific image from the OMIM database has also proved useful, in particular accuracy rate of 87%. Confirming the outcome of the previous sections, we establish that the 63% for predicting lung cancer was proposed in the publication [63]. According to the work in [64], DL models include six recommended ones, namely CNN, CNN GD, Inception V3, Resnet-50, CNN VGG16 and CNN VGG19, in the use of CT scans and histology images to diagnose lung cancer. From the results that we achieved we can conclude that the algorithm of CNN GD is better in precision, sensitivity, specificity, accuracy, and the F-Score compared to other algorithms. Alone it accomplishes 97 percent target for individual patient management, effectively handling the patient's condition. 86% accuracy, 96. 39% sensitivity, 96. 79% specificity, and 97. 40% sensitivity.

By utilizing 3DCNN and RNN, the authors in [65] offer a distinct approach for precisely detecting malignant lung nodules with a 95% accuracy rate. To enhance efficiency even more, upcoming enhancements could involve implementing cascading classifiers and utilizing big-data analytics. A CNN-based model for prior lung cancer diagnosis using CT scan imaging is presented in the study [66]. The model can distinguish between benign, malignant, and typical instances. For lung cancer survival rates to increase and treatment programs to start on time, early diagnosis is crucial. The model achieves an amazing accuracy rate of 99.45% while effectively reducing false positives. It was also suggested that radiomics and DL be used in tandem to detect and cure lung cancer [67]. The authors provide an example of how radiomics might be used to improve cancer diagnosis and prognosis by extracting quantitative data from medical imaging. After that, DL algorithms may examine the data.

Research on the application of DL to low-dose computed tomography (CT) image analysis—which is often used for lung cancer screening—is presented in [68]. This technique predicts the risk of cardiovascular disease. The researchers developed a DL model using data from lung CT scans to forecast the probability of cardiovascular disease based on a comprehensive and diverse set of cardiovascular risk factors. By using DL and hybrid dense clustering to facilitate quick neural network training, the technique published in [69] offered a more efficient way to identify lung cancer from CT data. The efficacy of the authors' approach in identifying lung nodules was assessed in comparison to other techniques for diagnosing lung cancer. A novel deep convolutional neural network (CNN) for 3D CT scan lung nodule detection is reported in reference number [70]. To demonstrate the CNN model's efficacy in identifying and classifying lung nodules, the researchers subjected it to a large set of CT images. In the research, an electronic nasal system based on weighted discriminative extreme learning machine was suggested for the detection of lung cancer [71]. They used an electronic nasal device to analyze breath samples from lung cancer patients and healthy controls and were able to distinguish between the two groups with accuracy. The author's 3D lung cancer detection method [72] utilizes multimodality attention guidance and more than 10 F-FDG PET/CT images.

C. Lung Tumour Diagnosis from Sequencing Data

Although it is recommended for high-risk individuals to undergo regular medical imaging tests, the high occurrence of false positives has made it difficult to put this into practice [73] effectively. New approaches to early detection of lung cancer are desperately needed. Advanced sequencing technology makes it possible to use a variety of techniques for lung cancer early detection. While waiting for the best course of action, accurately categorizing various forms of lung cancer is essential. Cancer cells are known to exhibit a wide range of genetic variants, and accumulating these differences may reveal the mutational patterns seen in many cancer types [74]. Because of this, current research has focused on obtaining improved genetic markers as input characteristics to improve the accuracy of their ML algorithms.

Blood-based liquid biopsy is considered as a valuable methodology for the early stage diagnosis. Circulating tumor markers, exosomes, methylation, circulating tumor cells (CTCs), circulating tumor DNA (ctDNA), microRNA (miRNA), and fragments of cell-free DNA (cfDNA) are employed in pursuing investigations of cell-free circulating tumor markers. Numerous researchers were investigated to analyze ct DNA, miRNA, methylation, exosome, and CTC as potential circulating markers for tumor [75]. Following these classifiers, the discriminative models which are SVM, RF, and LR is applied for detecting and diagnosing tumor from the signals obtained from liquid biopsy with high level of accuracy. It is also important to recognize SNVs, insertions, and deletions as the somatic mutations, which can also be distributed in a specific manner depending on the type of cancer, thus, it can be beneficial to categorize the subtypes of lung cancer [76]. Therefore, research has used somatic mutations as input characteristics to create classifiers that can distinguish between LUAD and LUSC [77]. Numerous mutations, particularly driver mutations, may alter the levels of gene expression, affecting how well the genes work and interfering with cellular signalling pathways. Therefore, the levels of specific proteins differ across various types of cancer [78]. ML models can categorize patient malignancy and subtypes (LUAD or LUSC) by analyzing the unique expression patterns of each cancer type using RNA sequencing data [79]. Likewise, it has been noted that cancer cells frequently exhibit copy number variation (CNV), which is closely linked to changes in gene expression [80]. Due to this reason, CNVs could also be utilized in studies on lung cancer to teach machine-learning algorithms for classifying cancer types [81]. Authors in similar type of article [82] implemented a recurrent hidden Markov model (HMM) that accurately classifies wide chromosomal regions with varying copy numbers. Bockmayr with other authors [83] utilized DNA methylation patterns as input features to differentiate primary lung cancer from metastasis in malignant nodules. If all genes generated were utilized as input features directly, there is a risk of overfitting [84]. To enhance their machine-learning models, many researchers chose multiple cancer-related genes through different computational methods. The models related to sequential data-based lung cancer diagnosis are evaluated from 2016 to 2023 and are listed in Table II.

