








## Research Article

# A Comprehensive Survey on the Progress, Process, and Challenges of Lung Cancer Detection and Classification

**M. F. Mridha** <sup>1</sup>, **Akibur Rahman Prodeep** <sup>2</sup>, **A. S. M. Morshedul Hoque** <sup>2</sup>,  
**Md. Rashedul Islam** <sup>3</sup>, **Aklima Akter Lima** <sup>2</sup>, **Muhammad Mohsin Kabir**,<sup>2</sup>  
**Md. Abdul Hamid** <sup>4</sup>, and **Yutaka Watanobe** <sup>5</sup>

<sup>1</sup>Department of Computer Science and Engineering, American International University Bangladesh, Dhaka 1229, Bangladesh

<sup>2</sup>Department of Computer Science and Engineering, Bangladesh University of Business and Technology, Dhaka 1216, Bangladesh

<sup>3</sup>Department of Computer Science and Engineering, University of Asia Pacific, Dhaka 1216, Bangladesh

<sup>4</sup>Department of Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

<sup>5</sup>Department of Computer Science and Engineering, University of Aizu, Aizuwakamatsu 965-8580, Japan

Correspondence should be addressed to M. F. Mridha; [firoz.mridha@aiub.edu](mailto:firoz.mridha@aiub.edu)

Received 27 July 2022; Revised 17 October 2022; Accepted 9 November 2022; Published 16 December 2022

Academic Editor: Rafael Morales

Copyright © 2022 M. F. Mridha et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Lung cancer is the primary reason of cancer deaths worldwide, and the percentage of death rate is increasing step by step. There are chances of recovering from lung cancer by detecting it early. In any case, because the number of radiologists is limited and they have been working overtime, the increase in image data makes it hard for them to evaluate the images accurately. As a result, many researchers have come up with automated ways to predict the growth of cancer cells using medical imaging methods in a quick and accurate way. Previously, a lot of work was done on computer-aided detection (CADe) and computer-aided diagnosis (CADx) in computed tomography (CT) scan, magnetic resonance imaging (MRI), and X-ray with the goal of effective detection and segmentation of pulmonary nodule, as well as classifying nodules as malignant or benign. But still, no complete comprehensive review that includes all aspects of lung cancer has been done. In this paper, every aspect of lung cancer is discussed in detail, including datasets, image preprocessing, segmentation methods, optimal feature extraction and selection methods, evaluation measurement matrices, and classifiers. Finally, the study looks into several lung cancer-related issues with possible solutions.

## 1. Introduction

Lung cancer is a significant obstacle to the survival of humans, and many people lose their lives every year because of lung cancer. Early detection of pulmonary nodules is essential for improving lung cancer patients' survival rates. Nodules are abnormal tissue growths that can occur anywhere in the body. They can also grow in in-depth skin tissues as well as internal organs. When a nodule forms in the lungs, it is referred to as a pulmonary nodule. A nodule with a diameter of three centimeters or less is called a tumor [1]. There are mainly two kinds of tumors. It can be either malignant or benign. Malignant tumors mean cancerous tumors. It can grow and spread all over the body. On the other hand, benign tumors are not cancerous. They either do not spread or grow very slowly or do so. They usually do not return after being

removed by a physician. Approximately 95% of lung nodules are benign [2]. But it can be malignant also. A larger lung nodule, such as 30 millimeters or more in diameter, has a higher risk of being cancerous than a smaller lung nodule [3].

Lung cancers are broadly divided into non-small-cell lung cancer (NSCLC) and small-cell lung cancer (SCLC) [4]. About 80%–85% of lung cancers are NSCLC, and 10%–15% of all lung cancers are SCLC. The survival rate of lung cancer is low. In 2008, there were 12.7 million cancer cases and 7.6 million cancer deaths, with 56% of patients and 64% of fatalities occurring in economically developing countries. Lung cancer is the most common cancer site in men, accounting for 17% of all new cancer cases and 23% of cancer deaths [5]. Lung cancer is diagnosed at an advanced stage in approximately 70% of patients, with a 5-year survival rate of approximately 16%. However, if lung cancer is detected early, it

has a better chance of being treated successfully, with a 5-year survival rate of 70% [6, 7]. One of the leading causes of lung cancer is smoking. It can even happen to those who have never smoked. It can be increased by exposure to secondhand smoking, arsenic, asbestos, radioactive dust, or radon.

Several attempts have been made since 1980 to develop a system that can detect, segment [8, 9], and diagnose pulmonary nodules from CT scans [10]. The detection of pulmonary nodules is complicated because their appearance varies depending on their type, whether they are malignant, and their size, internal structure, and location. Segmentation has become a big problem, and it now requires a lot of different methods to solve it. Each technique focuses on another part of the problem [11]. These systems are referred to as computer-aided diagnosis systems (CAD). They go beyond simple image processing to provide specific information about the lesion that can aid radiologists in making a diagnosis. The idea of CAD was initially presented in 1966 [12]. Researchers first thought about using computers to make automated diagnoses. There were no other ideas or technologies at the time, so CAD technology was still in its infancy until the 1980s when the concept moved from automatic computer diagnosis to CAD [13]. The relevant ideas and computer technology were also quickly evolving at the time. All of these factors contributed to the advancement of CAD technologies. The first study on lung cancer CAD systems based on CT scans was published in 1991 [14]. Several competitions, such as Lung Nodule Analysis 2016 (LUNA16) [15] and Kaggle Data Science Bowl (KDSB) [16], have attracted several professional teams who have created lung cancer CAD algorithms in recent years. By making it easier to compare alternative algorithms, these competitions have aided in advancing lung cancer CAD technology. Lung cancer CAD can detect lung nodules and predict the likelihood of malignancy, making it a handy tool for doctors. Computer-aided detection (CADe) and computer-aided diagnosis (CADx) systems are two types of CAD systems. The former can detect and locate pulmonary nodules, while the latter can classify them as benign or malignant.

Several researchers analyzed the existing articles previously for detecting and diagnosing lung nodules using CT images. Yang et al. [17] examined the use of deep learning techniques to detect and diagnose lung nodules in particular. Convolutional neural networks (CNNs) have been the most widely used deep learning methods in treating pulmonary nodules. CNNs have produced excellent results in lung cancer CAD systems. In the 2017 DSB competition, for example, the winning team's algorithm was a CNN model [18], and a CNN model developed by Google and published in Nature outperformed six professional radiologists [19]. The problem of pulmonary nodule application has been tackled using various deep learning methods. Poap et al. [20] introduced a heuristic and nature-inspired method for X-ray image segmentation-based detection over aggregated images. The proposed approach for automating medical exams delivers favorable results for detecting diseased and healthy tissues. A heuristic red fox heuristic optimization algorithm (RFOA) was also presented for medical image segmentation by Jaszcz et al. [21]. In addition, the operation of heuristics

was modified for the analysis of two-dimensional images, with an emphasis on equation modification and the development of a unique fitness function. Kumar et al. [22] were the first to employ an autoencoder (AE) to differentiate benign from malignant pulmonary nodules, while Chen et al. [23] were the first to use a deep belief network (DBN) in the context of pulmonary nodule CAD. To improve training efficiency, Wang and Chakraborty [24] proposed a sliced recurrent neural network (RNN) model. In their method, multiple layers of the RNN were taught simultaneously, which reduced training time. To train a deep learning model, a large amount of data is required. However, few labeled datasets are available for researchers due to the need for specialists and the time-consuming nature of the effort. A generative adversarial network (GAN) is based on the negative training paradigm and uses training to generate new images that are comparable to the original, which has piqued the interest of many medical imaging researchers [25]. Some researchers have chosen to generate lung nodule images with a GAN to increase the amount of data available [26]. Lung cancer detection has become more structured, making it more usable and reliable. This structure provides a basic workflow diagram for detecting lung cancer. However, the structure is not always the same, and there may be variations. When it comes to lung cancer detection, the process is divided into several steps, including collecting images or datasets, preprocessing the images, segmentation, feature extraction, feature selection and classification, and receiving the results. Figure 1 depicts the method for detecting cancer in images.

- (i) *Dataset.* Dataset collection is the initial step to starting the process. There are mainly 3 types of image datasets used for lung cancer detection: computed tomography (CT) scans, magnetic resonance imaging (MRI), and X-rays. CT scan images are mainly used because of their high sensitivity and low cost. Also, it is more available rather than MRI and X-ray. More about the dataset is discussed in Section 3.
- (ii) *Preprocessing.* Image preprocessing is used to improve the original image's quality and interpretability. The primary goal of CT image preprocessing is to remove noise, artifacts, and other irrelevant information from raw images, improve image quality, and detect relevant information. Section 5 has a brief discussion about it.
- (iii) *Segmentation.* The segmentation of CT images is an important step in detecting lung nodules and recognizing lung cancer. Pulmonary segmentation's main goal is to separate the pulmonary parenchyma from other tissues and organs accurately. It uses preprocessed medical images to calculate the volume of lung parenchyma. Section 6 discusses a variety of segmentation algorithms.
- (iv) *Feature Extraction.* The features of the segmented lung images are extracted and analyzed in this step. Feature extraction is a process in which a large amount of raw data is divided and reduced to more

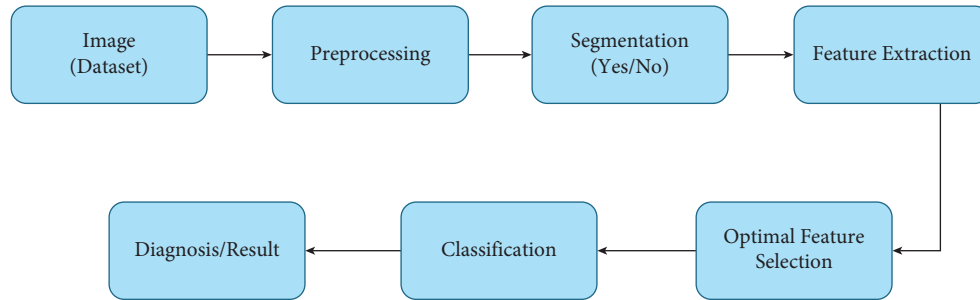


FIGURE 1: The workflow diagram of basic CAD system.

manageable groups after being initially collected. It makes the process a lot less complicated. Feature extraction methods are described in Section 7.

- (v) *Feature Selection.* Feature selection identifies and isolates the most consistent, non-redundant, and relevant features in model construction. Feature selection is primarily used to improve predictive model performance while lowering modeling computational costs. It is also a way to make the classification result more accurate. Section 8 describes the most commonly used feature selection methods.
- (vi) *Classification.* Classification is dividing a given set of data into groups of similar characteristics. It separates benign and malignant nodules based on the feature that has been selected. Well-known classification methods are discussed in Section 9.
- (vii) *Result.* Finally, the detection result of lung cancer shows us where the cancerous cell is in the lung. It is discussed in Section 10.

Figure 2 addresses the taxonomy of this survey. The lung nodule and cancer analysis were separated into two artificial intelligence plans applied in clinical imaging. This clinical imaging was divided into seven categories. We chose studies from various eras based on their popularity to conduct this survey. We upheld a systematic review methodology in this study, which will aid future researchers in determining the general skeleton of an artificial intelligence-based lung nodule and cancer analysis. This survey gives a reasonable perspective on ML and DL structures occupied with distinguishing lung cancer. This concentration likewise addresses the identification and characterization of lung nodules and malignant growth using imaging strategies. Finally, this survey coordinates a few open exploration challenges and opportunities for future scientists. We agree that this review serves as an essential guide for researchers who need to work with clinical image characterization using artificial intelligence-based lung nodules and cancer analysis while using various clinical images. Table 1 shows a correlation between the existing surveys and our survey. Table 2 provides a summary of recent surveys and reviews that have been conducted on various approaches for the detection, segmentation, and classification of lung cancer.

The survey discusses the findings of various related research work areas like nodule classification, nodule identification, lung cancer detection, lung cancer verification, and so

on. While looking at the present challenges, this study generates suggestions and recommendations for further research works. The total contributions of the research are as follows:

- (i) The article gives an intelligible review of detecting systems of lung nodules and cancer.
- (ii) The article inspects lung nodule and cancer-detecting procedures depending on the existing systems, datasets, image preprocessing, segmentation, feature extraction, and selection techniques. Further, the paper exploits the benefits and limitations of those systems.
- (iii) The article gives the procedures to detect lung nodules and cancer in a well-organized way.
- (iv) Finally, the survey adapts the present challenges of lung nodules and cancer detection systems, with further research on pathological diagnosis.

After going through this division, one should adapt how to get started with this topic.

The remaining sections of the paper are organized as follows. The methodology of the survey is described in Section 2. Various categorized datasets obtainable publicly are displayed in Section 3. Imaging modalities are briefly described in Section 4. Section 5 describes the preprocessing algorithm of the image dataset of lung cancer and nodules. Section 6 discusses the segmentation process and algorithms. Section 7 discusses the most commonly used algorithms for extracting features from CT scans, X-rays, and MRI images. Section 8 discusses the most commonly used methods for feature selection. Section 9 discusses the well-known classification and detection algorithms. A comprehensive exploration of the performance for lung cancer and nodule detection is discussed in Section 10. The challenges faced most commonly while detecting lung nodules and cancer are explained with their possible solutions in Section 11. Lastly, the conclusion of this article is given in Section 12.

## 2. Survey Methodology

The survey is analyzed following a process developed by Kitchenham [40, 41] called systematic literature review (SLR). This article divides the SLR processes into three different parts: the planning phase, the conducting phase, and the reporting phase. In the subsequent sections, the steps are discussed in detail.

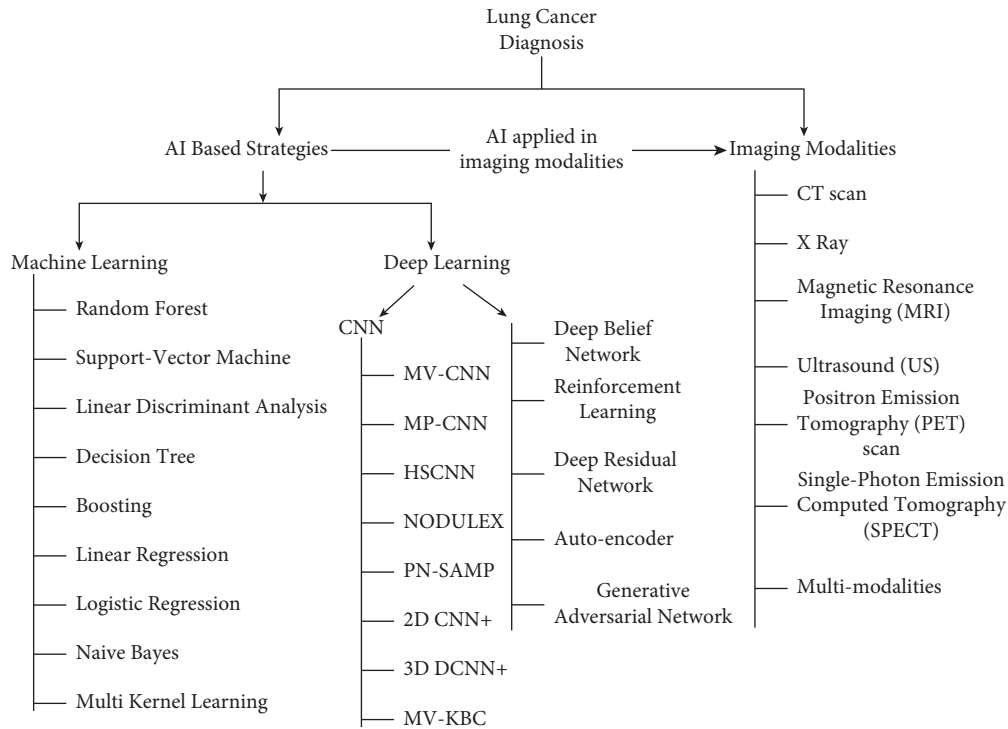


FIGURE 2: A taxonomy of AI-based lung nodule and cancer diagnosis.

TABLE 1: A comparison of different surveys based on lung nodules and cancer detection.

Survey	Ref Year	[27] 2018	[28] 2019	[29] 2019	[30] 2020	[31] 2020	[32] 2020	[33] 2020	[34] 2021	[35] 2021	Ours —
Taxonomy		✗	✗	✗	✗	✗	✗	✓	✗	✗	✓
Dataset		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Image preprocessing		✓	✗	✗	✗	✓	✗	✓	✓	✗	✓
Feature extraction		✗	✗	✗	✓	✓	✓	✓	✓	✓	✓
Segmentation		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Feature selection		✓	✗	✗	✗	✗	✗	✗	✓	✓	✓
Image modalities		✗	✓	✗	✓	✗	✓	✗	✓	✗	✓
Evaluation metrics		✗	✓	✓	✓	✓	✓	✗	✓	✗	✓
Challenges		✗	✓	✓	✗	✓	✗	✗	✗	✗	✓
AI based	Machine learning	✗	✓	✓	✓	✗	✓	✓	✓	✓	✓
	Deep learning										
	CNN	✗	✗	✗	✓	✗	✗	✗	✗	✗	✓
	Other	✓	✗	✗	✗	✓	✓	✓	✓	✓	✓

TABLE 2: A summary of recent surveys/reviews on various lung cancer detection, segmentation, and classification techniques.

Ref.	Purposes	Challenges
[36]	Deep learning techniques are used to detect, segment, and classify pulmonary nodules in CT scans	Generalization ability problem for learning-based methods. It happens because of the different training datasets and the methods.
[29]	A comprehensive analysis of deep learning with convolutional neural network (CNN) methods and their performances	Problems with the generalizability and explication of the detection results, lack of accurate clinical decision-making tools, and well-labeled medical datasets
[30]	The review of recent studies in lung nodule detection and classification provides an insight into technological advancements	Low sensitivity, high false positive rate, time-consuming, small database, poor performance rates, and so on
[4]	A comparison of various machine learning-based methods for detecting lung cancer has been presented	Mainly focuses on machine learning techniques for classification rather than other processes. Also avoid the MRI type data.

TABLE 2: Continued.

Ref.	Purposes	Challenges
[31]	Review of recent deep learning algorithms and architectures for lung cancer detection	The data and the unbalanced nature of it are the current limitations
[37]	Discussing the most recent developments in the field	The size of the target object within the image makes it difficult to implement a CNN; as the size of the target object varies, studies proposed training the model with images of varying scales to teach the model about this size variation
[38]	Providing an accurate diagnosis and prognosis is essential in lung cancer treatment selection and planning	Incorporating knowledge from clinical and biological studies into deep learning methods and utilizing and integrating multiple medical imaging methods
[27]	Algorithms used for each processing step are presented for some of the most current state-of-the-art CAD systems	Limitation of more interactive systems that allow for better use of automated methods in CT scan analysis
[33]	An overview of the current state-of-the-art deep learning-aided lung cancer detection methods, as well as their key concepts and focus areas	Limited datasets and high correlation of errors in handling large image sizes
[35]	A summary of existing CAD approaches for preprocessing, lung segmentation, false positive reduction, lung nodule detection, segmentation, classification, and retrieval using deep learning on CT scan data	Deficient data annotation, overfitting, lack of interpretability, and uncertainty quantification (UQ)
[39]	A survey of what CADe schemes are used to detect pulmonary nodules will help radiologists make better diagnoses	Slight increase in lung density and micronodules whose diameters are less than 3 mm are difficult to detect. For multimodality, clinical records and medical images are not combined.

**2.1. Planning the Review.** This section discusses the planning for creating this review article in detail. The following topics are elaborated upon in the next section. The first is the research topic, the second includes the review materials' sources, and the third includes the inclusion and exclusion criteria.

**2.1.1. Research Questions.** The basic research questions were as follows:

- (1) RQ1: what is the importance of lung cancer detection?
- (2) RQ2: what type of image modalities is used for lung cancer detection?
- (3) RQ3: which datasets are usually used in lung cancer detection?
- (4) RQ4: what are the most used algorithms for feature selection and extraction, segmentation, classification, and detection?
- (5) RQ5: which evaluation matrices are used for lung cancer detection to evaluate the system?
- (6) RQ5: what are the current challenges and limitations of the existing research and the scope of potential future research for lung cancer detection?

**2.1.2. Source of Review Materials.** The survey only looks at high-quality academic articles from MDPI, ScienceDirect, SpringerLink, IEEE Xplore, Hindawi, ACM Digital Library, etc. and papers from well-known conferences.

**2.1.3. Inclusion and Exclusion Criteria.** The most important information for this survey is collected using PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), which is shown in Figure 3. Table 3 shows

the criteria that PRISMA uses to choose which studies to include and which ones to leave out. In addition, this table shows how to select a paper based on certain criteria and standards, which criteria are used, and whether the article is initially accepted or rejected.

**2.2. Conducting the Review.** This section explains how the necessary information is extracted from the articles. Five subphases are addressed to get the most important information and conduct a structured literature review.

**2.2.1. Topical Relationship.** This section describes how the articles selected for this survey connect to the others. Figure 4 shows a word cloud comprised of the papers' keywords and the most important terms from their titles. It indicates how closely the selected articles are connected.

**2.2.2. Aims and Outcomes.** Objectives, contributions, and challenges of different useful articles are presented in Sections 1 and 11.

**2.2.3. Evaluation Metrics.** All the evaluation matrices used for evaluation are explained in Section 10.

**2.2.4. Research Type.** It indicates the type of documents, such as an academic journal, conference or workshop proceeding, book chapter, or thesis.

**2.2.5. Publication Year and Type.** At the start of this project, 610 papers were gathered from different sources, and 423 were chosen for the survey. More than 90% of these articles were published between 2010 and 2021. Therefore, we used more recent articles to update this review.

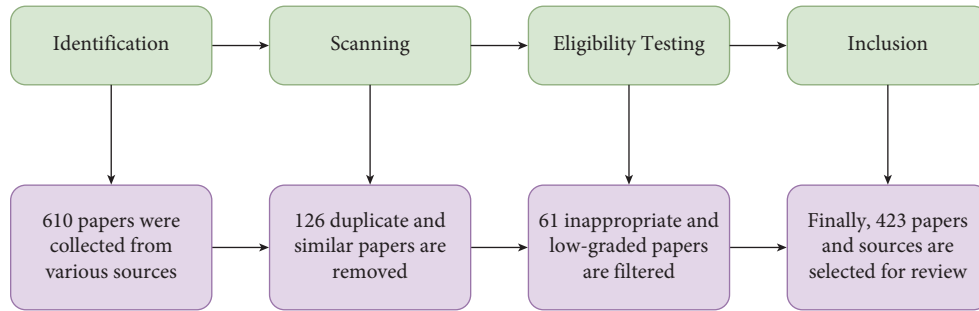


FIGURE 3: The PRISMA process that is followed in this article.

TABLE 3: The criteria used to choose which review articles to include and which ones to exclude.

Inclusion/ exclusion	Criteria
Inclusion	IC1: English language is used to write research papers
	IC2: all papers related to the lung cancer detection process
	IC3: publications in academic journals, book chapters, conference/workshop proceedings, and thesis dissertations
	IC4: published articles between the years 2000 and 2022 (in the survey, a few old papers are used for a specific purpose)
	IC5: articles being available with full text
Exclusion	EC1: not fitting with the theme of the review
	EC2: duplicate articles
	EC3: low-quality papers
	EC4: lack of enough information
	EC5: full text not available

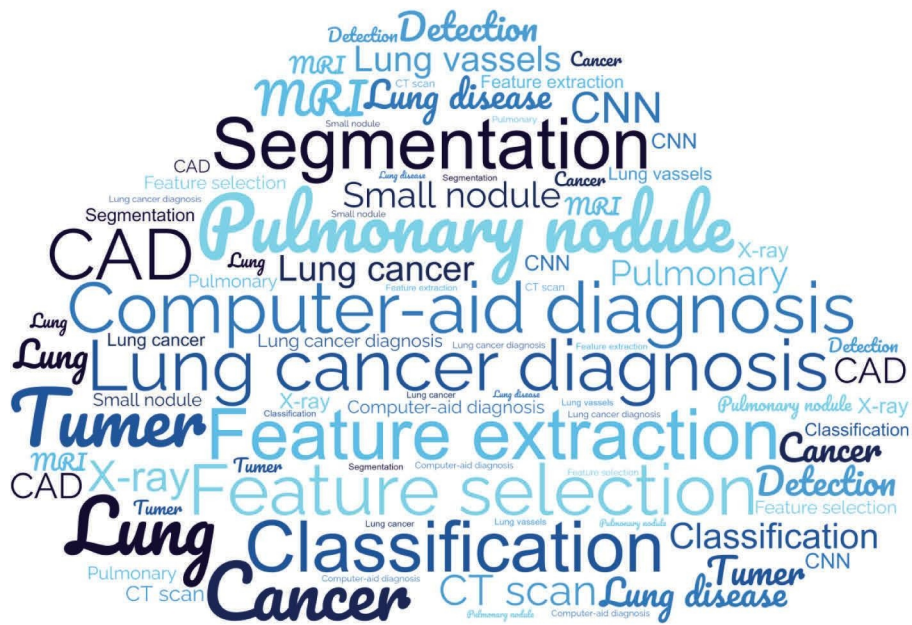


FIGURE 4: Word cloud of the title and selected articles on lung cancer.

2.3. *Outcome.* Finally, the obtained information is examined, existing issues and difficulties are addressed, and future research opportunities are presented.

### 3. Dataset

There are many frequently used datasets that researchers use for lung cancer diagnosis. From Table 4, it can be seen that the CT scan is currently the most reliable method for gathering data on nodule detection in lung cancer. X-rays and MRIs are also used to detect lung cancer and nodules. CT scan is used because it is a confined method that can handle most datasets well. CT scans provide a comprehensive approach for storing data for various reasons. First, the information must be procured and put away by some members or patients. It is unacceptable to have the same storing plan for different patients to get data. After the patients are prepared, individuals have to lie down on a table and go through a passage-like machine that will catch and gauge data. For some time, this data collection strategy has been in place, with a specific recording period dictated by the work's motivation. The data saved in these sessions and recordings are primarily lung nodule images estimated by blocks established in CT scans, X-rays, or MRI. CT scans, X-rays, or MRIs differ from one member to the next and from one session to the next. In this segment, the datasets are portrayed, just as the subjects and X-ray cylinder, indicators, and sessions.

### 4. Imaging Modalities

Imaging is vital for the analysis and treatment of lung nodules and cancer. Hence, this research exhibits that lung cancer analysis relies upon seven particular classifications of clinical imaging modalities. CT scan, X-ray, MRI, ultrasound (US), positron emission tomography (PET) scan, and single-photon emission computed tomography (SPECT) are the seven clinical imaging modalities, and their combination is known as multimodalities. The CT scan is the most basic and widely used imaging modality in lung imaging. As per Table 4, most of the work was done in computed tomography (CT) scan images. The second-highest number of studies delivered is for X-ray images and MRI [103–106]. Another imaging technique known as a chest radiograph is an expensive method with limited accessibility. These should be the reasons for the lower adaptivity of chest radiographs in research, as this imaging strategy was used in a small number of examinations [107, 108]. Ultrasound (US) and PET scan imaging strategies were utilized distinctly in a couple of studies [109–111]. The SPECT imaging strategy has acquired prevalence as of late in lung nodule classification and malignant growth recognition. Because the thermogram dataset is not publicly available, a couple of studies used this imaging strategy [112]. Unfortunately, none of the researchers used histopathology. The well-known imaging strategies are depicted in greater detail in the following section.

