

Review Article

Mustafa Mohammed Jassim* and Mustafa Musa Jaber

Systematic review for lung cancer detection and lung nodule classification: Taxonomy, challenges, and recommendation future works

<https://doi.org/10.1515/jisys-2022-0062>

received December 30, 2021; accepted May 27, 2022

Abstract: Nowadays, lung cancer is one of the most dangerous diseases that require early diagnosis. Artificial intelligence has played an essential role in the medical field in general and in analyzing medical images and diagnosing diseases in particular, as it can reduce human errors that can occur with the medical expert when analyzing medical image. In this research study, we have done a systematic survey of the research published during the last 5 years in the diagnosis of lung cancer classification of lung nodules in 4 reliable databases (Science Direct, Scopus, web of science, and IEEE), and we selected 50 research paper using systematic literature review. The goal of this review work is to provide a concise overview of recent advancements in lung cancer diagnosis issues by machine learning and deep learning algorithms. This article summarizes the present state of knowledge on the subject. Addressing the findings offered in recent research publications gives the researchers a better grasp of the topic. We checked all the characteristics, such as challenges, recommendations for future work were analyzed in detail, and the published datasets and their source were presented to facilitate the researchers' access to them and use it to develop the results achieved previously.

Keywords: deep learning, lung cancer, lung carcinoma, lung nodule classification, convolution neural network

1 Introduction

In 2020, an estimated 19.3 million new cancer cases and 10 million deaths occurred worldwide. Lung cancer ranked second in the number of patients with about 11.4% of the total cancer cases, with an estimated 2.2 million lung cancer cases only. Lung cancer is the number one cause of death among other cancers, with fatalities accounting for 18% of all cancers. Smoking is a major cause of lung cancer; in some countries, smoking has peaked or continues to increase. This suggests that lung cancer rates will increase for at least decades [1]. Figure 1 shows the incidence and mortality of ten common types of cancer in 2020.

It has been claimed that if the condition is found early, diagnosed, and treated effectively, the patient's chances of living a long life are increased.

The survival rate for lung cancer patients 5 years after diagnosis ranges from 10 to 20%. Low-dose computed tomography (CT) screening can help detect lung cancer early that makes the condition more responsive to treatment [1]. In general, It has been reported that if the cancer case is found early, diagnosed, and treated effectively, the patient's chances of living a long life increase [2]. Analyzing medical data and diagnosing diseases require a medical expert, and experts' opinions often differ when analyzing medical images because of their complexity. Artificial intelligence has played an essential role in the medical field.

* **Corresponding author: Mustafa Mohammed Jassim**, Department of Computer Science, Informatics Institute for Postgraduate Studies (IIPS), Iraqi Commission for Computers and Informatics (ICCI), Baghdad, Iraq, e-mail: ms202010619@iips.icci.edu.iq
Mustafa Musa Jaber: Department of Medical Instruments Engineering Techniques, Dijlah University College, Baghdad, 10021, Iraq; Department of Medical Instruments Engineering Techniques, Al-Farahidi University, Baghdad, 10021, Iraq

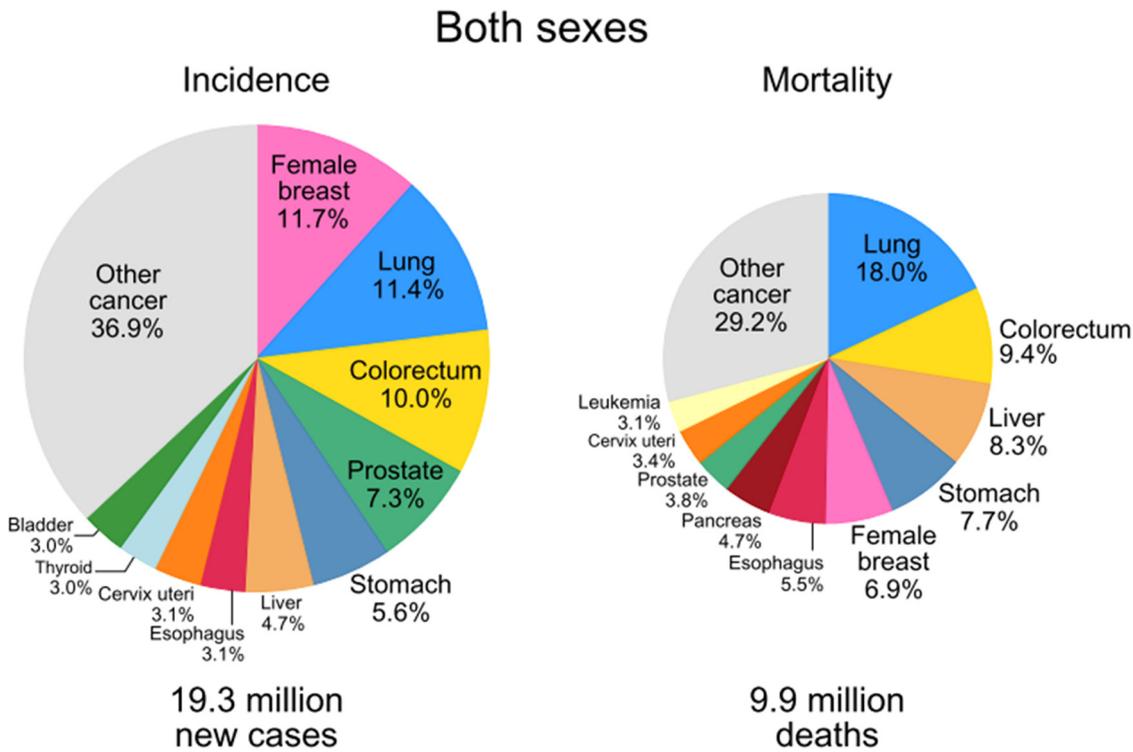


Figure 1: Case and mortality chart for ten most common types of cancer for both sexes [1].

Machine learning (ML) and deep learning (DL) algorithms have been used in analyzing and processing medical images and diagnosing diseases in recent years, as it provides exciting solutions for applications in medicine [3]. Providing a prediction system that makes accurate diagnoses is still challenging, and the field of research is still ongoing. This study aims to create a survey of what researchers have reached in developing the efficiency and accuracy of lung cancer diagnostic systems’ performance in the last 5 years. This study will give an accurate view of each research paper, starting with data acquisition. The pre-processing approach used, what methodologies are used to train the model, evaluate the model, review the results reached, examine the mentioned challenges, published dataset with its source, size, number of samples, and researchers’ suggestions for future work. This study was organized as follows: the second section explained the method used to collect the relevant literature for the study. The third section included the results of the technique used and divided the articles into groups to facilitate their understanding. The fourth section included the review and the survey, where the methods used in each article and their results were extracted. The fifth section is to discuss what was presented in the papers. The sixth section is to explain the limitations of the study. The seventh section is devoted to the conclusion of the study.

2 Method

The review style was followed by the systematic review (PRISMA). It is one of the strategies for evaluating intellectual output that focuses on a research topic and attempts to locate, assess, choose, and prepare all high-quality scientific evidence and research articles on specific issues [4]. This article used four digital databases, namely, Science direct, Scopus, Web of Science (WoS), and IEEE. Science direct provides access to research papers published in highly authoritative journals in the fields of technology and science; Scopus is a reliable source for all disciplines such as science, medicine, engineering, and technology. WoS is a highly specialized resource in the sciences, social sciences, humanities, engineering, arts, and interdisciplinary studies. IEEE Xplore delivers full-text access to the world’s highest quality technical literature in engineering and technology. The four databases cover the academic aspects for lung cancer detection and nodule classification; this review can contribute to saving lives by contributing to providing suggestions for the development of diagnostic systems using DL algorithms.

2.1 Search strategy

The search was conducted in the four previously mentioned databases for research published from 2017 to 2021 in the English language exclusively. These indicators were chosen because they can cover all developments in the early detection of lung cancer using DL algorithms. We searched using a query consisting of keywords related to improvement or development on detection systems (“improving” OR “improvement” OR “enhancement”), disease-specific keywords as defined in the literature (e.g., “lung cancer” OR “lung carcinoma”), keywords related to detection, diagnosis or prediction (“diagnosis” OR “detection” OR “predictive” OR “prediction), the type of algorithms used to make a prediction (“deep learning” OR “artificial intelligence” OR “machine learning”). The search and selection were made according to Figure 2.

2.2 Inclusion criteria

- 1 – Articles published in scientific journals in English only.
- 2 – Articles focused on developing different systems, applications, technologies, and algorithms to diagnose lung cancer using ML and DL.
- 3 – The articles that studied lung cancer detection or lung nodules classification.

