INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 6, June 2026 | ISSN: 2456-4184 | IJNRD.ORG

Automated Lung Cancer Detection Systems: Design Approaches for Nodule Segmentation and Classification

Ms. Rupali Patil, Dr. Neha Jain, Dr. Atul Narkhede
PhD Scholar, Guide, Co-Guide
Pacific University

Abstract

Lung cancer remains one of the most lethal cancers globally, with early detection being critical for improving survival rates. Automated lung cancer detection systems, particularly those utilizing machine learning and deep learning, have emerged as promising tools for accurate and efficient diagnosis. These systems rely on segmentation to isolate lung nodules and classification to determine whether they are benign or malignant. Recent advancements in image processing, particularly through Convolutional Neural Networks (CNNs) and U-Net, have significantly improved detection accuracy. However, challenges persist, such as limited annotated datasets, generalization across diverse imaging conditions, and the high computational costs associated with deep learning models. Despite these hurdles, automated detection systems offer substantial potential in transforming lung cancer diagnosis by aiding radiologists, particularly in low-resource settings. Future research will focus on improving segmentation accuracy, developing models that can handle heterogeneous datasets, and addressing the challenges of false positives and negatives. These systems may also integrate multimodal approaches, combining CT scans with other imaging modalities for enhanced diagnostic capabilities. Overall, automated lung cancer detection systems hold great promise but require continued advancements to meet clinical and practical challenges.

Keywords: *Lung Cancer, Automated Detection, Nodule Segmentation, Deep Learning, CT scans, Classification*

1. Introduction

Lung cancer is among the top causes of global mortality. This is largely due the fact that the disease is often diagnosed only in its late stages. Because the prognosis worsens significantly at later stages, obtaining a diagnosis of lung cancer early can mean a difference of life or death. A person can even develop lung cancer despite appearing healthy and exhibiting no symptoms. Because of the high risk of lung cancer in the general population, lung cancer screenings should be done regularly to ensure early detection, as lung cancer is exceedingly difficult to identify at its earliest stages. High risk individuals such as smokers and individuals with family histories of lung cancer often present no symptoms prior to diagnosis. Currently, the only tools that can aid in the early diagnosis of lung cancer include imaging tools such as X-ray and CT scans. Although screening tools have a high degree of effectiveness, traditional screening methods have numerous limitations that can is likely to lead to erroneous diagnosis due to the large volume of cases and the extensive detail of the medical images. Because of this, expenses related to social services and health care increase.

Healthcare is now more efficient thanks to the improvements in machine learning. Improved systems could assist in the diagnosis of diseases such as cancer. Diagnosis could be enhanced and human error reduced by the use of automated systems in conjunction with digitized health services, as they would make services quicker and more economical. Cad X-rays, among other medical images, could be analyzed by fully automated systems more efficiently than the average human doctor, and even better than the average human doctor, and even better, they would be able to identify trends in the images of human doctors which could be utilized to assist in the diagnosis of lung cancer. Moreover, the use of fully automated systems could even lighten the burden of medical workers, particularly in regions with inadequate physician coverage. Advanced systems would deliver automated, high-priority notifications to the physicians, helping them prioritize the evaluation of high-risk patient cases, which would streamline and optimize the overall evaluation process and save valuable time.

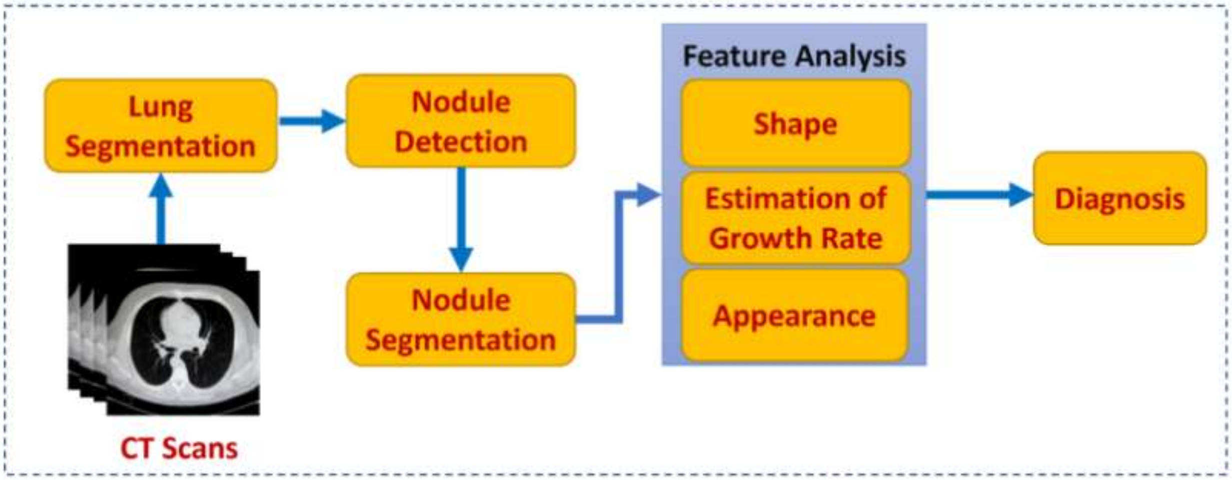

Figure Title: Block Diagram of an Automated Lung Cancer Detection System Using CT Images

An illustration is given of an automated process for detection of lung cancer. In processing CT images, the initial tasks are lung segmentation and tumor detection and segmentation. These scans are used by physicians for the assessment of lung cancer, and provide visual characteristics such as shape, growth rate, and other attributes. This review paper focuses on the automated detection of lung cancer, especially the processes of segmentation and clustering of lung nodules. Thus, the primary focus is to review the various methodologies for the segmentation of lung tumors in CT images and to analyze which are benign and which are malignant. Segmentation involves the removal of the tumour and the surrounding lung tissue. Classification involves describing the growth by its morphology. The objective of the study is to examine the evolution of segmentation and classification methods from basic machine learning to advanced methods. The study will also examine the challenges and issues related to segmentation and classification, such as the occurrence of false positives, insufficient computing power, and variations in data sets. The purpose of the study is to address the challenges and potential solutions in the increased automation of lung cancer diagnoses.