3. DISCUSSIONS ON DATASET AVAILABILITY FOR LUNG CANCER DIAGNOSIS

Many datasets were incorporated into the lung cancer diagnosis to evaluate how well DL methods work. Included in the research are the following datasets:

i Lung Image Database (LID): Utilized for developing and validating computer-aided detection systems for lung cancer [89].

ii LIDC-IDRI dataset: Utilized in lung cancer studies related to identifying and classifying nodules [89].

iii CT Lung Datasets: Research on nodule identification in lung-specific image examination [89].

iv National Lung Screening Trial dataset: DL models are evaluated in detecting early lung cancer [90].

v Dataset on immunotherapy: UtilizedDL methods, the relationship between lung cancer and responses to immunotherapy are explored [91].

vi PD-L1 expression dataset: DL has been investigated lung cancer patients via experiments [92].

vii Tianchi AI dataset: Utilized to develop and evaluate DL

TABLE II. Models Comparison using Sequential Data for Lung Cancer Diagnosis

Year	Ref	Model & Datatype	Pros	Cons
2016	[85]	KNN; NB normal distribution of attributes. SVM; C4.5 DT & RNA-seq	Evaluates several lung cancer subtype classification methods across various datasets.	Reduced overfitting may be achieved by using feature selection techniques.
2019	[83]	NN; SVM; RF & DNA methylation	Forecasting tumor metastasis using DNA methylation data.	Samples with poor tumor cellularity cannot be reliably predicted by the model using methylation data.
	[86]	LR & ctDNA	Creates a ML framework employing DNA methylation indicators to identify lung tumors early.	As there are only nine methylation indicators in the specified characteristics, the assay's performance is limited.
2020	[87]	Diet Networks with EIS Somatic mutation	Aids in maintaining stability in Diet Networks' training procedure.	Depending on the dataset, interpretable hidden meanings from EIS may be produced.
2021	[88]	LR model with a LASSO penalty & cfDNA fragment	Offers a framework for integrating additional indicators with cfDNA fragmentation characteristics to diagnose lung cancer.	DNA variants may impact the identification of cfDNA in late-stage diseases.

techniques for detecting lung cancer [93].

viii ImageNet: To detect lung tumour, Transfer learning techniques have been deployed [93].

ix Cancer Imaging Archive (CIA) Dataset: To develop and evaluate DL models for lung cancer detection. [94].

DL models may be made accurate and broadly applicable; however there are possibilities and obstacles due to the diversity of lung cancer cases, dataset size, imaging methodologies, and annotation correctness.

4. SHORTCOMINGS IN EXISTING MODELS

Preparation before applying DL models to CT scans for lung cancer detection and diagnosis brings several challenges, especially with varied datasets. These limitations could affect the accuracy and reliability of the segmentation and classification tasks. Highlighted are a few of the main shortcomings:

Variations in imaging techniques and equipment can lead to significant differences in the quality, thickness of slices, contrast, and noise levels of CT scans for lung cancer. Pre-processing techniques face difficulties due to the variability, requiring them to handle these variations to ensure trustworthy outcomes efficiently. Catastrophe to account for data variability can lead to poor performance and limited generalizability of the DL model.

Objects like steel, flash setting, and movement caused in CT scan can affect the quality of the images. Due to these disturbances and anomalies, the pre-processing task faces a critical challenge in acquiring the effective features and information of the dataset. Thus, effective methods need to be identified for fixing the anomalies caused by the artefacts and for the effective classification task.

There should be a required amount of labelled data to train the DL models for lung cancer identification. However, getting precise annotations for the scan images can be challenging and difficult, particularly when dealing with very complex segmentation chore. Scarce labelled data makes the DL models difficult to identify the lung cancer accurately and thus it leads to poor performance and less relevancy. Additionally, if several datasets are labeled using different standards and measures, lot of discrepancies can occur during the training phase. DL models should be able to simplify over an array of populaces and scan images, such that the preparation approaches should consider the various possible limitations and prejudices connected to the definite datasets. It is always important to identify the efficient algorithms and its strategies and systematically evaluate the dataset's topographies to overcome the limitations in the pre-processing phases.