2.3 Exclusion criteria

- 1 – Articles that detect lung cancer or nodules classification with another disease (e.g., detect lung cancer and covid-19).
- 2 – Articles that detect lung nodules without classification.

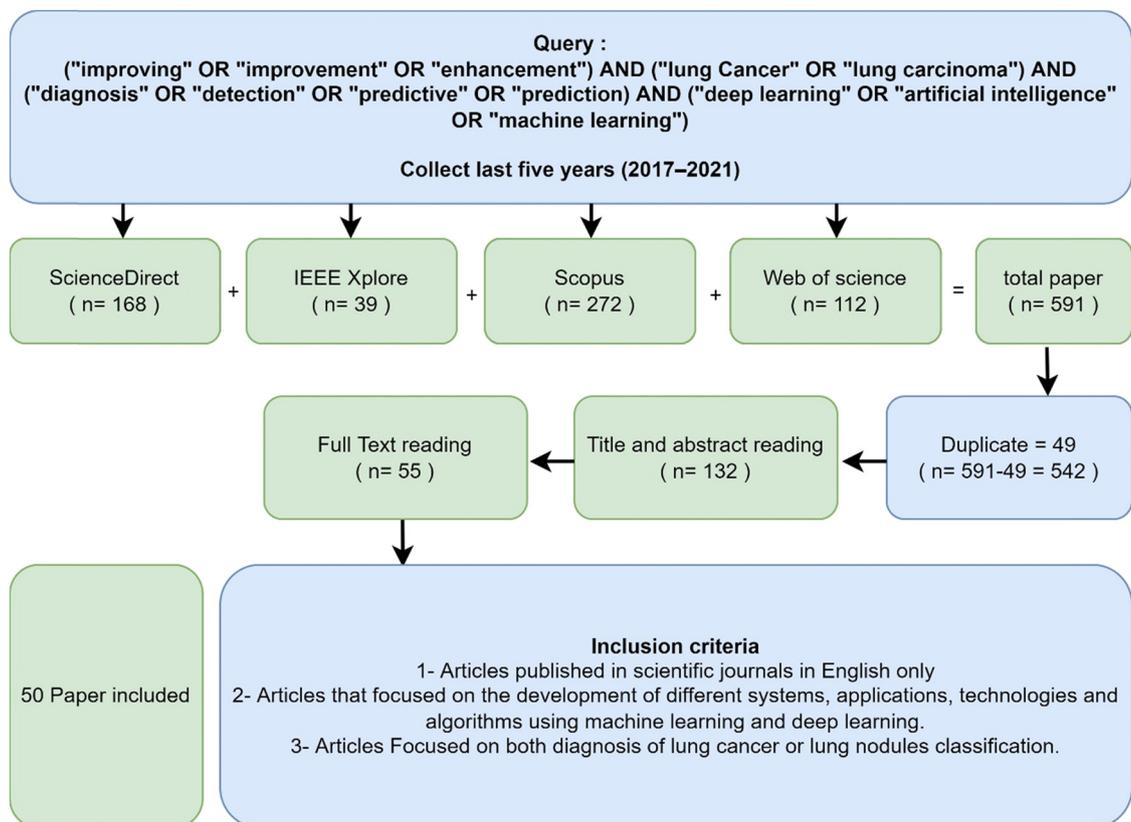


Figure 2: Study selection flowchart with exact query and inclusion criteria.

3 Result

The results of the query search were 591 articles from the four databases. After that, the duplicative articles ($n = 49$) were removed, which means the number of articles were $n = 542$. The second stage of the examination was done by scanning the title and abstract, where the number of articles became $n = 132$; in the final filter, we read the articles' full text, but we were unable to download 10 papers due to access issues, that resulted in 122 articles and finally based on inclusion criteria, we have selected 55 related articles. After that, a complete reading of all the selected articles was done, and a suggested classification was made. We found that researchers worked in two areas: the first field is lung cancer detection in general, and the second field is the classification of lung nodules as benign or malignant. In both fields, we found that some researchers used traditional ML algorithms and others who are the majority used DL algorithms; it was found that the convolutional neural network algorithm is the most widely used among other DL algorithms. Our relevant studies were subsequently classified based on the similarity of the methods and algorithms used, as in Figure 3.

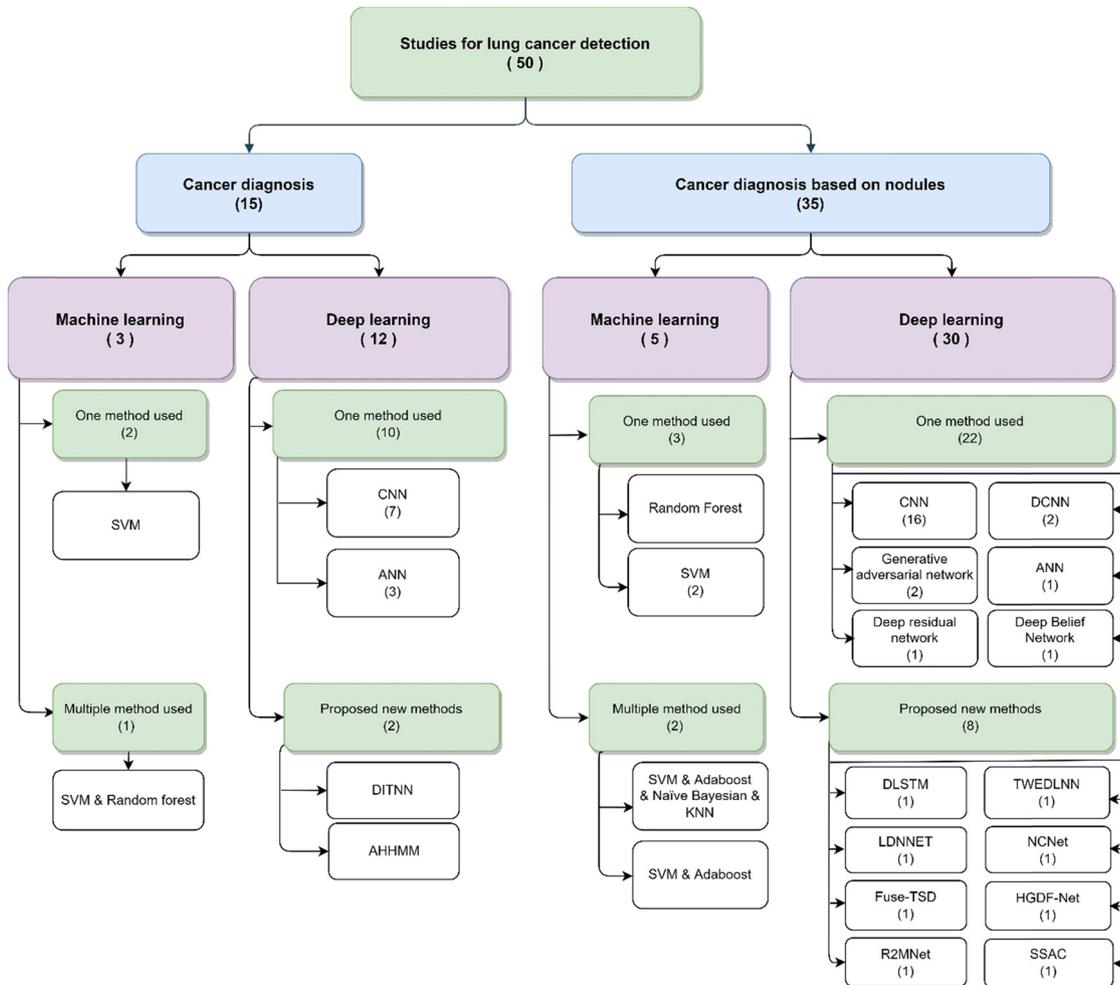


Figure 3: Taxonomy of studies of lung cancer diagnosis.

3.1 Cancer diagnosis

In this category, articles were placed that dealt with the detection of lung cancer in general without touching on the details of diagnosing lung nodules, and they were also divided into two sub-categories, the first sub-category that worked with traditional ML algorithms, numbering 2, and they were also divided into two parts for articles that worked in one method [4,5] and the other section for articles that used more than one method [7]. As for the second sub-category of articles that used DL algorithms, it was divided into two parts: the first section was for papers that used one traditional method and their number was 10 [7–16] and the other section was for articles that suggested new methods, there were two articles [17,18].

