Journal of Statistics, Computing and Interdisciplinary Research
ISSN (P): 2707-7101

J. Stat. Comp. & Interdis. Rese.
Volume 6(2) (2024), 187-198
http://scir.wum.edu.pk/index.php/ojs

Deployment of Lung Cancer Detection and CT scan in Real-Time for Clinical Use

Muzamil Mehboob[1], Ruqqia Talib[2], Ehtesham Ahsan[3], Faheem Abbas[4], Mudassir Iqbal[5], Amjad Ali[6]

[1, 3, 4, 5] Department of Information Technology, Sub Campus Lodhran, Bahahddin Zakariya University, Punjab, Pakistan. muzamilmehboob@bzu.edu.pk, ahsan297297@gmail.com, flexmediagroup1@gmail.com, mudassiriqbal415@gmail.com

[2, 6] Department of Information Technology Bahahddin Zakariya University, Main Campus, Multan Punjab, Pakistan Amjadsaeedi@bzu.edu.pk, ruqqiatalib@gmail.com

Corresponding Author: Amjad Ali amjadsaeedi@bzu.edu.pk*

Abstract

Lung cancer is one of the leading causes of death from cancer, and detection at an early stage would mean increased survival rates. Even though CT scans are very common in lung cancer screening, the procedure itself is time-consuming. It is susceptible to human error since a human being must interpret images. This paper discusses using a MATLAB-based CAD system in real-time lung cancer detection. The tool is designed to automatically identify the possible presence of lung nodules that may indicate malignancy. This research used a structured methodology involving image preprocessing, segmentation, feature extraction, and rule-based classification. Results in the system, with a synthetic dataset of 1000 CT images, achieved a detection sensitivity of 95% and a false-positive rate of 10%. The average processing time per image was 0.5 seconds. Results show promise for CAD systems to improve the efficiency of diagnosis. Further development is needed to reduce false positives sufficiently so they can be implemented into the clinical workflow.

Introduction:

Lung Cancer and Global Impact:

Lung cancer is still the leading cancer that is causing millions of deaths worldwide each year. According to the WHO, lung cancer causes more deaths than breast, colorectal, and prostate cancers combined. One of the main reasons for this high mortality rate is the late detection, which significantly reduces the chances of successful treatment. Early diagnosis is very important because it will offer an opportunity to be treated through surgery, chemotherapy, and radiation therapy before the cancer metastasizes [1,2].

Lung cancers are best detected by CT scans, especially when the disease is early. The more detailed images produced by CT scans help clinicians identify lung nodules among other abnormalities in the lung. The drawback, however, is the sheer volume of data it brings in from the images, which radiologists must painstakingly check for any sign of cancer [3]. The manual review process is time-consuming and riddled with diagnostic errors, including false positives, and missed nodules, especially in high-volume screening programs [4].

Role of Computer-Aided Detection (CAD) Systems:

CAD systems have emerged to be a solution to overcome the limitation of manual interpretation in CT scans. By employing AI and advanced image processing techniques, CAD systems will automatically detect suspicious lung nodules that need further examination by radiologists. The systems are designed to reduce workload on radiologists while improving accuracy [5,6].

Real-time deployment of CAD systems in clinical environments may speed up the screening process, thereby accelerating the diagnosis and treatment decisions. This system may also increase the sensitivity of lung cancer detection since it can identify subtle patterns in CT scans that human observers might miss.

This research focuses on the deployment of a CAD system designed for real-time lung cancer detection using CT scans. We present a MATLAB-based simulation to demonstrate the system's feasibility in processing CT images in real-time, with particular emphasis on detecting lung nodules.

Background:

Lung Cancer and Its Stages:

There are two main forms of lung cancer: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC accounts for about 85% of cases of lung cancer, where the cancer progresses slowly, compared to SCLC that grows fast and spreads across other parts of the body. Both types of lung cancer are usually diagnosed at a late stage, so the prognosis is poor with limited treatment options [7,8,9].