5. CONCLUSIONS

This study acknowledges the limits of the various lung cancer analysis algorithms and investigates them using different kinds of medical input data. The research also maintains that other input data, such as MRI and ultrasound images, are required in addition to CT scan images. In addition to the medical imaging as input, the individual's private information is required for evaluation and shared analysis. To distinguish between the early and late stages of the disease, the research will analyze several models for the diagnosis of lung cancer. Moreover, it is proposed that relevant personal information, such as genetic information and medical history, correlate with deep features extracted from lung scan images to improve the accuracy of automated tumor detection. This all-encompassing approach could lead to a more precise diagnosis of the illness. The authors recommend employing a variety of pre-processing techniques, such as filters, to improve picture quality. Edge-preserving methods and harmony search may be used to improve grayscale images. These techniques provide improved picture analysis and more precise diagnostic outcomes. The authors' successful recommendation for remote lung cancer detection supports their subsequent recommendation to investigate the use of AI technology for machine learning-based lung cancer diagnosis. Future research on early lung cancer detection that supports better patient treatment is suggested below.

i To improve the precision and uniformity of lung cancer diagnosis and concentrate on developing standardized pre-processing approaches that use DL methods and account for the variety of CT scans.

ii The research may also include the following areas: segmentation, integration, early detection, standardization, feature extraction, and picture quality improvement.

iii By concentrating on these problems, researchers may be able to improve the accuracy, efficacy, and reliability of lung cancer detection, which would eventually benefit people everywhere.

iv By focusing on the improvisation regards of the CT images that might have greater inferences for medical and cost-efficiency in various healthcare circumstances.

v Federated Learning medical datasets should be made available for researchers for comparing others algorithms.

References

[1] A. Leiter, R. R. Veluswamy, and J. P. Wisnivesky, "The global burden of lung cancer: current status and future trends," *Nature reviews. Clinical oncology*, vol. 20, no. 9, 2023.

[2] E. Svoboda, "Artificial intelligence is improving the detection of lung cancer," *Nature*, vol. 587, pp. S20 – S22, 2020.

[3] R. Javed, T. Abbas, A. H. Khan, A. Daud, A. Bukhari, and R. Alharbey, "Deep learning for lungs cancer detection: a review," *Artificial Intelligence Review*, vol. 57, no. 8, 2024.

[4] S. Zaira, L. Artem, L. R. Yates, and G. Moritz, "Spatial biology of cancer evolution," *Nature Reviews Genetics*, vol. 24, pp. 295–313, 2023.

[5] A. Murmu and B. Győrffy, "Artificial intelligence methods available for cancer research," *Front. Med.*, 2024.

[6] R. Sriram, S.-P. L. Nelson, H. Christopher M, C. Michael S, L. Yuan, and F. Todd A, "Artificial intelligence-based clinical decision support in pediatrics," *Pediatr Res*, vol. 93, pp. 334–341, 2023.

[7] S.-K. Tan, S.-C. Chong, K.-K. Wee, and L.-Y. Chong, "Personalized healthcare: A comprehensive approach for symptom diagnosis and hospital recommendations using ai and location services," *Journal of Informatics and Web Engineering*, vol. 3, no. 1, pp. 117–135, 2024.

[8] C. Chee Chiet, k. Wee How, P. Ying Han, and Y. Hui Yen, "A lung cancer detection with pre-trained cnn models," *Journal of Informatics and Web Engineering*, vol. 3, no. 1, pp. 41–54, 2024.

[9] U. Zia, Q. Yan, W. Long, S. Yiwei, Y. Qianqian, K. Saeed, Ullah, A. Rukhman, and Z. Juanjuan, "Effective lung nodule detection using deep cnn with dual attention mechanisms," vol. 14, 2024.

[10] K. Han, V. S. Sheng, Y. Song, Y. Liu, C. Qiu, S. Ma, and Z. Liu, "Deep semi-supervised learning for medical image segmentation: A review," *Expert Systems with Applications*, vol. 245, 2024.

[11] L. Wang, X. Zhang, H. Su, and J. Zhu, "A comprehensive survey of continual learning: Theory, method and application," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 8, pp. 5362–5383, 2024.

[12] M. Ling, W. Chuangye., H. Kexin., C. Annan., and L. Lizhi., "A novel fusion algorithm for benign-malignant lung nodule classification on ct images," *BMC Pulmonary Medicine*, vol. 23, 2023.