3.2 Cancer diagnosis based on nodules

In this category, articles were placed that defined the diagnosis of cancer based on the classification of lung nodules into benign or malignant, and this section was divided into two sub-categories; the first sub-category is about articles that used ML algorithms, which numbered 5, and was also divided into articles that used one method, numbering 3 [19,20], and the other that used more than one method, which numbered 2 [22,23]. The second subcategory is about articles that used DL algorithms, which numbered 30 articles and divided into articles that used one traditional method, numbering 22 [24,25], and another that suggested new methods, and their number was 8 [46,47]. All studies that depended on diagnosing lung nodules used CT scan data because it is the only one that can show lung nodules. Most of them used two (malignant cases and benign cases) or three (normal cases, malignant cases, and benign cases) class labels as shown in Figure 4; unlike X-ray images, they do not show pulmonary nodules.

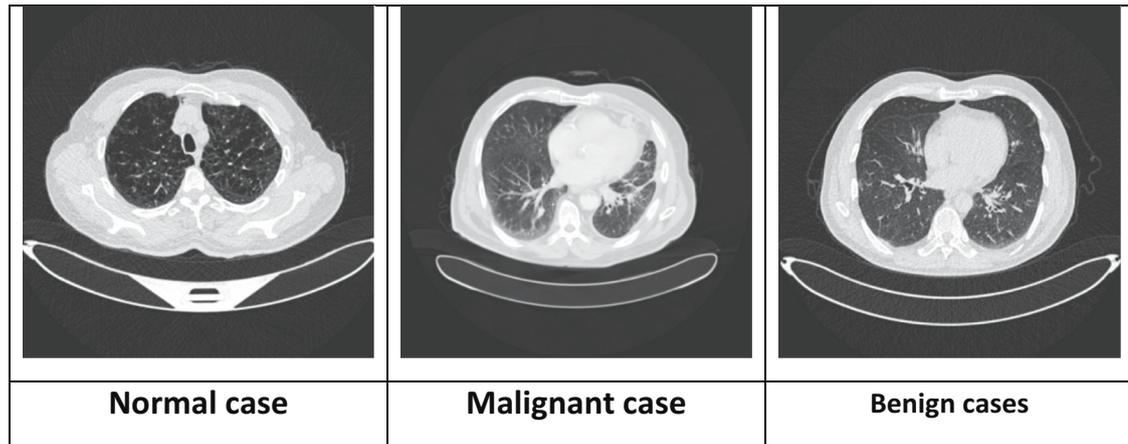


Figure 4: Sample of three types of CT SCAN of patients in classification operation.

4 Reviews and surveys

The main objective of this article and the laboratory articles on a particular topic is to understand the direction and ideas of the studies and to give suggestions for future work based on the previous articles. In the past, it was challenging to implement artificial intelligence applications in medicine due to the lack of medical data such as chest x-ray images because the accuracy of classification depends on the amount of training data, but now it is huge, which makes it easy to apply ML and DL algorithms [3]. For lung cancer, many data were available that were used to train DL algorithms and help them in the diagnosis process, such as CT Scan, Chest X-ray, magnetic resonance imaging (MRI), Contrast-Enhanced Computed

Tomography (CE-CT), Positron Emission Tomography (PET) Scan, etc. [55]. When we reviewed the 50 selected articles, we found that the articles in recent years used low-dose CT scan image dataset and moved away from chest X-ray image dataset because it did not contribute much to the early detection of lung tumors and a false negative in it also causes a significant delay in diagnosis. The low-dose CT scan (LDCT) images have been widely used because it can show areas where there is an abnormal growth and can be classified as malignant (cancerous) [55]. The National Lung Screening Trial (NLST) established in 2011 that those who were screened using LDCT had a lower risk of death by about (15–20%) than their peers who were screened with chest X-rays [56]. We found traditional ML algorithms in the selected 13 articles, and the largest share was for DL algorithms, as the number of articles that used the convolutional neural network algorithm reached 22 articles. It is the most widely used algorithm because it has been proven to be the best in classifying images [55]. The number of usages of the rest of the DL algorithms is displayed in Figure 5.

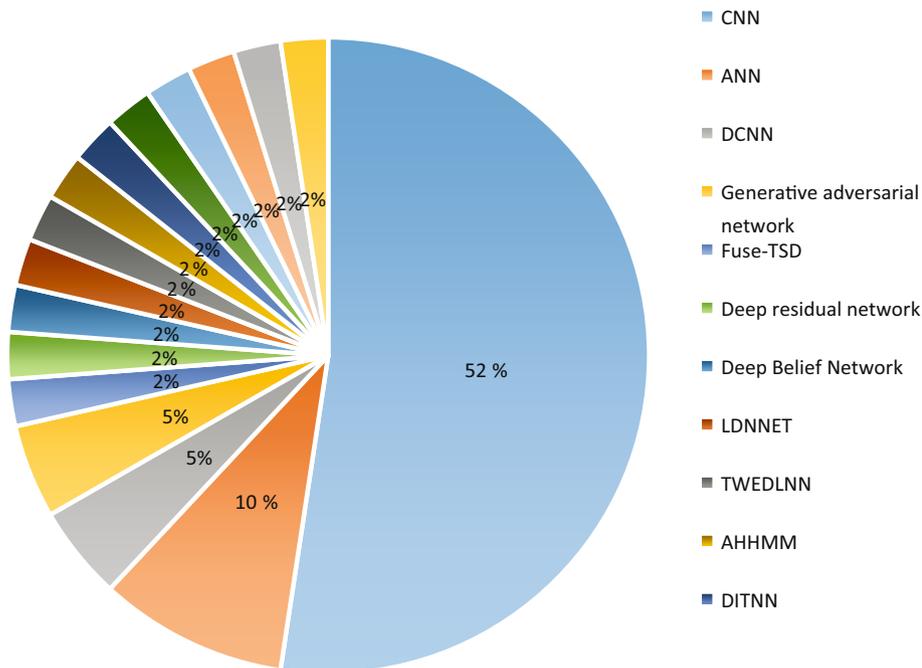


Figure 5: Distribution ratios for deep learning algorithms.

In this study, various data and information related to the topic were collected from related articles, such as the objectives of the studies and challenges and recommendations for future work that we will explain in detail. The distribution of the selected papers according to the years of publication is illustrated in Figure 6. As it turned out that the year 2020 was the most published year in which many articles related to our field of interest were published. Then comes the year 2021.

Table 1 gives the details of each article, such as the name of the dataset used in the article, the pre-processing methods, and the methods used in the diagnosis process. We reviewed the best result that the

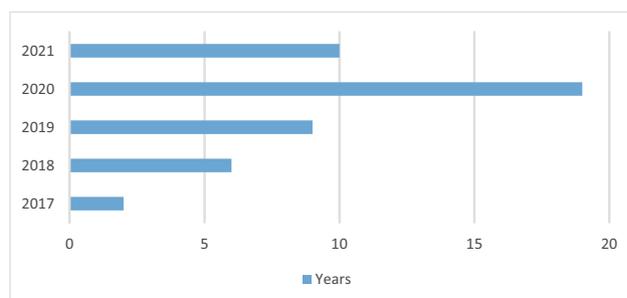


Figure 6: Distribution of articles based on years.

Table 1: Result and approaches of studies

Ref.	Dataset	Pre-processing Approaches	Methods used	Results (Metrics of CADx)	AUC	Precision	Specificity	Sensitivity
[23]	Lung Image Database Consortium image collection (LIDC-IDRI)	<ul style="list-style-type: none"> Optimal gray-level threshold Connected component labeling Hole filling and smoothing Segmented lung 	KNN Adaboost Naive Bayesian SVM 3D CNNs	99.20%	—	—	98.00%	98.30%
[11]	LIDC-IDRI for LUNA16 challenge	<ul style="list-style-type: none"> Clipping the scanned image to remove most of the variation due to bone and resampling the images Images were normalized to have a mean voxel value of 0 and variance of 1 To reduce overfitting and maximize the transfer learning across images taken with different setups Converting into gray scale Binarization Contourlet filter Laplacian filter Vessel tree segmentation The scans are clipped based on the HU Normalization of the images RFV-Net: A novel segmentation 	Deep neural network	94%	87%	—	—	93%
[7]	Private Data	<ul style="list-style-type: none"> Contrast stretching Median filter Otsu segmentation Volume filtering 3D median filter 	<ul style="list-style-type: none"> SVM Random forest Kernel-induced random forest 	94.50%	—	—	77.60%	74.20%
[20]	LIDC-IDRI	<ul style="list-style-type: none"> Anisotropic diffusion filtering Super pixel segmentation 	CNN	94%	—	93.30%	—	92.80%
[51]	LIDC-IDRI for LUNA16 challenge	<ul style="list-style-type: none"> Normalization of the images RFV-Net: A novel segmentation 	NCNet	98.21%	—	—	98.33%	98.38%
[9]	FDG-PET / CT imaging (Private dataset)	<ul style="list-style-type: none"> Contrast stretching Median filter Otsu segmentation Volume filtering 3D median filter 	SVM	97.10%	—	—	98.10%	95.90%
[57]	CT Scan and MRI images (private dataset)	<ul style="list-style-type: none"> Contrast stretching Median filter Otsu segmentation Volume filtering 3D median filter 	SVM	96.15%	—	—	—	—
[12]	—	<ul style="list-style-type: none"> Anisotropic diffusion filtering Super pixel segmentation 	CNN	—	91.30%	—	—	99.60%
[26]	LIDC-IDRI	<ul style="list-style-type: none"> Anisotropic diffusion filtering Super pixel segmentation 	CNN	97.60%	—	—	—	—

