International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

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, SRM Institute of Science and Technology, Kattankulathur, Tamil Nadu 6030203, India

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

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

Received ## Mon. 20##, Revised ## Mon. 20##, Accepted ## Mon. 20##, Published ## Mon. 20##

Abstract: In the current scenario of people's health, lung cancer is the principal cause of cancer-related losses, and its death rates are steadily rising. In any event, radiologists are understaffed and under pressure to work overtime, making it difficult to appropriately assess the increasing volume of image data. Consequently, 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 computer-aided 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. This study presents an overview of deep learning (DL) and machine learning (ML) approaches for applications centered on lung cancer diagnosis and the advancements of the methods being studied. This study focuses on segmentation and classification, which are the two primary deep learning methods for lung cancer detection and screening. The benefits and drawbacks of the deep learning models that are currently in use will also be covered. DL 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:

Every year, 2.20 million people are affected by lung cancer [1], and 75% of those patients pass away from the illness within five years after receiving a diagnosis [2]. Cancer treatment is becoming more and more difficult due to high intra-tumor heterogeneity (ITH) and intelligent cancer cells that develop drug resistance [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 -omics 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 Machine Learning (ML) and Deep Learning (DL) has helped in evaluating and interpreting medical images to accurately diagnose diseases [8]. The development of computer-aided 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 [9], which causes the effective employment of several medical image processing models [10]. 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 ML and DL methodologies used for 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.

A Paper Contributions

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 [11]. The model is reflected as an important tool for medical practitioners to discuss and provide second opinions that may change the patient's life [12].

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) [13] and logistic regression (LR) [14], 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, such as 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 tumor identification. [15,16]. These models usually have three stages: detecting and segmenting nodules, extracting features from nodules, and inferring clinical judgment [8]. Different from feature-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 three CNN 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].

A different study [25] classified H&E-stained histopathological whole-slide images into normal tissue, LUAD, and LUSC using the Inception-V3 network. The model's ability to identify somatic mutations in many lung cancer driver genes, including STK11, EGFR, FAT1, SETBP1, KRAS, and TP53, in each tissue is a significant contribution to this study. 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 sets.

Generative adversarial network (GAN)-based models 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. Conventional machine-learning algorithms cannot comprehend the relationships and patterns in data that these networks can recognize. It is possible to enhance DL algorithms to automatically recognize nodules, cancers, and lesions from CT images. They may also be used to divide the imageries into several structural areas, like the liver, lungs, and brain. This might make it easier for medical practitioners to identify abnormalities in specific bodily sections. There are several advantages to analyzing CT scans using DL. DL algorithms may be trained on vast volumes of data, and they are able to adapt to various input data sources, including CT scans with differing contrast, noise, and resolution levels. Additionally, they can interpret images instantly, which might improve the diagnostic process' effectiveness. DL algorithms increase the precision and consistency of CT imaging analysis, which may help reduce the likelihood of inaccurate findings, such as false positives and false negatives.

Many difficulties exist in applying DL to process CT images. Obtaining sufficient data for medical imaging applications is a major challenge as large, high-quality datasets are essential for training DL algorithms. It is essential to accurately tag and describe these datasets to ensure that algorithms correctly recognize important characteristics and trends. Understanding DL algorithms can be challenging because of their intricate 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.

.

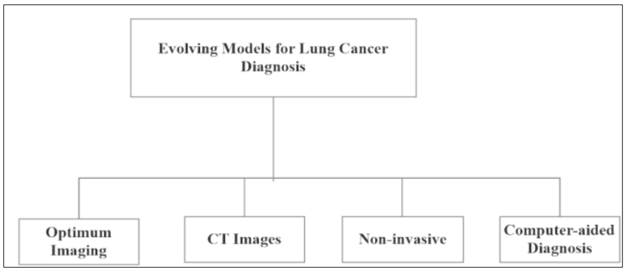


Figure 1. Inputs for Lung Cancer Diagnosis

Convolutional neural networks, or CNNs, are a specific kind of neural network that do very well in picture classification. A CNN filter emulates the function of a neuron 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 2D images from the CT scan, 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 1.

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

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 conventional machine-learning approaches cannot comprehend. As seen in Figure 2, because to its ability to effectively learn complex characteristics from a range of medical image components, including 2D, 3D, low-dose, and MR images, DL or ML algorithms are becoming more and more popular for the identification of lung cancer. CT scans are frequently employed to identify lung cancer because they provide accurate pictures of the lungs. Using CT image analysis to examine the nodules' sizes, shapes, grains, and powers, DL algorithms may identify the impacted areas. Due to its ability to give comprehensive imaging, 3D CT scans, as opposed to 2D CT scans, allow for a more complete examination of lung capacity. DL algorithms have been used in several studies to accurately identify nodules and other anomalies from 3D CT images. It is ideal to utilize image augmentation, denoising, and low-dose CT technologies to minimize radiation exposure during lung cancer screening. MRIs are more capable of revealing information about blood circulation and tissue density than CT images.