2. Background and Motivation

Traditional Methods of Lung Cancer Detection

Previously, locating lung cancer was done manually and relied on medical imaging, such as CT scans and chest X-rays. Radiologists would examine X-rays for abnormalities like masses and calcifications. Although advancements and steps have assisted physicians in making precise assessments, they remain flawed. Conversely, a physician's manual assessment can be time-consuming. There is a lot that radiologists need to pay attention to on the screen. This is very true when they don't have much time to work. A lot of scans are done every day in large hospitals and clinics. Because of the limited time, doctors do get tired and diagnoses are missing. In addition, the radiologist's skill is very important when it comes to figuring out what the tests mean. If the interpreter isn't very good, they might miss the signs of lung cancer, which would lead to a wrong diagnosis.

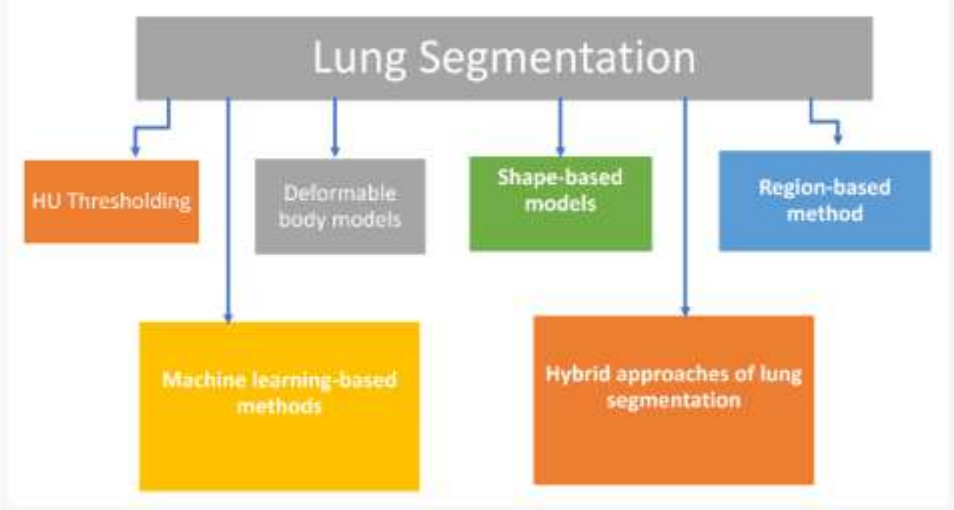

Figure 2. Main Categories of Lung Segmentation.

In medical image analysis, the lung segmentation techniques can be separated in the following main categories as presented in the Figure 2. These are HU thresholding for intensity separation, deformable models, and shape-based models, region growing for anatomical and spatial/region-based coherence, spatial region-based models, machine learning for data-based segmentation, and hybrid models that integrate multiple techniques for a more robust and accurate segmentation. Furthermore, the subjective interpretation of images can lead to inconsistency. Different radiologists may analyze the same imaging data in a variety of ways. Because of this, some may incorrectly identify a benign nodule as malignant, while others may entirely miss a malignant nodule. Another limitation is the possibility of false positives. These are the cases where benign nodules are mistakenly labeled as malignant, which drives unnecessary further investigations, carries the risks of over treatment, and increases the patients' worries. The traditional methods come with challenges that greatly decrease the effectiveness and the consistency with which lung cancer diagnosis is rendered.

Need for Automation

There is no doubt that with the challenges presented by traditional methods more automated systems are needed for lung cancer detection. Automation will help in the enhancement of lung cancer diagnosis and it will reduce the dependence on radiologists and the risks of human error. Automated systems are more efficient in the reviewing and evaluating of medical images in comparison to human specialists, and they will assist in the discovery of nodules and lesions that are often missed during a manual assessment.

There is an undeniable need to automate the detection of lung cancer due to the chronic understaffing of trained healthcare personnel, most importantly radiologists. Radiologists are one of the most underrepresented specialties in numerous resource-constrained countries and regions. They lack trained personnel to help interpret the medical imaging, which leads to delays in diagnosis and worse outcomes. Automating the interpretation of medical imaging will assist healthcare systems to improve the scope of work of radiologists to more sophisticated tasks and use automation to handle primary screenings and routine evaluations. The absence of qualified personnel or sophisticated medical imaging devices in developing countries and regions makes automation the most effective means of resolving inequities in healthcare. Such advanced equatos will improve the quality of diagnosis and medical imaging in regions that would otherwise lack the expertise of trained radiologists.

Advancements in Image Processing and Machine Learning

Recent advances in picture processing and machine learning could make it much easier to find lung cancer. Recent years have seen the improvement in the accuracy and efficiency of fully automated systems that find and classify lung cancer in its early stages. New techniques for processing lung cancer images have made big progress in feature extraction (finding out about the nodules' shape, size, and appearance) and segmentation (separating lung nodules from nearby tissues). To threshold and region-growing methods are added more complex algorithms like active contour models and edge recognition, which help CT scans show lung tumours more accurately.

Machine learning and deep learning have been very important in making automatic systems that can find lung cancer. Startups and study groups have used datasets with thousands to millions of X-rays, CT scans, and MRIs to teach different machine learning models. It has been taught that deep learning models, especially Convolutional Neural Networks (CNNs), can find and label lung tumours and other problems in medical pictures. Along with putting pictures into groups, machine learning models like Support Vector Machines (SVMs) and different grouping methods have been used with image processing to very accurately label lung tumours as either normal or cancerous. Classifier generalisability to other datasets and imaging methods has been a major area of study for making models that work well with a wide range of data sources. Recent studies using deep learning models have shown that they can find lung cancer very accurately, even when the picture quality, scanner type, or patient group changes. Lastly, advanced machine learning and picture processing techniques are being used together to make the identification of lung cancer automatic. In contrast to older methods, this is a significantly simpler approach. Particularly in areas with restricted healthcare data, these tools enhance the ability to evaluate, diagnose, and provide care to lung cancer patients. Consequently, this will lead to better patient outcomes.