(Continued)

Table 1: Continued

Ref.	Dataset	Pre-processing Approaches	Methods used	Results (Metrics of CADx)	AUC	Precision	Specificity	Sensitivity
	In-house clinical dataset (private dataset)			ACC				
[27]	LIDC-IDRI	—	CNN DBN SDAE	—	89.90%	—	—	—
[28]	LIDC-IDRI	— Slice thickness adjustment — Image pixel value adjustment — Feature data normalization — Lung parenchyma segmentation	3D CNN	94.40%	93.30%	—	92.60%	90.90%
[13]	Data science bowl 2017	—	Hybrid 3D CNN – SVM	91.81%	—	—	94.23%	88.53
[14]	Endobronchial ultrasonography (EBUS) images	—	Fine-tuned CaffeNet-SVM	85.40%	87.05%	90.4	82.10%	87.00%
[29]	Dataset from National Cancer Institute Data Access System	Rotation and flipping for augmentation data	CNN	90.29%	—	—	—	73%
[24]	LUNGx	— Contrast enhancement	SVM	90.46%	94.49%	—	92.56	90.10%
[18]	CT scan dataset form http://diagnijmegen.nl/	— Image histogram methods — K-mean algorithm for segmentation	AdaBoost AHHMM	98.90%	—	—	—	—
[15]	Seven independent datasets used in this study	This study used linear and nearest neighbor interpolations for the image	3D convolutional neural network	—	71%	—	—	—
[41]	(HarvardRT, Radboud, Moffitt, Maastr, MUMC, M-SPORE, RIDER) LIDC-IDRI for LUNA16 challenge	— Thresholding applied to highlight on pulmonary region — Applied linear metamorphosis — The mean is calculated for all images to zero-center the data — Other areas such as air, bones, water, and blood were hidden — 3D segmentation network	— Deep residual network — XGBoost classifier	Acc of diagnosing = 0.883 Acc of segmentation = 95.8%	—	90%	89.66%	87.10%

(Continued)

Table 1: Continued

Ref.	Dataset	Pre-processing Approaches	Methods used	Results (Metrics of CADx) ACC	AUC	Precision	Specificity	Sensitivity
[54]	LIDC-IDRI	– MC-CLAE – Segmentation using GOT	TWEDLNN	96%	–	94%	97%	94%
[16]	Private data from Zhongshan Hospital	–	CNN	70.30%	–	–	77.80%	62.80%
[30]	– LIDC-IDRI – SPIE- American Association of Physicians of Medicine (AAPM) for LUNGx challenge	Without any pre-processing	texture CNN	LIDC-IDRI 0.9669	–	–	–	LIDC-IDRI 0.9605
[48]	LIDC-IDRI	U-Net for segmentation	Fuse-TSD	89.53%	96.65%	–	92.02%	LUNGx = 88.76%
[21]	Data from Shanghai Pulmonary Hospital	Segmentation applied by four steps: 1) Pulmonary nodule enhancement and 2) Initial nodule segmentation 3) Vascular structures removal 4) Boundary reparation	SVM	–	95%	–	–	84.17%
[39]	LIDC-IDRI	– Standardized images – Convert to gray scale image	15-layer 2D DCNN	97.20%	–	–	97.30%	96.00%
[31]	LIDC-IDRI	For segmentation, the author used a proposed method that applied a threshold to separate air from tissues	3D CNN	88%	–	–	–	96%
[52]	LIDC-IDRI	– Normalization – Extraction: the foreground regions of normalized CT scans were extracted	R2MNet	94.74%	96%	–	–	–
[58]	LIDC-IDRI	– Resampling – Cropping nodule regions used to train and test A proposed new method for segmentation similar to 3D-UNet	End-to-end multi-task 3D CNN	97.58	–	–	–	–
[59]	LIDC-IDRI	Hyper-attention mechanism (HAM)	HGDF-Net	93.52%	–	–	94.62%	91.09%

(Continued)

Table 1: Continued

Ref.	Dataset	Pre-processing Approaches	Methods used	Results (Metrics of CADx) ACC	AUC	Precision	Specificity	Sensitivity
[50]	– LIDC-IDRI for LUNA16 challenge – Kaggle DSB 2017	– Median filter – Laplacian filter – Enhanced contrast	LDNNET	Both datasets = 98%	Luna16 = 98.8% Dsb 2017 = 0.999	Luna16 = 97.3% Dsb 2017 = 0.999	Luna16 = 99.4% Dsb 2017 = 0.999	Luna16 = 98.20% Dsb 2017 = 0.999
[19]	CT lung images collected from the Cancer imaging archive (CIA) dataset	– Noise removal	DITNN	94.50%	–	94%	97.20%	96.80%
[60]	Dataset of lung cancer from Aastha Hospital	– Histogram images – Segmentation using improved profuse clustering technique (IPCT) – Image enhancement using Gabor method – Median filter – Segmentation using watershed approach	ANN	93.30%	–	–	–	–
[17]	CT dataset images (private)	– Color images to gray conversion – Gray images to binary conversion – Edge drawing using edge detection – Thresholding segmentation	CNN	–	–	–	–	–
[22]	1-LIDC- IDRI 2-ELCAP	– Conversion to gray scale – Image enhancement (CLAHE) – Segmentation (OTSU)	SVM	1st = 97.3% 2nd = 96.5%	1st = 70% 2nd = 72%	–	1st = 91.3% 2nd = 90.8%	1st = 56.22% 2nd = 65.56%
[25]	SPIE-AAPM for LUNGx challenge	Segmenting the nodule and parenchyma were done semi-automatically using a modified pipeline version	ANN	–	96.50%	–	–	–
[33]	LIDC-IDRI	–	The author proposed a method called MSCS-DeepLN based on 3D CNN	1 = 92.64%	1 = 94%	1 = 90.39%	1 = 95.87%	1 = 85.58%
[34]	LIDC-IDRI	–	Learning framework	90.60%	93.90%	–	93.90%	83.70%
[35]	SPIE-AAPM for LUNGx challenge	Contrast normalization	Novel CNN architecture	90.38%	94.80%	–	92.40%	88.70%

(Continued)

Table 1: Continued

Ref.	Dataset	Pre-processing Approaches	Methods used	Results (Metrics of CADx) ACC	AUC	Precision	Specificity	Sensitivity
[38]	LIDC-IDRI	<ul style="list-style-type: none"> – Convert to gray images – Morphological filter – Thresholding technique 	Deep Belief Network	≈96.6%	–	–	≈93%	≈98.6%
[42]	EDTUC dataset (private dataset)	–	Generative adversarial network	69.24%	–	–	71.08%	76.54%
[53]	LIDC-IDRI	–	Proposed semi-supervised adversarial classification (SSAC)	92.53%	95.81%	–	96.28%	84.94%
[47]	1) NLST dataset	– Converts an image to Hounsfield used to remove the context outside the lung	Proposed new method (DLSTM)	1) 78.96%	1) 82.55%	1) 83.38%	–	1) 61.61%
	2) MCL datasets	–	–	2) 86.97%	2) 91.17%	2) 80.04%	–	2) 72.71%
	3) VLSP datasets	–	–	–	–	–	–	–
[40]	LIDC-IDRI for LUNA16 challenge	<ul style="list-style-type: none"> – Applied threshold value (0 HU) – Image segmentation with adaptive thresholds 	3D CNN	–	95%	–	89.70%	88%
[36]	LIDC-IDRI	<ul style="list-style-type: none"> Median filter – Gaussian Process (GP) Regression 	CNN	92.31%	–	–	–	–
[37]	1) LIDC-IDRI	–	2D CNN using Taguchi	1) 98.83%	–	–	1) 99.93%	5) 99.97%
	2) SPIE-AAPM	–	Parametric optimization	2) 99.97%	–	–	2) 99.94%	2) 99.94%
[43]	SPIE-AAPM	–	Generative adversarial networks with CNN (AlexNet with Taguchi method)	99.90%	–	–	100%	99.90%
[8]	–	<ul style="list-style-type: none"> – Median filter – Histogram equalization – Morphological operation – Segmentation for desired regions 	ANN	95.60%	–	–	–	–
[61]	IQ-OTHNCCD	–	AlexNet CNN	93.55%	–	97.10%	95%	95.71%
[44]	IQ-OTHNCCD	–	MobileNet	51%	–	42%	–	39%
		–	VGG16	46%	–	31%	–	34%

(Continued)

Table 1: Continued

Ref.	Dataset	Pre-processing Approaches	Methods used	Results (Metrics of CADx)	AUC	Precision	Specificity	Sensitivity
[6]	IQ-OTHNCDD	<ul style="list-style-type: none"> – Bit-plane slicing – Gaussian filter – Segment images 	VGG19 DenseNet ResNet Svm	56% 53% 48% 89.88%	— — — —	43% 43% 41% 98.50%	— — — 97.50%	42% 43% 37% 97.14%
[45]	IQ-OTHNCDD	Gabor filter, morphological operation, and ROI extraction	CNN GoogLeNet	94.38%	—	—	93.70%	95.08%

article reached by examining the metrics (accuracy, area under the curve (AUC), precision, specificity, and sensitivity).