4.1. *X-Ray.* A type of high-energy radiation, like electromagnetic waves, is called X-ray. An X-ray is also called X-radiation. X-ray imaging makes images of the inside of the human body. It shows the parts of the body in different shades of black and white [113]. The soft tissues in the human body, such as blood, skin, and muscle, absorb the majority of the X-ray and allow it to transit, resulting in dark gray areas on the film. However, a bone or tumor, which is thicker than soft tissue, prevents most X-rays from passing through and appears white on the film [104]. Gavelli and Giampalma [114] used X-ray images to detect lung cancer. They calculate the sensitivity and specificity for evaluating the outcome.

4.2. *CT Scan.* A computed tomography (CT) scan is a medical imaging method utilized in radiology to get comprehensive body images for diagnostic purposes. It merges a series of X-rays taken from various viewpoints around the body and makes cuts on the bones, veins, and delicate tissues inside the body [115]. CT scans point out a cross section of the body part like bones, organs, and soft tissues more clearly than standard X-rays because normal X-rays are done in two directions. It depicts the structure, size, and location of a tumor [116]. CT scans are more detailed than standard X-rays in identifying cross sections of body parts such as bones, organs, and delicate tissues [117]. In 2018, Makaju et al. [51] used CT scan images to detect lung cancer. Using their proposed model, they attempted to achieve 100% accuracy. Zheng et al. [118] also used CT images to detect lung cancer and inflammatory pseudo-tumor.

4.3. *Magnetic Resonance Imaging (MRI).* MRI is a clinical imaging technique that utilizes radiofrequency signals to create point-by-point images of the organs and tissues in the body. MRI scanners use solid magnetic fields, magnetic field gradients, and radio waves to generate images of the organs in the body [119]. MRI creates images of soft tissues in the human body that are often difficult to see with other imaging techniques. As a result, it is highly effective at detecting and locating cancers. It also generates images that allow specialists to see the location of a lung tumor and estimate its shape and size. A specific dye named a contrast medium is applied to create a better image before the scan [106]. Cervino et al. [120] tried to track lung tumors by performing ANN in MRI sagittal images. The mean error was 7.2 mm using only TM and 1.7 mm when the surrogate was combined with TM.

4.4. *Positron Emission Tomography (PET) Scan.* PET scan is a helpful imaging method that uses radioactive substances known as radiotracers to envision and measure changes in metabolic cycles and other physiological activities, including circulation system, regional compound course of action, and absorption [121]. In addition, PET scan is a diagnostic tool that helps doctors detect cancer in the body. The scan employs a unique shading technique that includes radioactive tracers. Depending on which part of the body is

TABLE 4: Different types of datasets of lung.

Dataset name	Image type	Used in	Unit	Link
LIDC-IDRI	CT scan images	[7, 19, 22, 23, 31, 42–60]	1018	[61]
LUNA16	CT scan (which was gathered from LIDC-IDRI with slice thickness less than 3 mm)	[18, 19, 42, 44, 49, 57, 60, 62–84]	888	[15]
NSLT	Low-dose CT images and chest radiographs	[85, 86]	3410	[87]
The Cancer Imaging Archive (TCIA)	All kind of CT scan and X-ray	[50, 88–96]	3.3 million images	[97]
Society of Radiological Technology (JSRT)	X-ray	[98–102]	154	[100]

being examined, these tracers are either swallowed, ingested, or implanted into a vein in the arm [122]. The PET scan utilizes a mildly radioactive medication to appear in spaces of the body where cells are more dynamic than regular cells. It is used to assist with diagnosing a few conditions, including malignant growth [123]. It can also help determine where the cancer has spread and whether or not it has spread. Because malignant growth cells have a higher metabolic rate than normal cells, they appear as bright spots on PET scans. Lung cancer is the bright spot in the chest that can be seen best on PET and PET-CT images [124]. Weder et al. [111] tried their model in PET scan and got a positive predictive value of 96%.

**4.5. Single-Photon Emission Computed Tomography (SPECT).** SPECT is an atomic medication tomographic imaging method utilizing gamma rays. It is similar to traditional nuclear medicine planar imaging with a gamma camera, but it can provide accurate 3D data. However, it can provide accurate 3D data [125]. A SPECT scan is a test that shows how bloodstreams connect to tissues and organs [126]. Antibodies (proteins that recognize and adhere to cancer cells) can be linked to radioactive substances. First, assuming a tumor is available, the antibodies will be attached to it. Then, at that point, a SPECT output should be possible to recognize the radioactive substance and uncover where the cancer is found [127].

**4.6. Multiple Modalities.** Multiple modalities are considered an educational approach used to relieve the stress of researchers [128]. It entails giving various introductions and experiences of the substance to utilize multiple senses and abilities in a single example. Numerous modalities frequently cater to different learning styles [129]. Modalities can be performed using a combination of chemotherapy and radiation therapy. Concurrent chemoradiotherapy is the simultaneous administration of chemotherapy and radiation therapy [130]. Farjah et al. [131] implemented single, double, and tri-modality in their research. They conducted a CT scan for single modality, CT scan or PET scan with invasive staging for bi-modality, CT scan, PET Scan, and invasive stage for tri-modality.

The advantages and disadvantages of these image modalities are described in Table 5.

## 5. Image Preprocessing

Image preprocessing organizes images before they are used in model preparation and induction. The goal of preprocessing is to improve the quality of the image so that it can be investigated more thoroughly [132]. It includes, but is not limited to, rectifications for resizing, arranging, and shading [133]. As a result, in some cases, a change that could be an expansion may be better served as a preprocessing step in others.

**5.1. Histogram Equalization.** There are two different ways to contemplate and carry out histogram leveling, either as picture change or as range change [134]. Much of the time range change is preferable because it protects the initial data [135]. It is employed in image analysis. To produce a high contrast image, the gray level intensities are expanded along the  $x$ -axis [136]. Asuntha and Srinivasan [137] used a histogram evening out to close the gap. Shakeel et al. [90] changed differences in their dataset. Ausawalaithong et al. [98] preprocessed their picture dataset with histogram balance. It enhances the CT scan's contrast; it spreads out the most frequent pixel intensity values or stretches out the intensity range of the scan. Let  $I$  be a given CT scan image represented as a  $I_x$  by  $I_y$  matrix of integer pixel intensities ranging from 0 to 256. Let  $N$  denote the normalized histogram bin of image  $I$  for available intensity.

$$I_N = \frac{\text{number of pixels with available intensity } n}{\text{total number of pixels}}, \quad (1)$$

where  $n = 0, 1, \dots, 255$ .

**5.2. Median Filter Mask.** The median filter is a non-straight computerized separating strategy, regularly used to eliminate roughness from an image or sign [138]. This type of noise reduction is a common prehandling step used to work on the aftereffects of later preparation. The median filter is a sifting procedure used to remove noise from images and signals [139]. The median filter is essential in image processing because it protects edges during clamor expulsion. It is broadly utilized as it is best at eliminating commotion while safeguarding borders [140]. Tun and Soe [141] used and claimed the median filter mask to be the best filter for their research. Shakeel et al. [142] and Ausawalaithong et al. [98] used a median filter mask in preprocessing their dataset.



TABLE 5: Advantages and disadvantages of imaging modality methods.

Methods	Advantages	Disadvantages
X-ray	Harmlessly and easily helps to analyze sickness and screen treatment	It can harm cells in the body, which thus can build the danger of creating malignant growth. CT scan is better than X-ray.
CT scan	It is easy, painless, and precise. It has the capacity to picture bone, delicate tissue, and veins all simultaneously. It gives point-by-point pictures of many kinds of tissue.	It requires breath holding and radiation which is difficult for a few patients
Magnetic resonance imaging (MRI)	It does not include radiation and is more averse to deliver an unfavorably susceptible response that might happen when iodine-based substances are utilized for X-beams and CT checks	The time required for an MRI is longer than CT. Additionally, MRI is typically less inclined to be quickly accessible than CT.
Positron emission tomography (PET) scan	It diminishes the quantity of examining meetings a patient should go through	Slow developing, less dynamic cancers may not assimilate a lot of tracer
Single-photon emission computed tomography (SPECT)	It tends to be seen in various planes and to isolate covering structures	It has significant expense and less accessibility

Asuntha and Srinivasan [137] reshaped and resized their data with a median filter. Sangamithraa and Govindaraju [143] used the median filter mask in image preprocessing to detect lung nodules. It moves through the lung images pixel by pixel, replacing each value with the median value of neighboring pixels. It can save sensitive components in a picture while filtering noise, and it is good at eliminating “salt and pepper” type noise.

**5.3. Gaussian Filter.** A Gaussian filter is a filter whose response is based on a Gaussian capacity [144]. This effect is widely used in design software, usually to smooth out images and reduce detail [145]. Gaussian noise, also known as Gaussian distribution, is a factual noise with a possible thickness equivalent to ordinary conveyance. This roughness is produced by combining irregular Gaussian capacity with image capacity [146]. This roughness can be eliminated by using a linear filter, as it is an ideal way of eliminating Gaussian roughness. Riquelme and Akhloufi [31], Teramoto et al. [147], and Rossetto and Zhou [148] used Gaussian filters to preprocess their image dataset. Ausawalathong et al. [98], Hosny et al. [149], and Shakeel et al. [150] also utilized this filter to reshape their dataset and use those from detecting lung nodules. Al-Tarawneh [151] and Avanzo et al. [152] used CT scans and preprocessed them with the Gaussian filter. Asuntha et al. [153], Wang et al. [154], Sang et al. [65], and Ozdemir et al. [42] smoothed and preserved edges with the Gaussian filter. Fang [155] and Song et al. [156] applied the Gaussian filter on the LUNA16 [15] dataset to detect lung cancer. The effect of Gaussian smoothing is to blur CT scans of the lung similar to the mean filter. The standard deviation of the Gaussian determines the degree of smoothing. Gaussian blurring the CT image minimizes the amount of noise and reduces speckles.

(i) In 1D:

$$g_{\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right). \quad (2)$$

(ii) In 2D:

$$G_{\sigma}(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \quad (3)$$

where  $\sigma$  is referred to as standard deviation of the distribution. The mean of the distribution is considered as 0.

**5.4. Wiener Filter.** The Wiener filter is the MSE-ideal fixed straight filter for images corrupted by added substance clamor and obscuring. Wiener filter works when the sign and roughness measures are assumed to be fixed [157]. Sangamithraa and Govindaraju [143] removed the added substance noise while also modifying the obscuring. In terms of the mean square error, Wiener filtering is ideal. It is used to measure a perfect or arbitrary target interaction by straight time-invariant sifting of a detected noisy cycle, expecting to know fixed sign and noise spectra, and adding substance noise. [158]. It restricts the overall mean square error during backward sifting and commotion smoothing. It removes the additive noise and inverts the blurring simultaneously in lung images. It eliminates the additive noise, transforms the obscuring, and limits the general mean square error during inverse filtering and noise smoothing. The Wiener sifting is a direct assessment of the first picture [159].

**5.5. Gabor Filter.** A Gabor filter is a straight filter used in image processing for surface analysis, which means it determines whether or not there is a specific recurrence content in the lung images in explicit terms in a restricted district surrounding the point or region of examination [160]. It investigates whether there is a particular recurrence content. It has gotten significant consideration as it takes after the human visual framework. It is a neighborhood operation in which the value of any given pixel in the output lung scan is determined by applying some algorithm to the

importance of the pixels in the neighborhood of the corresponding input pixel. To remove noise from the dataset, Mary and Dharma [161] used the Gabor filter.

$$F(u_1, u_2) = \exp\left(-\frac{(\hat{u}_1^2 + \gamma^2 \hat{u}_2^2)}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda} \hat{u}_1\right), \quad (4)$$

$$\hat{u}_1 = u_1 \cos \theta + u_2 \sin \theta,$$

$$\hat{u}_2 = -u_1 \sin \theta + u_2 \cos \theta,$$

where  $\lambda$  means the wavelength of the sinusoidal factor,  $\theta$  represents the orientation of the normal to the parallel stripes of a Gabor function,  $\sigma$  is the sigma/standard deviation of the Gaussian envelope, and  $\gamma$  is the spatial aspect ratio and specifies the ellipticity of the support of the Gabor function.

**5.6. Isotropic Voxel.** Voxel is short for volume pixel, the smallest recognizable box-formed piece of a 3D picture. It could be compared to the 2D pixel [162]. The voxel size on CBCT images is isotropic, meaning that all sides are of the same size and have a uniform goal in every direction. The voxel technique was used by Nagao et al. [163] and Wang et al. [164] to reduce sharp noises and classify lung cancer. This method was also used by Quattrocchi et al. [165] to reshape their dataset to detect breast cancer and lung cancer.

**5.7. Thresholding.** Thresholding is a non-linear operation that changes a grayscale image into a binary image in which the two levels are allocated to pixels that are either below or above the set threshold value. It mainly converts an image from shading or grayscale into a twofold picture [166]. Thresholding is used to convert a low-contrast lung scan to a high-contrast lung scan. Thresholding is also a very effective tool in image segmentation. Its purpose is to convert grayscale images to binary format [151]. It takes the colorful or grayscale lung scans and turns them into binary scans. It diminishes the intricacy, works on acknowledgment and grouping, and changes the pixels to simplify the picture.

**5.8. Binary Inversion.** High-contrast picture reversal is a picture handling strategy where light regions are planned to dim, and dull areas are scheduled to light. A rearranged high-contrast picture can be considered an advanced negative of the first picture. Sharma et al. [167] used binary inversion to reduce noise from image datasets.

**5.9. Interpolation.** Image interpolation happens when one resizes or contorts one's image, starting with a one-pixel grid and then onto the next. Zooming refers to increasing the number of pixels in an image so that the image's details can be seen more clearly [168]. Interpolation is a well-known method for surveying dark characteristics that lie between known characteristics [169]. Interpolation is a course of deciding the obscure qualities in the middle of the realized information focus. It smooths, enlarges, or averages CT

scans displayed with more pixels than that for which they have initially been reconstructed. It is used to foresee obscure qualities. It forecasts values for cubic in a raster. It is generally used to foresee the obscure qualities of any geographical information, such as commotion level, precipitation, rise, and so on. The most common way to use test points with known qualities to figure out prices at other unknown issues is by insertion [170]. It could be used to predict dark characteristics for any geographic point data, such as height, precipitation, substance obsessions, disturbance levels, and so on [171]. Several insertion strategies have previously been reported. The broadly utilized strategies are the nearest neighbor, bilinear, bicubic, b-splines, lanczos2, and discrete wavelet transform. Lehmann et al. [172] and Zhao et al. [173] used interpolation in their dataset to detect nodules in the lungs. Liu et al. [58] used interpolation in CT scans and cleared noise, and Cascio et al. [174] used interpolation in 3D images to reduce noise.

**5.10. Synthetic Minority Oversampling Technique (SMOTE).** SMOTE is an oversampling procedure that permits us to produce manufactured examples for our minority classes [175]. It is an oversampling method that creates fabricated models for the minority class. This computation aids in overcoming the overfitting problem caused by unpredictability in oversampling [176]. The imbalanced arrangement has the disadvantage of having too few instances of the minority class for a model to become comfortable with the choice limit [177]. Oversampling models from the minority class are regarded as one solution to this problem [178]. It randomly chooses a minority class instance and finds its  $k$  nearest minority class neighbors. The fabricated occasion is then created by arbitrarily selecting one of the  $k$  nearest neighbors  $b$  and coupling  $a$  and  $b$  to frame a line segment in the component space. The manufactured examples are made by mixing the two chosen occurrences,  $a$  and  $b$  [179]. While restructuring the information with SMOTE, Chen and Wu [180] found the risk factors. Patil et al. [181] utilized it to smooth textures and minimize noise. Wang et al. [182] employed SMOTE to remove borderlines.

**5.11. Contrast Limited Adaptive Histogram Equalization (CLAHE).** Contrast limited AHE (CLAHE) is a variation of versatile histogram in which the differentiation enhancement is restricted to diminish this issue of clamor intensification [183]. It is utilized further to develop hazy pictures or video ability levels. It works on little districts in images, called tiles. The adjacent tiles are then consolidated using bilinear insertion to remove the erroneous limits [184]. CLAHE calculation differs from standard HE in that CLAHE works on small areas of the image called tiles and registers a few histograms, each comparing to a specific segment of the image and using them to rearrange the advantages of the picture [185]. In CLAHE, the differentiation enhancement near given pixel value is provided by the incline of the change work [186]. Punithavathy et al. [187], Bhagyarekha and Pise [188], and Wajid et al. [189] used CLAHE as image preprocessing methodology. Technically, CLAHE does this

by setting a threshold. If some gray levels in the lung scan exceed the threshold, the excess is evenly distributed to all gray levels. After this processing, the lung scan will not be over-enhanced, and the problem of noise amplification can be reduced.

Table 6 shows the pros and cons of these image pre-processing techniques.

## 6. Segmentation

Lung nodule segmentation is a crucial process designed to make the quantitative assessment of clinical criteria such as size, shape, location, density, texture, and the CAD system more manageable and more efficient [196–198]. However, because of their solidity, location, or texture, lung nodules such as juxta-pleural (nodules directly attached to the pleura's surface), juxta-vascular (nodules connected to vessels), and ground-glass nodules can be challenging to remove. Deep learning-based segmentation is a pixel-by-pixel categorization technique used to calculate organ probability [30]. This method is divided into two stages: the first is the creation of the probability map using CNN and image patches and the second is the refinement of the probability map using the general background of images and the probability map [196].

**6.1. Watershed.** Watershed segmentation is a technique for segmenting watersheds that use image morphology [199]. It requires the selection of at least one marker (“seed” point) within each image object, including the background as a separate object. The markers are picked by an operator or provided by an automatic mechanism that considers the object's application-specific information. A morphological watershed transformation helps to grow them after marking the items [200]. After the lung image preprocessing, noise is removed, images are smooth, and features are enhanced. Watershed is used in lung segmentation to identify the various regional maxima and minima [201].

**6.2. U-Net.** The U-Net [202] architecture is the most used architecture for medical image segmentation, and it significantly improves process performance. The fundamental parts of the U-Net are association of convolution layers in the contracting path and deconvolution layers in the expansive direction. It includes a contraction method for capturing anatomical structure and an asymmetrical expansion method for precise localization [28]. U-Net has enabled the segmentation process to form a spatial context at several scales despite the challenges of collecting both global and local contexts. As a result, it may be trained from end to end using only a small quantity of training data [28]. Convolution layers with rectified linear units and max-pooling layers make up the contracting route, similar to the classic architecture of a convolutional neural network. On the other hand, the expanding method entails sampling the feature map, followed by up-convolution and convolution layers using ReLU. Because of the loss of border pixels at each convolution, the extracting path's matching feature map must be cropped and

concatenated with equivalent layers in the expensive direction [53]. The input photos and their respective masks are utilized for training the U-Net during the training phase. A lung image is supplied as input to generate the appropriate mask output during the testing phase. The mask is then applied to the relevant image to segment the area of interest, in this case, lung parenchyma [202].

**6.3. Multiview Deep Convolutional Neural Network (MV-CNN).** The multiview deep convolutional neural network (MV-CNN) [203] architecture for lung nodule segmentation is a CNN-based architecture that proposes to transform lung nodule segmentation into CT voxel classification. The MV-CNN comprises three branches that process voxel patches from CT images in axial, coronal, and sagittal views. To obtain the voxel label, the three branches all have identical structures, including six convolutional layers, two max-pooling layers, and one fully connected layer. In addition, a parametric rectified linear unit (PReLU) [204] is implemented as a non-linear activation function after each convolutional layer and the first fully connected layer, and batch normalization is used for training acceleration [205].

**6.4. Central Focused Convolutional Neural Network (CF-CNN).** The central focused convolutional neural network (CFCNN) [206] architecture includes three-dimensional and two-dimensional CT imaging views for lung nodules and cancer segmentation. It uses a CT image to extract a three-dimensional patch and a two-dimensional different plate patch self-contained on a single voxel as input to the CNN [207] model, which predicts whether the voxel belongs to the nodule or healthy tissue class. After feeding all voxels into this CNN model, a probability map assigns each voxel a probability of belonging to a nodule.

**6.5. Fuzzy C-Means (FCM).** The FCM algorithm [208] is one of the most extensively used fuzzy clustering methods. Data elements can belong to multiple clusters in fuzzy clustering, and each part has a set of membership levels associated with it. It uses a CT image to extract a three-dimensional patch and a two-dimensional different plate patch self-contained on a single voxel as input to the CNN [207] model, which predicts whether the voxel belongs to the nodule or healthy tissue class. After feeding all voxels into this CNN model, a probability map assigns each voxel a probability of belonging to a nodule.

**6.6. Hessian-Based Approaches.** Image enhancement is performed on voxels in Hessian-based strategies to acquire the 3D Hessian matrix for each voxel and calculate the relevant eigenvalues. These eigenvalues are used to locate and segment lung nodules in a subsequent step. To begin, multiscale smoothing is used to reduce noise in the image and make nodule segmentation easier. Following that, the 3D Hessian matrix and associated eigenvalues are computed, and the results of each method are combined to produce the segmentation masks [209].

TABLE 6: Advantages and disadvantages of image preprocessing methods.

Algorithms	Advantages	Disadvantages
Histogram equalization [190]	It is a versatile strategy to the picture and an invertible administrator. It can be recuperated and expands differentiation of pictures.	It is not the best technique for contrast improvement and is unpredictable. It expands the contrast of foundation noise.
Median filter mask [10]	It can save sharp components in a picture while filtering noise, and it is good at eliminating “salt and pepper” type noise	It separates picture edges and produces false noise edges and cannot smooth medium-tailed noise dissemination
Gaussian filter [191]	Its Fourier change has zero recurrence. It is broadly utilized to diminish picture noise and lessen detail.	It decreases subtleties and cannot deal with “salt and pepper” noise. It sometimes makes all parts blue and obscures the objects.
Wiener filter [192]	It eliminates the additive noise, transforms the obscuring, and limits the general mean square error during inverse filtering and noise smoothing	It is hard to acquire ideal rebuilding for the noise, relatively delayed to apply as working in the recurrence area
Gabor filter [151]	It investigates whether there is a particular recurrence content. It has gotten significant consideration as it takes after the human visual framework.	It requires huge investments. It has a high excess of provisions.
Isotropic voxel [193]	It is the fastest approach and a “precise” 3D structure block, as it copies particles and opens new reproduction procedures	It is hard to fabricate complex articles utilizing voxels. It does not have numerical accuracy.
Thresholding [142]	It diminishes the intricacy, works on acknowledgment and grouping, and changes the pixels to make the picture simpler	There is no assurance that the pixels distinguished by the thresholding system are bordering
Binary inversion [194]	CT scans were converted into black and white to detect the nodules as binary inversion will get the dark part as black which means 1	It is not a clear form to detect nodules and it has a huge chance to miss the nodules
Interpolation [195]	It is used to foresee obscure qualities. It forecasts values for cubic in a raster.	It obscures the edges when the decreased proportion is less
SMOTE [179]	It is an oversampling procedure and is powerful to handle class awkwardness. It assists with conquering the overfitting issue.	It can build the covering of classes and present extra commotion. Often it does not constrict the predisposition.
CLAHE [187]	The adjoining tiles are joined using bilinear expansion to take out incorrect representation incited bounds	Any commotion that might be accessible in the picture

6.7. *SegNet + Shape Driven Level Set.* SegNet [210], a deep, fully convolutional network architecture, is used for coarse segmentation because it is designed primarily for pixelwise semantic labeling. A high-level network model SegNet is a network composed of encoders and decoders. SegNet is a preconfigured segmentation solution for a variety of medical imaging applications [211, 212]. A batch of lung field images is used during the training phase to feed the deep network. The output of CNN is used to initialize the level set function for lung nodule segmentation. The authors [213] used shape information as the primary image feature to guide the evolving shape to the intended item border.

6.8. *Faster R-CNN.* Faster R-CNN [214] is an improvement on the previous Fast R-CNN [215]. As the name implies, Faster R-CNN is much faster than Fast R-CNN due to the region proposal network (RPN). The model comprises two parts: the RPN and the Fast R-CNN. The input image is first subjected to convolution and pooling operations via the basic feature extraction network to obtain the image’s feature map. After that, the feature map is transmitted to the RPN network, which performs preliminary border regression and classification judgment on the image. As the foundation for categorizing, the candidate frame is classified based on the background or object to be recognized. The RPN outputs the candidate frame’s position and score

information, and then they are sent to the Fast R-CNN for final processing by the fully connected layer. They are the final regression of the frame and the specific categorization of the object to be recognized in the final regression. First, ConvNet [216] is used to extract feature maps from lung pictures. Next, these are fed into RPN, which returns the candidate bounding boxes. The ROI pooling layer is then applied to reduce the size of the candidates. Finally, the proposals are transferred to a fully linked layer to obtain the final lung segmentation result [217].

6.9. *Mask R-CNN.* Mask R-CNN [80] is a compact and adaptable generic object instance segmentation system. It recognizes targets in images and provides high-quality segmentation results for each target. Mask R-CNN is divided into two sections, the first of which is RPN. It is a new network developed by Faster R-CNN [214] that replaces the previous R-CNN’s selective search approach [215], and Fast R-CNN [215] integrates all content into a single network, significantly improving detection speed. The second stage features two concurrent branches, one for detection and the other for classification and bounding box regression. The mask branch is used for segmentation. The preprocessing program receives raw lung image sequences and generates 2D images before processing basic images such as coordinate transformation, slice selection, mask generation, and

normalization. Then, it is used in the detection and segmentation module to detect and segment the locations and contours of expected pulmonary nodules [218].

**6.10. Robust Active Shape Model (RASM).** Biomedical photos typically feature complicated objects that fluctuate significantly in appearance from one image to the next. It can be challenging to measure or recognize the existence of specific structures in such photos. The RASM [219] is trained using hand-drawn contours in training images. It employs principal component analysis (PCA) to identify critical variances in the training data, allowing the model to automatically determine whether a contour is a potentially excellent object contour [220, 221]. It also includes matrices that describe the texture of lines perpendicular to the control point; these are utilized to rectify positions during the search stage. The contour is deformed by finding the best texture match for the control points when the RASM is created. The movement of the control points is limited by what the RASM perceives as a “normal” object contour based on the training data in this iterative procedure. Then, PCA determines the formation’s mean appearance (intensities) and variances in the training set. For example, the outline of the lungs is approximately segmented from lung images using a robust active shape model matching technique [222].

**6.11. Region Growing.** Growing a region is a bottom-up process that starts with a set of seed pixels [223]. The goal is for each seed to establish a uniformly connected zone. Intensity indicates that the measurement is used to grow a region from a seed point and to segment it. As each unallocated nearby pixel in the area is compared, the region’s size increases. To compute similarity, the difference between the intensity value of a pixel and the region’s mean is used. The pixel is assigned to the area, and the minor difference is calculated. The operation is terminated when the intensity difference between the region means and the new pixel exceeds a predetermined threshold. Each pixel’s intensity values are compared to those of its neighbors starting with the seed, and if they are within the threshold, the pixel is labeled as one [219]. Next, an image of a tumor-bearing lung is uploaded. The growth’s starting point (pixel) coordinate is established, and the base value stores the selected point’s color intensity. Next, the initial pixel is stored in an array’s coordinates. The process continues until all pixels are eligible and the queue is full. The tumor tissue refers to all pixels in the points array that create a surface. The outermost pixels are also introduced as the tumor boundary, which is curved [224].