DL systems can evaluate MR images to identify lumps and other defects based on their unique characteristics. To enhance image characteristic and minimize noise, filtering methods like Gaussian and Median filtering are commonly used for preprocessing 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 tumors using CT scans. CT images frequently show varying image attenuation patterns for scans of healthy and sick individuals. Segmenting the lungs involves techniques such as numerical methods, gray-level thresholding, and shape-based approaches 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 data. They automatically generated unintended rateable evaluations of individual lung performance that can't be obtained through traditional pulmonary function testing. 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-D CT data from eight healthy participants, a disparity of 0.8 pixels was found.

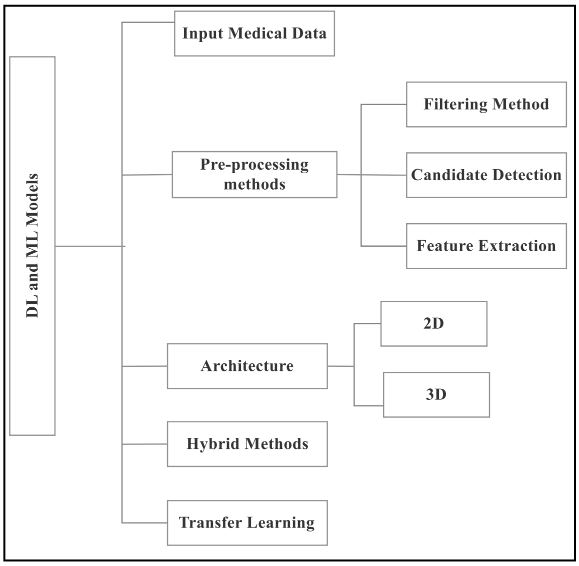


Figure 2. Structure of Procedures in Lung Cancer Diagnosis

By using two sets of classification requirements about size, circularity, and position data together with a pixel-value cutoff applied on a slice-by-slice basis, the study in reference [33] presented a completely automated approach for segmenting lungs. They achieved 94.0% segmentation accuracy using 2969 thick slice pictures and 97.6% segmentation accuracy using 1161 thin slice images in a test with 101 CT patients. Furthermore, an anisotropic filtering approach and a wavelet transform-based interpolation technique were suggested for segmenting and displaying lung volumes in the study 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. Using 403 juxta pleural nodules from the Lung Imaging Database Consortium (LIDC), the research showed a 92.6% reinclusion rate. Additionally, an automatic lung segmentation method and a hybrid geometric active contour model were created [38]. In regions with thin bands or indistinct boundaries, the method performs better when global region and edge information is used. The segmentation approach described in reference [39] begins with a sphere within the target lung and gradually deforms the lung as forces are applied to its boundaries. The algorithm's effectiveness was shown by the 40 CT scans' average F-measure of 99.22%.

Researchers have spent the past decade examining the robustness of CNNs for computer vision tasks. Several methods have been proposed using CNN for analyzing 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. An analogous artificial neural network (ANN) was used in the endeavor to detect lung cancer with an accuracy rate of 96.67% [41]. Additionally, the research revealed that the Instantaneously Trained Neural Networks (DITNN) approach combined with the Improved Profuse Clustering Technique (IPCT) improved lung image quality and raised the accuracy of lung cancer diagnosis to 98.42%. 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 double convolutional deep neural network (CDNN) and a conventional CDNN. 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 with 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. highlighted the significance of early detection 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 convolutional neural network 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 spatial pooling and cropping (SPC) layer.

For the automated classification of lung cancer, a unique Deep Convolutional Neural Network (DCNN) with convolutional, fully connected, and pooling layers has been presented in [49]. With just 76 cancer cases available for training, the DCNN's classification accuracy was only 71%. The research in [50] also suggested a 3D convolutional neural network for computer-assisted lung nodule diagnosis from volumetric CT data. Their model, which comprises of many groups of 3D convolutional layers, fully connected layers, max-pooling layers, and softmax layers, was tested using the LUNA16 dataset. 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. The study [59] demonstrated a DL method with a sensitivity of 93.55%

and a specificity of 91.5% for diagnosing lung cancer. The model developed in [60] assessed the effectiveness of immune checkpoint drugs and applied DL on CT scans to estimate PD-L1 levels in non-small cell lung cancer, with a focus on a case with less nodules. The research [61] demonstrated a DL method to automatically extract information on the stage of lung cancer from CT data, with an F1 score of 0.848. The authors of [62] who suggested a ML approach to detect benign, pre-invasive, and invasive lung nodules on CT scans, demonstrated the effectiveness of a DL-powered CAD system in identifying nodules on 1-mm-thick CT images. Furthermore, the publication [63] suggested a DL model for lung cancer prediction with an accuracy rate of 87.63%. The work in [64] suggests six DL models (CNN, CNN GD, Inception V3, Resnet-50, VGG-16, and VGG-19) for the use of CT scans and histology pictures in the diagnosis of lung cancer. In comparison to other algorithms, CNN GD performs better in terms of accuracy, F-Score, precision, sensitivity, and specificity. It attains 97.86% accuracy, 96.39% sensitivity, 96.79% specificity, and 97.40% sensitivity on an individual basis.