[13] Z. Jianwei. and S. Yonggang., "Personalized patch-based normality assessment of brain atrophy in alzheimer's disease," *Med Image Comput Comput Assist Interv.*, pp. 55–62, 2023.

[14] S. Ashwini, J. R. Arunkumar, R. T. Prabu, N. H. Singh, and N. P. Singh, "Diagnosis and multi-classification of lung diseases in cxr images using optimized deep convolutional neural network," *Soft Comput.*, vol. 28, no. 7–8, p. 6219–6233, 2023.

[15] K. L. Lew, C. Y. Kew, K. S. Sim, and S. C. Tan, "Adaptive gaussian wiener filter for ct-scan images with gaussian noise variance," *Journal of Informatics and Web Engineering*, vol. 3, no. 1, p. 169–181, 2024.

[16] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "Detrs beat yolos on real-time object detection," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 16 965–16 974.

[17] H. T. Gayap and M. A. Akhloufi, "Deep machine learning for medical diagnosis, application to lung cancer detection: A review," *BioMedInformatics*, vol. 4, no. 1, pp. 236–284, 2024.

[18] J. Shin, N. Kang, B. Kim, H. Hong, L. Yu, J. Kim, H. Kang, and J. S. Kim, "One-dimensional nanomaterials for cancer therapy and diagnosis," *Chem. Soc. Rev.*, vol. 52, pp. 4488–4514, 2023.

[19] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680–1716, 2023.

[20] W. Li, D. Liu, Y. Li, M. Hou, J. Liu, Z. Zhao, A. Guo, H. Zhao, and W. Deng, "Fault diagnosis using variational autoencoder gan and focal loss cnn under unbalanced data," *Structural Health Monitoring*, pp. 124–129, 2024.

[21] Y. Liu, D. Yang, Y. Wang, J. Liu, J. Liu, A. Boukerche, P. Sun, and L. Song, "Generalized video anomaly event detection: Systematic taxonomy and comparison of deep models," *ACM Comput. Surv.*, vol. 56, no. 7, 2024.

[22] E. Ghorani, C. Swanton, and S. A. Quezada, "Cancer cell-intrinsic mechanisms driving acquired immune tolerance," *Immunity*, vol. 56, no. 10, pp. 2270–2295, 2023.

[23] S. E. A. Raza, L. Cheung, M. Shaban, S. Graham, D. B. A. Epstein, S. Pelengaris, M. Khan, and N. M. Rajpoot, "Micro-net: A unified model for segmentation of various objects in microscopy images," *Medical Image Analysis*, vol. 52, p. 160–173, 2018.

[24] K. Sirinukunwattana, S. E. A. Raza, Y.-W. Tsang, D. R. J. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer

histology images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1196–1206, 2016.

[25] N. Coudray, A. L. Moreira, T. Sakellaropoulos, D. Feny, N. Razavian, and A. Tsirigos, "Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning," *nature medicine*, vol. 24, p. 1559–1567, 2017.

[26] Y. Q. amd Xiao L, L. P, and Z. SK, "Label-free segmentation of covid-19 lesions in lung ct," *IEEE Trans Med Imaging*, vol. 40, no. 19, pp. 2808–2819, 2021.

[27] J. Li, J. Jia, and D. Xu, "Unsupervised representation learning of image-based plant disease with deep convolutional generative adversarial networks," in *2018 37th Chinese Control Conference (CCC)*, 2018, pp. 9159–9163.

[28] C.-H. Lin, C.-J. Lin, Y.-C. Li, and S.-H. Wang, "Using generative adversarial networks and parameter optimization of convolutional neural networks for lung tumor classification," *Applied Sciences*, vol. 11, no. 2, 2021.

[29] Z. Ren, Y. Zhang, and S. Wang, "A hybrid framework for lung cancer classification," *Electronics*, vol. 11, no. 10, 2022.

[30] M. Jenita., M. Richa., and U. Eranga., "Pulmonary nodule detection in low dose computed tomography using a medical-to-medical transfer learning approach," *J Med Imaging (Bellingham)*, vol. 11, no. 4, 2024.

[31] T. Mohammad A., Z. Mohd Asyraf., A. M. Z. Muhammad Ammirrul., and A. Siti Raihanah., "Knowledge-based segmentation of thoracic computed tomography images for assessment of split lung function," *Diagnostics (Basel)*, vol. 13, no. 16, 2023.

[32] S. Pandey. and A. A. Bhandari., "Systematic review of modern approaches in healthcare systems for lung cancer detection and classification," *Arch Computat Methods Eng*, vol. 13, pp. 4359–4378, 2023.