5 Discussion

This study aims to explore and review the literature that was directed to the diagnosis of lung cancer. It also aims at the directions of the researchers' work in relation to the methods used and the challenges that the researchers worked to solve, as we show and clarify the methodologies for lung imaging, datasets used, nodules segmentation approaches, challenges of each article, and then review what has been suggested by the authors for future work.

5.1 Methodologies for lung imaging

Image lung data that are used to detect diseases are collected in different ways according to the method used for imaging. Table 2 explains lung imaging methods and the advantages and disadvantages based on diagnosis and problems for each one.

Table 2: Lung imaging methods [62]

Imaging Method	Application	Pros	Cons
X-ray	X-ray is ionizing kind of radiation that uses rays to capture a picture. X-rays often take pictures of bones in order to identify bone cancer, but they do not give medical data for tissues or organs.	<ul style="list-style-type: none"> – Based on information obtained from digital X-ray pictures, illness is classified. – It creates a picture that may be reviewed right away. 	Because of picture noise or blurring, an X-ray may not be able to diagnose the condition.
CT	To obtain cross-sectional pictures of the body, a computerized tomography (CT scan) employs spinning X-ray equipment and computers. CT scans include more information than standard X-ray images. It displays soft tissues, blood arteries, and bones in various bodily sections.	<ul style="list-style-type: none"> – It can be used to locate nodules and cancers. – Much faster than MRI scanning. – Provides more information than ultrasonography. 	<ul style="list-style-type: none"> – Extremely high radiation doses are involved. – Radiation is dangerous and can lead to cancer.
Ultrasound	Medical ultrasonography is a type of ultrasound that uses high-frequency sound waves to detect activities inside the human body. The soundwaves are produced and directed towards the region of interest using a transducer, and the results are presented on the integrated monitor in real-time. It is suitable for both thick and soft tissues.	<ul style="list-style-type: none"> – Easily accessible and widely available, noninvasive, rapid, and extremely sensitive. 	The individual's ability to do the scan has a significant impact on the image quality and interpretation.
MRI	MRI imaging creates pictures of bodily structures using magnetic, radiofrequency, and a computer.	<ul style="list-style-type: none"> – Produces comprehensive photos of organs and other inside body structures. – Accurate technique of illness detection throughout the body. 	

5.2 Datasets

Several lung cancer datasets are published from reliable sources that are designed to apply artificial intelligence algorithms to them to perform various tasks such as classification, segmentation, and detection of nodules. These data are shown in Table 3.

Table 3: Explanation datasets used in studies

Dataset name	Images/size	Type	About
Lung Image Database Consortium image collection (LIDC-IDRI)	244,527 images (12 GB)	CT, DX (digital radiography), and CR (computed radiography)	“The National Cancer Institute” created it, and the Foundation for the “National Institutes of Health” pushed it (FNIH). This data collection was created in collaboration with seven academic institutions and eight medical imaging firms. [63]
SPIE-AAPM lung CT dataset	22,489 image (124 GB)	CT	This data was generated as part of the 2015 Medical Imaging Conference and supported by the American Association of Physicians of Medicine (AAPM) and the National Cancer Institute (NCI) to create the “LUNGx” image analysis challenge for the diagnostic classification of malignant and benign lung nodules. [64]
PET/CT dataset	251,135 image (127.2 GB)	CT and PET	The images were taken from patients who had a suspicion of lung cancer, and then they underwent a biopsy. Adenocarcinoma, large cell carcinoma, small cell carcinoma, and squamous cell carcinoma were the three forms of cancer that were labeled. this dataset was collected with the assistance of the Beijing Municipal Administration of Clinical Medicine Development and the Computer Center and Cancer Institute at Harbin Medical University’s Second Affiliated Hospital.[65]
Data Science Bowl 2017	More than 1,000 images	CT	A dataset was used for a contest for researchers and programmers on Kaggle in 2017 to get the best diagnostic accuracy to find out if a patient had lung cancer or not. DICOM images were collected from patients at high risk
IQ-OTHNCCD	1,190 image (152 MB)	CT	Dataset collected by Iraq-Oncology Teaching Hospital/National Center for Cancer Diseases in 2019 for patients of both sexes from central and southern Iraq. The dataset is classified into three types: natural, malignant, and benign.
ELCAP dataset	More than 1,000 images	CT	The ELCAP database consists of an image set of low-doses lung CT scans for detection.
NLST	More than 1,200 images	CT	These data were collected by a randomized national lung screening trial to determine whether low-dose CT lung screening reduces lung cancer mortality rates.

5.3 Lung images segmentation

Lung nodule segmentation is a crucial step in classifying and detecting lung nodules. To check the nodule for malignancy, we should first segment the lung parenchyma, followed by the nodule segmentation. During the training and testing phases, the pulmonary nodule is segmented. Solid, semi-solid, non-solid, and calcified pulmonary nodules are the four kinds of nodules. The surface of a big solid nodule (>10 mm) has a different intensity range than smaller lesions, making the solid nodule identification technique challenging to capture. Separating the juxta-pleural and juxta-vascular nodules would be the most challenging stage since the contrast of a big solid nodule with the pleura is poor [64]. Segmentation is generally done after the pretreatment stage and has a significant impact on the system's efficiency. It is an essential process to locate the pulmonary lesion. The best five lung segmentation models according to their performance that were tested in different datasets are shown in Table 4.

Table 4: Top 5 lung nodules segmentation models

	Dataset	References	Model name	AUC	ACC
1	LUNA	[66]	BCDU-Net	0.9946	—
2	LIDC-IDRI	[67]	ModelGenesis	0.9849	—
3	NIH	[68]	U-Net + R + A4	0.962	—
4	Montgomery county	[69]	ET-Net	—	0.9865
5	Lung nodule	[66]	BCDU-net	0.994	—

5.4 Challenges

Lung cancer has been an important threat to human health and the prime cause of death worldwide in the last few decades. The diagnosis of lung cancer at early levels is the main approach to increase patients' survival averagely [5]. Many challenges are mentioned in the literature, and researchers are working to address them within the scope of a lung cancer diagnosis. The lung cancer prediction at an early stage based on CT images is very difficult for the radiologist which is time consuming, costly, tedious job, and error-prone task, screening process requires very high concentration and skill owing to factors such as low contrast variation, heterogeneity, and visual similarity between benign and malignant nodules [4,6,7,12,16,17,19,20,23–25,29,30,32,34,34,35,39,40,42,42,47–51]. Precise detection of lung nodules is a challenging task in the field of medical imaging. Especially the nodules are extremely unbalanced with high intra-class variance, the complex lung structure makes this problem even more difficult to be addressed [11,22,35]. In a previous study, authors demonstrated that machine learning algorithms are effective in diagnosing cancer with FDG-PET imaging in the setting of ultra-low-dose PET scans [9], analyzed the ability to extract automatically generated features using deep structured algorithms in lung nodule CT image diagnosis, and compared its performance with traditional computer-aided diagnosis (CADx) systems using hand-crafted features [27]. Acquisition of labeled samples is time-consuming and laborious in the medical field [28]. Reading EBUS images requires well-trained and experienced radiologists. Most chest physicians do not have sufficient training or experience to read EBUS images; there were no studies using the CNN to diagnose the EBUS images [14]. The ability to distinguish between benign nodules and nodules that can become malignant [29]. In another study, the author provided evidence that DL networks may be used for mortality risk stratification based on standard-of-care CT images from NSCLC patients [15]. Not being able to detect the cancerous nodule around the lobe or lung at an early stage [54]. Studies cannot tell how the CNN works in predicting the malignancy of the given nodule, e.g., it is hard to conclude whether the region within the nodule or the contextual information matters according to the output of the CNN [32]. Automatic lung disease detection is a critical, challenging task for researchers because of the noise signals getting included into creative signals amid the image capturing process, which may corrupt the cancer image

quality, thus bringing about the debased performance [19]. Plenty types of pulmonary nodules and the visual similarity between them and the surrounding tissues [34]. Few positive samples in the datasets [42]. Deep convolutional neural networks (DCNNs) always require many labeled training data, which are not available for most medical image analysis applications due to the work needed in image acquisition and particularly image annotation [53]. Clinical imaging acquisition may be irregularly sampled, and such sampling patterns may be commingled with clinical usages [47]. Lung cancer diagnosis in patients produces many false positives [10–36].