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 radionics 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 18 F-FDG PET/CT images.

C. Lung Cancer 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.

Liquid biopsy based on blood is thought to be a trustworthy technique for early diagnosis. To investigate possible circulating tumor markers, the research makes use of exosomes, methylation, circulating tumor cells (CTCs), microRNA (miRNA), cell-free DNA (cfDNA) fragments, circulating tumor DNA (ctDNA), and methylation. Cell-free DNA (cfDNA) fragments [75], circulating tumor DNA (ctDNA), microRNA (miRNA), methylation, exosomes, and circulating tumor cells (CTCs) are thought to be a valid approach for investigating putative circulating tumor indicators. These liquid biopsy signals have been combined with the help of many discriminative models (SVM, RF, and LR) to accurately diagnose tumors with a high detection rate. Somatic mutations, such as single-base variations (SNVs), insertions, and deletions, often exhibit distinct cancer-type patterns that help classify various lung cancer subtypes [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 signaling pathways. Therefore, the levels of specific proteins differ across various types of cancer [78]. ML models can categorize

ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

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]. Daemen et al. [82] proposed a recurrent hidden Markov model (HMM) that accurately classifies wide chromosomal regions with varying copy numbers. Jurmeister et al. [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 2.

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]	Ensemble model based on elastic net LR; SVM; hierarchical LR& RNA-seq of bronchial brushing samples	Increases risk prediction accuracy.	Small sample numbers in certain subgroups might lead to uneven training.
	[77]	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.

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: Utilized DL methods, the relationship between lung cancer and responses to immunotherapy is 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 techniques for detecting lung cancer [93].
viii. ImageNet: Transfer learning techniques have been applied for the detection of lung cancer [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. Failure to account for data variability can lead to poor performance and limited generalizability of the DL model.

Artifacts like metal, beam hardening, and motion in CT scan images can lower the quality of the image. Pre-processing methods might face challenges in effectively capturing characteristics and details because of the variations and alterations introduced by these anomalies within the dataset. Effective methods for identifying and fixing artifacts are essential for minimizing their impact on future classification tasks.

To train DL models for lung cancer diagnosis and detection, a substantial quantity of labeled data is often required. However, obtaining accurate annotations for CT images may be difficult and time-consuming, especially when dealing with intricate segmentation tasks. Inadequate labeled data might make it more difficult to evaluate and improve DL models, which would lead to worse performance and less generalizability. Furthermore, if multiple datasets are marked up using different criteria and procedures, biases and inconsistencies may appear during the training phase. Convolutional neural networks (CNNs) in DL models may be computationally expensive. DL models need an extensive quantity of processing power for both training and inference.

DL models must be able to generalize over a range of populations and imaging scenarios, which means that preparation strategies need to consider possible biases and limits related to specific datasets. It is vital to use efficient algorithmic strategies, collaborate across institutions, and thoroughly assess the dataset's features to get beyond these pre-processing constraints.

5. Conclusions

This study examines the various lung cancer diagnostic algorithms with varying medical data inputs and identifies their shortcomings. The authors stress the need for additional lung image data from different types of imaging, such as MRI and ultrasound imaging, as well as the necessity to disclose private information to enable comparison and joint investigation. The research aims to create a model for the detection of lung cancer that can distinguish between small malignant tumors and early benign nodules, which would greatly improve the accuracy of diagnosis and therapy.

Additionally, it is recommended that relevant patient data, including genetic reports and medical history, be linked with deep features extracted from lung scan image to improve the efficacy of automated tumor diagnosis. This comprehensive approach may result in a more accurate diagnosis of the ailment. To enhance image quality, the authors suggest using a range of pre-processing methods including filters. Grayscale images may be enhanced by using harmony search and edge-preserving algorithms. These methods provide more accurate diagnostic results and better image analysis. The authors' effective recommendation for remote lung cancer detection justifies their further recommendation to look into the use of AI technology for ML-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.