[33] B. Sahu., S. Saurabh., and T. Swarnakar., "Covid - 19 patient's ct images classification: Stackalexnet-19 a deep learning approach," *International Journal of Computing and Digital Systems*, vol. 12, no. 1, pp. 1453–1464, 2003.

[34] P. Swierczynski, B. W. Papież, J. A. Schnabel, and C. Macdonald, "A level-set approach to joint image segmentation and registration with application to ct lung imaging," *Computerized Medical Imaging and Graphics*, vol. 65, pp. 58–68, 2018, advances in Biomedical Image Processing.

[35] A. A. Farag, E. Munim Hossam, E. Abd, J. H. Graham, and A. A. Farag, "A novel approach for lung nodules segmentation in chest ct using level sets," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5202–5213, 2013.

[36] K. Sarkar, Nayan, M. Singh, Moirangthem, and U. Nandi, "Recent researches on image classification using deep learning approach," *International Journal of Computing and Digital Systems*, vol. 12(1), pp. 1357–1374, 2022.

[37] W. Zhang, X. Wang, P. Zhang, and J. Chen, "Global optimal hybrid geometric active contour for automated lung segmentation on ct images," *Comput. Biol. Med.*, vol. 91, p. 168–180, 2017.

[38] P. P. Rebouças Filho, P. C. Cortez, A. C. da Silva Barros, V. H. C. Albuquerque, and J. M. R. S. Tavares, "Novel and powerful 3d adaptive crisp active contour method applied in the segmentation of ct lung images," *Medical Image Analysis*, vol. 35, pp. 503–516, 2017.

[39] C. Zhang, X. Sun, K. Dang, K. Li, X.-W. Guo, J. Chang, Z.-Q. Yu, F.-Y. Huang, Y.-S. Wu, Z. Liang, Z.-Y. Liu, X.-G. Zhang, X.-L. Gao, S.-H. Huang, J. Qin, W.-N. Feng, T. Zhou, Y.-B. Zhang, W.-J. Fang, M.-F. Zhao, X.-N. Yang, Q. Zhou, Y.-L. Wu, and W.-Z. Zhong, "Toward an expert level of lung cancer detection and classification using a deep convolutional neural network," *The oncologist*, vol. 24, no. 9, p. 1159—1165, 2019.

[40] I. M. Nasser and S. S. Abu-Naser, "Lung cancer detection using artificial neural network," *International Journal of Engineering and Information Systems*, vol. 3, no. 3, pp. 17–23, 2019.

[41] Cifci and M. Akif, "Segchanet: A novel model for lung cancer segmentation in ct scans," *Applied Bionics and Biomechanics*, vol. 2022, no. 1, p. 1139587, 2022.

[42] G. Jakimovski and D. Davcev, "Using double convolution neural network for lung cancer stage detection," *Applied Sciences*, vol. 9, no. 3, 2019.

[43] J. Wang, J. Wang, Y. Wen, H. Lu, T. Niu, J. Pan, and D. Qian, "Pulmonary nodule detection in volumetric chest ct scans using cnns-based nodule-size-adaptive detection and classification," *IEEE Access*, vol. 7, pp. 46 033–46 044, 2019.

[44] C. Wang, D. Chen, L. Hao, X. Liu, Y. Zeng, J. Chen, and G. Zhang, "Pulmonary image classification based on inception-v3 transfer learning model," *IEEE Access*, vol. 7, pp. 146 533–146 541, 2019.

[45] H. Jiang, H. Ma, W. Qian, M. Gao, and Y. Li, "An automatic detection system of lung nodule based on multigroup patch-based deep learning network," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 4, pp. 1227–1237, 2018.

[46] Y. Liu, P. Hao, P. Zhang, X. Xu, J. Wu, and W. Chen, "Dense convolutional binary-tree networks for lung nodule classification," *IEEE Access*, vol. 6, pp. 49 080–49 088, 2018.

[47] L. Li, Z. Liu, H. Huang, M. Lin, and D. Luo, "Evaluating the performance of a deep learning-based computer-aided diagnosis (dl-cad) system for detecting and characterizing lung nodules: Comparison with the performance of double reading by radiologists," *Thoracic Cancer*, vol. 10, no. 2, pp. 183–192, 2019.

[48] 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.

[49] A. Teramoto, T. Tsukamoto, Y. Kiriyama, and H. Fujita, "Automated classification of lung cancer types from cytological images using deep convolutional neural networks," *BioMed Research International*, vol. 2017, no. 1, p. 4067832, 2017.

[50] J. Zhenge, C. Jianxu, X. Xiaowei, K. John, H. Jingtong, X. Han, P. Sui, S. H. Xiaobo, C. Danny, and Y. Shi, "The importance of resource awareness in artificial intelligence for healthcare," *Nat Mach Intell*, vol. 5, p. 687–698, 2023.