Table 7 shows the pros and cons of segmentation methods.

## 7. Feature Extraction

Feature extraction is a process that reduces an initial collection of raw data into more manageable groups that are easier to process [228]. It reduces the number of features in a dataset by creating new ones from existing ones. The feature extraction strategy provides new features that directly blend

with the existing elements. When compared to the first feature esteems, the new arrangement of elements will have various qualities [229]. The main point is that fewer features will be required to capture comparable data [230].

**7.1. Type of Features.** Some features need to be extracted and selected to detect lung nodules and cancer more efficiently. There are three kinds of features. If these features are removed, the outcome can be boosted.

**7.1.1. Shape-Based Feature.** Shape features are significant because they give an option in contrast to depicting an object, utilizing its many attributes, and diminishing how much data are put away. It is one of the most fundamental characteristics of a mass. The irregularity of the mass’s shape makes removal difficult [231]. It is classified into two types: region-based techniques and contour-based techniques. A curve estimation method, peak point characterization, and peak line following calculation are all used. Local procedures use the entire item region for its shape highlights, while form-based techniques use data in an article. Shape highlights are classifications of a morphological part. Figure 5 shows the shape-based features very clearly.

**7.1.2. Texture-Based Feature.** The texture is used to segment pictures into areas of interest and group those locales. It refers to all spatial area variations and the selection of general visual perfection or harshness of images. The texture is defined as the spatial distribution of force levels in a given area. They provide invaluable information about the underlying object arrangements of action in a picture, as well as their relationship to climate [231]. Texture-based features are shown in Figure 6.

**7.1.3. Intensity-Based Feature.** Intensity refers to how much light is emitted or the mathematical worth of a pixel. As demonstrated by image feature intensity, it first requests insights that rely upon individual pixel esteems. The intensity of the light varies from pixel to pixel [231]. Therefore, pixel intensity is the most easily accessible pattern recognition component. Shading is typically addressed by three or four-part intensities in a shading imaging system. The mode, median, standard deviation, and variance of image intensity can all be used to evaluate it. Figure 7 gives a clear view of intensity-based features.

**7.2. Feature Extraction Methods.** The feature extraction strategy gives us new elements, which are considered a straight mix of the current features. The new arrangement of features will have various qualities when contrasted with the first feature esteems. The fundamental point is that fewer features will be needed to catch similar data.

**7.2.1. Radiomics.** Radiomics is a strategy that separates an enormous number of provisions from clinical pictures utilizing information portrayal measurements [232].

TABLE 7: Advantages and disadvantages of segmentation methods.

Algorithms	Advantages	Disadvantages
Watershed [225]	Being able to divide an image into its components	Takes too long to run in order to meet the deadline, sensitivity to false edges and over-segmentation
U-Net [226]	Images can be segmented quickly and accurately	Redundancy occurs due to patch overlap, also relatively slow
MV-CNN [203]	No user-interactive parameters or assumptions about the shape of nodules are needed	The loss of gradients may have an effect
CF-CNN [206]	Gathered sensitive information about nodules from CT imaging data	Less adaptable for small nodules and cavitary nodules
FCM [188]	Ignored noise sensitivity limitation, successfully overcoming the PCM's clustering problem	Row sum constraints must be equal to one in order to work
Hessian-based approaches [209]	High robustness against noise and sensitivity to small objects	Performance decreases for large nodule
SegNet + shape driven level set [213]	Correct seed point initialization with no manual intervention in the level set	Segments the lung nodule partly occluded, also takes a longer time
Faster R-CNN [214]	The efficiency of detection is high	It could take a long time to reach convergence
Mask R-CNN [218]	Easy to train, generalizable to other tasks, effective, and only adds a minor overhead	Low-resolution motion blur detection typically fails to pick up on objects
RASM [219]	Well suited to large shape models and parallel implementation allowing for short computation times	Cannot segment areas with sharp angles and is not built to handle juxta-pleural nodules
Region growing [227]	The concept is simple, multiple criteria can be selected simultaneously, and it performs well in terms of noise	Computing is time-consuming. Noise or variation may result in holes or over-segmentation, making it difficult to distinguish the shading of real images.

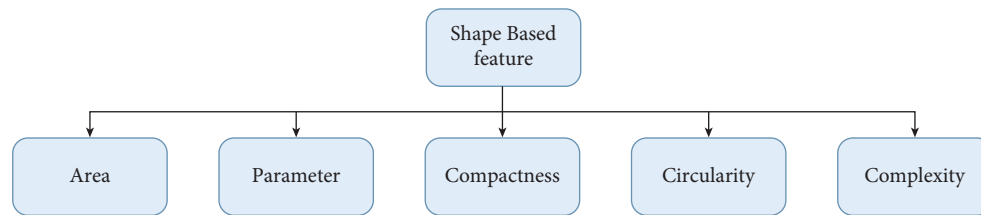


FIGURE 5: Shape-based features.

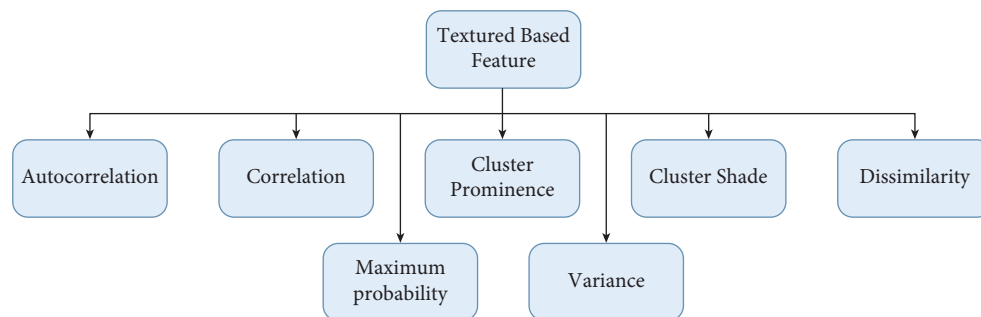


FIGURE 6: Texture-based features.

Radiomic highlights may reveal growth examples and qualities that the unaided eye does not recognize [233]. The standard radiomic investigation includes the evaluation of size, shape, and textural highlights that contain useful spatial data on pixel or voxel circulation and examples [234]. Echegaray et al. [235], Vial et al. [236], and Pankaj et al. [237] used the radiomics method for feature extraction. Mahon et al. [238] used radiomic radiology to extract features.

*7.2.2. Transfer Learning and Fine-Tuning.* It first trains a base network on a base informational index and undertakes transfer learning. Afterward, it exchanges the learned components to a subsequent objective organization to prepare for objective informational collection and errand. It trains a model on a dataset and uses it for preparing another dataset [239]. Nishio et al. [240], Sajja et al. [159], and da Nóbrega et al. [241] used transfer learning for lung cancer.

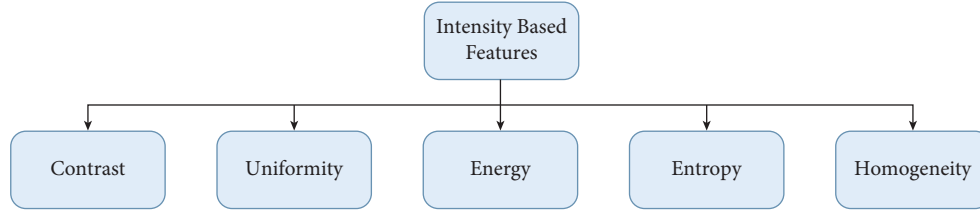


FIGURE 7: Intensity-based features.

Haarburger et al. [242], Marentakis et al. [94], Paul et al. [243], and Tan et al. [244] fine-tuned image to extract features. It takes the underlying patterns, and then a pre-trained model has learned and adjusted its outputs to be more suited to your problem. It saves preparation time, does better execution of neural organizations, does not require a great deal of data, and can prompt higher exactness.

**7.2.3. LSTM + CNN.** The LSTM strategy has turned into a fundamental structure square of neural NLP [245]. To strongly approve of moving examples, some use them as contributions to a value-based classification approximate to the first LSTM production [246]. The CNN long short-term memory network, or CNN LSTM for short, is LSTM engineering explicitly intended for grouping expectation issues with spatial information sources, similar to pictures or recordings. Concerning the improvement of the CNN LSTM model design for system expectations. Tekade and Rajeswari [247] used a layer of CNN LSTM for feature extraction in lung image datasets. Pictures can also be addressed with high-request statistical features processed from run-length matrices or frequent models. Statistics are basic measurements that help us for better comprehension of our pictures [248].

**7.2.4. Standard Deviation.** Standard deviation limits the ratio of reserves or dispersions of many properties. A low-quality deviation indicates that the properties will be close to the set average as a general rule. In contrast, an elite requirement deviation suggests that the properties will cover a large area [249].

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (S_i - \mu)^2}, \quad (5)$$

where  $\sigma$  is the population standard deviation,  $N$  means the size of items,  $S_i$  is each value from the set, and  $\mu$  is the mean of all the values.

**7.2.5. Variance.** Variance is the inconstancy in the model expectation—how much the ML capacity can change contingent upon the given informational collection [250]. In this technique, the modified term quantifies how far each number is from the mean and how far each unit number is from the mean [251].

$$\sum_{i=0}^{n-1} \sum_{j=0}^{n-1} (i - \mu)^2 \cdot p(i, j), \quad (6)$$

where  $\mu$  is the mean of all the values.

**7.2.6. Mean.** Mean is a method for executing feature extraction. It ascertains and takes away the mean for each component. A typical practice is similar to separate this worth by the reach or standard deviation.

$$\mu = \frac{1}{N} \sum_{i=1}^N S_i, \quad (7)$$

where  $\sigma$  is the population standard deviation,  $N$  is the total amount of pixel present in the segmented region,  $S_i$  is each value from the set, and  $\mu$  is the mean of all the values.

**7.2.7. Fourth-Moment Kurtosis.** The kurtosis  $k$  is characterized to be the normalized fourth focal second. The fourth second is kurtosis, which indicates the level of focal “peakedness” or, more accurately, the “largeness” of the external tails. Kurtosis denotes whether the data have been significantly or lightly followed by the traditional course [252].

$$ku = \left( \frac{1}{N * \sigma^4} * \sum_{i=1}^N (S_i - \mu)^4 \right)^{1/4}, \quad (8)$$

where  $\sigma$  is the population standard deviation,  $N$  is the total amount of pixel present in the segmented region,  $S_i$  is each value from the set, and  $\mu$  is the mean of all the values.

**7.2.8. Third-Moment Skewness.** Skewness is a proportion of the evenness of a circulation. It estimates the measure of likelihood in the tails [253]. The worth is frequently compared to the kurtosis of the average conveyance, which is equal to three. If the kurtosis is more remarkable than three, the dataset has heavier tails than a typical appropriation [254].

$$sk = \left( \frac{1}{N * \theta} * \sum_{i=1}^N (s_i - \mu)^3 \right)^{1/3}, \quad (9)$$

where  $\sigma$  is the population standard deviation,  $N$  is the total amount of pixel present in the segmented region,  $S_i$  is each value from the set, and  $\mu$  is the mean of all the values.

**7.2.9. Entropy.** Entropy is a substantial proportion of irregularity that can describe the surface of the info picture. In image processing, discrete entropy is a proportion of the number of pieces needed to encode picture data [255]. It distinguishes different communication signals by describing the signals' distribution state characteristics. It is utilized in any course of weight assurance. It is vigorous and computationally fundamental. The higher the entropy value is, the more detailed the image will be. Entropy is a proportion of haphazardness or confusion and thus a proportion of vulnerability [256]. Hussain et al. [257] used entropy to analyze lung cancer image data.

$$\sum_{i,j=0}^{n-1} -\ln (P_{ij})P_{ij}. \quad (10)$$

**7.2.10. Autoencoders.** Autoencoder is a sort of neural network that is utilized to gain proficiency with a compacted portrayal of unrefined information [258]. An autoencoder is made up of an encoder and a decoder submodel [259]. The encoder compresses the information, and the decoder attempts to reproduce the contribution from the encoder's compressed variant. Ahmed et al. [260], Z. Wang and Y. Wang [261], Z. Wang and Y. Wang [262], and Kumar et al. [22] used an autoencoder to extract the feature and classify lung nodules. The encoder compresses the input lung scan, and the decoder attempts to recreate the input lung scan from the compressed version provided by the encoder. It can be incredible to highlight extraction, conservativeness, and speed in using backpropagation.

**7.2.11. Wavelet.** Wavelet is a frequency-selective modulation technique [263]. The wavelet change can assist with changing over the sign into a structure that makes it a lot simpler for our pinnacle locator work. Sometime after the first ECG signal, the wave coefficient for each scale is plotted. Wavelet was used by Kumar et al. [22] to extract features. Soufi et al. [264] attempted to detect lung cancer using a wavelet. Park et al. [265] included and extracted a large number of wavelet features. A discrete wavelet transform (DWT) decomposes a signal into sets of numbers. Every set is a period series of coefficients portraying the time development of the signal in the corresponding frequency band (DWT). DWT is an effective tool for multiresolution analysis, and it is primarily pursued in signal processing, image analysis, and various classification systems [266]. It is broadly used in feature extraction as it is efficient, which can be declared by seeing its previous results.

**7.2.12. Histogram of Oriented Gradients (HOG) Features.** HOG, or histogram of oriented gradients, is a feature extractor that is frequently used to extract features from picture information [266]. Adetiba and Olugbara [267] used HOG to improve image clarity. Xie et al. [268] used a variety of feature extraction methods, including HOG. Firmino et al. [269] used HOG to extract features from lung image data to detect cancer.

(i) Mathematically, for a given vector  $V$ :

$$V = [a1, a2, a3, \dots, a36]. \quad (11)$$

(ii) We calculate root of the sum of squares:

$$k = \sqrt{(a1)^2 + (a2)^2 + (a3)^2 + \dots + (a36)^2}. \quad (12)$$

(iii) Divide all the values in the vector  $V$  with this value ( $K$ ):

$$\text{normalized vector} = \left( \frac{a_1}{k}, \frac{a_2}{k}, \frac{a_3}{k}, \dots, \frac{a_{36}}{k} \right). \quad (13)$$

**7.2.13. AlexNet, VGG16, and VGG19.** AlexNet is the name of a CNN that usually affects AI in a way that unequivocally selects some way of looking at a machine [270]. It joined ReLU initiation after each convolutional and completely associated layer. VGG16 is a CNN model that is represented in the paper by Zisserman from the University of Oxford in their survey [271]. The model achieved 92.7% of the top-5 test accuracy on ImageNet (a dataset of fourteen million+ images, including one thousand classes). The most striking feature of the VGG16 is that, unlike many other hyperboundaries, it consistently empties the convolution layers and uses the same cushioning and max pool [272]. VGG19 is a 19-level deep vascular neural entity. Creating more than 1,000,000 images from the Imagine information base can save an organization's pretrained presentation. Khan et al. [273] presented a pretrained VGG19-based automated segmentation and classification technique for analyzing lung CT images that achieved 97.83% accuracy.

Table 8 shows the pros and cons of feature extraction methods.

## 8. Feature Selection

Feature selection refers to reducing the number of input variables required to develop a predictive model. It would be preferable to reduce the number of input variables that can lower the overall computing cost of the model and, in some cases, improve its performance [281]. The primary advantage of feature selection is that it aids in determining the significance of the original feature set.

**8.1. Genetic Algorithm (GA).** GA is used to identify the most relevant features for lung nodule detection. The GA generates a binary chromosome of 4096 bits in length evaluated offline during the CADe system's training phase.

Logic "1" indicates that this feature is relevant, and logic "0" means irrelevant. As a result, it is removed from the test phase's optimized feature vector. The fitness function is then calculated for each of the population's chromosomes [282]. It uses an evolutionary approach to determine an efficient set from lung images. The initial stage in feature selection is to create a population based on subsets of the possible



TABLE 8: Advantages and disadvantages of feature extraction methods.

Algorithms	Advantages	Disadvantages
Radiomics [274]	It could extricate and distinguish many provisions and component types. It has a minimal expense.	For respiratory movement, it obscures data. It has restricted data of remade pictures.
Transfer learning and fine-tuning [244]	It saves preparation time, does better execution of neural organizations, does not require a great deal of data, and can prompt higher exactness	Transfer learning has the issue of negative exchange. Fine-tuning can at some point befuddle to sort out subclasses.
LSTM + CNN [94]	It is appropriate to separate compelling elements and group, process, and foresee time series given delays of obscure length	It is inclined to overfitting, and it is hard to apply as it requires 4 direct layers which require a lot of memory
Standard deviation [275]	It gives an exact thought of how the data are appropriated. It is detached by outrageous qualities.	It tends to be affected by anomalies, is hard to ascertain or comprehend, and works out all vulnerability as error
Autoencoder [276]	It can be incredible for highlight extraction, conservativeness, and speed in coding utilizing backpropagation	It cannot deal with adequate preparation information, prepares some unacceptable use cases, and is excessively lossy
Variance [277]	It treats all deviations from the mean and assists an association with being proactive in accomplishing targets	It gives added weight to anomalies, is not effectively deciphered, and does not offer wonderful precision
Fourth-moment kurtosis [50]	It will be in the positive structure, and conveyance about the mean gets tighter as the mean gets bigger	The weakness is that it will not have a negative or indistinct structure
Wavelet [278]	It offers a synchronous restriction on schedule and recurrence space. It is quick and can isolate the fine subtleties in a sign.	It has shift affectability, its directionality is poor, and it has absence of stage data
Entropy [279]	It is utilized in any course of weight assurance. It is vigorous and computationally basic.	It has restricted critical thinking part and relative disparity, contingent upon the given length and biasing
Histogram of oriented gradients [267]	It shows invariance to photometric changes by making a dark foundation with white molecules which sharpens the articles unmistakably	The last descriptor vector develops bigger to set more effort to extricate and to prepare utilizing a given classifier
Third-moment skewness [50]	It is smarter to gauge the presentation of the speculation returns, transforming the data point of high skewness into slanted conveyance	It is eccentric. The ascent and defeat of a network are best instances of the skewness.
AlexNet, VGG16, and VGG19 [280]	AlexNet has 8 layers that exceed the yield dissimilar to other enactment capacities. VGG is an incredible structure block for learning reasons.	AlexNet battles to examine all provisions accordingly delivering helpless performing models. VGGNet is agonizing to prepare and its loads itself are very huge.

characteristics derived through lung feature extraction. Then, the subsets of this population are evaluated using a predictive model for the target task.

**8.2. mRMR.** The minimum redundancy maximum relevance (mRMR) [93] algorithm is a filtering approach that attempts to minimize repetition between selected characteristics while also choosing the most linked attributes with class tags. First, the method determines a collection of features from lung images that have the highest correlation with the class (output) and the lowest correlation among themselves [283]. Then, it ranks features based on mutual information using the minimal-redundancy maximal-relevance criterion. Finally, a measure is used to eliminate redundancy between features, which is stated as follows:

$$\text{mRMR}(F_j) = \max_{F_j \in F \setminus S} [I(F_j; C_k) - \frac{1}{m-1} \sum_{F_i \in S} I(F_j; F_i)], \quad (14)$$

where  $I(F_j; C_k)$  represents the mutual correlation between feature  $X_j$  and class  $C_k$ ,  $I(F_j; F_i)$  represents the correlation between features  $F_j$  and  $F_i$ ,  $S$  denotes the selected feature set, and  $m$  means its size (i.e.,  $m = |S|$ ).

**8.3. Least Absolute Shrinkage and Selection Operator (LASSO).** The LASSO [284] is a method for modeling the relationship between one or more explanatory factors and a dependent variable by fitting a regularized least-squares model to the dependent variable. It can efficiently identify significant characteristics related to the dependent variable from a small number of observations with many features when used for compressed sensing. For example, it uses lung data by regularizing and selecting the most significant features simultaneously.

**8.4. Sequential Floating Forward Selection (SFFS).** The SFFS is a bottom-up search procedure that starts with the current feature set and adds new features by applying the basic SFS procedure. Then, if there is still room for improvement in the previous set, the worst feature in the new set is removed. It counts the number of backward steps taken after each forward step [285]. If an intermediate solution at the fundamental level cannot be improved upon, there are no backward steps. The procedure's inverse counterpart, on the other hand, can be described similarly. Because both algorithms provide "self-controlled backtracking," it is possible to find practical solutions by dynamically modifying the trade-off between forwarding and backward steps. They

analyze what they require in a way that does not rely on any parameters [286]. To begin, it starts with an empty set. Then, SFFS takes backward steps on lung images after each step as long as the objective function increases. It reduces the number of unnecessary features from lung images.

**8.5. PCA.** PCA is a dimensionality-reduction approach commonly used to reduce the dimensionality of data by lowering an extensive collection of variables into a smaller set of variables that retains the majority of the learning from the large set of variables [287]. In addition, smaller datasets are easier to analyze and visualize, making them more accessible. For example, it chooses characteristics from lung images based on the magnitude of their coefficients.

**8.6. Weight Optimized Neural Networks with Maximum Likelihood Boosting (WONN-MLB).** Newton and Raphson's MLMR preprocessing model and the boosted weighted optimized neural network ensemble classification algorithms are used to develop the WONN-MLB [288]. The additive combination approach is utilized in the WONN-MLB method to incorporate the highest relevancy with the least amount of redundancy. To achieve the goal of lung cancer detection accuracy with the least amount of time and error, an ensemble of WONN-MLB qualities is used [289]. It only overviewed the extracted features from the lung feature based on the probability.

**8.7. Hybrid Intelligent Spiral Optimization-Based Generalized Rough Set Approach (HSOGR).** The hybrid intelligent spiral optimization-based generalized rough set approach (HSOGR) [90] is used to select the features. The spiral optimization method [290] is based on spiral phenomena and aids in the resolution of the unconstrained optimization problem when picking features. The approach employs adequate settings such as convergence and periodic descent direction in the  $n$ -dimensional spiral model to achieve success. The approach predicts optimization characteristics according to the exploration (global solution) and exploitation (local key) phases with the help of the parameters (good solution). Rather than using a single gradient function when selecting an optimization process, this method employs several spiral points [291], which aid in the establishment of the current optimal fact at any given time. To determine whether the selected characteristics accurately aid in detecting lung cancer, the search space must be investigated using a generalized rough set procedure.

Table 9 shows the pros and cons of feature selection methods.

## 9. Classification and Detection

A classification algorithm is an algorithm that gauges the information included, so the yield isolates one class into positive qualities and the other into negative qualities [297]. The classification methodology is a supervised learning strategy used to recognize classes of novel perceptions based on information preparation [298].

Detection is a computer innovation connected with computer vision and image processing that arranges with recognizing occasions of semantic objects of a specific class in computerized pictures and recordings [299]. It is a computer vision strategy for finding objects in pictures or recordings. When humans look at pictures or videos, objects can be perceived and found in minutes. The objective of object detection is to reproduce this intelligence utilizing a computer [68]. In addition, well-informed areas of article recognition incorporate face location and passerby identification.

**9.1. Machine Learning (ML).** Machine learning is a subordinate part of artificial intelligence, which is comprehensively characterized as the ability of a machine to impersonate shrewd human conduct [300]. This implies machines that can perceive a visual scene, comprehend a text written in ordinary language, or play out an activity in the actual world [301]. In addition, machine learning calculations utilize computational techniques to "learn" data straightforwardly from information without depending on a foreordained condition as a model [302]. Table 10 describes various types of machine learning (ML) algorithms.

**9.2. Deep Learning (DL).** DL is a subfield of ML and AI that copies the path of individual achieving knowledge [313]. Deep learning uses both organized and disorganized information, like text and images, to train the models [314]. Deep learning methods are stored in a sequential pattern for complexity and abstraction, whereas established ML methods are linear [315]. Moreover, deep learning eliminates some data preprocessing techniques and can extract features automatically [316]. Several deep methods have gained tremendous results. They are described in Table 11.

**9.3. Convolutional Neural Network (CNN).** A convolutional neural network (CNN) is a methodology under DL that is capable of taking in input images, emphasizing different objects from the image, and distinguishing continuously [329]. In addition, CNNs are considered a type of neural network that allows for more feature extraction from captured images [330]. CNNs are classified into three categories: convolution, max-pooling, and activation [331]. In comparison to other classifiers, a CNN requires little preprocessing. Although the filters are hand-engineered in a primitive way, CNN can learn these filters/features through adequate training [332]. Table 12 describes the usage of CNN to detect lung nodules and cancer.

**9.4. Hybrid System.** A hybrid structure of CNN with LeNet and AlexNet is developed for analysis by combining the layer settings of LeNet with the parameter settings of AlexNet. It begins with the LeNet architecture, incorporates ReLU, LRN, and dropout layers into the framework, and finally develops the Agile CNN. In addition to two fully connected layers, the proposed CNN, based on LeNet, has two convolutional layers, two pooling layers, and two fully

TABLE 9: Advantages and disadvantages of feature selection methods.

Algorithms	Advantages	Disadvantages
GA [292]	Tries to avoid becoming stuck in a local optimal solution	GA does not guarantee an optimal solution and has high computational cost
mRMR [293]	Effectively reduces the redundant features while keeping the relevant features	Mutual information is incompatible with continuous data
LASSO [294]	Very accurate prediction, reduces overfitting, and improves model interpretability	In terms of independent risk factors, the regression coefficients may not be consistently interpretable
SFFS [295]	Reduces the number of nesting issues and unnecessary features	Difficult to detect all subsets
PCA [296]	Selects a number of important individuals from all the feature components, reduces the dimensionality of the original samples, and improves the classification accuracy	Only considers the linear relationships and interaction between variables at a higher level
WONN-MLB [288]	Integrates the maximum relevancy and minimum redundancy	Has certain amount of irrelevant attributes
HSOGR [90]	Effectively selects optimized features	Its execution is complex

TABLE 10: Most commonly utilized machine learning classifiers for classifying nodules and cancer.