References:

[1] Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lungcancer. Lancet 2021;398:535–54.
[2] Svoboda E. Artificial intelligence is improving the detection of lung cancer. Nature. 2020 Nov;587(7834):S20-S22.
[3] Ling, S., Hu, Z., Yang, Z., Yang, F., Li, Y., Lin, P., Chen, K., Dong, L., Cao, L., Tao, Y., Hao, L., Chen, Q., Gong, Q., Wu, D., Li, W., Zhao, W., Tian, X., Hao, C., Hungate, E. A., . . . Wu, C. (2015). Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. Proceedings of the National Academy of Sciences, 112(47), E6496-E6505.
[4] International Cancer Genome Consortium, Hudson TJ, Anderson W, Artez A, Barker AD, Bell C, et al. International network of cancer genome projects. Nature 2010;464:993–8.
[5] Pavlopoulou A, Spandidos DA, Michalopoulos I. Human cancer databases (review). Oncol Rep. 2015 Jan;33(1):3-18. doi: 10.3892/or.2014.3579. Epub 2014 Oct 31.
[6] Luo Y, Wang F, Szolovits P. Tensor factorization toward precision medicine. Brief Bioinform. 2017 May 1;18(3):511-514.

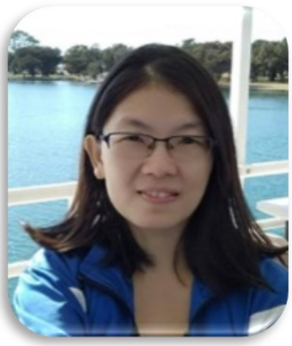
[7] Begum, Shemim, Sarkar, Ram, Chakraborty, Debasis and Maulik, Ujjwal. "Identification of Biomarker on Biological and Gene Expression data using Fuzzy Preference Based Rough Set" Journal of Intelligent Systems, vol. 30, no. 1, 2021, pp. 130-141.
[8] Buty, M., Xu, Z., Gao, M., Bagci, U., Wu, A., Mollura, D.J. (2016). Characterization of Lung Nodule Malignancy Using Hybrid Shape and Appearance Features. In: Ourselin, S., Joskowicz, L., Sabuncu, M., Unal, G., Wells, W. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. MICCAI 2016. Lecture Notes in Computer Science(), vol 9900. Springer, Cham.
[9] Hussein, S., Cao, K., Song, Q., Bagci, U. (2017). Risk Stratification of Lung Nodules Using 3D CNN-Based Multi-task Learning. In: Niethammer, M., et al. Information Processing in Medical Imaging. IPMI 2017. Lecture Notes in Computer Science(), vol 10265. Springer, Cham.
[10] Ciompi F, de Hoop B, van Riel SJ, Chung K, Scholten ET, Oudkerk M, de Jong PA, Prokop M, van Ginneken B. Automatic classification of pulmonary peri-fissural nodules in computed tomography using an ensemble of 2D views and a convolutional neural network out-of-the-box. Med Image Anal. 2015 Dec;26(1):195-202.
[11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60, 6 (June 2017), 84–90.
[12] Buty, M., Xu, Z., Gao, M., Bagci, U., Wu, A., & Mollura, D. J. (2016). Characterization of Lung Nodule Malignancy using Hybrid Shape and Appearance Features.
[13] Gu X, Wang Y, Chan TF, Thompson PM, Yau ST. Genus zero surface conformal mapping and its application to brain surface mapping. Inf Process Med Imaging 2003;18:172–84.
[14] Krizhevsky, Alex & Sutskever, Ilya & Hinton, Geoffrey. (2012). ImageNet Classification with Deep Convolutional Neural Networks. Neural Information Processing Systems. 25.
[15] Kiran Vaidhya Venkadesh, Arnaud A. A. Setio, Anton Schreuder, Ernst T. Scholten, Kaman Chung, Mathilde M. W. Wille, Zaigham Saghir, Bram van Ginneken, Mathias Prokop, and Colin Jacobs, 'Deep Learning for Malignancy Risk Estimation of Pulmonary Nodules Detected at Low-Dose Screening CT'Radiology 2021 300:2, 438-447.
[16] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. ArXiv. /abs/1512.03385.
[17] Gayap, H.T.; Akhloufi, M.A. Deep Machine Learning for Medical Diagnosis, Application to Lung Cancer Detection: A Review. *BioMedInformatics* 2024, *4*, 236-284.
[18] Ardila, D., Kiraly, A.P., Bharadwaj, S. et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 25, 954–961 (2019).
[19] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN.
[20] Lin, T., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal Loss for Dense Object Detection.
[21] Carreira, J., & Zisserman, A. (2017). Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset.