[51] C. Wang, J. Shao, J. Lv, Y. Cao, C. Zhu, J. Li, W. Shen, L. Shi, D. Liu, and W. Li, "Deep learning for predicting subtype classification and survival of lung adenocarcinoma on computed tomography," *Translational Oncology*, vol. 14, no. 8, p. 101141, 2021.

[52] Y. Gu, J. Chi, J. Liu, L. Yang, B. Zhang, D. Yu, Y. Zhao, and X. Lu, "A survey of computer-aided diagnosis of lung nodules from ct scans using deep learning," *Computers in Biology and Medicine*, vol. 137, p. 104806, 2021.

[53] D. Ardila, A. P. Kiraly, S. Bharadwaj, B. Choi, J. J. Reicher, L. H. Peng, D. Tse, M. Etemadi, W. Ye, G. S. Corrado, D. P. Naidich, and S. Shetty, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," *Nature Medicine*, vol. 25, pp. 954 – 961, 2019.

[54] J. Shao, G. Wang, L. Yi, C. Wang, T. Lan, X. Xu, J. Guo, T. Deng, D. Liu, B. Chen, Z. Yi, and W. Li, "Deep learning empowers lung cancer screening based on mobile low-dose computed tomography in resource-constrained sites," *Front Biosci (Landmark Ed)*, vol. 27, no. 7, 2022.

[55] C. Wang, X. Xu, J. Shao, K. Zhou, K. Zhao, Y. He, J. Li, J. Guo, Z. Yi, and W. Li, "Deep learning to predict egfr mutation and pd-l1 expression status in non-small-cell lung cancer on computed tomography images," *Journal of Oncology*, vol. 2021, pp. 1–11, 12 2021.

[56] R. Li, C. Xiao, Y. Huang, H. Hassan, and B. Huang, "Deep learning applications in computed tomography images for pulmonary nodule detection and diagnosis: A review," *Diagnostics*, vol. 12, no. 2, 2022.

[57] L. S.K., S. N. Mohanty, S. K., A. N., and G. Ramirez, "Optimal deep learning model for classification of lung cancer on ct images," *Future Generation Computer Systems*, vol. 92, pp. 374–382, 2019.

[58] S. Lee, J. B. Seo, J. Yun, Y.-H. Cho, J. Vogel-Claussen, M. Schiebler, W. Gefter, E. Beek, J. M. Goo, K. Lee, H. Hatabu, J. Gee, and N. Kim, "Deep learning applications in chest radiography and computed tomography: Current state of the art," *Journal of Thoracic Imaging*, vol. 34, 2019.

[59] S. Bhatia, Y. Sinha, and L. Goel, "Lung cancer detection: A deep learning approach," in *Soft Computing for Problem Solving*, 2019, pp. 699–705.

[60] P. Tian, B. He, W. Mu, K. Liu, L. Liu, H. Zeng, Y. Liu, L. Jiang, P. Zhou, Z. Huang, D. Dong, and W. Li, "Assessing pd-l1 expression in non-small cell lung cancer and predicting responses to immune checkpoint inhibitors using deep learning on computed tomography images," *Theranostics*, vol. 11, pp. 2098–2107, 2021.

[61] D. Hu, H. Zhang, S. Li, Y. Wang, N. Wu, and X. Lu, "Automatic extraction of lung cancer staging information from computed tomography reports: Deep learning approach (preprint)," *JMIR Medical Informatics*, vol. 9, 2021.

[62] S. F. Ashraf, K. Yin, C. X. Meng, Q. Wang, Q. Wang, J. Pu, and R. Dhupar, "Predicting benign, preinvasive, and invasive lung nodules on computed tomography scans using machine learning," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 163, no. 4, pp. 1496–1505.e10, 2022.

[63] R. R. Subramanian, R. N. Mourya, V. P. T. Reddy, B. N. Reddy, and S. Amara, "Lung cancer prediction using deep learning framework," *International Journal of Control and Automation*, vol. 13, no. 03, pp. 154–160, 2020.

[64] V. Rajasekar, M. Vaishnnave, S. Premkumar, V. Sarveshwaran, and V. Rangaraaj, "Lung cancer disease prediction with ct scan and histopathological images feature analysis using deep learning techniques," *Results in Engineering*, vol. 18, p. 101111, 2023.

[65] S. Wankhade and V. S., "A novel hybrid deep learning method for early detection of lung cancer using neural networks," *Healthcare Analytics*, vol. 3, p. 100195, 2023.

[66] S. Abunajm, N. Elsayed, Z. ElSayed, and M. Ozer, "Deep learning approach for early stage lung cancer detection," 2023.