2. Literature Review

Normal lung tissue looks less thick and has low Hounsfield Unit (HU) values on thorax CT studies compared to structures like the heart, blood vessels, and airways that are nearby. Using different methods and approaches, many studies have tried to find the best HU levels for separating the lung tissue into segments. The lung segmentation method that Hu et al. (2001) described has three main steps. In the first step, a certain greyscale level is used to separate the lung tissue[1]. A dynamic programming method is used to split the right and left lungs in the second step. As the last step, a structural operation is used to smooth out the edges of the lungs. The ideas in Ukil and Reinhardt (2009) and Van Rikxoort et al. (2009) were improved and built upon[2].

Amato et al. (2013) used grayscale thresholding two times: first, to separate the thoracic region from other chest structures, and second, to separate the lungs from the rest of the thoracic structures[3]. A 'rolling ball' algorithm was used to "soften" the lung boundary and retain juxta-pleural nodules, as well as to remove partial volume effects. In the same vein, Pu et al. (2009) used an adaptive border marching (ABM) algorithm to clean up the lung boundaries[4]. Using grayscale thresholding, Gao et al. (2018) utilized a four-step process to separate the pulmonary vessels and the airways from the lung parenchyma, as well as to distinguish between the right and left lungs[5]. More sophisticated techniques for determining thresholds, such as the histogram approach used by Sluimer et al. (2005) and the three-dimensional fuzzy adaptive thresholding proposed by Xu et al. (2006), have also been investigated[6], [7].

Table: Automated Lung Cancer Detection Using CT Images

Reference (APA short form)	Topic Name	Method Used	Key Findings
[8]	Deep Learning in Pulmonary Nodule Detection and Segmentation: A Systematic Review	Systematic review of CNNs, U-Net variants, 3D CNNs	Deep learning methods significantly outperform traditional approaches in detection accuracy; challenges include data imbalance, annotation cost, and generalization across datasets.
Archives of Computational Methods in Engineering, 2025[9]	Lung Cancer Detection Systems Applied to Medical Images: A State-of-the-Art Survey	Survey of CAD systems using ML and DL	Highlights evolution from handcrafted features to deep learning; identifies lack of clinical validation and explainability as major gaps.
Discover Oncology, 2025[10]	Multi-Objective Deep Learning for Lung Cancer Detection in CT Images	Multi-task CNN optimizing detection, classification, localization	Demonstrates improved classification and localization accuracy while reducing computational cost compared to single-task models.
[11]	Enhanced Pulmonary Nodule Detection with U-Net, YOLOv8, and Swin Transformer	Hybrid U-Net + YOLOv8 + Transformer architecture	Achieves superior nodule detection accuracy and robustness, especially for small and irregular nodules in CT scans.
BMC Cancer, 2025[12]	Deep Learning-Based Lung Cancer Classification of CT Images	CNN combined with Swin Transformer (DCSwinB)	Improves classification accuracy and feature representation over conventional CNN-based classifiers.
[13]	AI for Image-Based Lung Cancer Classification and Prognosis: A Meta-Analysis	Systematic review and meta-analysis	Confirms high diagnostic accuracy of AI models; emphasizes need for external validation and standardized evaluation metrics.
[14]	Lung Nodule Detection for CT-Guided Biopsy Images Using Deep Learning	Deep CNN for detection and ROI extraction	Improves nodule localization accuracy, supporting precise CT-guided biopsy procedures.
Intelligence-Based Medicine, 2024[15]	Improving Pulmonary Nodule Segmentation Using an Enhanced U-Net Model	Modified U-Net with attention mechanisms	Produces more accurate lung nodule boundaries and reduces over-segmentation compared to standard U-Net.
Expert Systems with Applications, 2024[16]	Deep Learning in Radiology for Lung Cancer Diagnostics: A Systematic Review	Review of CNNs, hybrid DL models, CAD systems	Reports consistent improvement in diagnostic performance but notes lack of interpretability and real-world deployment issues.
[17]	EMeRALDS: EMR-Driven Automated Lung Nodule Detection and Classification	Vision-Language Model (VLM)-assisted CAD pipeline	Integrates EMR data with CT images, enhancing classification accuracy and clinical decision support capability.

Table 1 outlines the studies focused on the automation of lung cancer detection for CT images, including areas of study, types of machine learning and deep learning techniques used, and major contributions. It offers a summary of the studies in relation to the recent progress, improvements in performance, and gaps in the literature.

3. LUNG NODULE SEGMENTATION TECHNIQUES

Lung tumour segmentation is a key part of the automatic methods used to find lung cancer. Correct division helps get the features of the lung tumour. These things are colour, shape, and size. Based on these characteristics, the machine can tell whether the tumours are normal or cancerous. If the segmentation is bad, the system will have a hard time finding the tumours and making a description that will help with the diagnosis. This makes it more likely to be inaccurate, wrong, and misdiagnosing. Automated systems that look for cancer often get it wrong, and they can rely on lung tumour segmentation to make a correct diagnosis. Systems can depend on lung tumour segmentation to make a correct diagnosis and cut down on fake positives and negatives.

Multiple segmentation methods exist for lung tumours, each with their own pros and cons. Otsu's Method is one standard example of segmentation. It computes the best threshold to separate the canters of images of nodules from background images of normal tissue. While this method is simple to compute, it is less frequently applied to more complicated situations, such as images of overlapping nodules or images of nodules that are poorly distinguished from the background tissue. Choosing a start point and expanding an area based on a pixel intensity criterion is another old way. To some extent, thresholding can be improved, but area growth depends a lot on the sowing and can lead to under or over segmentation if the settings are not set properly. Traditional segmentation methods also include edge-detection, which marks the edges of a region by changing the colour of the pixels in that area. Although the methods work well for finding the edges of clusters with clear edges, they don't work well for finding areas with fuzzy edges caused by a lot of pixel noise.

3.1 Machine Learning-Based Segmentation Approaches

The past few decades have processed multiple techniques for better segmentation of lung nodules with machine learning. Some of the supervised methods are K-means clustering and SVM-based segmentation. These methods see the segmented images as the background and the nodules as the two distinct components and assign a label to each. These methods, as a result, are reliant on a large quantity of segmentated images and fall short at the more sophisticated tasks of segmentation, such as spotting nodules at low contrast levels. Some of the most widely used un supervised implanted deep learning models, for instance CNN, have streamlined nodule segmentation to categorizing features at different levels of hierarchy. able to segment nodule images as a result of their self driven ways of categorizing nodule images and adapting to the different sizes, shapes, and appearances of nodules. U-Net, for example, is a model that has shown great success in medical image segmentation and lung nodule segmentation in particular as a result of its special architecture that successfully captures the high level and low level features that are needed for successful segmentation. More noteworthy is the fact that U-Net is a hybrid model since it combines its deep learning approaches with traditional processing methods.