[22] AbdulJabbar, K., Raza, S.E.A., Rosenthal, R. et al. Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. Nat Med 26, 1054–1062 (2020).
[23] Raza, S. E. A., Cheung, L., Shaban, M., Graham, S., Epstein, D., Pelengaris, S., Khan, M., & Rajpoot, N. M. (2019). Micro-Net: A unified model for segmentation of various objects in microscopy images. Medical Image Analysis, 52, 160-173.
[24] Sirinukunwattana K, Ahmed Raza SE, Yee-Wah Tsang, Snead DR, Cree IA, Rajpoot NM. Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images. IEEE Trans Med Imaging. 2016 May;35(5):1196-1206.
[25] Coudray, N., Ocampo, P.S., Sakellaropoulos, T. et al. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nat Med 24, 1559–1567 (2018).
[26] Yao Q, Xiao L, Liu P, Zhou SK. Label-Free Segmentation of COVID-19 Lesions in Lung CT. IEEE Trans Med Imaging. 2021 Oct;40(10):2808-2819.
[27] J. Li, J. Jia and D. Xu, "Unsupervised Representation Learning of Image-Based Plant Disease with Deep Convolutional Generative Adversarial Networks," 2018 37th Chinese Control Conference (CCC), Wuhan, China, 2018, pp. 9159-9163.
[28] Lin, C.-H.; Lin, C.-J.; Li, Y.-C.; Wang, S.-H. Using Generative Adversarial Networks and Parameter Optimization of Convolutional Neural Networks for Lung Tumor Classification. Appl. Sci. 2021, 11, 480.
[29] Ren Z, Zhang Y, Wang S. A Hybrid Framework for Lung Cancer Classification. Electronics (Basel). 2022 May 18;11(10):1614.
[30] Brown MS, McNitt-Gray MF, Mankovich NJ, Goldin JG, Hiller J, Wilson LS, Aberle DR. Method for segmenting chest CT image data using an anatomical model: preliminary results. IEEE Trans Med Imaging. 1997 Dec;16(6):828-39.
[31] Brown MS, Goldin JG, McNitt-Gray MF, Greaser LE, Sapra A, Li KT, Sayre JW, Martin K, Aberle DR. Knowledge-based segmentation of thoracic computed tomography images for assessment of split lung function. Med Phys. 2000 Mar;27(3):592-8.
[32] Hu S, Hoffman EA, Reinhardt JM. Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. IEEE Trans Med Imaging. 2001 Jun;20(6):490-8.
[33] Leader, J. K., Zheng, B., Rogers, R. M., Sciurba, F. C., Perez, A., Chapman, B. E., Patel, S., Fuhrman, C. R., & Gur, D. (2003). Automated lung segmentation in X-ray computed tomography: Development and evaluation of a heuristic threshold-based scheme1. Academic Radiology, 10(11), 1224-1236.
[34] Swierczynski, P., Papież, B. W., Schnabel, J. A., & Macdonald, C. (2018). A level-set approach to joint image segmentation and registration with application to CT lung imaging. Computerized Medical Imaging and Graphics, 65, 58-68.
[35] Farag AA, El Munim HE, Graham JH, Farag AA. A novel approach for lung nodules segmentation in chest CT using level sets. IEEE Trans Image Process. 2013 Dec;22(12):5202-13.
[36] Shen S, Bui AA, Cong J, Hsu W. An automated lung segmentation approach using bidirectional chain codes to improve nodule detection accuracy. Comput Biol Med. 2015 Feb;57:139-49.
[37] Zhang, W., Wang, X., Zhang, P., & Chen, J. (2017). Global optimal hybrid geometric active contour for automated lung segmentation on CT images. Computers in Biology and Medicine, 91, 168-180.
[38] Rebouças Filho PP, Cortez PC, da Silva Barros AC, C Albuquerque VH, R S Tavares JM. Novel and powerful 3D adaptive crisp active contour method applied in the segmentation of CT lung images. Med Image Anal. 2017 Jan;35:503-516.
[39] Zhang C, Sun X, Dang K, Li K, Guo XW, Chang J, Yu ZQ, Huang FY, Wu YS, Liang Z, Liu ZY, Zhang XG, Gao XL, Huang SH, Qin J, Feng WN, Zhou T, Zhang YB, Fang WJ, Zhao MF, Yang XN, Zhou Q, Wu YL, Zhong WZ. Toward an Expert Level of Lung Cancer Detection and Classification Using a Deep Convolutional Neural Network. Oncologist. 2019 Sep;24(9):1159-1165