Model name	Purpose	Data type	Result	Strength	Limitation
RF [303]	Using pretrained model to detect lung cancer accurately	CT	Acc 82.5%	Improves the capacity of lung nodule prediction	Limited dataset and result
SVM [300]	Classifying the lung nodules in four lung cancer stages	CT	Acc 84.58%	Predicts small-sized lung nodules, even in low density	The limited dataset affected their results
LDA [301]	Classifying cancer using ODNN and LDA	CT	Acc 94.56%	It is quick, easy to use, non-invasive, and inexpensive	Optimal feature selection with multiclassifier was missing
RF [304]	Automatic classification of pulmonary peri-fissural nodules (PFNs)	CT	Sens 86.8%	Pretrained CNNs are employed, which makes them faster than training CNNs	All kinds of nodules were not classified
SVM [78]	To increase the accurate prediction of lung cancer	CT	Acc 85.7%	Predicts lung cancer from low-resolution data images	The model sometimes fails to predict
RF [299]	To detect malignancy of nodules with self-built model NoduleX	CT	Pres 99%	Solid, part-solid, and non-solid nodule categorization is performed automatically	Big nodules were accurately detected
RF [305]	Classified the measured solidity or nodules	CT	Acc 95%	Avoids potential errors caused by inaccurate image processing	The description of their work is not described clearly
SVM [306]	An improved FP-reduction method is used to detect lung nodules in PET/CT images	CT	Spec 97.2%	Removes around half of the existing FPs	Only small cohort is used
Boosting [307]	Classification of nodules with fusion of texture, shape, and deep model-learned data	CT	F1 96.65%	Generates more accurate outcomes than three existing state-of-the-art techniques	The model only detects big nodules
Multikernel learning [302]	Distinguishing between the nodule and non-nodule classes with classification	CT	Acc 94.17%	Increases the efficacy of false positive reduction	Dataset name is unclear
SVM [308]	Extracting absolute information inherent in raw hand-crafted imaging components	CT	Acc 95.5%	Obtains promising classification outcomes	The reference is limited
Decision tree [22]	Using autoencoder with decision tree to detect nodule	CT	Sens 75.01%	Outperforms the state-of-the-art techniques on the overall accuracy measure, even after experimenting with nearly five times the data amount	The results are low
SVM [309]	Nodule classification with hybrid features	CT	Acc 99.3%	It extracts the representative image of lung nodule malignancy from chest CT images	The model cannot detect type, position, and size
Decision tree [310]	Discovering radiomics to detect lung cancer	CT	Sens 77.52%	Increases the accuracy of lung cancer prediction diagnostics	The reference is limited and results are low

TABLE 10: Continued.

Model name	Purpose	Data type	Result	Strength	Limitation
Boosting [66]	Identifying nodules from CT scan	CT	AUC 86.42%	Quickly finds the exact positions of latent lung nodule	The references of figure and table are accurately done
Multikernel learning [311]	To describe the algorithm for false positive reduction in lung nodule computer-aided detection (CAD)	CT	Jindex 91.39%	Automatically reduces unnecessary feature subsets to get a more discriminative feature set with promising classification performance	All false positive reduction is not done yet
Logistic regression [312]	Prediction of the malignancy of lung nodules in CT scans	CT	Sens 94.5%	Additional information based on nodule size has at best a mixed impact on classifier performance	It only takes large nodules
DBScan [68]	Detecting nodules with 3D DCNN	CT	Spec 79.67%	It can be expanded into other areas of medical image identification	FP reduction and automated classification are missing
Naïve Bayes [243]	A pretrained CNN to extract deep features from lung cancer images and train classifiers to predict all term survivors	CT	Acc 82.5%	The method's performance is such that adding nodule size information has only a mixed effect on classifier performance	The dataset was too small

TABLE 11: Most commonly utilized deep learning classifiers for classifying nodules and cancer.

Model name	Purpose	Data type	Result (%)	Strength	Limitation
DBN with RBM [317]	To detect nodules with deep networks	CT	Acc 92.83	No relative location information is ignored to extract features that express the original image better	The references were very limited with less info of method
DRL [318]	Detecting lung cancer with several potential deep reinforcement learning models	CT	Acc 80	Got promising results in tumor localization	The result of their work is not fully cleared
DRN [319]	Detecting lung cancer in FDG-PET imaging under ultra-low-dose PET scans	PET	Acc 97.1	Lung cancer detection is automated even at low effective radiation doses	The outcome is insufficient
DBN with RBM [320]	Testing the feasibility of using DL algorithms for lung cancer diagnosis	CT	Acc 79.40	It has shown very promising results	Accuracy was slightly less than CNN model
Deep denoising autoencoder [321]	A combination of deep-learned representations was employed to create a lengthy feature vector, which was then used to train the classification of nodules	CT	Acc 95.5	Increased the ability to differentiate between malignant and benign nodules, with a significant improvement in sensitivity	The dataset was not a benchmarked dataset
DRN [322]	Training model first and applying 3D ConvNet to detect lung nodule with hybrid loss learning	CT	Acc 86.7	It detects pulmonary nodules from low-dose CT scans	Detects small nodules and cannot classify malignant or benign nodules
DBN with RBM [23]	Comparing DL and CNN model on lung nodule detection	CT	Sens 73.4	It solves the longstanding challenge of classifying lung nodules as malignant or benign without computing morphological or textural data	The classification was very limited
DRN [323]	Identification of lung nodules from CT scans is efficient for lung cancer diagnosis, and false positive reduction is important, so it was the aim	CT	Acc 98	It is reliable and detects well. It may also be easily extended to detect 3D objects.	Figures and table are not referred clearly
DRL [77]	Developing and validating a reinforcement learning model for early identification of lung nodules in CT images	CT	Acc 99.1	Eliminated the major issue of false positives in CT lung nodule screening, saving unwanted tests and expenditures	Only the big nodules were detected

TABLE 11: Continued.

Model name	Purpose	Data type	Result (%)	Strength	Limitation
Deep denoising autoencoder [324]	A spherical harmonic expansion is used as it has ability to approximate the surfaces of tough shapes of the detected lung nodules	CT	Acc 96	It can show small or big lung nodule spatial inhomogeneities	Classification of nodule as malignant or benign was not done
Multilayer perceptron model [325]	To analyze the performance of several ML methods for detecting lung cancer	CT	Acc 88.55	The presented image preprocessing method detects cancerous bulk	The layers of the model were not discussed briefly
Deep stacked autoencoder [326]	The main purpose is to train a 3D CNN with data and convert it into a 3D fully convolutional network (FCN) that can generate the score map	CT	Sens 80	It can generate the score map for the whole volume in a single pass	The results were not compared with other models
Deep sparse autoencoder [327]	Analyzing the nodules of CT data and helping the experts to be more accurate with proposed analysis tool	CT	Acc 99.57	Improving the display of actual medical CT data may automatically extract pulmonary nodule features	The information of dataset is missing
GAN [328]	Building a 3D U-Net and CNN to segment and identify nodule and assist the radiologists understand CT images	CT	Acc 95.4	Malignant nodule detection is precise and effective	Detects large nodule more accurately than the small nodules
Deep stacked autoencoder [260]	To get an accurate diagnosis of the detected lung nodules	CT	Acc 92.20	It classified nodules using higher-order MGRF and geometric criteria	They did not mention any reshape or resize techniques

TABLE 12: Different types of CNN models.

Model name	Purpose	Data type	Result (%)	Strength	Limitation
MV-CNN [54]	Malignant nodule characterization	CT	Acc 92.31	It is a fast and reliable computer-aided system	A large amount of labeled data is needed for better accuracy
MP-CNN [333]	Automatic detection of lung cancer	CT	Acc 87.80, spec, 89.10, recall 87.40	It uses both local and global contextual variables to detect lung cancer	Different image size affects the accuracy
HSCNN [334]	To predict the malignancy of a pulmonary nodule seen on a computed tomography (CT) scan	CT	Acc 84.40, sens 70.50, spec 88.90, AUC 85.60	Model interpretability improves with prediction accuracy	No domain specialists can fine-tune it by prioritizing more discriminating features under challenging cases
NODULEX (CNN features + QIF features) [335]	Differentiate between malignant and benign nodule patterns with accuracy	CT	Acc 94.60, sens, 94.80, spec 94.30	Excellent accuracy in classifying nodule malignancy	Cross-validated results may be less accurate. Other datasets with significantly differing CT scan picture quality or criteria were not directly fit.
DENSEBTNET (centercrop operation) [336]	Identifying multiscale features in nodule candidates	CT	Acc 88.31, AUC 93.25	It has good parameter efficiency and is parameter light. It enhances DenseNet performance and classification accuracy over other approaches.	Its densely connected mechanism causes feature redundancy
PN-SAMP [337]	Accurately identifying the nodule areas, extracting semantic information from the detected nodules, and predicting the malignancy of the nodules	CT	Acc 97.58	It can predict the malignancy of lung nodules and offer high-level semantic features and nodule location	Only works on CT images

TABLE 12: Continued.

Model name	Purpose	Data type	Result (%)	Strength	Limitation
Dual-pathway CNN [338]	Predicting the nodule's malignancy	CT	Acc 86.84	It performs end-to-end lung nodule diagnostics with high classification accuracy. It can also handle smaller datasets using transfer learning.	A pulmonary nodule cannot be detected automatically
DeepLung (DUAL-path 3D DCNN+) [71]	Developing a fully automated lung CT cancer detection system	CT	Acc 90.44	It is smaller and more efficient than residual networks	Lung nodule annotation is not satisfactory
Ensemble learning of CNNs/multiview knowledge-based collaboration (MV-KBC) [268]	Differentiating between malignant and benign pulmonary nodules	CT	Acc 91.60, AUC 95.70	It uses an adaptive weighting system learned during error backpropagation to categorize lung nodules, allowing the MV-KBC model to be trained end-to-end	During training, there is a relatively high level of computational complexity

connected layers. Layer C1 contains 20 feature maps for each feature map in total. The input data for each unit are linked to a neighborhood. Therefore, a connection from the input cannot extend outside the confines of the feature map boundary. The first feature map in P1 is connected to the second feature map in C1 by 22 neighborhoods. Every unit in P1 is linked to the second feature map in C1. Then, on layer C2, there are 50 feature maps. The other options are the same as they were for the previous layers. F1 and F2 are the final two layers after layer P2. In terms of neuron units, F1 and F2 have 500 and 2 neuron units, respectively. The effect of the parameters of the kernel size, learning rate, and other aspects on the performance of the CNN model is explored by varying these parameters, and an optimized setup of the CNN model is obtained as a result [339]. There are various hybrid methods to detect lung cancer and nodules [340–343]. Figure 8 gives an overview of CNN's hybrid structure. In artificial intelligence, the image is commonly convolved with a particular filter (HOG or LBP) to enhance shapes and edges. Consequently, the first stage of CNN consists primarily of Gabor-like filters. Additionally, the scale-space method was initially designed to enhance the CNN method on which we based. We proposed a novel hybrid CNN model by incorporating standard features into CNN, considering complementary characteristics of the conventional texture method and CNN. This hybrid model's complex distinguishable higher-level features are made up of one-of-a-kind combinations of low-level features. The CNN filters have this hierarchy of simple elements to complex features: the first layer filters mostly have structures that look like Gabor. In contrast, the deep layer filters in the network have features that can be identified as objects. In this study, we combine CNN with texture features like LBP and HOG to improve the first layer filters, which are analogous to the human visual system's ability to decompose images into their oriented spatial frequencies. The data input layer, convolution layer, pooling layer, entire connection layer, and output layer are typically included in the structure of a CNN network. By combining data, our hybrid CNN model aims to make the data input layer easier. In contrast, the

primary objective of the training is to discover the optimal model parameters by minimizing a loss function. It has been found that HOG features and LBP features are fused with CNN in a specific way due to the significant differences in shape and texture between the benign and malignant nodules. However, CNN is believed to be able to extract lung nodules with possible distinguishing features.

*9.5. Transfer Learning.* Transfer learning refers to a methodology wherein the put-away information came about. Learning a model while settling a particular errand can address alternate undertaking of a related issue. Deep convolutional neural networks have accomplished amazing feats in normal picture examination. In any case, such incredible feats are exceptionally subject to the dataset. Transfer learning is a difficult substitute for examining nodules in clinical images using DCNN models, with the only aim of regulating deep CNN terrifying exposure due to the limited amount of clinical images. Some authors used transfer learning in their research [155, 159, 240, 241, 303, 309, 344, 345]. The basic framework of transfer learning is shown in Figure 9.

## 10. Performance Evaluation

It is hard to choose which metrics to use for various issues, and observational studies have shown yet assessed graphic elements to gauge different parts of the calculations [346]. It is often difficult to say which measurements are most appropriate for evaluating the analysis due to the frequent weight gain errors between the expected and actual values [347]. The interpretation of ML estimations is reviewed depending upon critical accuracy, which is routinely improper, assuming that there ought to emerge an occurrence of unequalled information and error cost shift strikingly [175]. ML execution evaluations include a degree of compromise between the true positive and accurate negative rate and between recall and precision. The receiver operating characteristic (ROC) curve depicts the compromise between the false negative and false positive rates for each possible cutoff.

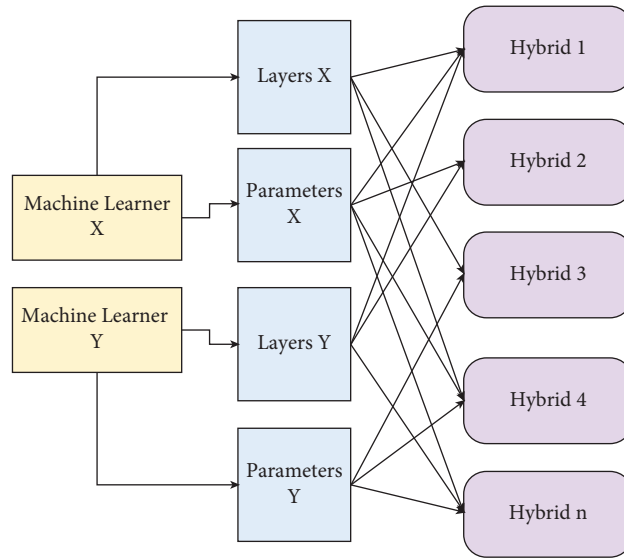


FIGURE 8: Brief overview of hybrid structure of CNN. Each yellow box is to train the machine with data, blue boxes have layers and parameters for individual machines, and purple boxes have layers for one to nth hybrid models.

10.1. *Generally Used Evaluation Metrics.* Evaluation metrics are considered a way of quantifying the effectiveness of a predictive model. Evaluation metrics are used to ensure the quality of a statistical or ML model. When evaluating your model, it is essential to use a variety of different evaluation metrics [348].

- (i) True positive (TP): TP is the correct classification of the positive class. For instance, if an image contains destructive cells and the model fragments the diseased part effectively, the result classifies cancer.
- (ii) True negative (TN): TN is the correct classification of the negative class, for example, when there is no malignant growth in the image. The model after classification declares that the cancer is absent.
- (iii) False positive (FP): FP is the erroneous prediction of the positives; for instance, the picture has carcinogenic cells, but the model classifies that the image does not contain cancer.
- (iv) False negative (FN): FN is the false expectation of the negatives. For instance, there is no malignancy in the picture except for the model that says an image is a carcinogenic one [349].

The effectiveness of any ML model is still up in the air utilizing measures like TP rate, FP rate, TN rate, and FN rate [350]. The sensitivity and specificity measures are commonly used to clarify demonstrative clinical tests as well as to assess how excellent and predictable the diagnostic test is [37]. The TP rate or positive class accuracy is the sensitivity measurement, while the TN rate or negative class accuracy refers to the specificity measurement [351]. There is frequently a compromise between the four measurements in “real-world” applications.

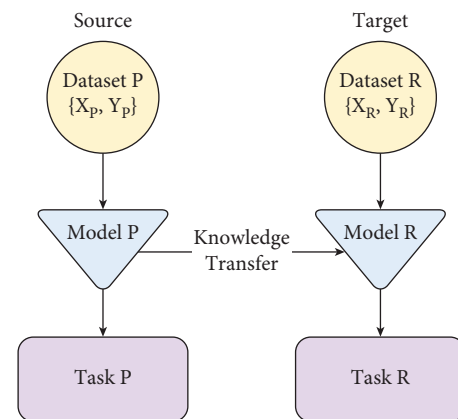


FIGURE 9: Basic structure of transfer learning.

10.2. *Classification Measurements.* There are a lot of methods used for the classification of lung nodule and lung cancer. The widely used metrics for classification problems are as follows.

10.2.1. *Precision.* Precision is the number of relevant reports recovered by a search isolated by the number of pieces retrieved. In short, precision is the number of pieces recovered that are important. It checks how exactly the model functions by actually taking a look at the correct, true positives from the anticipated ones [249].

$$\text{Prec.} = \frac{TP}{TP + FN}. \tag{15}$$

**10.2.2. Recall/Sensitivity.** Recall/sensitivity is the number of pertinent records recovered by a search isolated by existing significant archives. Sensitivity is another name for recall. The test's sensitivity reflects the likelihood that the screening test will be positive among unhealthy people. The number of applicable archives recovered is referred to as recall. It computes the number of true positives detected by the model and marks them as positives [352]. Finally, it estimates the capacity of a test to be positive when the condition is present. It is otherwise called false negative rate, review, Type II error,  $\beta$  error, error or oversight, or elective theory [69].

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (16)$$

**10.2.3. Accuracy.** Accuracy is the level of closeness to ground truth. For example, the accuracy of an estimation is a proportion of how close the deliberate worth is to the actual value of the amount. The estimation accuracy might rely upon a few factors, including the breaking point or the goal of the estimating instrument [353].

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \quad (17)$$

where TP, TN, FP, and FN mean true positive, true negative, false positive, and false positive, respectively.

Aside from this, there are other types of accuracy, such as predictive accuracy and average accuracy. Predictive accuracy should be estimated based on the difference between observed and predicted values [354, 355]. Average accuracy is the average of every accuracy per class (amount of accuracy for each class anticipated/number of classes) [356].

**10.2.4. F1-Score.** F-Measure or F1-score combines both precision and recall into a binary measure that catches the two properties, giving each similar weighting. The arithmetic mean of the two proportions is precision and recall [357]. The F-measure is used to fine-tune precision and recall. It is frequently used for evaluating data recovery frameworks, such as search engines, as well as some types of ML models, particularly in natural language processing [358]. F1-score is the function of precision and recall. It is evaluated when a balance between precision and recall is needed [359].

$$F1 = \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}. \quad (18)$$

**10.2.5. Specificity.** Specificity is the capacity of a test to distinguish individuals without illness effectively. The test's specificity reflects the likelihood that the screening test will be negative among people who do not have the illness. It estimates a test's ability to be harmful when the condition is not present. It is otherwise called FP rate, precision, Type I error,  $\alpha$  error, error of commission, or null hypothesis [278].

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \quad (19)$$

**10.2.6. Receiver Operating Characteristic Curve (ROC Curve) and Area under the ROC Curve (AUC).** A ROC curve is a graphical plot that outlines the symptomatic capacity of a twofold classifier framework as its separation edge is fluctuated [360]. ROC analysis provides methods for selecting ideal models and automatically removing imperfect ones from the expense setting or class conveyance. ROC analysis is directly and naturally linked to cost/benefit analysis of demonstrative dynamics [361]. ROC curves are considered a fantastic asset as an accurate display measure in location/characterization hypothesis and speculation testing. For a variety of reasons, AUC is often preferred over accuracy [362]. Indeed, since it is probably the most widely used performance metric, it is very uncomfortable to adjust how AUC works [363] properly.

**10.2.7. ROC Curve.** The ROC curve addresses the performance of the proposed model at all characterization limits [364]. The ROC curve summarizes classifier execution over a range of TP and FP error rates. It is a graph of the true positive rate versus the false positive rate (TPR versus FPR). A point on the ROC curve between (0, 100) would be ideal [365]. ROC helps investigate the compromises among various classifiers over a scope of situations, which is not great for circumstances with realized error costs [366–369].

$$\begin{aligned} \text{TPR} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{FPR} &= \frac{\text{FP}}{\text{FP} + \text{TN}}. \end{aligned} \quad (20)$$

**10.2.8. AUC.** AUC coordinates the region under the ROC curve from (0, 0) to (1, 1). It gives the total proportion of all conceivable characterization edges [370]. AUC has a range of 0 to 1. The AUC esteem for a correctly classified version will be 1.0, while it will be 0.0 in the case of a completely incorrect classification [371]. It is amazing for two reasons: first, it is scale-invariant, which means it examines how well the model is anticipated rather than the overall qualities; and second, it is grouping limit invariant, which means it examines the model's exhibition regardless of the chosen edge [372]. The region under the curve (AUC) is most favored because the bigger the region, the better the model. The AUC additionally has a decent translation as the likelihood that the classifier positions an arbitrarily picked positive occasion over a haphazardly picked negative one [373]. The AUC is a useful measurement for classifier execution because it is independent of the chosen standard and earlier probabilities [374]. AUC can be used to establish a predominance connection between classifiers. If the ROC curves cross, the absolute AUC is a normal comparison between models [375–380].

**10.3. Segmentation Measurements.** There are a lot of methods used for the segmentation of lung nodule and lung cancer. The widely used metrics for segmentation problems are as follows.



**10.4. Jaccard Index.** The Jaccard index, otherwise called the Jaccard similarity coefficient, is a measurement that checks the closeness and variety of test sets. It is defined as the width of the crossing point divided by the width of the association of two name sets. It is a proportion of comparability for the two information arrangements, ranging from 0% to 100% [382]. The higher the rate, the more comparable the two populaces.

$$\text{Jac}_{idx} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}}. \quad (21)$$

**10.5. Dice Coefficient.** The Dice similarity coefficient, otherwise called the Sorensen–Dice list or Dice coefficient, is a factual instrument that estimates the comparability between two arrangements of information [383]. The Dice coefficient should not be more noteworthy than 1. A Dice coefficient, for the most part, goes from 0 to 1 [384]. If the coefficient result is greater than 1, the execution may need to be rechecked [385]. It was used as a measurable approval metric to evaluate the reproducibility of manual divisions as well as the spatial crossover precision of robotized probabilistic partial division of MR images, as represented on two clinical models [386–388]. It is a substantial proportion of the comparability rate between two example sets:

$$\text{Dice}^{\text{cof}} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}}. \quad (22)$$

**10.6. Error Calculation.** The term “error” refers to a deviation from accuracy or correctness. Errors are considered a significant issue when anyone wants to evaluate the system’s performance. When the performance is evaluated, only the system’s efficiency is calculated. But the errors must be measured while calculating the performance. Many techniques are available to calculate the errors in lung cancer detection.

**10.6.1. Mean Absolute Error (MAE).** MAE is a model assessment metric utilized with relapse models. The mean outright error regarding a test set is the mean of the absolute values of the individual prediction errors on all examples in the test set. In insights, MAE is a proportion of errors between combined perceptions communicating a similar wonder [389, 390].

$$\text{MAE} = \frac{|p_1 - a_1| + \dots + |p_n - a_n|}{n}, \quad (23)$$

where  $p$  represents predicted target values ( $p_1, p_2, \dots, p_n$ ) while  $a$  represents actual value:  $a_1, a_2, \dots, a_n$ , in which  $n$  represents total number of data points.

**10.6.2. Root Mean Square Error (RMSE).** RMSE is the square root of the mean of the square of the entirety of the error. RMSE is a good proportion of accuracy, but it should only be used to analyze and compare prediction errors of different

models or model setups for a single variable, not between factors because it is scale-dependent [249].

$$\text{RMSE} = \sqrt{\frac{(p_1 - a_1)^2 + \dots + (p_n - a_n)^2}{n}}, \quad (24)$$

where  $p$  represents predicted target values ( $p_1, p_2, \dots, p_n$ ) while  $a$  represents actual value:  $a_1, a_2, \dots, a_n$ , in which  $n$  represents total number of data points.

**10.6.3. Relative Absolute Error.** RAE is a way of estimating the performance of a proactive model. RAE is a metric contrasting genuine figure error with the estimated error of a shortsighted (naive) model. A sensible model (which produces results that are superior to a trivial model) will bring about a proportion short of one [391, 392].

**10.6.4. Root Relative Squared Error (RRSE).** The RRSE is comparable with what it would have been if a straightforward indicator had been utilized. To put it bluntly, this specific indicator is only the average of the actual values. In this way, the relative squared error standardizes the total squared error by partitioning it by the absolute squared error of the forward indicators. By taking the square root of the relative squared error, one decreases the error to similar measurements as the amount being anticipated [393, 394].

$$\sqrt{\frac{(p_1 - \bar{a})^2 + \dots + (p_n - \bar{a})^2}{|a_1 - \bar{a}|^2 + \dots + |a_n - \bar{a}|^2}}, \quad (25)$$

where  $p$  represents predicted target values ( $p_1, p_2, \dots, p_n$ ) while  $a$  represents actual value:  $a_1, a_2, \dots, a_n$ .

## 11. Challenges and Research Direction

Lung cancer detection techniques are improving day by day. Currently, available lung cancer detection techniques are quite good in terms of performance, but there are many more limitations that researchers have encountered. Many issues have been resolved, but some remain.

Some of them are mentioned below.

**11.1. Insufficient Number of Annotated Medical Datasets with Cases.** Most of the significant successes of deep learning techniques in general, and convolutional neural networks in particular, have been achieved using large amounts of data. Large annotated datasets of lung CT images are in high demand, but obtaining such datasets in medical imaging remains challenging due to various factors, such as the time-consuming nature of clinician annotation tasks, the need for privacy, and ethical considerations, among others. Expert radiologists must construct and annotate large datasets, which is costly and time consuming. As a result, the insufficiency of datasets with a large number of samples is a significant barrier to the application of deep learning to the study of medical data [17].

*11.2. Accurate Segmentation.* Accurate segmentation of the lung fields is necessary to efficiently reduce the search space for lung nodules. Due to inhomogeneities within the lung region and similar density pulmonary components such as arteries, veins, bronchi, and bronchioles, technical issues concerning lung segmentation techniques should be researched further. These technical difficulties include the technique's automation level, sensitivity to scanning parameters, an algorithm's ability to work with multiple image modalities (e.g., CT, LDCT, or CE-CT), and the algorithm's ability to provide proper lung segmentation.

*11.3. Nodule Types.* Most nodules are harmless, indicating a more severe health issue. Among other tissues, parenchymal tissues are distinct and difficult to segment. On the other hand, solitary and large solid nodules are easy for segmentation. But the problem occurs when these types of nodules are targeted.

*11.3.1. Small Nodules.* Small-nodule segmentation is critical for the early identification of lung cancer [395]. Thin-slice high-resolution computed tomography (HRCT) has enabled the visibility of tiny nodules less than 5 mm in diameter, which was previously invisible using previous-generation CT technology. Accurate segmentation of such small nodules is required to assess the malignancy of the lesions. A partial-volume effect is the primary technical concern when dealing with tiny nodules (PVE). The spatial discrimination used in CT imaging allows a single voxel to represent multiple tissue types by averaging their intensity values. This induces PVE and picture blur, particularly near lesion margins, challenging segmentation. When dealing with smaller lesions, PVE becomes more pronounced since the fraction of mistakes over the lesion volume increases. This makes measuring the area/volume of tiny nodules more difficult. The partial-volume approach (PVM) [396] is presented for calculating nodule volume based on the consistency of the average attenuation quantities. PVM outperforms other thresholding algorithms in volumetric accuracy, according to their phantom study. SPVA (segmentation-based partial-volume analysis) [397] is proposed to extend the PVM approach to include VOI segmentation into the nodule core, parenchyma area, and partial-volume region. A histogram from the partial volume region was used to estimate the volume of the nodule near its boundary. Finally, the proposed RAGF [398] yields an elliptical approximation of the lesion boundary.

*11.3.2. Nodules Attached to Vessels.* Lung nodules are frequently connected to other pulmonary structures such as the airways, blood vessels, parenchymal walls, and diaphragm. Because the CT values of nodules and these non-target objects are frequently extremely similar, determining the extent of the nodule from these structures becomes a difficult technical issue. Juxta-vascular nodules are nodules that connect to blood vessels. Morphological filtering is a systematic strategy for this purpose

[397, 399–403]. Because the proportion of nodules that attach to vessels/airways is often minimal compared to the entire extent of the 3D nodule surface, basic MOs such as erosion, dilatation, and opening are frequently effective in most juxta-vascular situations [400, 402]. These fundamental operators were combined with convex-hull operations [397, 404] and 3D moment analysis [405] to refine the segmentation process after it was completed. Geometric/shape constrained segmentation is another prominent strategy in this context [398, 403, 406–408]. This method incorporates shape-based prior information into the segmentation process to bias the results toward a spherical/nodular shape. It suppresses elongated non-target components linked to the target.