In the context of nodule segmentation in the algorithm for lung cancer detection systems, over-segmentation occurs when the algorithm recognizes nodules and non-nodules regions, leading to false positive results. On the contrary, under-segmentation describes the case when nodules are missing or only partially segmented, leading to false negative results. Some issues are characterized by the variety of nodule sizes and shapes, and the presence of additional anatomical components that can mislead the algorithm. Robust methods which are able to address overlapping and irregular nodule cases are needed. Managing false positive and negative results continues to be one of the most important matters concerning the implementation of automated lung cancer detection systems.

4. INTEGRATION OF SEGMENTATION AND CLASSIFICATION

Pipeline of Automated Systems

The segmentation and integration classification methods in automatic systems for finding lung cancer make it easier for users to make accurate evaluations and good treatment choices. While the segmentation step is happening, lung tumours are surrounded by tissue to create a Region of Interest (ROI) that can be analysed. Sorting begins after nodules have been separated into groups using segments. By determining its size, shape, structure, and position, amongst other things, it is put into a category and labelled as either normal or cancerous. The identification of tumours in lung CT pictures, as is done at the very core of any automatic cancer detection system, cuts down on the need for doctors to read the films by hand. A detecting system's pipeline starts with picture preparation, which includes segmentation. Step-by-step instructions separate nodules, and then the information gathered from these separated nodules is sent to a classification model, which decides whether the information is safe or cancerous. When used in the ways described, this method speeds up the process of finding lung cancer while lowering the risk of human mistake.

The system depends on both steps working at the same time, though, because mistakes in classification can hurt its performance and cause fake positives and false negatives.

End-to-End Systems

New developments in building uniform frameworks for segmentation and classification are complete methods for finding lung cancer. Such systems are made with deep learning models and convolutional neural networks (CNNs). Both of these jobs can be trained on the same CNN at the same time. Using the same learnt features for both segmentation and classification, U-Net systems can do both tasks automatically. This situation uses a CNN model that is taught to both find the lung tumours on the CT scan (segmentation) and look at their traits to figure out whether they are cancerous or not (classification). With systems like this one, there is no need for middle steps, which is their main benefit. Overall, this makes the model easier to understand. It's not necessary to build and design features that split and group picture parts because these models learn clinically important names (like "benign," "malignant," etc.) that are linked to the pixels themselves. There is a lot of time pressure in clinical settings, so the systems' ability to quickly and efficiently handle big amounts of data will help doctors find lung cancer faster and more accurately.

Challenges in Integration

Integrating segmentation with classification, despite advancements in end-to end systems, continues to pose unique challenges. One of the most substantial of these challenges is the performance of segmentation in relation to classification, as the segmentation component will invariably produce suboptimal classifications. An example of this is the segmentation of a nodule that is either incomplete or inaccurate. In this case, there can be a classification flip, where a nodule that is clinically benign is incorrectly classified as malignant, and vice versa. Reliable and accurate segmentation can be difficult to achieve, particularly with small, overlapping, and/or irregularly shaped nodules. Fully automated systems also pose challenges, particularly with the performance speed of the system. In clinical scenarios, speed is of the essence and the radiologist requires a prompt answer. There is a need to combine both segmentation and classification in such a manner that there is an optimal balance between accuracy and fast processing. Deep learning models that can work on standard medical imaging system hardware without requiring any special processing needs, and without relying on cloud computing, should be easily accessible for integration. Lastly, while attempting to combine both steps, there is a need to consider the availability of computation resources. Although complex deep learning models are highly accurate, they also require a great amount of computing resources; this is especially the case for high res CT scans.

Creating systems that achieve high levels of accuracy while managing resource consumption is challenging. Consequently, the majority of contemporary systems are unable to provide optimal results for real-time applications, particularly in resource-constrained environments. To summarize, the merging of segmentation and classification into integrated, end-to-end automated systems to improve lung cancer detection proposes considerable promise, but the challenges of accuracy, real-time performance, and efficient levels of computation remain significant. These challenges will continue to prevail until there are notable improvements in deep learning, hardware, and systems optimization to make the systems more streamlined and available for clinical use.

5. LUNG NODULE CLASSIFICATION TECHNIQUES

Finding out if a lung tumour is cancerous or not is a key part of fully automated methods that find lung cancer. This labelling is an important part of customising a treatment plan because normal nodules don't need any action, but cancerous nodules might contain cancer and do need action. Patients don't have to go through treatments they don't need to, and this lowers their worry. It saves lives by correctly finding a cancer based on a tumour, which could be in an early state that can be treated. So, tumour labelling is a key part of figuring out if someone has lung cancer and how accurate that diagnosis is.

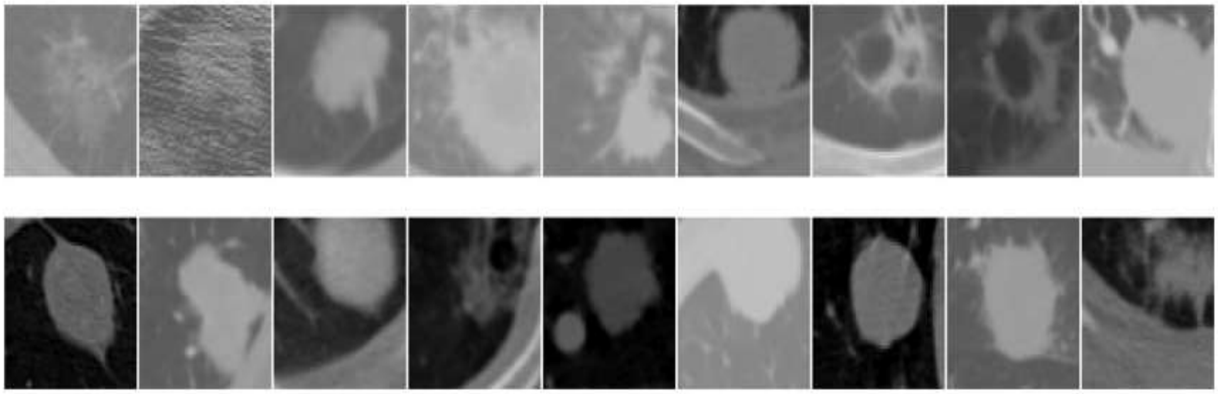

Figure 3. The example of lung nodules

Source: https://www.nature.com/articles/s41598-023-38350-z[18]

Figure 3 illustrates multiple CT scans demonstrating a range of pulmonary nodules. The images seem to indicate different stages and/or types of nodules, which from appearance could be either benign or malignant, thus demonstrating diversity in the form, dimensions, and composition. These attributes are essential for segmentation and classification in automated systems for pulmonary cancer detection.