[40] Nasser, Ibrahim & Abu-Naser, Samy. (2019). Lung Cancer Detection Using Artificial Neural Network. 3. 17-23..
[41] Mehmet Akif Cifci, "SegChaNet: A Novel Model for Lung Cancer Segmentation in CT Scans", Applied Bionics and Biomechanics, vol. 2022, Article ID 1139587, 16 pages, 2022.
[42] Jakimovski, G., & Davcev, D. (2018). Using Double Convolution Neural Network for Lung Cancer Stage Detection. Applied Sciences, 9(3), 427.
[43] J. Wang et al., "Pulmonary Nodule Detection in Volumetric Chest CT Scans Using CNNs-Based Nodule-Size-Adaptive Detection and Classification," in IEEE Access, vol. 7, pp. 46033-46044, 2019..
[44] C. Wang et al., "Pulmonary Image Classification Based on Inception-v3 Transfer Learning Model," in IEEE Access, vol. 7, pp. 146533-146541, 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," in IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 4, pp. 1227-1237, July 2018.
[46] Y. Liu, P. Hao, P. Zhang, X. Xu, J. Wu and W. Chen, "Dense Convolutional Binary-Tree Networks for Lung Nodule Classification," in IEEE Access, vol. 6, pp. 49080-49088, 2018.
[47] Li L, Liu Z, Huang H, Lin M, Luo D. 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. Thorac Cancer. 2019 Feb;10(2):183-192.
[48] Jin, H., Li, Z., Tong, R. and Lin, L. (2018), A deep 3D residual CNN for false-positive reduction in pulmonary nodule detection. Med. Phys., 45: 2097-2107.
[49] Teramoto A, Tsukamoto T, Kiriyama Y, Fujita H. Automated Classification of Lung Cancer Types from Cytological Images Using Deep Convolutional Neural Networks. Biomed Res Int. 2017;2017:4067832.
[50] Dou Q, Chen H, Yu L, Qin J, Heng PA. Multilevel Contextual 3-D CNNs for False Positive Reduction in Pulmonary Nodule Detection. IEEE Trans Biomed Eng. 2017 Jul;64(7):1558-1567.
[51] Wang, C., Shao, J., Lv, J., Cao, Y., Zhu, C., Li, J., Shen, W., Shi, L., Liu, D., & Li, W. (2021). Deep learning for predicting subtype classification and survival of lung adenocarcinoma on computed tomography. Translational Oncology, 14(8), 101141.
[52] Gu Y, Chi J, Liu J, Yang L, Zhang B, Yu D, Zhao Y, Lu X. A survey of computer-aided diagnosis of lung nodules from CT scans using deep learning. Comput Biol Med. 2021 Oct;137:104806.
[53] Ardila, D., Kiraly, A.P., Bharadwaj, S. et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med 25, 954–961 (2019).
[54] Shao J, Wang G, Yi L, Wang C, Lan T, Xu X, Guo J, Deng T, Liu D, Chen B, Yi Z, Li W. Deep Learning Empowers Lung Cancer Screening Based on Mobile Low-Dose Computed Tomography in Resource-Constrained Sites. Front Biosci (Landmark Ed). 2022 Jul 4;27(7):212.
[55] Wang C, Xu X, Shao J, Zhou K, Zhao K, He Y, Li J, Guo J, Yi Z, Li W. Deep Learning to Predict EGFR Mutation and PD-L1 Expression Status in Non-Small-Cell Lung Cancer on Computed Tomography Images. J Oncol. 2021 Dec 31;2021:5499385.
[56] Li, R.; Xiao, C.; Huang, Y.; Hassan, H.; Huang, B. Deep Learning Applications in Computed Tomography Images for Pulmonary Nodule Detection and Diagnosis: A Review. Diagnostics 2022, 12, 298.
[57] S.K., L., Mohanty, S. N., K., S., N., A., & Ramirez, G. (2019). Optimal deep learning model for classification of lung cancer on CT images. Future Generation Computer Systems, 92, 374-382.
[58] Lee, S.M.; Seo, J.B.; Yun, J.; Cho, Y.H.; Vogel-Claussen, J.; Schiebler, M.L.; Gefter,W.B.; van Beek, E.J.; Goo, J.M.; Lee, K.S.; et al.Deep learning applications in chest radiography and computed tomography. J. Thorac. Imaging 2019, 34, 75–85.
[59] Bhatia, S., Sinha, Y., Goel, L. (2019). Lung Cancer Detection: A Deep Learning Approach. In: Bansal, J., Das, K., Nagar, A., Deep, K., Ojha, A. (eds) Soft Computing for Problem Solving. Advances in Intelligent Systems and Computing, vol 817. Springer, Singapore.
[60] Tian P, He B, Mu W, Liu K, Liu L, Zeng H, Liu Y, Jiang L, Zhou P, Huang Z, Dong D, Li W. 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 Jan 1;11(5):2098-2107...
[61] Hu D, Zhang H, Li S, Wang Y, Wu N, Lu X. Automatic Extraction of Lung Cancer Staging Information From Computed Tomography Reports: Deep Learning Approach. JMIR Med Inform. 2021 Jul 21;9(7):e27955..
[62] Ashraf SF, Yin K, Meng CX, Wang Q, Wang Q, Pu J, Dhupar R. Predicting benign, preinvasive, and invasive lung nodules on computed tomography scans using machine learning. J Thorac Cardiovasc Surg. 2022 Apr;163(4):1496-1505.
[63] Subramanian, R. Raja & Mourya, Ravella & Prudhvi, V & Reddy, Teja & Reddy, B & Amara, Srikar. (2020). Lung Cancer Prediction Using Deep Learning Framework. International Journal of Control and Automation. 13. 154-160..
[64] Rajasekar, V., Vaishnnave, M., Premkumar, S., Sarveshwaran, V., & Rangaraaj, V. (2023). Lung cancer disease prediction with CT scan and histopathological images feature analysis using deep learning techniques. Results in Engineering, 18, 101111.
[65] Wankhade, S., & S., V. (2023). A novel hybrid deep learning method for early detection of lung cancer using neural networks. Healthcare Analytics, 3, 100195.
[66] Abunajm, S., Elsayed, N., ElSayed, Z., & Ozer, M. (2023). Deep Learning Approach for Early Stage Lung Cancer Detection.
[67] Avanzo M, Stancanello J, Pirrone G, Sartor G. Radiomics and deep learning in lung cancer. Strahlenther Onkol. 2020 Oct;196(10):879-887.
[68] Chao, H., Shan, H., Homayounieh, F. et al. Deep learning predicts cardiovascular disease risks from lung cancer screening low dose computed tomography. Nat Commun 12, 2963 (2021).
[69] Shakeel, P. M., Burhanuddin, M., & Desa, M. I. (2019). Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks. Measurement, 145, 702-712.
[70] Zhang, Q., Wang, H., Yoon, S. W., Won, D., & Srihari, K. (2018). Lung Nodule Diagnosis on 3D Computed Tomography Images Using Deep Convolutional Neural Networks. Procedia Manufacturing, 39, 363-370.
[71] L. Zhao et al., "A Weighted Discriminative Extreme Learning Machine Design for Lung Cancer Detection by an Electronic Nose System," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-9, 2021.
[72] L. Chen et al., "Multimodality Attention-Guided 3-D Detection of Nonsmall Cell Lung Cancer in 18F-FDG PET/CT Images," in IEEE Transactions on Radiation and Plasma Medical Sciences, vol. 6, no. 4, pp. 421-432, April 2022.