*11.3.3. Nodules Attached to Parenchymal Wall and Diaphragm.* Juxta-pleural nodules are cases that are attached to the parenchymal wall or the diaphragm. These nodules are connected to the chest wall and pleural surface. Many automated measurement algorithms struggle with these nodules because they need to determine where the nodule ends and the chest wall begins. Solitary nodules, on the other hand, that do not border any other structures, such as airways or blood arteries, are much easier to segment [409].

*11.3.4. Ground-Glass Opacity Nodules.* The ground-glass opacity (GGO) nodule is a nodule with subsolid CT values that are much lower than usual solid nodules. They are classified into two types based on whether or not solid components are present: non-solid/pure and partially solid/mixed. GGO nodule segmentation is a technological issue because it is difficult to distinguish their tiny boundaries and model their uneven appearances. In clinical practice, modern CT technology's more excellent picture resolution has enabled the investigation of small GGO nodules. Although their growth is frequently slow [410], such GGO nodules, particularly mixed ones, have been linked to a high risk of malignancy [411]. Recent clinical studies are part of the histological spectrum of peripheral adenocarcinomas, which encompass premalignant atypical adenomatous hyperplasia (AAH) and malignant bronchioloalveolar carcinoma (BAC) [412]. Over ten years, a tiny non-solid GGO representing AAH or BAC can gradually grow into an invasive lung adenocarcinoma [410]. In this method, segmentation is accomplished by labeling each voxel with a nodule/background label based on a probabilistic decision rule established from training data.

*11.4. Article Selection Bias.* A measurement of association, such as a risk ratio, that is distorted as a result of sample selection that does not accurately reflect the target population is known as selection bias. The selection of individuals, groups, or data for analysis in such a way that proper randomization is not achieved, failing to ensure that the obtained sample is representative of the intended population, is known as selection bias. On the other hand,

TABLE 13: Best articles and their details.

Author info	Patient group	Outcomes	Key results	Comments
Raz et al. [417](USA) (retrospective cohort study (level 4, good))	37 patients identified with isolated adrenal metastases from NSCLC 20 underwent surgical resection	5-year survival	34% in the adrenalectomy group versus 0% in the non-operative group ( $P = 0.002$ ) 83% for ipsilateral tumors versus 0% for contralateral tumors ( $P = 0.003$ )	The selection process for operative and non-operative management was inconsistent Adrenalectomy patients were on average 10 years younger 50% of patients in the adrenalectomy group (and 70% in the non-operated group) had N2 or T4 diseases; therefore, the adrenal metastasis was not truly isolated
	17 underwent non-operative management  Maximum follow-up period of 16 years		67% in case of lower lobe NSCLC versus 27% in cases of upper lobe tumors ( $P = 0.29$ )  27% synchronous metastasis versus 41% metachronous metastases ( $P = 0.81$ ) 52% with N0 or N1 disease versus 0% with N2 diseases ( $P = 0.008$ )	
Luketich and Burt [418] (USA) (retrospective cohort study (level 4, good))	14 patients with isolated synchronous adrenal metastasis from NSCLC	Medium survival	Medium survival of 8.5 months in the chemotherapy alone group versus 31 months in the chemotherapy + surgery group	Small study, but no significant differences were seen in preoperative characteristics, tumor size, or cell type to otherwise explain the improved survival  The authors recommend that surgery should be advocated after ensuring that curative resection of the lung primary can be achieved
	8 patients had neoadjuvant chemotherapy followed by concomitant lung resection and adrenalectomy 6 patients had only 3 cycles of chemotherapy (mitomycin, cisplatin, and vinblastine) 5-year follow-up		In the surgically resected group, the 3-year actuarial survival was 38%  Longest survivor at end of follow-up was 61 months	
Higashiyama et al. [416] (retrospective cohort study (level 4, good))	9 patients with isolated adrenal metastases from surgically resected lung cancer (4 non-curative and 5 curative)	Survival	Adrenalectomy group: 2/5 alive at 24 and 40 months, respectively, and 3/5 died at 9, 17, and 20 months, respectively	All patients in the palliative group had a disease-free interval of 7 months. This selection bias may explain some of the observed difference in survival in addition to the influence of treatment strategy.  The authors concluded that short FDIs are probably due to lymphatic spread and probably signify a more aggressive tumor
	5 treated with adrenalectomy followed by adjuvant chemo or radiotherapy 4 treated with palliative chemo or radiotherapy Maximum follow-up of 40 months		Palliative group: all died within 6 months	

selection bias might be an issue: the sociodemographic profile of DLCST participants was better. They had greater psychological fortitude than the general population of people who smoked a lot [413]. As a result, selection bias could lead to underestimating the actual psychosocial effects [413]. According to a psychometric analysis of survey data and qualitative focus group interviews, abnormal and false positive LCS results can have a wide range of psychosocial effects that can be adequately quantified with PROMs [414, 415]. The finest articles and specifics of each are described in Table 13.

**11.5. Efficient CADe System.** Developing an efficient computer-aided detection (CADe) system for detecting lung

nodules is a difficult task. The level of automation, speed, and ability to recognize nodules of varying shapes, such as irregularly shaped nodules rather than only spherical ones, as well as the CADe system's ability to detect cavity nodules, nodules attached to the lung borders, and small nodules, are all critical considerations to consider (e.g., less than 3 mm).

**11.6. Volumetric Measurements.** Volumetric measurements are essential because various sizes in different situations make the system more accurate. When calculating the growth rate in the volumetric unit, the global movement of patients caused by their actions and the local activity of the entire lung tissue caused by respiration and heartbeat should be considered. It is impossible to distinguish between

TABLE 14: Challenges and research directions for lung nodule and cancer diagnosis.

Name	Challenges	Research direction
Insufficient number of annotated medical datasets with cases	All datasets are not publicly available	All datasets need to be available openly. Additionally, research should be conducted utilizing such imaging modalities using unpublished datasets. All datasets should be disclosed for future research works and implementations.
Accurate segmentation	Segmentation models are not properly executed	All segmentation models need to be implementing in various modalities which may uplift the lung nodule and cancer detection results
Nodule size and types	Small nodules are needed to be detected more efficiently	All kinds of nodules need to be investigated. Implementing feature extraction and selection can detect most of the nodules. Nodules can be identified by feature and classifier selection.
Efficient CAde system	Nodules and cancer detection need to be more accurate using all architectures	Random forest, SVM, DBN with RM, and CNNs are mostly used for lung cancer diagnosis. ML and DL networks of other kinds should be analyzed in this field.
Volumetric measurements	All lung image shapes are not the same. So, all datasets need to be extracted.	When patients are breathing, their lung shape changes and it varies from patient to patient. We recommend investigating all datasets and measuring different shapes of lungs.

changes caused by the direct application of global and local registration to the segmented nodule and changes in the shape of the nodule caused by breathing and heartbeat.

The research directions that should be inspected to uplift the lung nodule and cancer detection outcomes are described here. Through profound investigation on this topic, the recommendations for the study are described below.

Table 14 represents challenges and limitations in lung nodule and cancer diagnosis, as well as research directions in terms of the dataset, architectures, and so on.

- (i) Datasets focused on CT scans are available openly. Ultrasound, PET scans, and SPECT datasets, on the other hand, are not publicly available. Furthermore, studies utilizing such imaging modalities use unpublished datasets. These datasets should be made public for future research and implementations.
- (ii) Like U-Net and SegNet, segmentation models have provided sophisticated segmentation results across various image datasets. Furthermore, implementing these techniques involving different modalities may improve lung nodule and cancer detection results.
- (iii) All kinds of nodules need to be investigated. Implementing feature extraction and selection can detect any nodule. The selection of features and classifiers can be used to identify nodules. The most common methods for selecting features are genetic algorithms, Wonn-MLB, and HSOGR. Feature extraction, on the other hand, is critical for detecting nodules. Most of the time, radiomic methods extract features from lung images. HOG, autoencoders, and wavelets should also be investigated to be more accurate.
- (iv) Random forest, SVM, DBN with RM, and CNNs are primarily used for lung cancer diagnosis. ML techniques such as boosting, decision trees, and DL networks of various types such as GANs and clustering should be analyzed. CNN is widely used to detect lung nodules and cancer because it can extract essential features from images. CNN can

identify and classify lung cancer types with greater accuracy in a shorter period. But as CNN is a DL model, it needs a massive amount of data, so if the dataset is insufficient, it will not give benchmark accuracy. We recommend that strategies based on different CNN architectures and CNN+ and other dimensional CNN must be inquired.

- (v) When patients are breathing, their lung shape changes, and it varies from patient to patient. The patient's lung cancer cells appeared in large numbers, and there were more irregular shapes than in healthy lungs. Availability of all datasets is needed to measure all kinds of lungs. We recommend investigating all datasets and measuring different shapes of lungs. The authors in [419, 420] have already started working on this idea.

## 12. Conclusions

Lung cancer is the most widely recognized disease-related reason for death among people. Early detection of pulmonary nodules and lung cancer saves lives because it is known that the chances of surviving cancer are higher if it is found, diagnosed, and treated quickly. Several methods and systems have been proposed for analyzing pulmonary nodules in medical images. Additionally, the domain covers biological, engineering, computer science, and histological research. However, this article provides a comprehensive overview of the lung cancer detection interface. It is intended for novices interested in learning about the present state of lung cancer detection methods and technologies. The essential concepts of lung cancer detection methods are fully explored. The article focuses on many aspects of the research domain, including image preprocessing, feature extraction, segmentation, feature selection methodologies, performance measurements, and challenges and limitations along with the possible solutions. The article endorses a summary of current methods to help new researchers quickly understand the research domain conceptions. The study also looks into the various types of datasets available to lung cancer

detection systems. The fundamental principles of lung cancer detection and nodule classification procedures are thoroughly explored using CT scan, MRI, or X-ray imaging. Furthermore, the article combines current cancer-detecting systems, describing a preliminary review based on previous works. The article also describes the challenges and limitations that will help explore the inconvenience of lung cancer detection technologies. The majority of lung cancer detection methods are now in the primary stages of development. Still, there are many things that could be changed to make the system work better. The combined efforts of scientific researchers and the tech sectors are required to commercialize this vast area for the benefit of ordinary people.

## Data Availability

No data were used to support this study.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] M. M. Wahidi, J. A. Govert, R. K. Goudar, and D. Mc. Crory, "Evidence for the treatment of patients with pulmonary nodules: when is it lung cancer?: accp evidence-based clinical practice guidelines," *Chest*, vol. 132, no. 3, pp. 94S–107S, 2007.
- [2] P. J. Mazzone and L. Lam, "Evaluating the patient with a pulmonary nodule: a review," *JAMA*, vol. 327, no. 3, pp. 264–273, 2022.
- [3] Mayo Clinic, "What to know if you have lung nodules," <https://www.mayoclinic.org/diseases-conditions/lung-cancer/%20expert-answers/lung-nodules/faq-20058445>.
- [4] D. M. Abdullah N. S. Ahmed et al., "A review of most recent lung cancer detection techniques using machine learning," *International Journal of Science and Business*, vol. 5, no. 3, pp. 159–173, 2021.
- [5] Cancer, "What is lung cancer?: types of lung cancer," <https://www.cancer.org/cancer/lung-cancer/about/what-is.html>.
- [6] D R Baldwin, "Prediction of risk of lung cancer in populations and in pulmonary nodules: significant progress to drive changes in paradigms," *Lung Cancer*, vol. 89, no. 1, pp. 1–3, 2015.
- [7] Q. Z Song, L. Zhao, X. K. Luo, and X. C. Dou, "Using deep learning for classification of lung nodules on computed tomography images," *Journal of healthcare engineering*, vol. 2017, pp. 1–7, 2017.
- [8] Z. Ma, R. S. T. João Manuel, and R. M. Natal Jorge, "A review on the current segmentation algorithms for medical images," in *Proceedings of the 1st International Conference on Imaging Theory and Applications*, IMAGAPP, Lisboa, Portugal, February 2009.
- [9] Z. Ma, J. M. R. Tavares, R. N. Jorge, and T. Mascarenhas, "A review of algorithms for medical image segmentation and their applications to the female pelvic cavity," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 13, no. 2, pp. 235–246, 2010.
- [10] K. Senthil Kumar, K. Venkatalakshmi, and K. Karthikeyan, "Lung cancer detection using image segmentation by means of various evolutionary algorithms," *Computational and Mathematical Methods in Medicine*, vol. 2019, Article ID 4909846, 16 pages, 2019.
- [11] P. Badura and E. Pietka, "Soft computing approach to 3d lung nodule segmentation in ct," *Computers in Biology and Medicine*, vol. 53, pp. 230–243, 2014.
- [12] G. S. Lodwick, "Computer-aided diagnosis in radiology: a research plan," *Investigative Radiology*, vol. 1, no. 1, pp. 72–80, 1966.
- [13] K. Doi, "Computer-aided diagnosis in medical imaging: historical review, current status and future potential," *Computerized Medical Imaging and Graphics*, vol. 31, no. 4–5, pp. 198–211, 2007.
- [14] K. T. Bae, M. L. Giger, H. MacMahon, and K. Doi, "Computer-aided detection of pulmonary nodules in ct images," *Radiology*, vol. 181, no. 1, p. 144, 1991.
- [15] Paperswithcode, "Papers with code - luna16 dataset," <https://paperswithcode.com/dataset/luna16>.
- [16] Kaggle, "Data-science-bowl," 2017, <https://www.kaggle.com/c/data-science-bowl-2017>.
- [17] Y. Yang, X. Feng, W. Chi et al., "Deep learning aided decision support for pulmonary nodules diagnosing: a review," *Journal of Thoracic Disease*, vol. 10, no. S7, pp. S867–S875, 2018.
- [18] F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, "Evaluate the malignancy of pulmonary nodules using the 3-d deep leaky noisy-or network," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3484–3495, 2019.
- [19] D. Ardila, A. P. Kiraly, S. Bharadwaj et al., "End-to-end lung cancer screening with three-dimensional deep learning on lowdose chest computed tomography," *Nature Medicine*, vol. 25, no. 6, pp. 954–961, 2019.
- [20] D. Poap, M. Wozniak, R. Damaševicius, and W. Wei, "Chest radiographs segmentation by the use of nature-inspired algorithm for lung disease detection," in *Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 2298–2303, IEEE, Bangalore, India, November 2018.
- [21] A. Jaszcz, D. Połap, and R. Damaševicius, "Lung x-ray image segmentation using heuristic red fox optimization algorithm," *Scientific Programming*, vol. 2022, Article ID 4494139, 8 pages, 2022.
- [22] D. Kumar, A. Wong, and D. A. Clausi, "Lung nodule classification using deep features in ct images," in *Proceedings of the 2015 12th Conference on Computer and Robot Vision*, pp. 133–138, IEEE, Halifax, Canada, June 2015.
- [23] Y. J. Chen, K. L. Hua, C. H. Hsu, W.-H. Cheng, and S. C. Hidayati, "Computer-aided classification of lung nodules on computed tomography images via deep learning technique," *OncoTargets and Therapy*, vol. 8, pp. 2015–2022, 2015.
- [24] W. Wang and G. Chakraborty, "Evaluation of malignancy of lung nodules from ct image using recurrent neural network," in *Proceedings of the 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pp. 2992–2997, IEEE, Bari, Italy, October 2019.
- [25] X. Yi, E. Walia, and P. Babyn, "Generative adversarial network in medical imaging: a review," *Medical Image Analysis*, vol. 58, Article ID 101552, 2019.
- [26] Y. Onishi, A. Teramoto, M. Tsujimoto et al., "Multiplanar analysis for pulmonary nodule classification in ct images using deep convolutional neural network and generative adversarial networks," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 1, pp. 173–178, 2020.

- [27] S. A. El-Regaily, M. A. Salem, M. H. Abdel Aziz, and M. I. Roushdy, "Survey of computer aided detection systems for lung cancer in computed tomography," *Current Medical Imaging*, vol. 14, no. 1, pp. 3–18, 2018.
- [28] J. Ma, Y. Song, Xi Tian, Y. Hua, R. Zhang, and J. Wu, "Survey on deep learning for pulmonary medical imaging," *Frontiers of Medicine*, vol. 14, no. 4, pp. 450–469, 2019.
- [29] P. Monkam, S. Qi, H. Ma, W. Gao, Y. Yao, and W. Qian, "Detection and classification of pulmonary nodules using convolutional neural networks: a survey," *IEEE Access*, vol. 7, Article ID 78075, 2019.
- [30] R. Mastouri, N. Khlifa, H. Neji, and S. Hantous-Zannad, "Deep learning-based cad schemes for the detection and classification of lung nodules from ct images: a survey," *Journal of X-Ray Science and Technology*, vol. 28, no. 4, pp. 591–617, 2020.
- [31] D. Riquelme and M. A. Akhloufi, "Deep learning for lung cancer nodules detection and classification in ct scans," *A&I*, vol. 1, no. 1, pp. 28–67, 2020.
- [32] T. G. Debelee, S. R. Kebede, F. Schwenker, and Z. M. Shewarega, "Deep learning in selected cancers' image analysis—a survey," *Journal of Imaging*, vol. 6, no. 11, p. 121, 2020.
- [33] S. T. H. Kieu, A. Bade, M. H. A. Hijazi, and H. Kolivand, "A survey of deep learning for lung disease detection on medical images: state-of-the-art, taxonomy, issues and future directions," *Journal of Imaging*, vol. 6, no. 12, p. 131, 2020.
- [34] A. Naik and D. R. Edla, "Lung nodule classification on computed tomography images using deep learning," *Wireless Personal Communications Communications*, vol. 116, no. 1, pp. 655–690, 2021.
- [35] Yu Gu, J. Chi, J. Liu et al., "A survey of computer-aided diagnosis of lung nodules from ct scans using deep learning," *Computers in Biology and Medicine*, vol. 137, Article ID 104806, 2021.
- [36] J. Wu and T. Qian, "A survey of pulmonary nodule detection, segmentation and classification in computed tomography with deep learning techniques," *J. Med. Artif. Intell.*, vol. 2, no. 8, pp. 8–12, 2019.
- [37] K. Munir, H. Elahi, A. Ayub, F. Frezza, and A. Rizzi, "Cancer diagnosis using deep learning: a bibliographic review," *Cancers*, vol. 11, no. 9, p. 1235, 2019.
- [38] S. Wang, D. M. Yang, R. Rong et al., "Artificial intelligence in lung cancer pathology image analysis," *Cancers*, vol. 11, no. 11, p. 1673, 2019.
- [39] J. Zhang, Y. Xia, H. Cui, and Y. Zhang, "Pulmonary nodule detection in medical images: a survey," *Biomedical Signal Processing and Control*, vol. 43, pp. 138–147, 2018.
- [40] S. Keele, "Guidelines for performing systematic literature reviews in software engineering," ebsc, Technical report, Technical report, ver. 2.3 ebsc technical report, 2007.
- [41] B. Kitchenham, *Procedures for Performing Systematic Reviews*, vol. 33, pp. 1–26, Keele University, Keele, UK, 2004.
- [42] O. Ozdemir, R. L. Russell, and A. A. Berlin, "A 3d probabilistic deep learning system for detection and diagnosis of lung cancer using low-dose ct scans," *IEEE Transactions on Medical Imaging*, vol. 39, no. 5, pp. 1419–1429, 2020.
- [43] J. Tan, Y. Huo, Z. Liang, and L. Li, "A comparison study on the effect of false positive reduction in deep learning based detection for juxtapleural lung nodules: cnn vs dnn," in *Proceedings of the Symposium on Modeling and Simulation in Medicine*, pp. 1–8, Virginia Beach, VA, USA, April 2017.
- [44] N. Nasrullah, J. Sang, M. S. Alam, M. Mateen, B. Cai, and H. Hu, "Automated lung nodule detection and classification using deep learning combined with multiple strategies," *Sensors*, vol. 19, no. 17, p. 3722, 2019.
- [45] W. Sun, B. Zheng, and W. Qian, "Automatic feature learning using multichannel roi based on deep structured algorithms for computerized lung cancer diagnosis," *Computers in Biology and Medicine*, vol. 89, pp. 530–539, 2017.
- [46] P. Monkam, S. Qi, M. Xu, F. Han, X. Zhao, and W. Qian, "Cnn models discriminating between pulmonary micronodules and non-nodules from ct images," *BioMedical Engineering Online*, vol. 17, no. 1, pp. 96–16, 2018.
- [47] G. L. F. da Silva, T L A Valente, A. C. Silva et al., "Convolutional neural network-based pso for lung nodule false positive reduction on ct images," *Computer Methods and Programs in Biomedicine*, vol. 162, pp. 109–118, 2018.
- [48] P. Rao, N. A. Pereira, and R. Srinivasan, "Convolutional neural networks for lung cancer screening in computed tomography (ct) scans," in *Proceedings of the 2016 2nd International Conference on Contemporary Computing and Informatics*, pp. 489–493, IEEE, Greater Noida, India, December 2016.
- [49] G. S. Tran, T. P. Nghiem, V T Nguyen, C M. Luong, J. C. Burie, and J.-C. Burie, "Improving accuracy of lung nodule classification using deep learning with focal loss," *Journal of healthcare engineering*, vol. 2019, Article ID 5156416, 9 pages, 2019.
- [50] P. M. Shakeel, M. A. Burhanuddin, and M. I. Desa, "Lung cancer detection from ct image using improved profuse clustering and deep learning instantaneously trained neural networks," *Measurement*, vol. 145, pp. 702–712, 2019.
- [51] S. Makaju, P. W. C. Prasad, A. Alsadoon, A. K. Singh, and A. Elchouemi, "Lung cancer detection using ct scan images," *Procedia Computer Science*, vol. 125, pp. 107–114, 2018.
- [52] S. Bhatia, Y. Sinha, and L. Goel, "Lung cancer detection: a deep learning approach," in *Soft Computing for Problem Solving*, pp. 699–705, Springer, Singapore, 2019.
- [53] B. Ait Skourt, A. El Hassani, and A. Majda, "Lung ct image segmentation using deep neural networks," *Procedia Computer Science*, vol. 127, pp. 109–113, 2018.
- [54] S. Hussein, R. Gillies, K. Cao, S. Qi, and U. Bagci, "Tumornet: lung nodule characterization using multi-view convolutional neural network with Gaussian process," in *Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pp. 1007–1010, IEEE, Melbourne, Australia, April 2017.
- [55] A. O. de Carvalho Filho, A. C. Silva, A. C. de Paiva, R. A. Nunes, and M. Gattass, "Lung-nodule classification based on computed tomography using taxonomic diversity indexes and an svm," *Journal of Signal Processing Systems*, vol. 87, no. 2, pp. 179–196, 2017.
- [56] W. Shen, Mu Zhou, F. Yang et al., "Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification," *Pattern Recognition*, vol. 61, pp. 663–673, 2017.
- [57] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, "Multilevel contextual 3-d cnns for false positive reduction in pulmonary nodule detection," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1558–1567, 2017.
- [58] X. Liu, F. Hou, H. Qin, and A. Hao, "Multi-view multi-scale cnns for lung nodule type classification from ct images," *Pattern Recognition*, vol. 77, pp. 262–275, 2018.
- [59] M. J. Cha, M. J. Chung, J. H. Lee, and K. S. Lee, "Performance of deep learning model in detecting operable lung cancer with chest radiographs radiographs," *Journal of Thoracic Imaging*, vol. 34, no. 2, pp. 86–91, 2019.

- [60] A. A. A. Setio, F. Ciompi, G. Litjens et al., "Pulmonary nodule detection in ct images: false positive reduction using multi-view convolutional networks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1160–1169, 2016.
- [61] S. G. Armato III, G. McLennan, L. Bidaut et al., "The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans," *Medical Physics (Woodbury)*, vol. 38, no. 2, pp. 915–931, 2011.
- [62] P. Sahu, D. Yu, M. Dasari, F. Hou, and H. Qin, "A light-weight multi-section cnn for lung nodule classification and malignancy estimation," *IEEE journal of biomedical and health informatics*, vol. 23, no. 3, pp. 960–968, 2019.
- [63] D. Jia, A. Li, Z. Hu, and L. Wang, "Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 559–567, Springer, Quebec, Canada, September 2017.
- [64] S. Krishnamurthy, G. Narasimhan, and U. Rengasamy, "An automatic computerized model for cancerous lung nodule detection from computed tomography images with reduced false positives," in *Proceedings of the International conference on recent trends in image processing and pattern recognition*, pp. 343–355, Springer, Bidar, India, December 2016.
- [65] J. Sang, M. S. Alam, H. Xiang et al., "Automated detection and classification for early stage lung cancer on ct images using deep learning," in *Pattern Recognition and Tracking XXX*, vol. 10995, Campus Saint-Christophe – Europa, SPIE, 2019 International Society for Optics and Photonics, Article ID 109950S.
- [66] H. Xie, D. Yang, N. Sun, Z. Chen, and Y. Zhang, "Automated pulmonary nodule detection in ct images using deep convolutional neural networks," *Pattern Recognition*, vol. 85, pp. 109–119, 2019.
- [67] J. Gong, Ji-yu Liu, Li-jia Wang, Xw Sun, B. Zheng, and Sd Nie, "Automatic detection of pulmonary nodules in ct images by incorporating 3d tensor filtering with local image feature analysis," *Physica Medica*, vol. 46, pp. 124–133, 2018.
- [68] Yu Gu, X. Lu, L. Yang et al., "Automatic lung nodule detection using a 3d deep convolutional neural network combined with a multiscale prediction strategy in chest cts," *Computers in Biology and Medicine*, vol. 103, pp. 220–231, 2018.
- [69] C. Albert, N. Balachandar, and P. Lu, *Deep Convolutional Neural Networks for Lung Cancer Detection*, Stanford University, Stanford, CA, USA, 2017.
- [70] K. Kuan, M. Ravaut, G. Manek et al., "Deep learning for lung cancer detection: tackling the kaggle data science bowl 2017 challenge," 2017, <https://arxiv.org/abs/1705.09435>.
- [71] W. Zhu, C. Liu, W. Fan, and X. Xie, "Deeplung: deep 3d dual path nets for automated pulmonary nodule detection and classification," in *Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 673–681, IEEE, Lake Tahoe, NV, USA, March 2018.
- [72] G. Bansal, V. Chamola, P. Narang, S. Kumar, and S. Raman, "Deep3dscan: deep residual network and morphological descriptor based framework for lung cancer classification and 3d segmentation," *IET Image Processing*, vol. 14, no. 7, pp. 1240–1247, 2020.
- [73] W. J. Sori, J. Feng, A. W. Godana, S. Liu, and D. J. Gelmecha, "Dfd-net: lung cancer detection from denoised ct scan image using deep learning," *Frontiers of Computer Science*, vol. 15, no. 2, Article ID 152701, 2021.
- [74] X. Huang, W. Sun, T. L. B. Tseng, C. Li, and W. Qian, "Fast and fully-automated detection and segmentation of pulmonary nodules in thoracic ct scans using deep convolutional neural networks," *Computerized Medical Imaging and Graphics*, vol. 74, pp. 25–36, 2019.
- [75] E. Lopez Torres, E. Fiorina, F. Pennazio et al., "Large scale validation of the m5l lung cad on heterogeneous ct datasets," *Medical Physics (Woodbury)*, vol. 42, no. 4, pp. 1477–1489, 2015.
- [76] W. Alakwaa, M. Nassef, Amr, and Badr, "Lung cancer detection and classification with 3d convolutional neural network (3d-cnn)," *Lung Cancer*, vol. 8, no. 8, p. 409, 2017.
- [77] I. Ali, G. R. Hart, G. Gunabushanam et al., "Lung nodule detection via deep reinforcement learning," *Frontiers in Oncology*, vol. 8, no. 108, p. 108, 2018.
- [78] A. Gupta, S. Das, T. Khurana, and K. Suri, "Prediction of lung cancer from low-resolution nodules in ct-scan images by using deep features," in *Proceedings of the 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 531–537, IEEE, Bangalore, India, September 2018.
- [79] O. Ozdemir, B. Woodward, and A. A. Berlin, "Propagating uncertainty in multi-stage bayesian convolutional neural networks with application to pulmonary nodule detection," 2017, <https://arxiv.org/abs/1712.00497>.
- [80] M. Liu, J. Dong, X. Dong, Yu Hui, and L. Qi, "Segmentation of lung nodule in ct images based on mask r-cnn," in *Proceedings of the 2018 9th International Conference on Awareness Science and Technology (iCAST)*, pp. 1–6, IEEE, Fukuoka, Japan, September 2018.
- [81] N. Khosravan and U. Bagci, "Semisupervised multi-task learning for lung cancer diagnosis," in *Proceedings of the 2018 40th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pp. 710–713, IEEE, Honolulu, HI, USA, July 2018.
- [82] Y. Qin, H. Zheng, Y.-M. Zhu, and J. Yang, "Simultaneous accurate detection of pulmonary nodules and false positive reduction using 3d cnns," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1005–1009, IEEE, Calgary, Canada, April 2018.
- [83] C. Zhang, X. Sun, K. Dang et al., "Toward an expert level of lung cancer detection and classification using a deep convolutional neural network," *The Oncologist*, vol. 24, no. 9, pp. 1159–1165, 2019.
- [84] A. A. A. Setio, A. Traverso, T. De Bel et al., "Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge," *Medical Image Analysis*, vol. 42, pp. 1–13, 2017.
- [85] The National Lung Screening Trial Research Team, "Reduced lung-cancer mortality with low-dose computed tomographic screening," *New England Journal of Medicine*, vol. 365, no. 5, pp. 395–409, 2011.
- [86] S. Trajanovski, D. Mavroeidis, C. L. Swisher et al., "Towards radiologist-level cancer risk assessment in ct lung screening using deep learning," *Computerized Medical Imaging and Graphics*, vol. 90, Article ID 101883, 2021.
- [87] Cdas, "Datasets - nlst - the cancer data access system," <https://cdas.cancer.gov/datasets/nlst/>.
- [88] Md Z. Islam, Md M. Islam, and A.nullah Asraf, "A combined deep cnn-lstm network for the detection of novel coronavirus (covid-19) using x-ray images," *Informatics in Medicine Unlocked*, vol. 20, Article ID 100412, 2020.