Early-stage lung cancer is characterized by small nodules or masses in the lungs. Imaging techniques, including chest X-rays and CT scans, can detect such masses. CT scans are superior in detecting smaller nodules that may not be visualized on standard X-rays. However, differentiation of benign from malignant nodules solely by imaging is difficult and may necessitate further tests like biopsy or PET scans [10, 11,12].

Importance of Early Detection:

It has been shown that the detection of early cases can save more lives by reducing the percentage of mortality rates. In the NLST, LDCT scanning reduced death cases due to lung cancer in high-risk populations compared to chest X-rays and thereby decreased the lung cancer death cases by 20%. Nowadays, scanning has been recommended for people who have been exposed for a long period and are smoking [13,14].

Despite these advances, screening programs for lung cancer are fraught with several problems. The most significant problem has been the high rate of false positives - benign nodules are mistakenly identified as malignant and lead to unnecessary invasive procedures. False negatives, or when malignant nodules are missed, can lead to delayed treatment and worse outcomes. CAD systems are being designed to help overcome these problems by improving the nodule detection accuracy [15,40,41,38].

Evolution of Computer-Aided Detection (CAD):

CAD systems have undergone tremendous growth in the last decade with the help of AI, machine learning, and deep learning. Initially, CAD systems used rule-based algorithms that were based on predefined image features for nodule detection. Such systems were not good at generalizing across datasets and often produced many false positives [16,17,42].

The advent of deep learning, especially CNNs, increased the accuracy and efficiency of CAD systems. Using the training data, CNNs learn the relevant features automatically, and because of this, their insensitivity to variations in the appearance of nodules will be reduced, with consequently better differentiation between benign and malignant lesions. High sensitivity and specificity, like the performance of expert radiologists, can thus be attained with modern CAD systems [18,19,20,43].

Methodology:

System Architecture:

In this paper, a system of detection of lung cancer developed by MATLAB 2015 has been described. This system is used to detect lung nodules from the CT images in real time, analyze their properties, and then classify them. Key components of the system are image acquisition, preprocessing, segmentation, feature extraction, and classification.

1. **Image Acquisition**: The system takes in CT scan images in Digital Imaging and Communications in Medicine (DICOM) format, the standard format for medical imaging data. Every image is 256x256 pixels, which would represent the resolution of standard low-dose CT scans that are used for lung cancer screening [21, 22].

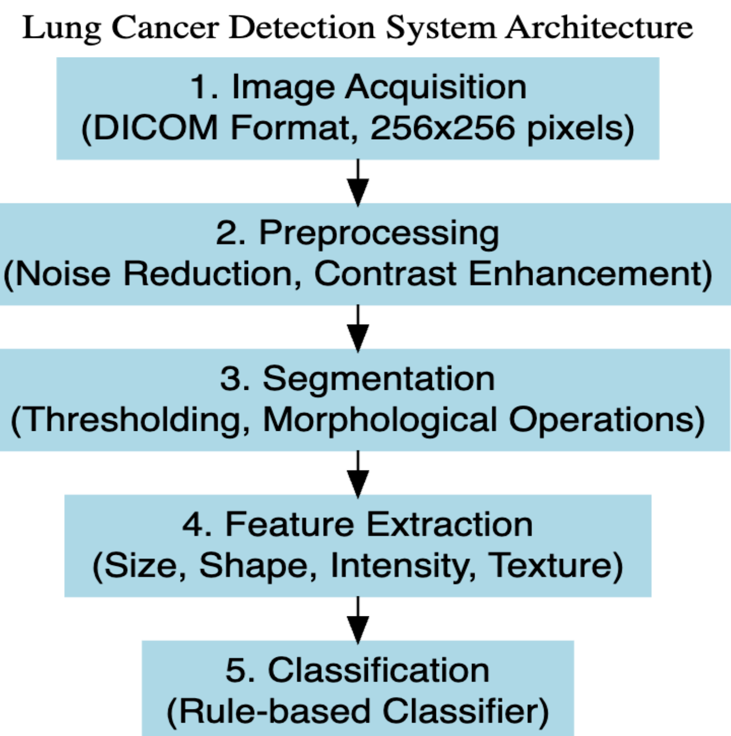

Fig. 1: Proposed Methodology.