[73] National Lung Screening Trial Research Team, Church TR, Black WC, Aberle DR, Berg CD, Clingan KL, et al. Results of initial low-dose computed tomographic screening for lungcancer. N Engl J Med 2013;368:1980–91.
[74] Alexandrov, L., Nik-Zainal, S., Wedge, D. et al. Signatures of mutational processes in human cancer. Nature 500, 415–421 (2013).
[75] Herath, S., Rad, H. S., Radfar, P., Ladwa, R., Warkiani, M., & Kulasinghe, A. (2021). The Role of Circulating Biomarkers in Lung Cancer. Frontiers in Oncology, 11.
[76] Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular biology of the cell. 4th ed. New York: GarlandScience; 2002.
[77] Kobayashi K, Bolatkan A, Shiina S, Hamamoto R. Fully-Connected Neural Networks with Reduced Parameterization for Predicting Histological Types of Lung Cancer from Somatic Mutations. Biomolecules. 2020 Aug 28;10(9):1249.
[78] Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhori G, Benfeitas R, Arif M, Liu Z, Edfors F, Sanli K, von Feilitzen K, Oksvold P, Lundberg E, Hober S, Nilsson P, Mattsson J, Schwenk JM, Brunnström H, Glimelius B, Sjöblom T, Edqvist PH, Djureinovic D, Micke P, Lindskog C, Mardinoglu A, Ponten F. A pathology atlas of the human cancer transcriptome. Science. 2017 Aug 18;357(6352):eaan2507.
[79] Aliferis, Constantin F. et al. "Machine Learning Models for Classification of Lung Cancer and Selection of Genomic Markers Using Array Gene Expression Data." The Florida AI Research Society (2003)..
[80] Shao, X., Lv, N., Liao, J. et al. Copy number variation is highly correlated with differential gene expression: a pan-cancer study. BMC Med Genet 20, 175 (2019).
[81] Aliferis CF, Hardin D, Massion PP. Machine learning models for lung cancer classification using array comparative genomic hybridization. Proc AMIA Symp. 2002:7-11.
[82] Daemen A, Gevaert O, Leunen K, Legius E, Vergote I, De Moor B. Supervised classification of array CGH data with HMM-based feature selection. Pac Symp Biocomput. 2009:468-79.
[83] Jurmeister P, Bockmayr M, Seegerer P, Bockmayr T, Treue D, Montavon G, Vollbrecht C, Arnold A, Teichmann D, Bressem K, Schüller U, von Laffert M, Müller KR, Capper D, Klauschen F. Machine learning analysis of DNA methylation profiles distinguishes primary lung squamous cell carcinomas from head and neck metastases. Sci Transl Med. 2019.
[84] Luo Y, Riedlinger G, Szolovits P. Text mining in cancer gene and pathway prioritization. Cancer Inform. 2014 Oct 13;13(Suppl 1):69-79.
[85] Podolsky MD, Barchuk AA, Kuznetcov VI, Gusarova NF, Gaidukov VS, Tarakanov SA. Evaluation of Machine Learning Algorithm Utilization for Lung Cancer Classification Based on Gene Expression Levels. Asian Pac J Cancer Prev. 2016;17(2):835-8.
[86] Liang W, Zhao Y, Huang W, Gao Y, Xu W, Tao J, Yang M, Li L, Ping W, Shen H, Fu X, Chen Z, Laird PW, Cai X, Fan JB, He J. Non-invasive diagnosis of early-stage lung cancer using high-throughput targeted DNA methylation sequencing of circulating tumor DNA (ctDNA). Theranostics. 2019 Apr 6;9(7):2056-2070.
[87] Choi, Y., Qu, J., Wu, S. et al. Improving lung cancer risk stratification leveraging whole transcriptome RNA sequencing and machine learning across multiple cohorts. BMC Med Genomics 13 (Suppl 10), 151 (2020).
[88] Mathios D, Johansen JS, Cristiano S, Medina JE, Phallen J, Larsen KR, et al. Detection and characterization of lung cancer using cell-free DNA fragmentomes. Nat Commun 2021;12:5060.
[89] Ozdemir O, Russell RL, Berlin AA (2020) A 3D probabilistic deep learning system for detection and diagnosis of lung cancer using low-dose ct scans. IEEE Transactions Medical Imaging 39(5):1419,pp.395–409.
[90] The National Lung Screening Trial Research Team. Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening. N. Engl. J. Med. 2011, 365, 395–409.
[91] Ye Y, Zhang Y, Yang N, Gao Q, Ding X, Kuang X, Bao R, Zhang Z, Sun C, Zhou B, Wang L, Hu Q, Lin C, Gao J, Lou Y, Lin SH, Diao L, Liu H, Chen X, Mills GB, Han L. Profiling of immune features to predict immunotherapy efficacy. Innovation (Camb). 2021 Dec 2;3(1):100194.
[92] Tian, P., He, B., Mu, W., Liu, K., Liu, L., Zeng, H., Liu, Y., Jiang, L., Zhou, P., Huang, Z., Dong, D., & Li, W. (2021). 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, 11(5), 2098-2107.
[93] Cheng Jin, Weixiang Chen, Yukun Cao, Zhanwei Xu, Zimeng Tan, Xin Zhang, Lei Deng, Chuansheng Zheng, Jie Zhou, Heshui Shi & Jianjiang Feng, Development and evaluation of an artificial intelligence system for COVID-19 diagnosis, Nature Communications, 2020.