Feature Extraction for Classification

Different characteristics are used in the classification of nodules extracted from CT scans. These characteristics are used to distinguish between nodules that are either malignant or benign.

- Shape features of the nodules include size, roundness, and eccentricity. Malignant or cancerous nodules have irregular sizes and shapes (more variety in shape and size), which means that benign nodules are more likely to have a consistent shape and size. Eccentricity is a characteristic of the nodule and its measure of elongation or how circular it is. Typically, malignant nodules are more irregular and have eccentric features.
- Several texture features are involved. These include the contrast, smoothness, and texture entropy. Malignant or cancerous nodules usually contain a contrast and a texture heterogeneity that is a lot higher. In comparison, benign nodules tend to have smooth and uniform textures. These traits differentiate the nodules and also improve their classification.

Machine Learning Models for Classification

Lung nodule classification is essential for the detection of lung cancer because it helps distinguish between normal and malignant nodules. Typically, machine learning algorithms such as Support Vector Machines (SVM) and Random Forests are used for this type of classification task. Based on the size, shape, and appearance of the nodule, SVM develops hyperplanes to determine which side is malignant and which side is normal. Random Forests produces multiple decision trees and combines them for improved classification. It can identify relationships between features which are not linear. This occurs when the features are not obscure and the data is not complex.

Deep learning has been employed for some of the most new and innovative approaches for classifying lung cancer over the last few years. Convolutional Neural Networks are some of the most commonly employed approaches for classifying lung cancer into different categories using deep learning approaches. These approaches are highly adept at extracting features directly from CT scans, and as a result, there is no requirement for feature extraction. Additionally, it is possible for these approaches to learn different features for classifying these scans as per the requirement for classifying the pictures themselves. Deep Belief Networks are another example of deep learning approaches for classifying pictures. They are made up of several layers of limited Boltzmann machines that are set up in a hierarchical way to describe the depth of the data. Some mixed models have been made that blend CNNs with Recurrent Neural Networks (RNNs) or Support Vector Machines (SVMs) to get better results. The good things about both deep learning and traditional machine learning are kept in these mixed models.

Challenges in Nodule Classification

Even the newest, most advanced machine learning models still have trouble putting lung tumours into groups. When tumours have traits that aren't clear, models can put them in the wrong category. The classification is wrong if the growth is classified wrong. Sometimes the models get it wrong, even if they are very sensitive and detailed. It takes a high level of sensitivity to find cancerous nodules and a high level of precision to avoid marking normal nodules as positive. It is hard to find this balance, especially when the lumps have traits that aren't clear. The model won't be able to generalize to different datasets even if it can mix sensitivity and precision. The information that a model is trained on determine how it can be used in the real world. If you train a model on a set of datasets, it will probably not work when you use it on a new set of datasets because the new datasets will have different picture quality, scanner types, and patient demographics. Different kinds of information are needed for the models to be taught, otherwise they won't work well in a wide range of clinical situations. In the real world, this is a very important thing for lung cancer detection tools to be able to do in order to correctly evaluate different datasets.

5. Integration of Segmentation and Classification

In lung cancer detection systems, the combination of segmentation and classification serves to optimize the performance of the system as a reliable diagnostic tool. The process starts with segmentation during which lung nodules are differentiated from surrounding tissues in the CT images. The subsequent stage of classification is dependent on the accuracy of segmentation. This is mainly because the system must first identify and isolate the needed features, which may include the size, form, and texture, among others, to ensure proper classification of the nodules. Once the system has completed the classification of all the nodules, it can then go to the final process known as segmentation, where in this case it aims to establish whether the nodules are benign or malignant in nature. The system can combine the processes in a step-by-step procedure to ensure that it maximizes the lung cancer identification system while minimizing any possibilities of human errors and the time it takes to come up with the diagnosis. Each process within the system is designed to build on the success achieved by the previous process, thus improving the process of segmentation and diagnosis.

End-to-End Systems

End-to-end systems that integrate segmentation and classification in a single system have gained popularity these days with the advent of advanced technologies in the domain of deep learning. These end-to-end systems allow the combination of both segmentation and classification in a single task, thereby avoiding the involvement of a human in the classification process after segmentation. A lot of research work has been done in the domain of deep learning with technologies like Convolutional Neural Networks to automate the process of segmentation as well as classification . These models have the capability to simultaneously learn the two tasks. End-to-end systems that utilize a single deep learning model for segmentation and classification significantly process the detection tasks more efficiently and accurately than traditional methods as they do not rely on separate algorithms for each task. This is especially useful when more data is made available to the system since it is able to self-tune based on the data that is provided.

There are some issues that might show up when segmentation and classification systems are put together. The most important thing is that both methods work. If there are problems with segmentation, classification doesn't work very well. Segmentation mistakes can happen and not be found. These mistakes can change how the ranking turns out. There is a chance that the rating results are wrong. In order for these methods to work, the findings can't be changed. The second worry is that results need to be given right away. When treating a patient, doctors often have to make quick clinical choices. For these kinds of systems to work, they need to be able to look at and understand pictures. The last main worry is the tools that are needed to get results in real time. For deep learning to work, your computer system needs to be better. Because of this, the device can only be used in certain places. It takes a lot of technical design and planning to find the right mix between these three main issues. More study needs to be done on how to combine segmentation and classification systems so that useful automatic systems can be made to find lung cancer. We don't just need these tools in theory; we need them right away.