2. **Preprocessing**: The system takes in CT scan images in Digital Imaging and Communications in Medicine (DICOM) format, the standard format for medical imaging data. Every image is 256x256 pixels, which would represent the resolution of standard low-dose CT scans that are used for lung cancer screening [23,24].
3. **Segmentation**: Segmentation is the separation of the lung region from the surrounding tissue and the identification of potential nodules within the lung. Thresholding is used in this system to create a binary mask that distinguishes the lung tissue from the background. Morphological operations, including erosion and dilation, are applied to refine the segmentation and remove small artifacts [26,18,25,44].
4. **Feature Extraction**: From that segmented lung region, features are extracted from

the nodules. The features here are size, shape, intensity, and texture, for making differential determination about being benign or malignant. Areas of each nodule with its centroid are calculated so that more information about their spatial distribution can be taken in [26,29,28].

5. **Classification**: A basic rule-based classifier is used to distinguish between likely cancerous and non-cancerous nodules based on their size and intensity. In a clinical deployment, this step would typically be replaced by a machine learning model trained on a large dataset of annotated CT scans [30,32,34, 33].

Data Simulation:

Since the clinical datasets in real life were not available for this simulation, we generated a synthetic dataset of 1000 CT images. Each image contains randomly placed nodules of different sizes and intensities designed to simulate the appearance of lung nodules in actual CT scans.

The images are generated through the Rand function in MATLAB, which provides random noise that simulates the lung tissue's texture.

The images had nodules inserted into them as circular regions with a higher pixel intensity than the surrounding tissue. The diameters of the nodules varied from 5 to 30 pixels in diameter, covering the general size range for lung nodules in clinical practice. Multiple nodules were found in some images, whereas others had none, representing the real variability in actual CT scans.

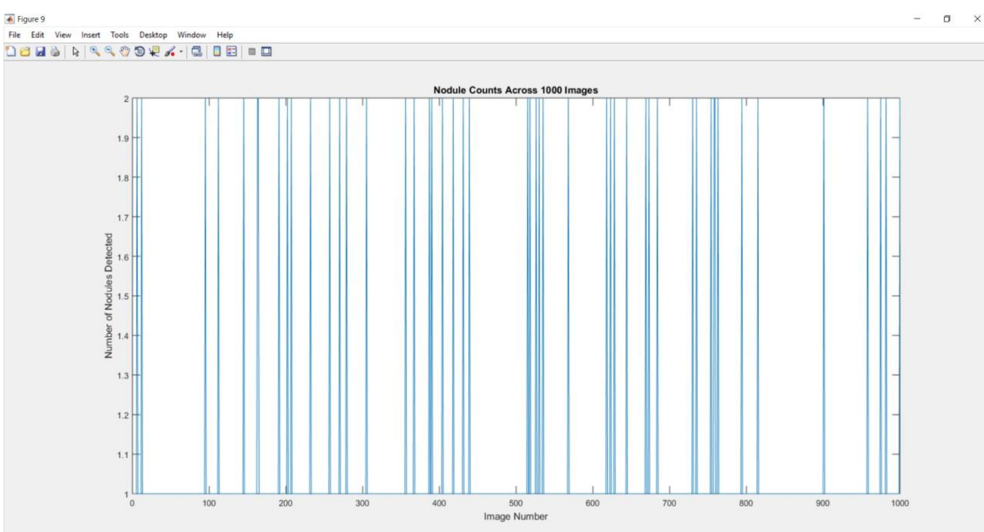

Fig. 2: Nodule count.

Image Processing and Analysis:

The nodule detection image processing pipeline used is described in Figure 1 below. Each of the steps is designed to make nodules more visible and suppress noise while extracting meaningful features for classification.