5.5 Recommendation feature work

This section will explain the future work found in the articles, and the authors suggested ways to develop efficient diagnostic systems for future work. As potential future work, we would like to incorporate patient referral (or reject option) as part of the training strategy and learn models that would automatically reject the most uncertain decisions. We would also like to visually analyze learned feature representations to assess whether they could be used as informative biomarkers and help radiologists better understand and interpret CADe/CADx results [11]. Future research can focus on the limitation that many objects are generated in the initial segmentation process, which increases the computation time [23]. The FSVT-KIRF-based lung cancer nodule detection method can be extended for the early detection and diagnosis of lung cancer. A semantic segmentation can be developed by using a CNN model and optimization algorithm that would maintain high efficiency, while obtaining even better results in detecting lung cancer nodules [20]. In a study, the author proposed Hybridized Heuristic Mathematical Model for predicting lung cancer at an earlier stage for future work [18]. Exploring classification network architectures that accept inputs of simultaneous multi-scale resolutions or variable sizes, an approach common to fully convolutional networks is used in image segmentation [15]. Author of ref. [8] suggested enhancement work by using MRI and Ultrasound images. Proposed future work will focus on discovering more effective deep learning methods for characterizing lung nodules [48]. In another study, the author used Ant colony optimization with ANN for better results [10]. Pulmonary nodule detection should be added to perform medical imaging interpretation without requiring human-cropped nodule regions [35]. Automatic detection of high-level nodule attributes and their use for malignancy determination and use of imaging modalities such as PET could be considered for diagnostic imaging of lung cancer and treatment planning within the TumorNet framework [36]. Focusing on medical image classification, lung CT image classification scenarios use more effective network structures to extract common features from images of different types and perform cross-type matching to improve the performance [50]. Finding a approach to detect the accurate location of cavitory and Juxta vascular nodules [26]. Standardizing image quality and reducing contrast in medical laboratories can increase classification accuracy [5]. The reinforcement learning is applied to optimize automatically the architecture of the reconstruction network and classification network and the settings of hyper-parameters, such as the weighting factor λ , batch size, and learning rate, aiming to make the proposed MK-SSAC model more accurate and more efficient [53]. Authors have planned, as future work, to regularly update the LCRM as more data will be collected. We intend to package the model as software and combine its predictions with medical expertise using Bayes reasoning. Expert's experience can help improve the interpretability of the LCRM and encourage its better performance when examples of positive cases are limited. In addition, the LCPM model can take clinical test variables, imaging data, and genetic data into consideration, so as to improve the prediction performance and have a prognostic evaluation function [42]. In future work, we will focus on lung nodule malignancy prediction with pathology data and unsure labeled data. Moreover, we plan to analyze neural networks to further increase model interpretability [58]. In future work, the extraction of more advanced features that could be useful, along with malignancy, to improve cancer prediction is planned [52]. The author would recommend using the convex function as the backbone since the DLSTM1 achieves the most robust performances across different settings and metrics [47]. Nodules adjacent to blood vessels and the dura cavity can be detected accurately by

diagnostic methods [23]. Author recommends that in the future, more focus is required on promoting the performance in clinical practice and on constructing new lighter effective subnetworks [33]. The most common future work proposal was to increase the volume of data of various types and collect them from different sources to increase the accuracy of lung cancer diagnosis [13,16,20,26,36,38,50]. A future study was suggested to increase the capacity to classify numerous types of pulmonary nodules by gathering data such as pulmonary nodule-connected vessels and lung walls for network model training [28]. In future work, the authors recommend an end-to-end explainable CAD system for lung cancer diagnosis that integrates nodule detection, segmentation, and malignancy prediction, which will be of extensive clinical application value.

6 Limitation

We were unable to reach about 15 research papers that could have helped in developing this review work.

7 Conclusion

This research aimed to analyze the literature related to the field of lung cancer diagnosis. The authors' suggestions in the field of lung cancer diagnosis were either to suggest new methods of solution or to change the methods of pretreatment and segmentation for improving the accuracy of the diagnosis. Some of them have proved their ability for lung cancer diagnosis based on new types of dataset and it was reviewed based on the authors' suggestions (dataset used, pre-processing approaches, and method used) with the results (area under the ROC curve, accuracy, precision, specificity, and sensitivity) obtained for diagnosis. This study reviewed the published datasets that are available to researchers, their sources, and size. The idea of lung nodule segmentation was illustrated and five of the best models for this purpose have been listed. The challenges identified in the literature are registered and suggestions are given for future work that could help researchers in the field of research.

Our suggestions to researchers regarding future work are several directions that this research scope lacks, and they are as follows:

1. Focusing on the triple and quadrant classification based on the type of nodule or the type of malignant tumor such as squamous cell carcinoma, adenocarcinoma, and large cell carcinoma.
2. The need of more lung image data from various types of imaging and publishing private datasets so that researchers can compare with each other and focus on making available MRI and Ultrasound datasets. Ultrasound imaging has proven its high ability to identify many diseases, for example identifying a complete gynecological abnormality [70].
3. Segmentation of large solid nodules because of their difficulty and lack of search in it.
4. Develop a lung cancer detection model that can distinguish between a tiny malignant lesion and a benign nodule at an early stage.
5. In order to increase the efficiency of automated tumor diagnosis, additional patient information such as medical history and genetic reports may be examined and combined with deep features derived from lung scan pictures.
6. Use different pre-processing methodologies and filters to improve image quality such as harmony search to improve gray scale image quality [71] and images with Edge preserving [72].
7. The use of cloud computing technologies to make a system based on the remote diagnosis of lung cancer using ELM, where this proposal was applied for the remote detection of breast cancer [73].
8. To extract features from lung medical pictures, employ a cat swarm-optimized deep belief network [74].

Acknowledgment: We would gratefully acknowledge the support of Dr. Ahmed Shihab Albahri, in making this study a success.

Author contributions: Mustafa Mohammed Jassim: Main Author; Mustafa Musa Jaber: Supervision of study.

Conflict of interest: The authors state no conflict of interest.