Table 2: Summary of Techniques for Lung Nodule Segmentation and Classification

Stage	Techniques	Advantages	Challenges	Reference
Segmentation	Thresholding (Otsu's method)	- Simple and fast	- Over-segmentation or under-segmentation	[19]
	Region Growing	- Can handle some complex shapes	- Sensitive to initial seed and parameters	[20]
	Edge Detection	- Effective for detecting well-defined nodules	- Struggles with irregular or poorly defined nodules	[21]
	Deep Learning (U-Net, CNN)	- High accuracy, automatic feature extraction	- Requires large annotated datasets, computationally intensive	[22]
Classification	- Support Vector Machines (SVM)	- Strong performance in high-dimensional spaces	- Requires careful tuning and feature selection	[23]
	Random Forests	- Robust against overfitting	- May not handle very complex feature patterns	[24]
	Convolutional Neural Networks (CNN)	- End-to-end feature extraction and classification, deep learning	- Requires high computational power, need large datasets	[25]
	Hybrid Models (CNN + SVM)	- Combines the strengths of different models for better accuracy	- Complexity in training and tuning multiple models	[26]

Challenges in Integration	Error Propagation between Segmentation and Classification	- Integration ensures continuous data flow between stages	- Segmentation errors (over/under-segmentation) affect classification accuracy	[27]
	Real-time Performance	- Fast and efficient diagnostic tools in clinical practice	- Balancing speed with accuracy, especially in real-time settings	[28]
	Computational Efficiency	- Optimized systems for practical deployment in resource-constrained settings	- High computational requirements for deep learning-based systems	[29]

The techniques for segmentation and classification of lung nodules in automated lung cancer detection systems have been summarized in Table 2. It describes the pros and cons, as well as applicable citations, for traditional machine learning models, deep learning methods, and the integration barriers in clinical practice.

6. EVALUATION METRICS AND PERFORMANCE

Assessing how well newly developed automatic systems for finding lung cancer work requires using metrics. Using certain measures, you can rate how well a model can find lung tumours and tell whether they are normal or cancerous. Common measures include F1 score, area under the curve (AUC), positive predictive value (precision), negative predictive value (recall), and accuracy. By correctly identifying tumours as either normal or cancerous, the model's accuracy is a basic measure. Unfortunately, the accuracy measure by itself doesn't show how well the model works when most of the data is made up of harmless lumps. Detecting lung cancer is all about sensitivity, which measures how well the model can find cancerous tumours. The ability to correctly identify normal tumours, on the other hand, is measured by sensitivity. This gets rid of fake positive cases. By comparing the number of true positive cases to the positive cases defined by the model, accuracy helps you stop worrying too much about patients and doing too many treatments on them. Furthermore, and just as significantly, the sensitivity level needs to be high to make sure that the vast majority of cancerous tumours are found, even if some false positive cases are not followed up on.

The F1 Score tells you where precision and memory stand and lets you do a more thorough evaluation of both precision and recall if false positives and false negatives are more of a worry. Additionally, the area under the receiver operating characteristic (ROC) graph (AUC) shows how well the model can tell the difference between cancerous and noncancerous tumours in all of the different layers. For each category, the score that is closest to 1 is better; 1 is the best number for clearly separating categories. Comparing the results of various data-driven methods for finding lung cancer is necessary to figure out which ones work best. An overall better accuracy and sensitivity of deep learning techniques, especially Convolutional Neural Networks (CNNs), compared to methods like Support Vector Machines (SVMs) and Random Forests. Deep learning methods are best at finding nodules and figuring out what kind they are because they can learn which traits are most useful for finding and classifying nodules. Recent research has shown that mixed techniques that combine deep learning with traditional methods also work well. These techniques have the great virtue of getting high levels of accuracy and memory while also lowering the cost of processing. The models are still very hard to compute, and a lot of data is still needed to teach them.

Performance review has some problems, mainly with the automatic CT methods for finding lung cancer. Having inconsistent quality and size information is one problem. Creating machine learning models involves knowing and being good at a lot of different, sometimes strange ideas. As a result, not having enough good datasets can cause overfitting, which is when machine learning models work well on training data but not on real-world cases. Rating performance is also hard because automatic systems are compared to traditional ways of diagnosing problems. Although automated systems have some benefits, they will never be able to match the highly accurate diagnoses made by clinically expert radiologists. Although computerised systems are very efficient and can be scaled up easily, the most important problem that needs to be fixed is how different hospital environments can be in real life. Automated spotting systems will always need more varied datasets, stronger models, and more expert comparison testing.

7. CURRENT CHALLENGES AND LIMITATIONS

The setup and operation of automatic systems for finding lung cancer are affected by a number of problems and restrictions. Data clarity and availability is one of the main problems. A lot of well-annotated material is needed for machine learning models to work well. Not many of these datasets exist in the medical field, though, because manually annotating them takes a lot of time and money. Additionally, scans with different quality levels (for example, changes in pixels, colors, and noise) can affect how well the model works. The inability to apply generalizations to different datasets is another major problem. Unfortunately, machine learning models probably won't work well with data that isn't related to the CT scan methods, model, or type of patient that the model was taught on. When used in new scenarios, models may not work well because of differences in imaging, tumour types, and scan procedures. Failure to generalize in models shows how important it is to have image tools that are made for a variety of situations. Costly computers, especially those used for deep learning, are another big problem. Both the training and inference steps of these techniques use a lot of resources, so you need fast GPUs and processors that can handle using a lot of memory.

Advances in artificial intelligence (AI) and machine learning have made it possible to create systems that can handle tasks that were once hard to do, like keeping an eye on people who are at risk of getting lung cancer all the time. Real-time deployment of these kinds of tools is not possible, though. Despite efforts to create automated systems that track people who are at risk of getting lung cancer, the problem of lung cancer being misclassified still exists. These methods mistakenly label harmless bumps as cancerous ones. Numerous cancerous tumours are wrongly labelled as normal when they are not. Their actions have bad results because they delay care and make things worse for the patient. To the contrary, the mistake happens because some tumours are over-classified as cancerous. Over-classification leads to more investigations, treatments, and stress for patients, which makes the system more expensive. Altering the division and classification models is needed to make the system work better by computing instead of feeling the system.