- [89] D. Moitra and R. Kr Mandal, "Automated AJCC (7th edition) staging of non-small cell lung cancer (NSCLC) using deep convolutional neural network (CNN) and recurrent neural network (RNN)," *Health Information Science and Systems*, vol. 7, no. 1, pp. 14–12, 2019.
- [90] P. M. Shakeel, M. A. Burhanuddin, and M. I. Desa, "Automatic lung cancer detection from ct image using improved deep neural network and ensemble classifier," *Neural Computing & Applications*, vol. 34, no. 12, pp. 9579–9592, 2020.
- [91] E. Kurniawan, P. Prajitno, and D. S. Soejoko, "Computer-aided detection of mediastinal lymph nodes using simple architectural convolutional neural network," *Journal of Physics: Conference Series*, vol. 1505, Article ID 12018, 2020.
- [92] T. L. Chaunzwa, A. Hosny, Y. Xu et al., "Deep learning classification of lung cancer histology using ct images," *Scientific Reports*, vol. 11, no. 1, pp. 5471–12, 2021.
- [93] M. Togaçar, B. Ergen, and Z. Cömert, "Detection of lung cancer on chest ct images using minimum redundancy maximum relevance feature selection method with convolutional neural networks," *Biocybernetics and Biomedical Engineering*, vol. 40, no. 1, pp. 23–39, 2020.
- [94] P. Marentakis, P. Karaiskos, V. Kouloulis et al., "Lung cancer histology classification from ct images based on radiomics and deep learning models," *Medical, & Biological Engineering & Computing*, vol. 59, no. 1, pp. 215–226, 2021.
- [95] J. Tan, L. Jing, Y. Huo, L. Li, O. Akin, and Y. Tian, "Lgan: lung segmentation in ct scans using generative adversarial network," *Computerized Medical Imaging and Graphics*, vol. 87, Article ID 101817, 2021.
- [96] F. Kanavati, G. Toyokawa, S. Momosaki et al., "Weakly-supervised learning for lung carcinoma classification using deep learning," *Scientific Reports*, vol. 10, no. 1, pp. 9297–11, 2020.
- [97] Cancerimagingarchive, "Welcome to the cancer imaging archive," <https://www.cancerimagingarchive.net/>.
- [98] W. Ausawalaitong, A. Thirach, S. Marukatat, and T. Wilaiprasitporn, "Automatic lung cancer prediction from chest xray images using the deep learning approach," in *Proceedings of the 2018 11th Biomedical Engineering International Conference (BMEICON)*, pp. 1–5, IEEE, Chiang Mai, Thailand, November 2018.
- [99] G. Yu, G. Peng, H. Jiang et al., "Deep learning with lung segmentation and bone shadow exclusion techniques for chest x-ray analysis of lung cancer," in *Proceedings of the International Conference on Computer Science, Engineering and Education Applications*, pp. 638–647, Springer, Kiev, Ukraine, January 2018.
- [100] J. Shiraishi, S. Katsuragawa, J. Ikezoe et al., "Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules," *American Journal of Roentgenology*, vol. 174, no. 1, pp. 71–74, 2000.
- [101] B. B. Ogul, P. Kos, ucu, A. Ízšam, and S. R. D. Kanik, "Lung nodule detection in x-ray images: a new feature set," in *Proceedings of the 6th European Conference of the International Federation for Medical and Biological Engineering*, pp. 150–155, Springer, Dubrovnik, Croatia, September 2015.
- [102] Qi Gong, Q. Li, M. A. Gavrielides, and N. Petrick, "Data transformations for statistical assessment of quantitative imaging biomarkers: application to lung nodule volumetry," *Statistical Methods in Medical Research*, vol. 29, no. 9, pp. 2749–2763, 2020.
- [103] A. Chaudhary and S. S. Singh, "Lung cancer detection on ct images by using image processing," in *Proceedings of the 2012 International Conference on Computing Sciences*, pp. 142–146, IEEE, Phagwara, India, September 2012.
- [104] J. K. Gohagan, P. M. Marcus, R. M. Fagerstrom et al., "Final results of the lung screening study, a randomized feasibility study of spiral ct versus chest x-ray screening for lung cancer," *Lung Cancer*, vol. 47, no. 1, pp. 9–15, 2005.
- [105] K. Yokoi, N. Kamiya, H. Matsuguma et al., "Detection of brain metastasis in potentially operable non-small cell lung cancer: a comparison of ct and mri," *Chest*, vol. 115, no. 3, pp. 714–719, 1999.
- [106] B. Hochegger, E. Marchiori, O. Sedlaczek et al., "Mri in lung cancer: a pictorial essay," *The British Journal of Radiology British journal of radiology*, vol. 84, no. 1003, pp. 661–668, 2011.
- [107] M. M. Oken, P. M. Marcus, P. Hu et al., "PLCO Project Team Baseline chest radiograph for lung cancer detection in the randomized Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial. and ovarian cancer screening trial," *Journal of the National Cancer Institute (1988)*, vol. 97, no. 24, pp. 1832–1839, 2005.
- [108] J. Gohagan, P. Marcus, R. Fagerstrom, P. Pinsky, B. Kramer, and P. Prorok, "Baseline findings of a randomized feasibility trial of lung cancer screening with spiral ct scan vs chest radiograph: the lung screening study of the national cancer institute," *Chest*, vol. 126, no. 1, pp. 114–121, 2004.
- [109] D. P. Steinfors, Y. H. Khor, R. L. Manser, and L. B. Irving, "Radial probe endobronchial ultrasound for the diagnosis of peripheral lung cancer: systematic review and meta-analysis," *European Respiratory Journal*, vol. 37, no. 4, pp. 902–910, 2011.
- [110] K. Yasufuku, T. Nakajima, K. Motoori et al., "Comparison of endobronchial ultrasound, positron emission tomography, and ct for lymph node staging of lung cancer," *Chest*, vol. 130, no. 3, pp. 710–718, 2006.
- [111] W. Weder, R. A. Schmid, H. Bruchhaus, S. Hillinger, G. K. von Schulthess, and H. C. Steinert, "Detection of extrathoracic metastases by positron emission tomography in lung cancer," *The Annals of Thoracic Surgery*, vol. 66, no. 3, pp. 886–892, 1998.
- [112] C.-H. Kao, J. F. Hsieh, S. C. Tsai, Y. J. Ho, and J. K. Lee, "Quickly predicting chemotherapy response to paclitaxel-based therapy in non-small cell lung cancer by early technetium-99m methoxyisobutylisonitrile chest single-photon-emission computed tomography," *Clinical Cancer Research: An Official Journal of the American Association for Cancer Research*, vol. 6, no. 3, pp. 820–824, 2000.
- [113] Medlineplus, "X-rays," 2021, <https://medlineplus.gov/xrays.html>.
- [114] G. Gavelli and E. Giampalma, "Sensitivity and specificity of chest x-ray screening for lung cancer," *Cancer*, vol. 89, no. S11, pp. 2453–2456, 2000.
- [115] S. J. Swensen, J. Jett, J. A. Sloan et al., "Screening for lung cancer with low-dose spiral computed tomography," *American Journal of Respiratory and Critical Care Medicine*, vol. 165, no. 4, pp. 508–513, 2002.
- [116] Z. Xie and H. Zhang, "Analysis of the diagnosis model of peripheral non-smallcell lung cancer under computed tomography images," *Journal of Healthcare Engineering*, vol. 2022, Article ID 3107965, 13 pages, 2022.
- [117] P. B. Bach, R. J. James, P. Ugo, M. S. Tockman, S. J. Swensen, and C. B. Begg, "Computed tomography screening and lung cancer outcomes," *JAMA*, vol. 297, no. 9, pp. 953–961, 2007.



- [118] S. Zheng, J. Shu, J. Xue, and C. Ying, "Ct signs and differential diagnosis of peripheral lung cancer and inflammatory pseudotumor: a meta-analysis," *Journal of Healthcare Engineering*, vol. 2022, Article ID 3547070, 11 pages, 2022.
- [119] J. Biederer, Y. Ohno, H. Hatabu et al., "Screening for lung cancer: does mri have a role?" *European Journal of Radiology*, vol. 86, pp. 353–360, 2017.
- [120] L. I. Cervino, J. Du, and S. B. Jiang, "Mri-guided tumor tracking in lung cancer radiotherapy," *Physics in Medicine and Biology*, vol. 56, no. 13, pp. 3773–3785, 2011.
- [121] D. Lardinois, W. Weder, T. F. Hany et al., "Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography," *New England Journal of Medicine*, vol. 348, no. 25, pp. 2500–2507, 2003.
- [122] R. M. Pieterman, J. W. G. van Putten, J. J. Meuzelaar et al., "Preoperative staging of non-small-cell lung cancer with positron emission tomography," *New England Journal of Medicine*, vol. 343, no. 4, pp. 254–261, 2000.
- [123] Y. E. Erdi, K. Rosenzweig, A. K. Erdi et al., "Radiotherapy treatment planning for patients with non-small cell lung cancer using positron emission tomography (pet)," *Radiotherapy & Oncology*, vol. 62, no. 1, pp. 51–60, 2002.
- [124] R. C. Viney, M. J. Boyer, M. T. King et al., "Randomized controlled trial of the role of positron emission tomography in the management of stage i and ii non-small-cell lung cancer," *Journal of Clinical Oncology*, vol. 22, no. 12, pp. 2357–2362, 2004.
- [125] S. K. Das, M. M. Miften, S. Zhou et al., "Feasibility of optimizing the dose distribution in lung tumors using fluorine-18-fluorodeoxyglucose positron emission tomography and single photon emission computed tomography guided dose prescriptions," *Medical Physics (Woodbury)*, vol. 31, no. 6, pp. 1452–1461, 2004.
- [126] S. Katyal, E. L. Kramer, M. E. Noz, D. McCauley, A. Chachoua, and A. Steinfeld, "Fusion of immunoscintigraphy single photon emission computed tomography (spect) with ct of the chest in patients with non-small cell lung cancer," *Cancer Research*, vol. 55, no. 23, Article ID 5759s, 1995.
- [127] O. Schillaci, "Single-photon emission computed tomography/computed tomography in lung cancer and malignant lymphoma," *Seminars in Nuclear Medicine*, vol. 36, pp. 275–285, 2006.
- [128] S. Sone and S. Yano, "Molecular pathogenesis and its therapeutic modalities of lung cancer metastasis to bone," *Cancer and Metastasis Reviews*, vol. 26, no. 3-4, pp. 685–689, 2007.
- [129] C. P. Belani, H. Choy, P. Bonomi et al., "Combined chemoradiotherapy regimens of paclitaxel and carboplatin for locally advanced non-small-cell lung cancer: a randomized phase ii locally advanced multi-modality protocol," *Journal of Clinical Oncology*, vol. 23, no. 25, pp. 5883–5891, 2005.
- [130] S. M. Yoon, T. Shaikh, and M. Hallman, "Therapeutic management options for stage iii non-small cell lung cancer," *World Journal of Clinical Oncology*, vol. 8, no. 1, p. 1, 2017.
- [131] F. Farjah, D. R. Flum, S. D. Ramsey, P. J. Heagerty, R. G. Symons, and D. E. Wood, "Multi-modality mediastinal staging for lung cancer among medicare beneficiaries," *Journal of Thoracic Oncology*, vol. 4, no. 3, pp. 355–363, 2009.
- [132] S.-T. Bow, *Pattern Recognition and Image Preprocessing*, CRC Press, Boca Raton, FL, USA, 2002.
- [133] S. Bhattacharyya, "A brief survey of color image pre-processing and segmentation techniques," *Journal of Pattern Recognition Research*, vol. 6, no. 1, pp. 120–129, 2011.
- [134] S. M. Pizer, E. P. Amburn, J. D. Austin et al., "Adaptive histogram equalization and its variations," *Computer Vision, Graphics, and Image Processing*, vol. 39, no. 3, pp. 355–368, 1987.
- [135] S. S. Bagade and K. S. Vijaya, "Use of histogram equalization in image processing for image enhancement," *International Journal of Software Engineering Research and Practices*, vol. 1, no. 2, pp. 6–10, 2011.
- [136] R. P. Singh and M. Dixit, "Histogram equalization: a strong technique for image enhancement," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 8, no. 8, pp. 345–352, 2015.
- [137] A. Asuntha and A. Srinivasan, "Deep learning for lung cancer detection and classification," *Multimedia Tools and Applications*, vol. 79, no. 11-12, pp. 7731–7762, 2020.
- [138] J. C. Church, Y. Chen, and S. V. Rice, "A spatial median filter for noise removal in digital images," in *Proceedings of the IEEE SoutheastCon 2008*, pp. 618–623, IEEE, Huntsville, AL, USA, April 2008.
- [139] S.-W. Hong and N.-Ho Kim, "A study on median filter using directional mask in salt & pepper noise environments," *Journal of the Korea Institute of Information and Communication Engineering*, vol. 19, no. 1, pp. 230–236, 2015.
- [140] Ji-Y. Kim and Y. Lee, "Preliminary study of improved median filter using adaptively mask size in light microscopic image," *Microscopy*, vol. 69, no. 1, pp. 31–36, 2020.
- [141] K. M. Tun and K. A. Soe, "Feature extraction and classification of lung cancer nodule using image processing techniques," *International Journal of Engineering Research and Technology*, vol. 3, no. 3, 2014.
- [142] P. M. Shakeel, M. I. Desa, and M. A. Burhanuddin, "Improved watershed histogram thresholding with probabilistic neural networks for lung cancer diagnosis for cbmir systems," *Multimedia Tools and Applications*, vol. 79, no. 23-24, Article ID 17115, 2020.
- [143] P. B. Sangamithraa and S. Govindaraju, "Lung tumour detection and classification using ek-mean clustering," in *Proceedings of the 2016 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, pp. 2201–2206, IEEE, Chennai, India, March 2016.
- [144] I. T. Young and L. J. van Vliet, "Recursive implementation of the Gaussian filter," *Signal Processing*, vol. 44, no. 2, pp. 139–151, 1995.
- [145] G. Deng and L. W. Cahill, "An adaptive Gaussian filter for noise reduction and edge detection," in *Proceedings of the 1993 IEEE conference record nuclear science symposium and medical imaging conference*, pp. 1615–1619, IEEE, San Francisco, CA, USA, November 1993.
- [146] F. Neyenssac, "Contrast enhancement using the laplacian-of-a-Gaussian filter," *CVGIP: Graphical Models and Image Processing*, vol. 55, no. 6, pp. 447–463, 1993.
- [147] A. Teramoto, T. Tsukamoto, Y. Kiriya, and H. Fujita, "Automated classification of lung cancer types from cytological images using deep convolutional neural networks," *BioMed Research International*, vol. 2017, Article ID 4067832, 6 pages, 2017.
- [148] A. M. Rossetto and W. Zhou, "Deep learning for categorization of lung cancer ct images," in *Proceedings of the 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies*

- (CHASE), pp. 272-273, IEEE, Philadelphia, PA, USA, July 2017.
- [149] A. Hosny, C. Parmar, T. P. Coroller et al., "Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study," *PLoS Medicine*, vol. 15, no. 11, Article ID e1002711, 2018.
- [150] P. M. Shakeel, M. Burhanuddin, and M. I. Desa, "Lung cancer detection from ct image using improved profuse clustering and deep learning instantaneously trained neural networks," *Measurement*, vol. 145, pp. 702-712, 2019.
- [151] M. S. Al-Tarawneh, "Lung cancer detection using image processing techniques," *Leonardo Electronic Journal of Practices and Technologies*, vol. 11, no. 21, pp. 147-158, 2012.
- [152] M. Avanzo, J. Stancanello, G. Pirrone, and G. Sartor, "Radiomics and deep learning in lung cancer," *Strahlentherapie und Onkologie*, vol. 196, no. 10, pp. 879-887, 2020.
- [153] A. Asuntha, A. Brindha, S. Indirani, and A. Srinivasan, "Lung cancer detection using svm algorithm and optimization techniques," *J. Chem.Pharm. Sci*, vol. 9, no. 4, pp. 3198-3203, 2016.
- [154] Xi Wang, H. Chen, C. Gan et al., "Weakly supervised deep learning for whole slide lung cancer image analysis," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 3950-3962, 2020.
- [155] T. Fang, "A novel computer-aided lung cancer detection method based on transfer learning from googlenet and median intensity projections," in *Proceedings of the 2018 IEEE international conference on computer and communication engineering technology (CCET)*, pp. 286-290, IEEE, Beijing, China, August 2018.
- [156] Z. Song, S. Zou, W. Zhou et al., "Clinically applicable histopathological diagnosis system for gastric cancer detection using deep learning," *Nature Communications*, vol. 11, no. 1, pp. 4294-4299, 2020.
- [157] A. Khireddine, K. Benmahammed, and W. Puech, "Digital image restoration by wiener filter in 2d case," *Advances in Engineering Software*, vol. 38, no. 7, pp. 513-516, 2007.
- [158] R. Hardie, "A fast image super-resolution algorithm using an adaptive wiener filter," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2953-2964, 2007.
- [159] T. Sajja, R. Devarapalli, and H. Kalluri, "Lung cancer detection based on ct scan images by using deep transfer learning," *Traitement du Signal*, vol. 36, no. 4, pp. 339-344, 2019.
- [160] R. Mehrotra, K. Namuduri, and N. Ranganathan, "Gabor filter-based edge detection," *Pattern Recognition*, vol. 25, no. 12, pp. 1479-1494, 1992.
- [161] N. A. B. Mary and D. Dharma, "A novel framework for real-time diseased coral reef image classification," *Multimedia Tools and Applications*, vol. 78, no. 9, Article ID 11387, 2019.
- [162] Ch Brücker, D. Hess, and J. Kitzhofer, "Singleview volumetric piv via high-resolution scanning, isotropic voxel restructuring and 3d least-squares matching (3d-lsm)," *Measurement Science and Technology*, vol. 24, no. 2, Article ID 24001, 2012.
- [163] M. Nagao, N. Miyake, and Y. Yoshino, "Detection of abnormal candidate regions on temporal subtraction images based on dcnn," in *Proceedings of the 2017 17th International Conference on Control, Automation and Systems (ICCAS)*, pp. 1444-1448, IEEE, Jeju, Republic of Korea, October 2017.
- [164] H. Wang, Z. Zhou, Y. Li et al., "Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18 f-fdg pet/ct images," *EJNMMI Research*, vol. 7, no. 1, p. 11, 2017.
- [165] C. C. Quattrocchi, Y. Errante, C. Gaudino et al., "Spatial brain distribution of intraaxial metastatic lesions in breast and lung cancer patients," *Journal of neuro-oncology*, vol. 110, no. 1, pp. 79-87, 2012.
- [166] P. K. Sahoo, S. A. K. C. Soltani, and A. K. C. Wong, "A survey of thresholding techniques," *Computer Vision, Graphics, and Image Processing*, vol. 41, no. 2, pp. 233-260, 1988.
- [167] M. Sharma, S B Jignesh, and M. V. Joshi, "Early detection of lung cancer from ct images: nodule segmentation and classification using deep learning," in *Proceedings of the Tenth International Conference on Machine Vision (ICMV 2017)*, vol. 10696, SPIE, Campus Saint-Christophe - Europa, April 2018, International Society for Optics and Photonics, Article ID 106960W.
- [168] N. Plaziac, "Image interpolation using neural networks," *IEEE Transactions on Image Processing*, vol. 8, no. 11, pp. 1647-1651, 1999.
- [169] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transactions on Acoustics, Speech, & Signal Processing*, vol. 29, no. 6, pp. 1153-1160, 1981.
- [170] E. Meijering, "A chronology of interpolation: from ancient astronomy to modern signal and image processing," *Proceedings of the IEEE*, vol. 90, no. 3, pp. 319-342, 2002.
- [171] S. Fadnavis, "Image interpolation techniques in digital image processing: an overview," *International Journal of Engineering Research in Africa*, vol. 4, no. 10, pp. 70-73, 2014.
- [172] T. M. Lehmann, C. Gonner, and K. Spitzer, "Survey: interpolation methods in medical image processing," *IEEE Transactions on Medical Imaging*, vol. 18, no. 11, pp. 1049-1075, 1999.
- [173] J. Zhao, Z. Zhang, and J. Yang, "An automatic detection model of pulmonary nodules based on deep belief network," *International Journal of Wireless and Mobile Computing*, vol. 16, no. 1, pp. 7-13, 2019.
- [174] D. Cascio, R. Magro, F. Fauci, M. Iacomi, and G. Raso, "Automatic detection of lung nodules in ct datasets based on stable 3d mass-spring models," *Computers in Biology and Medicine*, vol. 42, no. 11, pp. 1098-1109, 2012.
- [175] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321-357, 2002.
- [176] T. Zhu, Y. Lin, and Y. Liu, "Synthetic minority oversampling technique for multiclass imbalance problems," *Pattern Recognition*, vol. 72, pp. 327-340, 2017.
- [177] T. Deepa and M. Punithavalli, "An e-smote technique for feature selection in high-dimensional imbalanced dataset," vol. 2, pp. 322-324, in *Proceedings of the 2011 3rd International Conference on Electronics Computer Technology*, vol. 2, pp. 322-324, IEEE, Kanyakumari, India, April 2011.
- [178] X. Wang, B. Yu, A. Ma, C. Chen, B. Liu, and Q. Ma, "Protein-protein interaction sites prediction by ensemble random forests with synthetic minority oversampling technique," *Bioinformatics*, vol. 35, no. 14, pp. 2395-2402, 2019.
- [179] M. Naseriparsa and M. M. Riahi Kashani, "Combination of pca with smote resampling to boost the prediction rate in lung cancer dataset," 2014, <https://arxiv.org/abs/1403.1949>.
- [180] S. Chen and S. Wu, "Identifying lung cancer risk factors in the elderly using deep neural networks: quantitative analysis of web-based survey data," *Journal of Medical Internet Research*, vol. 22, no. 3, Article ID e17695, 2020.