[94] Prior, F., Smith, K., Sharma, A. et al. The public cancer radiology imaging collections of The Cancer Imaging Archive. Sci Data 4, 170124 (2017).

Jayapradha J is an Assistant Professor at department of Computing Technologies, SRMIST. She had completed her Ph.D(CSE) in Privacy Preservation. and has published many research articles on ML, data mining, and privacy. She has 6 patents published and 2 patents granted in the field of ML and data privacy. Her research interests include ML, databases, and privacy.



Su-Cheng Haw is Professor at Faculty of Computing and Informatics, Multimedia University, where she leads several funded projects on the XML databases. Her research interests include XML databases, query optimization, data modeling, semantic web, and recommender system.

International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. #, No.# (Mon-20..)

Palanichamy Naveen joined the Faculty of Computing and Informatics, Multimedia University after receiving Ph.D from Curtin University, Malaysia. She received her Bachelor of Engineering (CSE) and Master of Engineering (CSE) from Anna University, India. Her research interest includes Smart Grid, Cloud Computing, ML, DL and Recommender system.

Elham Abdulwahab Anaam is a Postdoctoral at Faculty of Computing and Informatics, Multimedia University. Her research focuses on E-CRM adoption, multi-criteria decision-making, hybrid recommendation system and blockchain technology.

Senthil Kumar T graduated with a bachelor’s degree in Electronics and Communication Engineering and a master’s degree in computer science and engineering in 2004 and 2006, respectively. He completed his Ph.D. in CSE from SRMIST, Chennai, 2019. His areas of expertise include wireless communication, network security, ML, the Internet of Things, and vehicular networks.