8. APPLICATIONS AND FUTURE DIRECTIONS

Automated systems for finding lung cancer could be useful in many situations, especially in places where medical services aren't well developed and there aren't enough trained doctors and diagnostic specialists. It is possible for automated recognition systems to help with diagnosis in these situations. They could make a diagnosis quickly and correctly without creating extra work. People who work on systems for places with few resources often try to use less computing power while still keeping diagnostic accuracy so that the systems can be used on simple medical imaging devices in developing countries. In busy clinical situations, hospital information systems, medical picture files, and real-time tests need to be mixed with automatic monitoring systems in order to give the best care to patients. In this way, tools can be added to the doctors' work flow, which makes the clinical decision support process more efficient. Real-time monitoring tools in these clinical settings can help doctors focus on the most seriously ill patients, find lung cancer earlier, reduce the need for medical workers, give patients an accurate second opinion, and do all of this while lowering the total workload. Using tools that combine a number of different image methods can also be helpful. A combination of CT scans, MRI, and Positron Emission Tomography, for example, can make them more precise. There is a potential of gaining new insights into lung cancer with the help of a multimodal imaging system, which can solve issues of its classification, detection, among other challenges.

For example, the images obtained from a PET scan can reveal information regarding the metabolism of the body, and the images obtained from the MRI scan reveal more information in the soft tissues of the body, meaning it can provide a more accurate piece of information regarding the diagnosis of cancer in the tumour. There are a lot of key areas in the future that require additional research and consideration. In order to increase the accuracy of the classification, the key focus would be to improve the efficiency of the automatic systems in the following areas: dealing with issues of over-segmentation and under-segmentation, particularly in the clusters that are small, irregular, and meet at the corners of the image, and the handling of irregular clusters with complex shapes, specifically when the size of the clusters is relatively small compared to the image size in the dataset, and there is a limited availability of training samples, and the consideration of . Another key research aim is to produce models which are able to process different kinds of data. It is essential for medical data to be sourced from different origins with different scanning approaches, patient characteristics, and scanning scenarios. This implies different models are able to generalize to different kinds of datasets to be applicable on a larger scale. It will also be essential to examine the progression of deep learning over time, specifically with regard to the concept of transfer and unsupervised learning. With regards to transfer learning models, models which have already been trained on large datasets are applied to a smaller and more focused set. This implies models are applicable to different scenarios, and there is no need for additional labelled data. Unsponsored learning will be applicable to make predictions when labelled data is not available. In medical images, labelled data is not typically an issue.

By providing the computer with multi-class and multi-phase classification solutions for cancer staging, they will be able to detect not only lung cancer but also provide insight into its progression in a more complex manner.

CONCLUSION

The proposed two-part segmentation and classification has been of major help in making automatic systems for the identification of lung cancer better. The majority of research has been focused on dividing lung tumours into different parts and labelling them as either normal or cancerous. The latest research seems to focus on the change from older, more traditional methods like thresholding and edge recognition to more advanced deep learning models like CNNs and U-Net. The clearest edge up until now has come from ways that have been the most accurate and easiest to change. While switching to new datasets. Still, adding the two-part segmentation and classification model has made the biggest difference in the progress of automatic systems for finding lung cancer. The majority of research has been focused on dividing lung tumours into different parts and labelling them as either normal or cancerous. The latest research seems to focus on the change from older, more traditional methods like thresholding and edge recognition to more advanced deep learning models like CNNs and U-Net. Methods that have been the most accurate and easiest to change have clearly had the upper hand up until now. While switching to new datasets. It becomes even more crucial to conduct more advanced studies to make these models work effectively in areas like dataset variety and complexity and their running speed. Medical applications could drastically reduce wrong diagnoses and time consumption, which is really critical in many parts of the world where doctors are scarce. There are several technical and practical issues that have to be explicitly sorted out to reach a clinical level of acceptance. Computer-assisted detection of lung cancer may mark a revolution in many parts of the world with poor medical care.

References

[1] S. Hu, E. A. Hoffman, and J. M. Reinhardt, “Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images,” *IEEE Trans. Med. Imaging*, vol. 20, no. 6, pp. 490–498, Jun. 2001, doi: 10.1109/42.929615.

[2] S. Ukil and J. M. Reinhardt, “Anatomy-Guided Lung Lobe Segmentation in X-Ray CT Images,” *IEEE Trans. Med. Imaging*, vol. 28, no. 2, pp. 202–214, Feb. 2009, doi: 10.1109/TMI.2008.929101.

[3] F. Amato, A. López, E. M. Peña-Méndez, P. Vaňhara, A. Hampl, and J. Havel, “Artificial neural networks in medical diagnosis,” *J Appl Biomed*, vol. 11, no. 2, pp. 47–58, Jul. 2013, doi: 10.2478/v10136-012-0031-x.

[4] Jiantao Pu *et al.*, “Pulmonary Lobe Segmentation in CT Examinations Using Implicit Surface Fitting,” *IEEE Trans. Med. Imaging*, vol. 28, no. 12, pp. 1986–1996, Dec. 2009, doi: 10.1109/TMI.2009.2027117.

[5] M. Gao *et al.*, “Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks,” *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, no. 1, pp. 1–6, Jan. 2018, doi: 10.1080/21681163.2015.1124249.

[6] I. Sluimer, M. Prokop, and B. Van Ginneken, “Toward automated segmentation of the pathological lung in CT,” *IEEE Trans. Med. Imaging*, vol. 24, no. 8, pp. 1025–1038, Aug. 2005, doi: 10.1109/TMI.2005.851757.

[7] Y. Xu, E. J. R. Van Beek, Y. Hwanjo, J. Guo, G. McLennan, and E. A. Hoffman, “Computer-aided Classification of Interstitial Lung Diseases Via MDCT: 3D Adaptive Multiple Feature Method (3D AMFM),” *Academic Radiology*, vol. 13, no. 8, pp. 969–978, Aug. 2006, doi: 10.1016/j.acra.2006.04.017.

[8] C. Gao *et al.*, “Deep learning in pulmonary nodule detection and segmentation: a systematic review,” *Eur Radiol*, vol. 35, no. 1, pp. 255–266, Jul. 2024, doi: 10.1007/s00330-024-10907-0.