References

- [1] Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 Countries. *CA Cancer J Clin.* 2021;71(3):209–49. doi: 10.3322/caac.21660.
- [2] Begum S, Sarkar R, Chakraborty D, Maulik U. Identification of biomarker on biological and gene expression data using fuzzy preference based rough set. *J Intell Syst.* 2021;30(1):130–41.
- [3] Razzak MI, Naz S, Zaib A. Deep learning for medical image processing: Overview, challenges and the future. *Lect Notes Comput Vis Biomech.* 2018;26:323–50. doi: 10.1007/978-3-319-65981-7_12.
- [4] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ.* 2009;339(7716):332–6. doi: 10.1136/bmj.b2535.
- [5] Hoque A, Farabi AKMA, Ahmed F, Islam MZ. Automated detection of lung cancer using CT scan images. 2020 IEEE Region 10 Symposium TENSYP 2020; 2020 June. p. 1030–3. doi: 10.1109/TENSYP50017.2020.9230861.
- [6] Kareem HF, AL-Husieny MS, Mohsen FY, Khalil EA, Hassan ZS. Evaluation of SVM performance in the detection of lung cancer in marked CT scan dataset. *Indones J Electr Eng Comput Sci.* 2021;21(3):1731–8.
- [7] Roy K, Chaudhury SS, Burman M, Ganguly A, Dutta C, Banik S, et al. A Comparative study of lung cancer detection using supervised neural network. 2019 International Conference on Opto-Electronics and Applied Optics (Optronix); 2019. p. 1–5. doi: 10.1109/OPTRONIX.2019.8862326.
- [8] Bhise SS, Khot SR. Early stage lung cancer diagnosis using ANN classifier. In 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS); 2021. p. 22–7. doi: 10.1109/ICAIS50930.2021.9395952.
- [9] Schwyzer M, Ferraro DA, Muehlemaier UJ, Curioni-Fontecedro A, Huellner MW, von Schulthess GK, et al. Automated detection of lung cancer at ultralow dose PET/CT by deep neural networks – Initial results. *Lung Cancer.* 2018;126:170–3. doi: 10.1016/j.lungcan.2018.11.001.
- [10] Kaur L, Sharma M, Dharwal R, Bakshi A. Lung cancer detection using CT scan with artificial neural network. In 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE); 2018. p. 1624–9. doi: 10.1109/ICRIEECE44171.2018.9009244.
- [11] Ozdemir O, Russell RL, Berlin AA. A 3D probabilistic deep learning system for detection and diagnosis of lung cancer using low-dose CT scans. *IEEE Trans Med Imaging.* 2020;39(5):1419–29. doi: 10.1109/TMI.2019.2947595.
- [12] Perez G, Arbelaez P. Automated lung cancer diagnosis using three-dimensional convolutional neural networks. *Med Biol Eng Comput.* 2020;58(8):1803–15. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85086128473&doi=10.1007%2Fs11517-020-02197-7&partnerID=40&md5=8e137ba0f8d2090c795b3053bf989c92>
- [13] Polat H, Mehr HD. Classification of pulmonary CT images by using hybrid 3D-deep convolutional neural network architecture. *Appl Sci (Switz).* 2019;9(5):940. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85063694846&doi=10.3390%2Fapp9050940&partnerID=40&md5=eedaf30424fc7b28c2cc18e2fd9e49a0>.
- [14] Chen CH, Lee YW, Huang YS, Lan WR, Chang RF, Tu CY, et al. Computer-aided diagnosis of endobronchial ultrasound images using convolutional neural network. *Comput Methods Prog Biomed.* 2019;177:175–82. doi: 10.1016/j.cmpb.2019.05.020.
- [15] Hosny A, Parmar C, Coroller TP, Grossmann P, Zeleznik R, Kumar A, et al. Deep learning for lung cancer prognostication: A retrospective multi-cohort radiomics study. *PLOS Med.* 2018;15(11):1002711. doi: 10.1371/journal.pmed.1002711.
- [16] Zhao L, Bai C-X, Zhu Y. Diagnostic value of artificial intelligence in early-stage lung cancer. *Chin Med J (Engl).* 2020;133(4):503–4. doi: 10.1097/CM9.0000000000000634.
- [17] Mukherjee S, Bohra SU. Lung cancer disease diagnosis using machine learning approach. In 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS); 2020. p. 207–11. doi: 10.1109/ICISS49785.2020.9315909.
- [18] Yu H, Zhou Z, Wang Q. Deep learning assisted predict of lung cancer on computed tomography images using the adaptive hierarchical heuristic mathematical model. *IEEE Access.* 2020;8:86400–10. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85085177762&doi=10.1109%2FACCESS.2020.2992645&partnerID=40&md5=80ec325d1f74ddb30711403e605a06d9>.
- [19] Shakeel PM, Burhanuddin MA, Desa MI. Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks. *Measurement.* 2019;145:702–12. doi: 10.1016/j.measurement.2019.05.027.

- [20] Deepa P, Suganthi M. A fuzzy shape representation of a segmented vessel tree and kernel-induced random forest classifier for the efficient prediction of lung cancer. *J Supercomput.* 2020;76(8):5801–24. doi: 10.1007/s11227-019-03002-5.
- [21] Gong J, Liu J, Jiang Y, Sun X, Zheng B, Nie S. Fusion of quantitative imaging features and serum biomarkers to improve performance of computer-aided diagnosis scheme for lung cancer: A preliminary study. *Med Phys.* 2018;45(12):5472–81. doi: 10.1002/mp.13237.
- [22] Abbas W, Khan KB, Aqeel M, Azam MA, Ghouri MH, Jaskani FH. Lungs nodule cancer detection using statistical techniques. in 2020 IEEE 23rd International Multitopic Conference (INMIC); 2020. p. 1–6. doi: 10.1109/INMIC50486.2020.9318181.
- [23] Naqi SM, Sharif M, Lali IU. A 3D nodule candidate detection method supported by hybrid features to reduce false positives in lung nodule detection. *Multimed Tools Appl.* 2019;78(18):26287–6311. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85067083278&doi=10.1007%2F511042-019-07819-3&partnerID=40&md5=652b4118f9efae22c482f5c3dc06f702>.
- [24] Ali I, Muzammil M, Haq IU, Khaliq AA, Abdullah S. Deep feature selection and decision level fusion for lung nodule classification. *IEEE Access.* 2021;9:18962–73. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85100509780&doi=10.1109%2FACCESS.2021.3054735&partnerID=40&md5=113eac49a5774c03e3b4a90003bda1b>.
- [25] Uthoff J, Stephens MJ, JrNewellJD, Hoffman EA, Larson J, Koehn N, et al. Machine learning approach for distinguishing malignant and benign lung nodules utilizing standardized perinodular parenchymal features from CT. *Med Phys.* 2019;46(7):3207–16. doi: 10.1002/mp.13592.
- [26] Mathews AB, Jeyakumar MK. Automatic detection of segmentation and advanced classification algorithm. In 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC); 2020. p. 358–62. doi: 10.1109/ICCMC48092.2020.ICCMC-00067.
- [27] Sun W, Zheng B, Qian W. Automatic feature learning using multichannel ROI based on deep structured algorithms for computerized lung cancer diagnosis. *Comput Biol Med.* 2017;89:530–9. doi: 10.1016/j.combiomed.2017.04.006.
- [28] Tang S, Ma R, Li Q, Bai Y, Chen S. Classification of Benign and Malignant Pulmonary Nodules Based on the Multiresolution 3D DPSECN Model and Semisupervised Clustering. *IEEE Access.* 2021;9:43397–410. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85101743287&doi=10.1109%2FACCESS.2021.3060178&partnerID=40&md5=58b5f8dc9d8cdc5993b28d520b3b9a83>.
- [29] Paul R, Schabath M, Gillies R, Hall L, Goldgof D. Convolutional Neural Network ensembles for accurate lung nodule malignancy prediction 2 years in the future. *Computers Biol Med.* 2020;122:103882.
- [30] Ali I, Muzammil M, Haq IU, Khaliq AA, Abdullah S. Efficient lung nodule classification using transferable texture convolutional neural network. *IEEE Access.* 2020;8:175859–70. doi: 10.1109/ACCESS.2020.3026080.
- [31] Bonavita I, Rafael-Palou X, Ceresa M, Piella G, Ribas V, González Ballester MA. Integration of convolutional neural networks for pulmonary nodule malignancy assessment in a lung cancer classification pipeline. *Comput Methods Prog Biomed.* 2020;185:105172. doi: 10.1016/j.cmpb.2019.105172.
- [32] Wu B, Zhou Z, Wang J, Wang Y. Joint learning for pulmonary nodule segmentation, attributes and malignancy prediction. in 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018); 2018. p. 1109–13. doi: 10.1109/ISBI.2018.8363765.
- [33] Xu X, Wang C, Guo J, Gan Y, Wang J, Bai H, et al. MSCS-DeepLN: Evaluating lung nodule malignancy using multi-scale cost-sensitive neural networks. *Med Image Anal.* 2020;65:101772. doi: 10.1016/j.media.2020.101772.
- [34] Liu H, Cao H, Song E, Ma G, Xu X, Jin R, et al. Multi-model ensemble learning architecture based on 3D CNN for lung nodule malignancy suspiciousness classification. *J Digit Imaging.* 2020;33(5):1242–56. doi: 10.1007/s10278-020-00372-8.
- [35] Wang Y, Zhang H, Chae KJ, Choi Y, Jin GY, Ko S-B. Novel convolutional neural network architecture for improved pulmonary nodule classification on computed tomography. *Multidimens Syst Signal Process.* 2020;31(3):1163–83. doi: 10.1007/s11045-020-00703-6.
- [36] Hussein S, Gillies R, Cao K, Song Q, Bagci U. TumorNet: Lung nodule characterization using multi-view Convolutional Neural Network with Gaussian Process. In 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017); 2017. p. 1007–10. doi: 10.1109/ISBI.2017.7950686.
- [37] Lin C-J, Jeng S-Y, Chen M-K. Using 2D CNN with Taguchi parametric optimization for lung cancer recognition from CT images. *Appl Sci (Switz).* 2020;10(7):2591.
- [38] Priya MMA, Jawhar DSJ, Geisa DJM. Optimal deep belief network with opposition based pily beetle algorithm for lung cancer classification: A DBNOPBA approach. *Comput Methods Prog Biomed.* 2021;199:105902. doi: 10.1016/j.cmpb.2020.105902.
- [39] Tran GS, Nghiem TP, Nguyen VT, Luong CM, Burie J-C. Improving accuracy of lung nodule classification using deep learning with focal loss. *J Healthcare Eng.* 2019;2019:5156416. doi: 10.1155/2019/5156416.
- [40] Zhang C, Sun X, Dang K, Li K, Guo XW, Chang J, et al. Toward an expert level of lung cancer detection and classification using a deep convolutional neural network. *Oncologist.* 2019;24(9):1159–65. doi: 10.1634/theoncologist.2018-0908.
- [41] Bansal G, Chamola V, Narang P, Kumar S, Raman S. Deep3DSCan: Deep residual network and morphological descriptor based framework for lung cancer classification and 3D segmentation. *IET IMAGE Process.* 2020;14(7):1240–7. doi: 10.1049/iet-ipr.2019.1164.
- [42] Li J, Tao Y, Cai T. Predicting lung cancers using epidemiological data: A generative-discriminative framework. *IEEE/CAA J Autom Sin.* 2021;8(5):1067–78. doi: 10.1109/JAS.2021.1003910.