- [181] R. Patil, G. Mahadevaiah, and A. Dekker, "An approach toward automatic classification of tumor histopathology of non-small cell lung cancer based on radiomic features," *Tomography*, vol. 2, no. 4, pp. 374–377, 2016.
- [182] K-J Wang, B Makond, and K.M Wang, "Modeling and predicting the occurrence of brain metastasis from lung cancer by bayesian network: a case study of taiwan," *Computers in Biology and Medicine*, vol. 47, pp. 147–160, 2014.
- [183] G. Yadav, S. Maheshwari, and A. Agarwal, "Contrast limited adaptive histogram equalization based enhancement for real time video system," in *Proceedings of the 2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pp. 2392–2397, IEEE, Delhi, India, September 2014.
- [184] E. D. Pisano, S. Zong, B. M. Hemminger et al., "Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms," *Journal of Digital Imaging*, vol. 11, no. 11, pp. 193–200, 1998.
- [185] A. M Reza, "Realization of the contrast limited adaptive histogram equalization (clahe) for realtime image enhancement," *The Journal of VLSI Signal Processing-Systems for Signal, Image, and Video Technology*, vol. 38, no. 1, pp. 35–44, 2004.
- [186] S. Sahu, S. Sahu, A. K. Singh, S. Ghrera, and M. Elhoseny, "An approach for denoising and contrast enhancement of retinal fundus image using clahe," *Optics & Laser Technology*, vol. 110, pp. 87–98, 2019.
- [187] K. Punithavathy, M. M. Ramya, and S. Poobal, "Analysis of statistical texture features for automatic lung cancer detection in pet/ct images," in *Proceedings of the 2015 International Conference on Robotics, Automation, Control and Embedded Systems (RACE)*, pp. 1–5, IEEE, Chennai, India, February 2015.
- [188] U D Bhagyarekha and A. C. Pise, "Lung cancer detection using bayasein classifier and fcm segmentation," in *Proceedings of the 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICADOT)*, pp. 170–174, IEEE, Pune, India, September 2016.
- [189] S. K. Wajid, A. Hussain, K. Huang, and W. Boulila, "Lung cancer detection using local energy-based shape histogram (lesh) feature extraction and cognitive machine learning techniques," in *Proceedings of the 2016 IEEE 15th International Conference on Cognitive Informatics & Cognitive Computing (ICCI\* CC)*, pp. 359–366, IEEE, Palo Alto, CA, USA, August 2016.
- [190] P. Kaur and R Bhatia, "A review on lung cancer detection using PET/CT scan," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 7, no. 5, pp. 977–981, 2017.
- [191] S. Satish Kanitkar, N. D. Thombare, and S. S. Lokhande, "Detection of lung cancer using marker-controlled watershed transform," in *Proceedings of the 2015 international conference on pervasive computing (ICPC)*, pp. 1–6, IEEE, Pune, India, January 2015.
- [192] D. Sharma and Gagandeep Jindal, "Identifying lung cancer using image processing techniques," vol. 17, pp. 872–880, in *Proceedings of the International Conference on Computational Techniques and Artificial Intelligence (ICCTAI)*, vol. 17, pp. 872–880, Citeseer, Barcelona, Spain, August 2011.
- [193] M. Shafiq-ul Hassan, K. Latifi, G. Zhang, G. Ullah, R. Gillies, and E. Moros, "Voxel size and gray level normalization of ct radiomic features in lung cancer," *Scientific Reports*, vol. 8, no. 1, Article ID 10545, 2018.
- [194] R. Wulandari, R. Sigit, and S. Wardhana, "Automatic lung cancer detection using color histogram calculation," in *Proceedings of the 2017 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC)*, pp. 120–126, IEEE, East Java, Indonesia, September 2017.
- [195] L. A. Hunter, S. Krafft, F. Stingo et al., "High quality machine-robust image features: identification in nonsmall cell lung cancer computed tomography images," *Medical Physics (Woodbury)*, vol. 40, no. 12, Article ID 121916, 2013.
- [196] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into imaging*, vol. 9, no. 4, pp. 611–629, 2018.
- [197] A. Halder, D. Dey, and A. K. Sadhu, "Lung nodule detection from feature engineering to deep learning in thoracic ct images: a comprehensive review," *Journal of Digital Imaging*, vol. 33, no. 3, pp. 655–677, 2020.
- [198] A. El-Baz, G. M. Beache, G. Gimel'farb et al., "Computer-aided diagnosis systems for lung cancer: challenges and methodologies," *International Journal of Biomedical Imaging*, vol. 2013, Article ID 942353, 46 pages, 2013.
- [199] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, Cengage Learning, Boston, MA, USA, 2014.
- [200] S. Beucher, "Segmentation tools in mathematical morphology," in *Image algebra and morphological image processing*, vol. 1350, pp. 70–84, SPIE, International Society for Optics and Photonics, 1990.
- [201] S. Kaur and G. Jindal, "Watershed segmentation of lung ct scanimages for early diagnosis of cancer," *International journal of computer and electrical engineering*, vol. 3, no. 6, pp. 850–852, 2011.
- [202] O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, Toronto Canada, September 2015.
- [203] S. Wang, Mu Zhou, O. Gevaert et al., "A multi-view deep convolutional neural networks for lung nodule segmentation," in *Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 1752–1755, IEEE, Jeju Island, Korea, July 2017.
- [204] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034, NW Washington, DC, December 2015.
- [205] Sergey Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the International conference on machine learning*, pp. 448–456, PMLR, Lille, France, July 2015.
- [206] S. Wang, Mu Zhou, Z. Liu et al., "Central focused convolutional neural networks: developing a data-driven model for lung nodule segmentation," *Medical Image Analysis*, vol. 40, pp. 172–183, 2017.
- [207] H.-C. Shin, H. R. Roth, M. Gao et al., "Deep convolutional neural networks for computer-aided detection: cnn architectures, dataset characteristics and transfer learning," *IEEE*

- Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [208] H. Min, W. Jia, X.-F. Wang et al., “An intensity-texture model based level set method for image segmentation,” *Pattern Recognition*, vol. 48, no. 4, pp. 1547–1562, 2015.
- [209] L. Gonçalves, J. Novo, and A. Campilho, “Hessian based approaches for 3d lung nodule segmentation,” *Expert Systems with Applications*, vol. 61, pp. 1–15, 2016.
- [210] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: a deep convolutional encoder-decoder architecture for image segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.
- [211] F. Liu, Z. Zhou, H. Jang, A. Samsonov, G. Zhao, and R. Kijowski, “Deep convolutional neural network and 3d deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging,” *Magnetic Resonance in Medicine*, vol. 79, no. 4, pp. 2379–2391, 2018.
- [212] I. Nathan, Z. Ma, J. Li et al., “Semantic segmentation for prostate cancer grading by convolutional neural networks,” in *Medical Imaging 2018: Digital Pathology*, vol. 10581, Houston, Texas, US, SPIE, 2018 International Society for Optics and Photonics, Article ID 105811B.
- [213] R. Roy, T. Chakraborti, and A. S. Chowdhury, “A deep learning-shape driven level set synergism for pulmonary nodule segmentation,” *Pattern Recognition Letters*, vol. 123, pp. 31–38, 2019.
- [214] Y. Su, D. Li, and X. Chen, “Lung nodule detection based on faster r-cnn framework,” *Computer Methods and Programs in Biomedicine*, vol. 200, Article ID 105866, 2021.
- [215] R. Girshick, “Fast r-cnn,” in *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, Santiago, Chile, December 2015.
- [216] Du Tran, J. Ray, S. Zheng, S.-Fu Chang, and M. Paluri, “Convnet architecture search for spatiotemporal feature learning,” 2017, <https://arxiv.org/abs/1708.05038>.
- [217] Analyticsvidhya, “Image segmentation python: implementation of mask r-cnn,” 2021, <https://www.analyticsvidhya.com/blog/2019/07/computer-vision-implementing-mask-r-cnn-image-segmentation/>.
- [218] L. Cai, T. Long, Y. Dai, and Y. Huang, “Mask r-cnn-based detection and segmentation for pulmonary nodule 3d visualization diagnosis,” *IEEE Access*, vol. 8, Article ID 44400, 2020.
- [219] S. Sun, C. Bauer, and R. Beichel, “Automated 3-d segmentation of lungs with lung cancer in ct data using a novel robust active shape model approach,” *IEEE Transactions on Medical Imaging*, vol. 31, no. 2, pp. 449–460, 2012.
- [220] K. Kothavari and S. N. Deepa, “Segmentation of lung on ct images using robust active shape model (rasm) and tumour location using morphological processing,” *Academic Journal of Cancer Research*, vol. 7, no. 2, pp. 73–80, 2014.
- [221] J. H. Moltz, L. Bornemann, J.-M. Kuhnigk et al., “Advanced segmentation techniques for lung nodules, liver metastases, and enlarged lymph nodes in ct scans,” *IEEE Journal of selected topics in signal processing*, vol. 3, no. 1, pp. 122–134, 2009.
- [222] S. Sun, C. Bauer, and R. Beichel, “Robust active shape model based lung segmentation in ct scans,” in *Proceedings of the Fourth International Workshop on Pulmonary Image Analysis*, pp. 213–224, Toronto, Canada, September 2011.
- [223] S. A. Hojjatoleslami and J. Kittler, “Region growing: a new approach,” *IEEE Transactions on Image Processing*, vol. 7, no. 7, pp. 1079–1084, 1998.
- [224] J. Soltani-Nabipour, A. Khorshidi, and B. Noorian, “Lung tumor segmentation using improved region growing algorithm,” *Nuclear Engineering and Technology*, vol. 52, no. 10, pp. 2313–2319, 2020.
- [225] S. Avinash, K. Manjunath, and S. S. Kumar, “An improved image processing analysis for the detection of lung cancer using gabor filters and watershed segmentation technique,” in *Proceedings of the 2016 International Conference on Inventive Computation Technologies (ICICT)*, vol. 3, pp. 1–6, Coimbatore, India, August 2016.
- [226] H. Shaziya, K. Shyamala, and R. Zaheer, “Automatic lung segmentation on thoracic ct scans using u-net convolutional network,” in *Proceedings of the 2018 International conference on communication and signal processing (ICCSIP)*, pp. 643–647, IEEE, Tamilnadu, India, April 2018.
- [227] S. S. Parveen and C. Kavitha, “Detection of lung cancer nodules using automatic region growing method,” in *Proceedings of the 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, pp. 1–6, IEEE, Tiruchengode, India, July 2013.
- [228] I. Guyon, S. Gunn, M. Nikravesh, and L. A. Zadeh, *Feature Extraction: Foundations and Applications*, vol. 207, Springer, Singapore, 2008.
- [229] R. Nevatia and K. Ramesh Babu, “Linear feature extraction and description,” *Computer Graphics and Image Processing*, vol. 13, no. 3, pp. 257–269, 1980.
- [230] Ø Due Trier, A. K. Jain, and T. Taxt, “Feature extraction methods for character recognition—a survey,” *Pattern Recognition*, vol. 29, no. 4, pp. 641–662, 1996.
- [231] S. Rani Jena, T. George, and N. Ponraj, “Feature extraction and classification techniques for the detection of lung cancer: a detailed survey,” in *Proceedings of the 2019 International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1–6, IEEE, Tamil Nadu, India, January 2019.
- [232] I. Fornaçon-Wood, H. Mistry, C. J. Ackermann et al., “Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform,” *European Radiology*, vol. 30, no. 11, pp. 6241–6250, 2020.
- [233] P. Lambin, E. Rios-Velazquez, R. Leijenaar et al., “Radiomics: extracting more information from medical images using advanced feature analysis,” *European Journal of Cancer*, vol. 48, no. 4, pp. 441–446, 2012.
- [234] A. Vial, D. Stirling, M. Field et al., “The role of deep learning and radiomic feature extraction in cancer-specific predictive modelling: a review,” *Translational Cancer Research*, vol. 7, no. 3, pp. 803–816, 2018.
- [235] S. Echegaray, V. Nair, M. Kadoch et al., “A rapid segmentation-insensitive “digital biopsy” method for radiomic feature extraction: method and pilot study using ct images of non-small cell lung cancer,” *Tomography*, vol. 2, no. 4, pp. 283–294, 2016.
- [236] A. Vial, D. Stirling, M. Field et al., “Assessing the prognostic impact of 3d ct image tumour rind texture features on lung cancer survival modelling,” in *Proceedings of the 2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pp. 735–739, IEEE, Montreal, QC, Canada, November 2017.
- [237] N. Pankaj, S. Kumar, A. K. Luhach et al., “Detection and analysis of lung cancer using radiomic approach,” in *Smart Computational Strategies: Theoretical and Practical Aspects*, pp. 13–24, Springer, Berlin/Heidelberg, Germany, 2019.
- [238] R. N. Mahon, M. Ghita, G. D. Hugo, and E. Weiss, “Combat harmonization for radiomic features in independent phantom and lung cancer patient computed tomography

- datasets," *Physics in Medicine and Biology*, vol. 65, no. 1, Article ID 15010, 2020.
- [239] C. O. Eric and O. Beijbom, "Transfer learning and deep feature extraction for planktonic image data sets," in *Proceedings of the 2017 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1082–1088, IEEE, Santa Rosa, CA, March 2017.
- [240] M. Nishio, O. Sugiyama, M. Yakami et al., "Computer-aided diagnosis of lung nodule classification between benign nodule, primary lung cancer, and metastatic lung cancer at different image size using deep convolutional neural network with transfer learning," *PLoS One*, vol. 13, no. 7, Article ID e0200721, 2018.
- [241] R. V. M. da Nóbrega, P. P. Reboucas Filho, M. B. Rodrigues, S. P. P. da Silva, C. M. J. M. Dourado Junior, and V. H. C. de Albuquerque, "Lung nodule malignancy classification in chest computed tomography images using transfer learning and convolutional neural networks," *Neural Computing & Applications*, vol. 32, no. 15, Article ID 11065, 2020.
- [242] C. Haarbuerger, P. Weitz, O. Rippel, and D. Merhof, "Image-based survival prediction for lung cancer patients using cnns," in *Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 1197–1201, IEEE, Venice, Italy, April 2019.
- [243] R. Paul, S. H. Hawkins, L.O Hall, D. B Goldgof, and R. J. Gillies, "Combining deep neural network and traditional image features to improve survival prediction accuracy for lung cancer patients from diagnostic ct," in *Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2570–2575, IEEE, Budapest, October 2016.
- [244] T. Tan, Z. Li, H. Liu et al., "Optimize transfer learning for lung diseases in bronchoscopy using a new concept: sequential finetuning," *IEEE journal of translational engineering in health and medicine*, vol. 6, pp. 1–8, 2018.
- [245] Goutham Swapna, S. Kp, Ravi, and Vinayakumar, "Automated detection of diabetes using cnn and cnn-lstm network and heart rate signals," *Procedia Computer Science*, vol. 132, pp. 1253–1262, 2018.
- [246] S. Lee and J. Shin, "Hybrid model of convolutional lstm and cnn to predict particulate matter," *International Journal of Information and Electronics Engineering*, vol. 9, no. 1, pp. 34–38, 2019.
- [247] R. Tekade and K. Rajeswari, "Lung cancer detection and classification using deep learning," in *Proceedings of the 2018 Fourth International Conference on Computing Communication Control and Automation (ICCCUBEA)*, pp. 1–5, IEEE, Pune, India, August 2018.
- [248] N. Varish and A. K. Pal, "Content based image retrieval using statistical features of color histogram," in *Proceedings of the 2015 3rd international conference on signal processing, communication and networking (ICSCN)*, pp. 1–6, IEEE, Chennai, India, March 2015.
- [249] C. Bhuvaneswari, P. Aruna, and D. Loganathan, "A new fusion model for classification of the lung diseases using genetic algorithm," *Egyptian Informatics Journal*, vol. 15, no. 2, pp. 69–77, 2014.
- [250] M. A. Gavrielides, R. Zeng, L. M. Kinnard, K. J. Myers, and N. Petrick, "Information-theoretic approach for analyzing bias and variance in lung nodule size estimation with ct: a phantom study," *IEEE Transactions on Medical Imaging*, vol. 29, no. 10, pp. 1795–1807, 2010.
- [251] A. Okita, M. Yamashita, K. Abe et al., "Variance analysis of a clinical pathway of video-assisted single lobectomy for lung cancer," *Surgery Today*, vol. 39, no. 2, pp. 104–109, 2009.
- [252] A. Kamiya, S. Murayama, H. Kamiya, T. Yamashiro, Y. Oshiro, and N. Tanaka, "Kurtosis and skewness assessments of solid lung nodule density histograms: differentiating malignant from benign nodules on ct," *Japanese Journal of Radiology*, vol. 32, no. 1, pp. 14–21, 2014.
- [253] A.-N. Frix, F. Cousin, T. Refaee et al., "Radiomics in lung diseases imaging: state-of-the-art for clinicians," *Journal of Personalized Medicine*, vol. 11, no. 7, p. 602, 2021.
- [254] R. Sturm, "Aerosol bolus inhalation as technique for the diagnosis of various lung diseases—a theoretical approach," *Comp Math Biol*, vol. 3, no. 2, 2014.
- [255] N. R. Pal and S. K. Pal, "Entropy: a new definition and its applications," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 21, no. 5, pp. 1260–1270, 1991.
- [256] Researchgate, "What is the significance of image entropy (plain image and cipher image) in image processing?," 2005, <https://www.researchgate.net/post/What-is-the-significance-of-image-entropy-plain-image-and-cipher-image-in-image-processing>.
- [257] L. Hussain, W. Aziz, A. A. Alshdadi, M. S. Ahmed Nadeem, I. R. Khan, and Q. U. A. Chaudhry, "Analyzing the dynamics of lung cancer imaging data using refined fuzzy entropy methods by extracting different features," *IEEE Access*, vol. 7, Article ID 64704, 2019.
- [258] J. Zabalza, J. Ren, J. Zheng et al., "Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging," *Neuro-computing*, vol. 185, pp. 1–10, 2016.
- [259] Q. Meng, D. Catchpoole, D. Skillicom, and P. J. Kennedy, "Relational autoencoder for feature extraction," in *Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 364–371, IEEE, Anchorage, AK, USA, May 2017.
- [260] S. Ahmed, S. Ahmed, and M. Ghazal, "A novel autoencoder-based diagnostic system for early assessment of lung cancer," in *Proceedings of the 2018 25th IEEE international conference on image processing (ICIP)*, pp. 1393–1397, IEEE, Athens, Greece, October 2018.
- [261] Z. Wang and Y. Wang, "Extracting a biologically latent space of lung cancer epigenetics with variational autoencoders," *BMC Bioinformatics*, vol. 20, no. S18, pp. 568–577, 2019.
- [262] Z. Wang and Y. Wang, "Exploring dna methylation data of lung cancer samples with variational autoencoders," in *Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1286–1289, IEEE, Madrid, Spain, December 2018.
- [263] S. An, M. Bodruzzaman, and M. J. Malkani, "Feature extraction using wavelet transform for neural network based image classification," in *Proceedings of Thirtieth Southeastern Symposium on System Theory*, pp. 412–416, IEEE, Morgantown, WV, USA, March 1998.
- [264] M. Soufi, H. Arimura, and N. Nagami, "Identification of optimal mother wavelets in survival prediction of lung cancer patients using wavelet decomposition-based radiomic features," *Medical Physics*, vol. 45, no. 11, pp. 5116–5128, 2018.
- [265] S. Park, S. M. Lee, K. H. Do et al., "Deep learning algorithm for reducing ct slice thickness: effect on reproducibility of radiomic features in lung cancer," *Korean Journal of Radiology*, vol. 20, no. 10, pp. 1431–1440, 2019.

- [266] L. Nanni and A. Lumini, "Wavelet selection for disease classification by dna microarray data," *Expert Systems with Applications*, vol. 38, no. 1, pp. 990–995, 2011.
- [267] E. Adetiba and O. O. Olugbara, "Lung cancer prediction using neural network ensemble with histogram of oriented gradient genomic features," *The Scientific World Journal*, vol. 2015, Article ID 786013, 17 pages, 2015.
- [268] Y. Xie, Y. Xia, J. Zhang et al., "Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest ct," *IEEE Transactions on Medical Imaging*, vol. 38, no. 4, pp. 991–1004, 2019.
- [269] M. Firmino, G. Angelo, H. Morais, M. R. Dantas, and R. Valentim, "Computer-aided detection (cade) and diagnosis (cadx) system for lung cancer with likelihood of malignancy," *Biomedical Engineering Online*, vol. 15, no. 1, pp. 1–17, 2016.
- [270] F. N. Iandola, S. Han, and M W Moskewicz, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size," 2016, <https://arxiv.org/abs/1602.07360>.
- [271] K. Simonyan and A. Zisserman, "Very deep convolutional networks for largescale image recognition," 2014, <https://arxiv.org/abs/1409.1556>.
- [272] H. Qassim, A. Verma, and D. Feinzimer, "Compressed residual-vgg16 cnn model for big data places image recognition," in *Proceedings of the 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 169–175, IEEE, Nevada, USA, January 2018.
- [273] M. A. Khan, V. Rajinikanth, S. C. Satapathy et al., "Vgg19 network assisted joint segmentation and classification of lung nodules in ct images," *Diagnostics*, vol. 11, no. 12, p. 2208, 2021.
- [274] B. Chen, R. Zhang, Y. Gan, L. Yang, and W. Li, "Development and clinical application of radiomics in lung cancer," *Radiation Oncology*, vol. 12, no. 1, p. 154, 2017.
- [275] N. K. Oswald, J. Halle-Smith, R. Mehdi, P. Nightingale, B. Naidu, and A. M. Turner, "Predicting postoperative lung function following lung cancer resection: a systematic review and meta-analysis," *EClinicalMedicine*, vol. 15, pp. 7–13, 2019.
- [276] M. Astaraki, I. Toma-Dasu, Örjan Smedby, and C. Wang, "Normal appearance autoencoder for lung cancer detection and segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 249–256, Springer, Singapore, September 2019.
- [277] C. Zhuo, H. Zhuang, X. Gao, and P. T. Triplett, "Lung cancer incidence in patients with schizophrenia: metaanalysis," *British Journal of Psychiatry*, vol. 215, no. 6, pp. 704–711, 2019.
- [278] R. Arulmurugan and H. Anandakumar, "Early detection of lung cancer using wavelet feature descriptor and feed forward back propagation neural networks classifier," in *Computational Vision and Bio Inspired Computing*, pp. 103–110, Springer, Cham, 2018.
- [279] K. Kayser, G. Kayser, and K. Metzke, "The concept of structural entropy in tissue-based diagnosis," *Analytical and Quantitative Cytology and Histology*, vol. 29, no. 5, pp. 296–308, 2007.
- [280] A. Bhandary, G. A. Prabhu, V. Rajinikanth et al., "Deep-learning framework to detect lung abnormality—a study with chest x-ray and lung ct scan images," *Pattern Recognition Letters*, vol. 129, pp. 271–278, 2020.
- [281] J. Brownlee, "Autoencoder feature extraction for classification," 2020, <https://machinelearningmastery.com/autoencoder-for-classification/>.
- [282] E. Ahmed, M. A. Hanan, and F. E. Z. Abou-Chadi, "Early lung cancer detection using deep learning optimization," 2020.
- [283] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of Bioinformatics and Computational Biology*, vol. 03, no. 02, pp. 185–205, 2005.
- [284] Oneapi, "Onedal documentation," [https://oneapi-src.github.io/oneDAL/daal/algorithms/lasso\\_elastic\\_net/lasso.html](https://oneapi-src.github.io/oneDAL/daal/algorithms/lasso_elastic_net/lasso.html).
- [285] P. Pudil, J. Novovičová, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters*, vol. 15, no. 11, pp. 1119–1125, 1994.
- [286] P. Somol, J. Novovicova, and P. Pudil, "Efficient feature subset selection and subset size optimization," *Pattern Recognition Recent Advances*, vol. 1, 2010.
- [287] Z. Jaadi, "A step-by-step explanation of principal component analysis (pca)," <https://builtin.com/data-science/step-step-explanation-principal-component-analysis>.
- [288] J. A. Alzubi, B. Bharathikannan, S. Tanwar, R. Manikandan, A. Khanna, and C. Thaventhiran, "Boosted neural network ensemble classification for lung cancer disease diagnosis," *Applied Soft Computing*, vol. 80, pp. 579–591, 2019.
- [289] O. Obulesu, S. Kallam, G. Dhiman et al., "Adaptive diagnosis of lung cancer by deep learning classification using wilcoxon gain and generator," *Journal of Healthcare Engineering*, vol. 2021, Article ID 5912051, 13 pages, 2021.
- [290] C.-W. Tsai, Bo-C. Huang, and M.-C. Chiang, "A novel spiral optimization for clustering," in *Mobile, Ubiquitous, and Intelligent Computing*, pp. 621–628, Springer, Berlin Heidelberg, 2014.
- [291] K. Tamura and K. Yasuda, "Spiral multipoint search for global optimization," vol. 1, pp. 470–475, in *Proceedings of the 2011 10th International Conference on Machine Learning and Applications and Workshops*, vol. 1, pp. 470–475, IEEE, Washington, DC, US, December 2011.
- [292] N. Maleki, Y. Zeinali, and S. T. A. Niaki, "A k-nn method for lung cancer prognosis with the use of a genetic algorithm for feature selection," *Expert Systems with Applications*, vol. 164, Article ID 113981, 2021.
- [293] Z. Cai, D. Xu, Q. Zhang, J. Zhang, S. M. Ngai, and J. Shao, "Classification of lung cancer using ensemble-based feature selection and machine learning methods," *Molecular BioSystems*, vol. 11, no. 3, pp. 791–800, 2015.
- [294] Y. Zhang and S. Biswas, "An improved version of logistic bayesian lasso for detecting rare haplotype-environment interactions with application to lung cancer," *Cancer Informatics*, vol. 14s2, 2015.
- [295] G. Valdes, T. D. Solberg, M. Heskell, L. Ungar, and C. B. Simone, "Using machine learning to predict radiation pneumonitis in patients with stage i non-small cell lung cancer treated with stereotactic body radiation therapy," *Physics in Medicine and Biology*, vol. 61, no. 16, pp. 6105–6120, 2016.
- [296] E. Kaznowska, J. Depciuch, K. Łach et al., "The classification of lung cancers and their degree of malignancy by ftir, pcalda analysis, and a physics-based computational model," *Talanta*, vol. 186, pp. 337–345, 2018.
- [297] M. Aly, "Survey on multiclass classification methods," *Neural Netw*, vol. 19, pp. 1–9, 2005.
- [298] P. C. Sen, M. Hajra, and M. Ghosh, "Supervised classification algorithms in machine learning: a survey and review," in