[9] S. L. Tan, G. Selvachandran, R. Paramesran, and W. Ding, “Lung Cancer Detection Systems Applied to Medical Images: A State-of-the-Art Survey,” *Arch Computat Methods Eng*, vol. 32, no. 1, pp. 343–380, Jan. 2025, doi: 10.1007/s11831-024-10141-3.

[10] A. F. Abdulqader *et al.*, “Multi-objective deep learning for lung cancer detection in CT images: enhancements in tumor classification, localization, and diagnostic efficiency,” *Discov Onc*, vol. 16, no. 1, p. 529, Apr. 2025, doi: 10.1007/s12672-025-02314-8.

[11] X. Wang *et al.*, “Enhanced pulmonary nodule detection with U-Net, YOLOv8, and swin transformer,” *BMC Med Imaging*, vol. 25, no. 1, p. 247, Jul. 2025, doi: 10.1186/s12880-025-01784-0.

[12] B. Ar, “Deep Learning-based Lung Cancer Classification of CT Images using Augmented Convolutional Neural Networks,” *ELCVIA*, vol. 21, no. 1, Sep. 2022, doi: 10.5565/rev/elcvia.1490.

[13] X. Yuan *et al.*, “Systematic review and meta-analysis of artificial intelligence for image-based lung cancer classification and prognostic evaluation,” *npj Precis. Onc.*, vol. 9, no. 1, p. 300, Aug. 2025, doi: 10.1038/s41698-025-01095-1.

[14] B. Prashanthi and S. P. A. Claret, “Lung Nodule Detection For CT-Guided Biopsy Images Using Deep Learning,” *JAETS*, vol. 5, no. 2, pp. 909–924, Jun. 2024, doi: 10.37385/jaets.v5i2.3716.

[15] A. Sadremomtaz and M. Zadnorouzi, “Improving the quality of pulmonary nodules segmentation using the new proposed U-Net neural network,” *Intelligence-Based Medicine*, vol. 10, p. 100166, 2024, doi: 10.1016/j.ibmed.2024.100166.

[16] A. Atmakuru *et al.*, “Deep learning in radiology for lung cancer diagnostics: A systematic review of classification, segmentation, and predictive modeling techniques,” *Expert Systems with Applications*, vol. 255, p. 124665, Dec. 2024, doi: 10.1016/j.eswa.2024.124665.

[17] H. Eman, F. Shaukat, M. H. Zafar, and S. M. Anwar, “EMeRALDS: Electronic Medical Record Driven Automated Lung Nodule Detection and Classification in Thoracic CT Images,” 2025, *arXiv*. doi: 10.48550/ARXIV.2509.11714.

[18] H. Wang, “A diagnostic classification of lung nodules using multiple-scale residual network,” p. 2023.

[19] X. Yang, X. Shen, J. Long, and H. Chen, “An Improved Median-based Otsu Image Thresholding Algorithm,” *AASRI Procedia*, vol. 3, pp. 468–473, 2012, doi: 10.1016/j.aasri.2012.11.074.

[20] S. A. Hojjatoleslami and J. Kittler, “Region growing: a new approach,” *IEEE Trans. on Image Process.*, vol. 7, no. 7, pp. 1079–1084, Jul. 1998, doi: 10.1109/83.701170.

[21] G. M. H. Amer and A. M. Abushaala, “Edge detection methods,” in *2015 2nd World Symposium on Web Applications and Networking (WSWAN)*, Sousse, Tunisia: IEEE, Mar. 2015, pp. 1–7. doi: 10.1109/WSWAN.2015.7210349.

[22] D. Gertsvolf, M. Horvat, D. Aslam, A. Khademi, and U. Berardi, “A U-net convolutional neural network deep learning model application for identification of energy loss in infrared thermographic images,” *Applied Energy*, vol. 360, p. 122696, Apr. 2024, doi: 10.1016/j.apenergy.2024.122696.

[23] M. O. Oyediran, O. S. Ojo, I. A. Raji, A. E. Adeniyi, and O. J. Aroba, “An optimized support vector machine for lung cancer classification system,” *Front. Oncol.*, vol. 14, p. 1408199, Dec. 2024, doi: 10.3389/fonc.2024.1408199.

[24] L. C, P. S, A. H. Kashyap, A. Rahaman, S. Niranjan, and V. Niranjan, "Novel Biomarker Prediction for Lung Cancer Using Random Forest Classifiers," *Cancer Inform*, vol. 22, p. 11769351231167992, Jan. 2023, doi: 10.1177/11769351231167992.

[25] D. Jain, P. Singh, A. K. Pandey, M. Singh, H. Singh, and A. Singh, "Lung Cancer Detection Using Convolutional Neural Network," in *2022 3rd International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, Ghaziabad, India: IEEE, Nov. 2022, pp. 1–4. doi: 10.1109/ICICT55121.2022.10064513.

[26] S. Mehta and S. Choudhary Chitkara, "AI-Powered Lung Cancer Detection: CNN-SVM Hybrid Model with Statistical Insights," in *2025 International Conference on Electronics and Renewable Systems (ICEARS)*, Tuticorin, India: IEEE, Feb. 2025, pp. 1651–1655. doi: 10.1109/ICEARS64219.2025.10940698.

[27] J. Garcia Manrrique, L. Rada Mota, and J. Ruiz, "Systematic Review of The Challenges in Implementing Data Segmentation with Machine Learning," in *Proceedings of the 23rd LACCEI International Multi-Conference for Engineering, Education and Technology (LACCEI): "Engineering, Artificial Intelligence, and Sustainable Technologies in service of society,"* Latin American and Caribbean Consortium of Engineering Institutions, 2025. doi: 10.18687/LACCEI2025.1.1.1779.

[28] M. F. Arshad *et al.*, "Artificial Intelligence and Machine Learning in Lung Cancer: Advances in Imaging, Detection, and Prognosis," *Cancers*, vol. 17, no. 24, p. 3985, Dec. 2025, doi: 10.3390/cancers17243985.

[29] R. Jain, P. Singh, M. Abdelkader, and W. Boulila, "Efficient lung cancer detection using computational intelligence and ensemble learning," *PLoS ONE*, vol. 19, no. 9, p. e0310882, Sep. 2024, doi: 10.1371/journal.pone.0310882.