- [43] 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 (Switz)*. 2021;11(2):1–17. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85099218394&doi=10.3390%2Fapp11020480&partnerID=40&md5=bad116c59b9eda039eb1e84949793af1>.
- [44] Mohite A. Application of transfer learning technique for detection and classification of lung cancer using CT images. *Int J Sci Res Manag*. 2021;9(11):621–34.
- [45] al-Huseiny MS, Sajit AS. Transfer learning with GoogLeNet for detection of lung cancer. *Indones J Electr Eng Comput Sci*. 2021;22(2):1078–86.
- [46] Al-Yasriy HF, Al-Husieny MS, Mohsen FY, Khalil EA, Hassan ZS. Diagnosis of lung cancer based on CT Scans Using CNN. *IOP Conf Ser Mater Sci Eng*. 2020;928(2):22035.
- [47] Gao R, Tang Y, Xu K, Huo Y, Bao S, Antic SL, et al. Time-distanced gates in long short-term memory networks. *Med Image Anal*. 2020;65:101785. doi: 10.1016/j.media.2020.101785.
- [48] Yutong X, Jianpeng Z, Yong X, Fulham M, Yanning Z. Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest CT. *Inf Fusion*. 2018;42:102–10. doi: 10.1016/j.inffus.2017.10.005.
- [49] Xu W, Wang K, Lin J, Lu Y, Huang S, Zhang X. Knowledge-guided and hyper-attention aware joint network for benign-malignant lung nodule classification. In 2020 IEEE International Conference on Image Processing (ICIP); 2020. p. 310–4. doi: 10.1109/ICIP40778.2020.9190653.
- [50] Chen Y, Wang Y, Hu F, Feng L, Zhou T, Zheng C. LDNNET: Towards robust classification of lung nodule and cancer using lung dense neural network. *IEEE Access*. 2021;9:50301–20. doi: 10.1109/ACCESS.2021.3068896.
- [51] Dodia S, Basava A, Padukudru Anand M. A novel receptive field-regularized V-net and nodule classification network for lung nodule detection. *Int J Imaging Syst Technol*. 2022;32:88–101. doi: 10.1002/ima.22636.
- [52] Zheng S, Shen Z, Pei C, Ding W, Lin H, Zheng J, et al. Interpretative computer-aided lung cancer diagnosis: From radiology analysis to malignancy evaluation. *Comput Methods Prog Biomed*. 2021;210:106363. doi: 10.1016/j.cmpb.2021.106363.
- [53] Xie Y, Zhang J, Xia Y. Semi-supervised adversarial model for benign–malignant lung nodule classification on chest CT. *Med Image Anal*. 2019;57:237–48. doi: 10.1016/j.media.2019.07.004.
- [54] Tiwari L, Raja R, Awasthi V, Miri R, Sinha GR, Alkinani MH, et al. Detection of lung nodule and cancer using novel Mask-3 FCM and TWEDLNN algorithms. *Measurement*. 2021;172:108882. doi: 10.1016/j.measurement.2020.108882.
- [55] Naik A, Edla DR. Lung nodule classification on computed tomography images using deep learning. Vol. 116, Issue 1. US: Springer; 2021.
- [56] Team NLSTR. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365(5):395–409.
- [57] Hoque A, Farabi AKMA, Ahmed F, Islam MZ. Automated detection of lung cancer using CT scan images. In 2020 IEEE Region 10 Symposium (TENSYPMP); 2020. p. 1030–3. doi: 10.1109/TENSYPMP50017.2020.9230861.
- [58] Wu B, Zhou Z, Wang J, Wang Y. Joint learning for pulmonary nodule segmentation, attributes and malignancy prediction Nat’l engineering laboratory for video technology cooperative medianet innovation center key laboratory of machine perception (MoE) Sch’l of EECS, peking univer. In 2018 International Symposium Biomedical Imaging (ISBI 2018); 2018. p. 1109–13.
- [59] Xu W, Wang K, Lin J, Lu Y, Huang S, Zhang X. “Knowledge-guided and hyper-attention aware joint network for benign-malignant lung nodule classification School of Big Data and Software Engineering. Chongqing University, Chongqing 401331, China Key Laboratory of Dependable Service Computing in Cyber.”; 2020.
- [60] Kaur L, Sharma M, Dharwal R, Bakshi A. Lung cancer detection using CT scan with artificial neural network. In 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE); 2018. p. 1624–9. doi: 10.1109/ICRIEECE44171.2018.9009244.
- [61] Al-Yasriy HF, Al-Husieny MS, Mohsen FY, Khalil EA, Hassan ZS. Diagnosis of lung cancer based on CT scans using CNN. *IOP Conf Ser Mater Sci Eng*. 2020;928(2):022035. doi: 10.1088/1757-899X/928/2/022035.
- [62] Kaur C, Garg U. Artificial intelligence techniques for cancer detection in medical image processing: A review. *Mater Today Proc*. 2021;51:764–9.
- [63] McLennan SG, Bidaut L, McNitt-Gray MF, Meyer CR, Reeves AP, et al. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans. *Med Phys*. 2011;38(2):915–31. doi: 10.1118/1.3528204.
- [64] Kirby JS, Armato SG, Drukker K, Li F, Hadjiiski L, Tourassi GD, et al. LUNGx challenge for computerized lung nodule classification. *J Med Imaging*. 2016;3(4):044506. doi: 10.1117/1.jmi.3.4.044506.
- [65] Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The cancer imaging archive (TCIA): Maintaining and operating a public information repository. *J Digit Imaging*. 2013;26(6):1045–57. doi: 10.1007/s10278-013-9622-7.
- [66] Azad R, Asadi-Aghbolaghi M, Fathy M, Escalera S. Bi-directional ConvLSTM U-Net with densely connected convolutions. in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*; 2019.
- [67] Zhou Z, Sodha V, Rahman Siddiquee MM, Feng R, Tajbakhsh N, Gotway MB, et al. Models genesis: Generic autodidactic models for 3d medical image analysis. *International Conference on Medical Image Computing and Computer-assisted Intervention*; 2019. p. 384–93.

- [68] Tang Y-B, Tang Y-X, Xiao J, Summers RM. Xlsor: A robust and accurate lung segmentor on chest x-rays using criss-cross attention and customized radiorealistic abnormalities generation. *International Conference on Medical Imaging with Deep Learning*; 2019. p. 457–67.
- [69] Zhang Z, Fu H, Dai H, Shen J, Pang Y, Shao L. Et-net: A generic edge-attention guidance network for medical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*; 2019. p. 442–50.
- [70] Hussein IJ, Burhanuddin MA, Mohammed MA, Benameur N, Maashi MS, Maashi MS. Fully-automatic identification of gynaecological abnormality using a new adaptive frequency filter and histogram of oriented gradients (HOG). *Expert Syst.* 2022;39(3):1–18. doi: 10.1111/exsy.12789.
- [71] Al-Betar MA, Alyasseri ZAA, Khader AT, Bolaji AL, Awadallah MA. Gray image enhancement using harmony search. *Int J Comput Intell Syst.* 2016;9(5):932–44. doi: 10.1080/18756891.2016.1237191.
- [72] Zaid Abdi Alkareem YA, Venkat I, Al-Betar MA, Khader AT. Edge preserving image enhancement via harmony search algorithm. *Conference on Data Mining and Optimization*; September 2012. p. 47–52. doi: 10.1109/DMO.2012.6329797.
- [73] Lahoura V, Singh H, Aggarwal A, Sharma B, Mohammed MA, Damaševičius R, et al. Cloud computing-based framework for breast cancer diagnosis using extreme learning machine. *Diagnostics.* 2021;11(2):1–19. doi: 10.3390/diagnostics11020241.
- [74] Elameer AS, Jaber MM, Abd SK. Radiography image analysis using cat swarm optimized deep belief networks. *J Intell Syst.* 2022;31(1):40–54. doi: 10.1515/jisys-2021-0172.