[67] S. Avanzo, J. Stancanello, G. Pirrone, and G. Sartor, "Radiomics and deep learning in lung cancer," *Strahlentherapie und Onkologie*, vol. 196, pp. 879 – 887, 2020.

[68] H. Chao, H. Shan, F. Homayounieh, R. Singh, R. Khera, H. Guo, T. Su, G. Wang, M. Kalra, and P. Yan, "Deep learning predicts cardiovascular disease risks from lung cancer screening low dose computed tomography," *Nature Communications*, vol. 12, 2021.

[69] P. M. Shakeel, M. Burhanuddin, and M. I. Desa, "Lung cancer detection from ct image using improved profuse clustering and deep learning instantaneously trained neural networks," *Measurement*, vol. 145, pp. 702–712, 2019.

[70] A. E. Sebastian and D. Dua, "Lung nodule detection via optimized convolutional neural network: Impact of improved moth flame algorithm," *Sensing and Imaging*, vol. 24, pp. 702–712, 2023.

[71] L. Zhao, J. Qian, F. Tian, R. Liu, B. Liu, S. Zhang, and M. Lu, "A weighted discriminative extreme learning machine design for lung cancer detection by an electronic nose system," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–9, 2021.

[72] L. Chen, K. Liu, H. Shen, H. Ye, H. Liu, L. Yu, J. Li, K. Zhao, and W. Zhu, "Multimodality attention-guided 3-d detection of nonsmall cell lung cancer in 18f-fdg pet/ct images," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 6, no. 4, pp. 421–432, 2022.

[73] R. P. Young, R. C. Ward, R. J. Scott, G. D. Gamble, and G. Silvestri, "Airflow limitation and mortality during cancer screening in the national lung screening trial: why quantifying airflow limitation matters," *Thorax*, vol. 78, no. 7, pp. 690–697, 2023.

[74] B. Otlu, M. Díaz-Gay, I. Vermes, E. Bergstrom, M. Zhivagui, M. Barnes, and L. Alexandrov, "Topography of mutational signatures in human cancer," *Cell Rep*, vol. 42(8), 2023.

[75] S. Herath, H. S. Rad, P. Radfar, R. Ladwa, M. Warkiani, K. O'Byrne, and A. Kulasinghe, "The role of circulating biomarkers in lung cancer," *Frontiers in Oncology*, vol. 11, 2021.

[76] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Robert, and P. Walter, *Molecular Biology of the Cell*, 4th ed., 2002.

[77] K. Kobayashi, A. Bolatkan, S. Shiina, and R. Hamamoto, "Fully-connected neural networks with reduced parameterization for predicting histological types of lung cancer from somatic mutations," *Biomolecules*, vol. 10, p. 1249, 2020.

[78] K. M. Garland, T. L. Sheehy, and J. T. Wilson, "Chemical and biomolecular strategies for sting pathway activation in cancer immunotherapy," *Chemical Reviews*, vol. 122, no. 6, pp. 5977–6039, 2022.

[79] A. P. Kumar and P. Jagadeesh, "Innovative detection of lung cancer using decision tree classifier and comparison with K-nearest neighbor classifier," *AIP Conference Proceedings*, vol. 2853, no. 1, 2024.

[80] X. Shao, N. Lv, J. Liao, J. Long, R. Xue, N. Ai, D. Xu, and X. Fan, "Copy number variation is highly correlated with differential gene expression: A pan-cancer study," *BMC Medical Genetics*, vol. 20, 2019.

[81] M. Ruiz-Fresneda, A. Gijón, and P. Morales-Álvarez, "Bibliometric analysis of the global scientific production on machine learning applied to different cancer types," *Environ Sci Pollut Res*, vol. 30, p. 96125–96137, 2023.

[82] A. Daemen, O. Gevaert, K. Leunen, E. Legius, I. Vergote, and B. D. Moor, *Supervised Classification Of Array Cgh Data With Hmm-Based Feature Selection*, pp. 468–479.

[83] P. Jurmeister, M. Bockmayr, P. Seegerer, T. Bockmayr, D. Treue, G. Montavon, C. Vollbrecht, A. Arnold, D. Teichmann, K. Bressem, U. Schüller, M. von Laffert, K.-R. Müller, D. Capper, and F. Klauschen, "Machine learning analysis of dna methylation profiles distinguishes primary lung squamous cell carcinomas from head and neck metastases," *Science Translational Medicine*, vol. 11, 2019.

[84] S. C. da Silva Rosa, A. B. Behrooz, S. Guedes, R. Vitorino, and S. Ghavami, "Prioritization of genes for translation: a computational approach," *Expert Review of Proteomics*, vol. 21, no. 4, pp. 125–147, 2024.