- Emerging Technology in Modelling and Graphics*, pp. 99–111, Springer, Berlin/Heidelberg, Germany, 2020.
- [299] R. Paul, S. H. Hawkins, Y. Balagurunathan et al., “Deep feature transfer learning in combination with traditional features predicts survival among patients with lung adenocarcinoma,” *Tomography*, vol. 2, no. 4, pp. 388–395, 2016.
- [300] A. Masood, B. Sheng, P. Li et al., “Computer-assisted decision support system in pulmonary cancer detection and stage classification on ct images,” *Journal of Biomedical Informatics*, vol. 79, pp. 117–128, 2018.
- [301] L. Sk, S. N. Mohanty, N. Arunkumar, G. Ramirez, and G. Ramirez, “Optimal deep learning model for classification of lung cancer on ct images,” *Future Generation Computer Systems*, vol. 92, pp. 374–382, 2019.
- [302] P. Cao, X. Liu, J. Yang et al., “A multi-kernel based framework for heterogeneous feature selection and oversampling for computeraided detection of pulmonary nodules,” *Pattern Recognition*, vol. 64, pp. 327–346, 2017.
- [303] R. Paul, L. Hall, D. Goldgof, M. Schabath, and R. Gillies, “Predicting nodule malignancy using a cnn ensemble approach,” in *Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, Rio de Janeiro, Brazil, July 2018.
- [304] F. Ciompi, B. de Hoop, S. J. van Riel et al., “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,” *Medical Image Analysis*, vol. 26, no. 1, pp. 195–202, 2015.
- [305] X. Tu, M. Xie, J. Gao et al., “Automatic categorization and scoring of solid, part-solid and non-solid pulmonary nodules in ct images with convolutional neural network,” *Scientific Reports*, vol. 7, no. 1, pp. 8533–8610, 2017.
- [306] A. Teramoto, H. Fujita, O. Yamamuro, and T. Tamaki, “Automated detection of pulmonary nodules in pet/ct images: ensemble false-positive reduction using a convolutional neural network technique,” *Medical Physics*, vol. 43, no. 6Part1, pp. 2821–2827, 2016.
- [307] Y. Xie, J. Zhang, Y. Xia, M. Fulham, and Y. Zhang, “Fusing texture, shape and deep model-learned information at decision level for automated classification of lung nodules on chest ct,” *Information Fusion*, vol. 42, pp. 102–110, 2018.
- [308] J. Yuan, X. Liu, F. Hou, H. Qin, and A. Hao, “Hybrid-feature-guided lung nodule type classification on ct images,” *Computers & Graphics*, vol. 70, pp. 288–299, 2018.
- [309] S. Alves Peixoto, “Lung nodule classification via deep transfer learning in ct lung images,” in *Proceedings of the 2018 IEEE 31st international symposium on computer-based medical systems (CBMS)*, pp. 244–249, IEEE, Karlstad, Sweden, June 2018.
- [310] D. Kumar, A. G. Chung, M. J. Shaifee et al., “Discovery radiomics for pathologically-proven computed tomography lung cancer prediction,” in *Proceedings of the International Conference Image Analysis and Recognition*, pp. 54–62, Springer, Póvoa de Varzim, Portugal, August 2017.
- [311] P. Cao, X. Liu, J. Zhang et al., “A 2, 1 norm regularized multi-kernel learning for false positive reduction in lung nodule cad,” *Computer Methods and Programs in Biomedicine*, vol. 140, pp. 211–231, 2017.
- [312] Z.H Luo, M A Brubaker, and M Brudno, “Size and texture-based classification of lung tumors with 3d cnns,” in *Proceedings of the 2017 IEEE winter conference on applications of computer vision (WACV)*, pp. 806–814, IEEE, Santa Rosa, CA, March 2017.
- [313] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT press, Cambridge, Massachusetts, 2016.
- [314] Li Deng and Yu Dong, “Deep learning: methods and applications,” *Foundations and Trends in Signal Processing*, vol. 7, no. 3–4, pp. 197–387, 2014.
- [315] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [316] Y. Bengio, I. Goodfellow, and A. Courville, *Deep Learning*, Vol. 1, MIT press, Massachusetts, USA, 2017.
- [317] X. Jin, C. Ma, Y. Zhang, and L. Li, “Classification of lung nodules based on convolutional deep belief network,” vol. 1, pp. 139–142, in *Proceedings of the 2017 10th International Symposium on Computational Intelligence and Design (ISCID)*, vol. 1, pp. 139–142, IEEE, Hangzhou, China, December 2017.
- [318] Z. Liu, C. Yao, H. Yu, and T. Wu, “Deep reinforcement learning with its application for lung cancer detection in medical internet of things,” *Future Generation Computer Systems*, vol. 97, no. 1–9, pp. 1–9, 2019.
- [319] M. Schwyzer, D. A. Ferraro, U. J. Muehlematter et al., “Automated detection of lung cancer at ultralow dose pet/ct by deep neural networks—initial results,” *Lung Cancer*, vol. 126, pp. 170–173, 2018.
- [320] W. Sun, B. Zheng, and W. Qian, “Computer aided lung cancer diagnosis with deep learning algorithms,” in *Medical Imaging 2016: Computer-Aided Diagnosis*, vol. 9785, Campus Saint-Christophe – Europa, SPIE, Article ID 97850Z, 2016.
- [321] B. C. Kim, S. S. Yu, and H. Il Suk, “Deep feature learning for pulmonary nodule classification in a lung ct,” in *Proceedings of the 2016 4th International winter conference on brain-computer interface (BCI)*, pp. 1–3, IEEE, Gangwon-do, Korea, February 2016.
- [322] D. Qi, H. Chen, Y. Jin et al., “Automated pulmonary nodule detection via 3d convnets with online sample filtering and hybrid-loss residual learning,” in *Proceedings of the International conference on medical image computing and computer-assisted intervention*, pp. 630–638, Springer, Singapore, September 2017.
- [323] H. Jin, Z. Li, R. Tong, and L. Lin, “A deep 3d residual cnn for falsepositive reduction in pulmonary nodule detection,” *Medical Physics*, vol. 45, no. 5, pp. 2097–2107, 2018.
- [324] S. Ahmed, S. Ahmed, M. Ghazal et al., “A new framework for incorporating appearance and shape features of lung nodules for precise diagnosis of lung cancer,” in *Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP)*, pp. 1372–1376, IEEE, Beijing, China, September 2017.
- [325] G. A. P. Singh and P. K. Gupta, “Performance analysis of various machine learningbased approaches for detection and classification of lung cancer in humans,” *Neural Computing and Applications*, vol. 31, no. 10, pp. 6863–6877, 2019.
- [326] S. Hamidian, B. Sahiner, N. Petrick, and A. Pezeshk, “3D convolutional neural network for automatic detection of lung nodules in chest CT,” in *Medical Imaging 2017: Computer-Aided Diagnosis*, S. G. Armato and N. A. Petrick, Eds., vol. 10134, SPIE, International Society for Optics and Photonics, pp. 54–59, 2017.
- [327] B. Sun, C.-H. Ma, X.-Yu Jin, and Ye Luo, “Deep sparse auto-encoder for computer aided pulmonary nodules ct diagnosis,” in *Proceedings of the 2016 13th international computer conference on wavelet active media technology and information processing (ICCWAMTIP)*, pp. 235–238, IEEE, Chengdu, China, December 2016.

- [328] C. Zhao, J. Han, Y. Jia, and F. Gou, "Lung nodule detection via 3d u-net and contextual convolutional neural network," in *Proceedings of the 2018 International Conference on Networking and Network Applications (NaNA)*, pp. 356–361, Xi'an, China, October 2018.
- [329] S. Saha, "A comprehensive guide to convolutional neural networks — the eli5 way," 2018, <https://towardsdatascience.com/acomprensive-guide-to-convolutional-neuralnetworks-the-eli5-way-3bd2b1164a53>.
- [330] A. Wadhwa and S. S. Roy, "10 - driver drowsiness detection using heart rate and behavior methods: a study," in *Data Analytics in Biomedical Engineering and Healthcare*, K. Chang Lee, S. Sekhar Roy, P. Samui, and V. Kumar, Eds., pp. 163–177, Academic Press, Cambridge, Massachusetts, 2021.
- [331] R. M. Tharsanee, R. S. Soundariya, A. S. Kumar, M. Karthiga, and S. Sountharajan, "Deep convolutional neural network-based image classification for covid-19 diagnosis," *Data Science for COVID-19*, pp. 117–145, 2021, <https://www.sciencedirect.com/topics/engineering/convolutional-neural-network>.
- [332] W. Jiang, G. Zeng, S. Wang, X. Wu, and C. Xu, "Application of deep learning in lung cancer imaging diagnosis," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–12, 2022.
- [333] W. J. Sori, J. Feng, and S. Liu, "Multi-path convolutional neural network for lung cancer detection," *Multidimensional Systems and Signal Processing*, vol. 30, no. 4, pp. 1749–1768, 2019.
- [334] S. Shen, S. X. Han, D. R. Aberle, A. A. Bui, and W. Hsu, "An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification," *Expert Systems with Applications*, vol. 128, pp. 84–95, 2019.
- [335] J. L. Causey, J. Zhang, S. Ma et al., "Highly accurate model for prediction of lung nodule malignancy with ct scans," *Scientific Reports*, vol. 8, no. 1, pp. 9286–9312, 2018.
- [336] Y. Liu, P. Hao, P. Zhang, X. Xu, J. Wu, and W. Chen, "Dense convolutional binary-tree networks for lung nodule classification," *IEEE Access*, vol. 6, pp. 49080–49088, Article ID 49080, 2018.
- [337] B. Wu, Z. Zhou, J. Wang, and Y. Wang, "Joint learning for pulmonary nodule segmentation, attributes and malignancy prediction," in *Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 1109–1113, IEEE, Washington, DC, USA, April 2018.
- [338] R. Dey, Z. Lu, and H. Yi, "Diagnostic classification of lung nodules using 3d neural networks," in *Proceedings of the 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pp. 774–778, IEEE, Washington, DC, USA, April 2018.
- [339] X. Zhao, L. Liu, S. Qi, Y. Teng, J. Li, and W. Qian, "Agile convolutional neural network for pulmonary nodule classification using ct images," *International Journal of Computer Assisted Radiology and Surgery*, vol. 13, no. 4, pp. 585–595, 2018.
- [340] P. Nanglia, S. Kumar, A. N. Mahajan, P. Singh, and D. Rathee, "A hybrid algorithm for lung cancer classification using svm and neural networks," *ICT Express*, vol. 7, no. 3, pp. 335–341, 2021.
- [341] F. Mottaghitlab, M. Farokhi, Y. Fatahi, F. Atyabi, and R. Dinarvand, "New insights into designing hybrid nanoparticles for lung cancer: diagnosis and treatment," *Journal of Controlled Release*, vol. 295, pp. 250–267, 2019.
- [342] M. R. Daliri, "A hybrid automatic system for the diagnosis of lung cancer based on genetic algorithm and fuzzy extreme learning machines," *Journal of Medical Systems*, vol. 36, no. 2, pp. 1001–1005, 2012.
- [343] T. Tunç, "A new hybrid method logistic regression and feedforward neural network for lung cancer data," *Mathematical Problems in Engineering*, vol. 2012, pp. 1–10, 2012.
- [344] S. Wang, L. Dong, X. Wang, and X. Wang, "Classification of pathological types of lung cancer from ct images by deep residual neural networks with transfer learning strategy," *Open Medicine*, vol. 15, no. 1, pp. 190–197, 2020.
- [345] B.S Shah, L. Hakim, M. Kavitha, H. W. Kim, and T. Kurita, "Transfer learning by cascaded network to identify and classify lung nodules for cancer detection," in *Proceedings of the International Workshop on Frontiers of Computer Vision*, pp. 262–273, Springer, Kagoshima, Japan, February 2020.
- [346] X. Li, G. C. Nsofor, and L. Song, "A comparative analysis of predictive data mining techniques," *International Journal of Rapid Manufacturing*, vol. 1, no. 2, pp. 150–172, 2009.
- [347] K. Aftarczuk, "Evaluation of selected data mining algorithms implemented in medical decision support systems," 2007.
- [348] DeepAI, "Evaluation metrics," 2019, <https://deepai.org/machinelearning-glossary-and-terms/evaluation-metrics>.
- [349] J. Kuruvilla and K. Gunavathi, "Lung cancer classification using neural networks for ct images," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 1, pp. 202–209, 2014.
- [350] S. D. S. Deepa and V. S. Bharathi, "Textural feature extraction and classification of mammogram images using cccm and pnn," *IOSR Journal of Computer Engineering (IOSR-JCE)*, vol. 10, no. 6, pp. 7–13, 2013.
- [351] V. Krishnaiah, G. Narsimha, and N. S. Chandra, "Diagnosis of lung cancer prediction system using data mining classification techniques," *International Journal of Computer Science and Information Technologies*, vol. 4, no. 1, pp. 39–45, 2013.
- [352] D. M. Ibrahim, N. M. Elshennawy, and A. M. Sarhan, "Deepchest: multiclassification deep learning model for diagnosing covid-19, pneumonia, and lung cancer chest diseases," *Computers in Biology and Medicine*, vol. 132, Article ID 104348, 2021.
- [353] I. M. Nasser and S. S. Abu-Naser, "Lung cancer detection using artificial neural network," *International Journal of Engineering and Information Systems (IJEAIS)*, vol. 3, no. 3, pp. 17–23, 2019.
- [354] C. J. Ryerson, E. Vittinghoff, B. Ley et al., "Predicting survival across chronic interstitial lung disease: the ild-gap model," *Chest*, vol. 145, no. 4, pp. 723–728, 2014.
- [355] A. Potti, S. Mukherjee, R. Petersen et al., "A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer," *New England Journal of Medicine*, vol. 355, no. 6, pp. 570–580, 2006.
- [356] Q. Ke, J. Zhang, W. Wei et al., "A neuro-heuristic approach for recognition of lung diseases from x-ray images," *Expert Systems with Applications*, vol. 126, pp. 218–232, 2019.
- [357] T. Rahman, A. Khandakar, M. A. Kadir et al., "Reliable tuberculosis detection using chest x-ray with deep learning, segmentation and visualization," *IEEE Access*, vol. 8, pp. 191586–191601, Article ID 191586, 2020.
- [358] J.-X. Wu, Pi-Y. Chen, C.-M. Li, Y. C. Kuo, N. S. Pai, and C. H. Lin, "Multilayer fractional-order machine vision classifier for rapid typical lung diseases screening on digital chest x-ray images," *IEEE Access*, vol. 8, pp. 105886–105902, 2020.
- [359] C.-H. Lin, J.-X. Wu, C.-M. Li, P. Y. Chen, N. S. Pai, and Y. C. Kuo, "Enhancement of chest x-ray images to improve screening accuracy rate using iterated function system and



- multilayer fractional-order machine learning classifier," *IEEE Photonics Journal*, vol. 12, no. 4, pp. 1–18, 2020.
- [360] Z. Hui Hoo, J. Candlish, and D. Teare, "What is an roc curve?" 2017.
- [361] C. Marzban, "The roc curve and the area under it as performance measures," *Weather and Forecasting*, vol. 19, no. 6, pp. 1106–1114, 2004.
- [362] T. K. Burki, "Predicting lung cancer prognosis using machine learning," *The Lancet Oncology*, vol. 17, no. 10, p. e421, 2016.
- [363] G. A. Silvestri, A. Vachani, D. Whitney et al., "A bronchial genomic classifier for the diagnostic evaluation of lung cancer," *N Engl J Med*, vol. 373, no. 3, pp. 243–251, 2015.
- [364] D. K. McClish, "Analyzing a portion of the roc curve," *Medical Decision Making*, vol. 9, no. 3, pp. 190–195, 1989.
- [365] F. Bach, D. Heckerman, and E. Horvitz, "On the path to an ideal roc curve: considering cost asymmetry in learning classifiers," in *Proceedings of the International Workshop on Artificial Intelligence and Statistics*, pp. 9–16, PMLR, Barbados, January 2005.
- [366] N. Asada, K. Doi, H. MacMahon et al., "Potential usefulness of an artificial neural network for differential diagnosis of interstitial lung diseases: pilot study," *Radiology*, vol. 177, no. 3, pp. 857–860, 1990.
- [367] G. Schmalisch, S. Wilitzki, and R. R. Wauer, "Differences in tidal breathing between infants with chronic lung diseases and healthy controls," *BMC Pediatrics*, vol. 5, no. 1, pp. 36–10, 2005.
- [368] Q. Yang, P. Zhang, R. Wu, K. Lu, and H. Zhou, "Identifying the best marker combination in cea, ca125, cy211, nse, and scc for lung cancer screening by combining roc curve and logistic regression analyses: is it feasible?" *Disease Markers*, vol. 2018, pp. 1–12, 2018.
- [369] K. Ashizawa, T. Ishida, H. MacMahon, C. J. Vyborny, S. Katsuragawa, and K. Doi, "Artificial neural networks in chest radiography: application to the differential diagnosis of interstitial lung disease," *Academic Radiology*, vol. 6, no. 1, pp. 2–9, 1999.
- [370] J. M. Lobo, A. Jiménez-Valverde, and R. Real, "Auc: a misleading measure of the performance of predictive distribution models," *Global Ecology and Biogeography*, vol. 17, no. 2, pp. 145–151, 2008.
- [371] A. Jiménez-Valverde, "Insights into the area under the receiver operating characteristic curve (auc) as a discrimination measure in species distribution modelling," *Global Ecology and Biogeography*, vol. 21, no. 4, pp. 498–507, 2012.
- [372] J. Huang and C. X. Ling, "Using auc and accuracy in evaluating learning algorithms," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 3, pp. 299–310, 2005.
- [373] C. X. Ling, J. Huang, H. Zhang et al., "Auc: a statistically consistent and more discriminating measure than accuracy," *Ijcai*, vol. 3, pp. 519–524, 2003.
- [374] O. J. O'Connell, F. A. Almeida, M. J. Simoff et al., "A prediction model to help with the assessment of adenopathy in lung cancer: Hal," *American Journal of Respiratory and Critical Care Medicine*, vol. 195, no. 12, pp. 1651–1660, 2017.
- [375] J. Nowak, B. Hudzik, D. Jastrzbski et al., "Pulmonary hypertension in advanced lung diseases: echocardiography as an important part of patient evaluation for lung transplantation," *The Clinical Respiratory Journal*, vol. 12, no. 3, pp. 930–938, 2018.
- [376] S. C. Park, J. Tan, X. Wang et al., "Computer-aided detection of early interstitial lung diseases using low-dose ct images," *Physics in Medicine & Biology*, vol. 56, no. 4, p. 1139, 2011.
- [377] W. Lee, W. S. Chung, Ki-S. Hong, and J. Huh, "Clinical usefulness of bronchoalveolar lavage cellular analysis and lymphocyte subsets in diffuse interstitial lung diseases," *Annals of Laboratory Medicine*, vol. 35, no. 2, pp. 220–225, 2015.
- [378] G. Sozzi, D. Conte, L. Mariani et al., "Analysis of circulating tumor dna in plasma at diagnosis and during follow-up of lung cancer patients," *Cancer Research*, vol. 61, no. 12, pp. 4675–4678, 2001.
- [379] L. Zong, Q. Sun, H. Zhang et al., "Increased expression of circrna\_102231 in lung cancer and its clinical significance," *Biomedicine & Pharmacotherapy*, vol. 102, pp. 639–644, 2018.
- [380] M. Kirienko, M. Sollini, G. Silvestri et al., "Convolutional neural networks promising in lung cancer t-parameter assessment on baseline fdg-pet/ct," *Contrast Media & Molecular Imaging*, vol. 2018, pp. 1–6, 2018.
- [381] S. FletcherMd Z. Islam et al., "Comparing sets of patterns with the jaccard index," *Australasian Journal of Information Systems*, vol. 22, 2018.
- [382] L. Hamers, Y. Hemeryck, G. Herweyers et al., "Similarity measures in scientometric research: the jaccard index versus salton's cosine formula," *Information Processing & Management*, vol. 25, no. 3, pp. 315–318, 1989.
- [383] R. R. Shamir, Yuval Duchin, J. Kim, G. Sapiro, and N. Harel, "Continuous dice coefficient: a method for evaluating probabilistic segmentations," 2019, <https://arxiv.org/abs/1906.11031>.
- [384] N. J. Tustison and J. C. Gee, "Introducing dice, jaccard, and other label overlap measures to itk," *Insight J*, vol. 2, 2009.
- [385] V. Thada and V. Jaglan, "Comparison of jaccard, dice, cosine similarity coefficient to find best fitness value for web retrieved documents using genetic algorithm," *International Journal of Innovations in Engineering and Technology*, vol. 2, no. 4, pp. 202–205, 2013.
- [386] E. Yip, J. Yun, K. Wachowicz, Z. Gabos, S. Rathee, and B. Fallone, "Sliding window prior data assisted compressed sensing for mri tracking of lung tumors," *Medical Physics*, vol. 44, no. 1, pp. 84–98, 2017.
- [387] V. K. Pattisapu, I. Daunhawer, T. Weikert et al., "Pet-guided attention network for segmentation of lung tumors from pet/ct images," *Pattern Recognition*, vol. 12544, p. 445, 2020.
- [388] H. Hu, Q. Li, Y. Zhao, and Ye Zhang, "Parallel deep learning algorithms with hybrid attention mechanism for image segmentation of lung tumors," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2880–2889, 2021.
- [389] I. Bukovsky, N. Homma, K. Ichiji et al., "A fast neural network approach to predict lung tumor motion during respiration for radiation therapy applications," *BioMed Research International*, vol. 2015, pp. 1–13, 2015.
- [390] M. Maspero, A. C. Houweling, M. H. F. Savenije et al., "A single neural network for cone-beam computed tomography-based radiotherapy of head-and-neck, lung and breast cancer," *Physics and Imaging in Radiation Oncology*, vol. 14, no. 24–31, pp. 24–31, 2020.
- [391] T. Uemura, M. Matsuhira, C. Watari et al., "Deep radiomic precision ct imaging for prognostic biomarkers for interstitial lung diseases," in *Medical Imaging 2019: Imaging Informatics for Healthcare, Research, and Applications*, vol. 10954, Campus Saint, Christophe,

- EuropaInternational Society for Optics and Photonics, Article ID SPIE, 2019.
- [392] C. Bhuvaneshwari, P. Aruna, and D. Loganathan, "Classification of lung diseases by image processing techniques using computed tomography images," *International Journal of Advanced Computer Research*, vol. 4, no. 1, p. 87, 2014.
- [393] C. Bhuvaneshwari, P. Aruna, and D. Loganathan, "Advanced segmentation techniques using genetic algorithm for recognition of lung diseases from ct scans of thorax," *International Journal of Engineering Research and Applications*, vol. 3, no. 4, pp. 2517–2524, 2013.
- [394] C. Bhuvaneshwari, P. Aruna, and D. Loganathan, "Classification of the lung diseases from ct scans by advanced segmentation techniques using genetic algorithm," *International Journal of Computer Applications*, vol. 77, no. 16, pp. 21–27, 2013.
- [395] C. I. Henschke, D. I. McCauley, D. F. Yankelevitz et al., "Early lung cancer action project: overall design and findings from baseline screening," *The Lancet*, vol. 354, no. 9173, pp. 99–105, 1999.
- [396] J. P. Ko, H. Rusinek, E. L. Jacobs et al., "Small pulmonary nodules: volume measurement at chest ct—phantom study," *Radiology*, vol. 228, no. 3, pp. 864–870, 2003.
- [397] J.-M. Kuhnigk, V. Dicken, L. Bornemann et al., "Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic ct scans," *IEEE Transactions on Medical Imaging*, vol. 25, no. 4, pp. 417–434, 2006.
- [398] K. Okada, D. Comaniciu, and A. Krishnan, "Robust anisotropic Gaussian fitting for volumetric characterization of pulmonary nodules in multislice ct," *IEEE Transactions on Medical Imaging*, vol. 24, no. 3, pp. 409–423, 2005.
- [399] W. J. Kostis, A. Reeves, D. Yankelevitz, and C. Henschke, "Three-dimensional segmentation and growth-rate estimation of small pulmonary nodules in helical ct images," *IEEE Transactions on Medical Imaging*, vol. 22, no. 10, pp. 1259–1274, 2003.
- [400] I F Catalin, F. Preteux, C. B. Aubry, and P. Grenier, "3d automated lung nodule segmentation in hrct," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 626–634, Springer, London UK, September 2003.
- [401] Li Zhang, M. Fang, D. P. Naidich, and L. N. Carol, "Consistent interactive segmentation of pulmonary ground glass nodules identified in ct studies," in *Medical Imaging 2004: Image Processing*, vol. 5370, pp. 1709–1719, SPIE, International Society for Optics and Photonics, 2004.
- [402] A. P. Reeves, A. P. Reeves, D. Yankelevitz, and C. I. Henschke, "Three-dimensional multi-criterion automatic segmentation of pulmonary nodules of helical computed tomography images," *Optical Engineering*, vol. 38, no. 8, pp. 1340–1347, 1999.
- [403] A. Reeves, A. Chan, D. Yankelevitz, C. Henschke, B. Kressler, and W. Kostis, "On measuring the change in size of pulmonary nodules," *IEEE Transactions on Medical Imaging*, vol. 25, no. 4, pp. 435–450, 2006.
- [404] T. Kubota, A. K. Jerebko, M. Dewan, M. Salganicoff, and A. Krishnan, "Segmentation of pulmonary nodules of various densities with morphological approaches and convexity models," *Medical Image Analysis*, vol. 15, no. 1, pp. 133–154, 2011.
- [405] W. A. Browder, A. P. Reeves, and V A Tatiyana, "Automated volumetric segmentation method for growth consistency of nonsolid pulmonary nodules in high-resolution ct," in *Medical Imaging 2007: Computer-Aided Diagnosis* vol. 6514, Campus Saint, Christophe, Europa, SPIE, 2007 International Society for Optics and Photonics, Article ID 65140Y.
- [406] L. R. Goodman, M. Gulsun, L. Washington, P. G. Nagy, and K. L. Piacsek, "Inherent variability of ct lung nodule measurements in vivo using semiautomated volumetric measurements," *American Journal of Roentgenology*, vol. 186, no. 4, pp. 989–994, 2006.
- [407] J. Dehmeshki, H. Amin, M. Valdivieso, and X. Ye, "Segmentation of pulmonary nodules in thoracic ct scans: a region growing approach," *IEEE Transactions on Medical Imaging*, vol. 27, no. 4, pp. 467–480, 2008.
- [408] S. Diciotti, G. Picozzi, M. Falchini, M. Mascacchi, N. Villari, and G. Valli, "3-d segmentation algorithm of small lung nodules in spiral ct images," *IEEE Transactions on Information Technology in Biomedicine*, vol. 12, no. 1, pp. 7–19, 2008.
- [409] A. C. Jirapatnakul, Y. D. Mulman, A. P. Reeves, D. F. Yankelevitz, and C. I. Henschke, "Segmentation of juxtapleural pulmonary nodules using a robust surface estimate," *International Journal of Biomedical Imaging*, vol. 2011, pp. 1–14, 2011.
- [410] Ji H. Min, Ho Y. Lee, K. S. Lee et al., "Stepwise evolution from a focal pure pulmonary ground-glass opacity nodule into an invasive lung adenocarcinoma: an observation for more than 10 years," *Lung Cancer*, vol. 69, no. 1, pp. 123–126, 2010.
- [411] C. I. Henschke, D. F. Yankelevitz, R. Mirtcheva, G. McGuinness, D. McCauley, and O. S. Miettinen, "Ct screening for lung cancer: frequency and significance of part-solid and nonsolid nodules," *American Journal of Roentgenology*, vol. 178, no. 5, pp. 1053–1057, 2002.
- [412] M. C. B. Godoy and D. P. Naidich, "Subsolid pulmonary nodules and the spectrum of peripheral adenocarcinomas of the lung: recommended interim guidelines for assessment and management," *Radiology*, vol. 253, no. 3, pp. 606–622, 2009.
- [413] M. S. Hestbech, V. Siersma, A. Dirksen, J. H. Pedersen, and J. Brodersen, "Participation bias in a randomised trial of screening for lung cancer," *Lung Cancer*, vol. 73, no. 3, pp. 325–331, 2011.
- [414] J. Brodersen, H. Thorsen, and S. Kreiner, "Consequences of screening in lung cancer: development and dimensionality of a questionnaire," *Value in Health*, vol. 13, no. 5, pp. 601–612, 2010.
- [415] P. Sastry, A. Tockock, and A. S. Coonar, "Adrenalectomy for isolated metastasis from operable non-small-cell lung cancer," *Interactive Cardiovascular and Thoracic Surgery*, vol. 18, no. 4, pp. 495–497, 2014.
- [416] M. Higashiyama, O. Doi, K. Kodama, H. Yokouchi, S. Imaoka, and H. Koyama, "Surgical treatment of adrenal metastasis following pulmonary resection for lung cancer: comparison of adrenalectomy with palliative therapy," *International Surgery*, vol. 79, no. 2, pp. 124–129, 1994.
- [417] D. J. Raz, M. Lanuti, H. C. Gaissert, C. D. Wright, D. J. Mathisen, and J. C. Wain, "Outcomes of patients with isolated adrenal metastasis from non-small cell lung carcinoma," *The Annals of Thoracic Surgery*, vol. 92, no. 5, pp. 1788–1793, 2011.
- [418] J. D. Luketich and M. E. Burt, "Does resection of adrenal metastases from non-small cell lung cancer improve

survival?" *The Annals of Thoracic Surgery*, vol. 62, no. 6, pp. 1614–1616, 1996.

- [419] S. Hiratsuka, S. Goel, W. S. Kamoun et al., "Endothelial focal adhesion kinase mediates cancer cell homing to discrete regions of the lungs via e-selectin up-regulation," *Proceedings of the National Academy of Sciences*, vol. 108, no. 9, pp. 3725–3730, 2011.
- [420] N. A. Memon, A. M. Mirza, and S. A. M. Gilani, "Segmentation of lungs from ct scan images for early diagnosis of lung cancer," *International Journal of Medical and Health Sciences*, vol. 2, no. 8, pp. 297–302, 2008.