[85] M. D. Podolsky, A. A. Barchuk, V. I. Kuznetcov, N. F. Gusarova, V. S. Gaidukov, and S. A. Tarakanov, "Evaluation of machine learning algorithm utilization for lung cancer classification based on gene expression levels," *Asian Pacific journal of cancer prevention APJCP*, vol. 17, pp. 835–838, 2016.

[86] W. J. Locke, D. Guanzon, C. Ma, Y. J. Liew, K. R. Duesing, K. Y. C. Fung, and J. P. Ross, "Dna methylation cancer biomarkers: Translation to the clinic," *Front Genet*, vol. 10, pp. 1–22, 2019.

[87] Y. Choi, J. Qu, S. Wu, Y. Hao, J. Zhang, J. Ning, X. Yang, L. Lofaro, D. G. Pankratz, J. Babiarz, P. S. Walsh, E. Billatos, M. E. Lenburg, G. C. Kennedy, J. McAuliffe, and J. Huang, "Improving lung cancer risk stratification leveraging whole transcriptome rna sequencing and machine learning across multiple cohorts)," *BMC Med Genomics*, vol. 13, no. 151, 2020.

[88] D. Mathios, J. S. Johansen, S. Cristiano, J. E. Medina, J. Phallen, K. R. Larsen, D. C. Bruhm, N. Niknafs, L. Ferreira, V. Adleff, J. Y. Chiao, A. Leal, M. Noe, J. R. White, A. S. Arun, C. Hruban, A. V. Annapragada, S. Østrup Jensen, M.-B. W. Ørntoft, A. H. Madsen, B. Carvalho, M. de Wit, J. Carey, N. C. Dracopoli, T. Maddala, K. C. Fang, A.-R. Hartman, P. M. Forde, V. Anagnostou, J. R. Brahmer, R. J. A. Fijneman, H. J. Nielsen, G. A. Meijer, C. L. Andersen, A. Mellemgaard, S. E. Bojesen, R. B. Scharpf, and V. E. Velculescu, "Detection and characterization of lung cancer using cell-free dna fragmentomes," *Nature Communications*, vol. 12, 2021.

[89] O. Ozdemir, R. L. Russell, and A. A. Berlin, "A 3d probabilistic deep learning system for detection and diagnosis of lung cancer using low-dose ct scans," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1419–1429, 2020.

[90] H. J. de Koning, C. M. van der Aalst, P. A. de Jong, E. T. Scholten, K. Nackaerts, M. A. Heuvelmans, J.-W. J. Lammers, C. Weenink, U. Yousaf-Khan, N. Horeweg, S. van 't Westeinde, M. Prokop, W. P. Mali, F. A. M. Hoesein, P. M. van Ooijen, J. G. Aerts, M. A. den Bakker, E. Thunnissen, J. Verschakelen, R. Vliegenthart, J. E. Walter, K. ten Haaf, H. J. Groen, and M. Oudkerk, "Reduced lung-cancer mortality with volume ct screening in a randomized trial," *New England Journal of Medicine*, vol. 382, no. 6, pp. 503–513, 2020.

[91] Y. Ye, Y. Zhang, N. Yang, Q. Gao, X. Ding, X. Kuang, R. Bao, Z. Zhang, C. Sun, B. Zhou, L. Wang, Q. Hu, C. Lin, J. Gao, Y. Lou, S. H. Lin, L. Diao, H. Liu, X. Chen, G. B. Mills, and L. Han, "Profiling of immune features to predict immunotherapy efficacy," *The Innovation*, vol. 3, no. 1, p. 100194, 2022.

[92] P. Tian, B. He, W. Mu, K. Liu, L. Liu, H. Zeng, Y. Liu, L. Jiang, P. Zhou, Z. Huang, D. Dong, and W. Li, "Assessing pd-l1 expression in non-small cell lung cancer and predicting responses to immune checkpoint inhibitors using deep learning on computed tomography images," *Theranostics 2021*, vol. 11, no. 5, pp. 2098–2107, 2021.

[93] C. Jin, W. Chen, Y. Cao, Z. Xu, Z. Tan, X. Zhang, L. Deng, C. Zheng, J. Zhou, H. Shi, and J. Feng, "Development and evaluation of an ai system for covid-19 diagnosis," *Nature Communications*, vol. 11, 2020.

[94] S. Khandakar, M. A. A. Mamun, M. M. Islam, K. Hossain, M. M. H. Melon, and M. S. Javed, "Unveiling early detection and prevention of cancer: Machine learning and deep learning approaches:," *Educational Administration: Theory and Practice*, vol. 30, no. 5, p. 14614–14628, 2024.