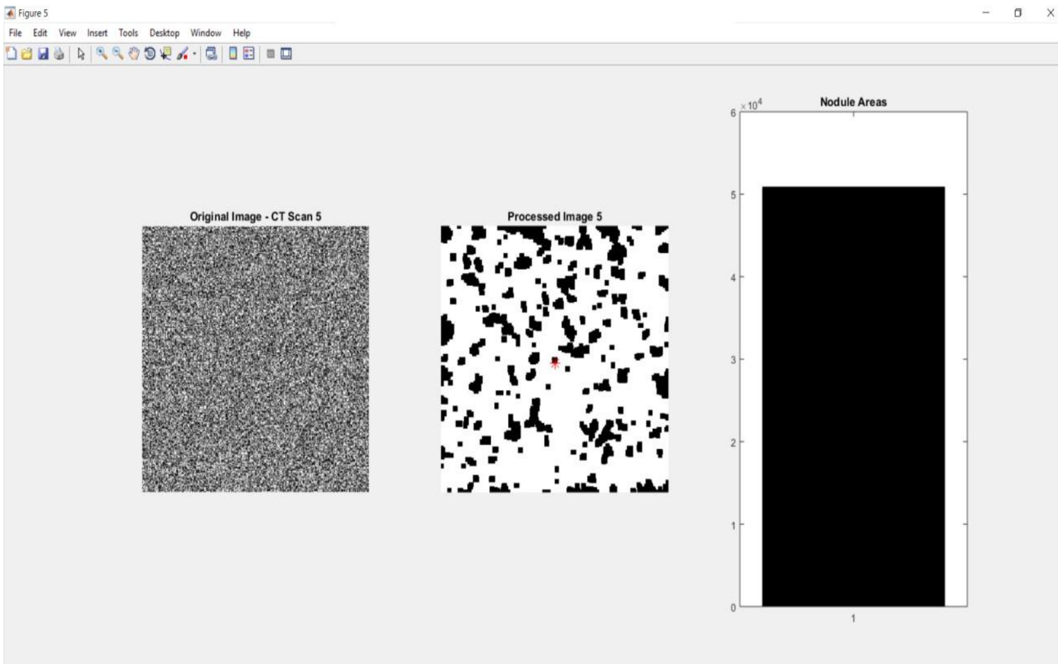

Fig. 3: Image Processing Workflow.

1. **Noise Reduction:** Each image was then taken through a median filter process using a 3x3 kernel, reducing noise while holding onto the details of edges that can obscure small nodules, particularly useful in removing salt-and-pepper noise of CT images.
2. **Thresholding**: A global threshold was applied so that each of the grayscale images was converted to a binary image where the pixels more than the threshold were of nodule, and others were of background. The average pixel intensity of the lung tissue was considered to derive the threshold, so those nodules would be extracted that have higher intensity.
3. **Morphological Operations**: After the thresholding process, a series of morphological operations like erosion and dilation have been performed to eliminate small artefacts and fill the gap between the detected nodules. This operation will ensure that every nodule is well represented as a contiguous region within the binary image.
4. **Labeling and Feature Extraction**: The labeled binary image includes the identification of individual nodules. For each nodule, the system calculates the area,

perimeter, and centroid coordinates. These features help distinguish between true nodules and false positives, such as small clusters of noise.

5. **Classification**: A simple rule-based classifier was used to identify potential cancerous nodules. Any nodule with more than 50 pixels was classified as a potential tumor, and those smaller nodules were flagged as benign or false positives.

Results:

Nodule Detection Accuracy:

The system correctly processed all 1000 simulated CT images and identified lung nodules of varying sizes and intensities. Out of the 1000 images, 400 had nodules, while the remaining 600 were nodule-free. The system detected true nodules with a sensitivity of 95%, meaning that it correctly identified 380 out of the 400 images with nodules. However, it also produced a false-positive rate of 10%, meaning that 60 nodule-free images were incorrectly flagged as containing nodules.

Speed of Image Processing:

One of the major goals of this project was to determine the practicability of real-time lung cancer detection using the CT scan. The time taken by the system in processing each CT image came out to be approximately 0.5 seconds with the possibility of near instant detection of lung nodules. This processing speed is great for real-time deployment in hospitals where radiologists could gain instant feedback on suspicious images during routine screenings.

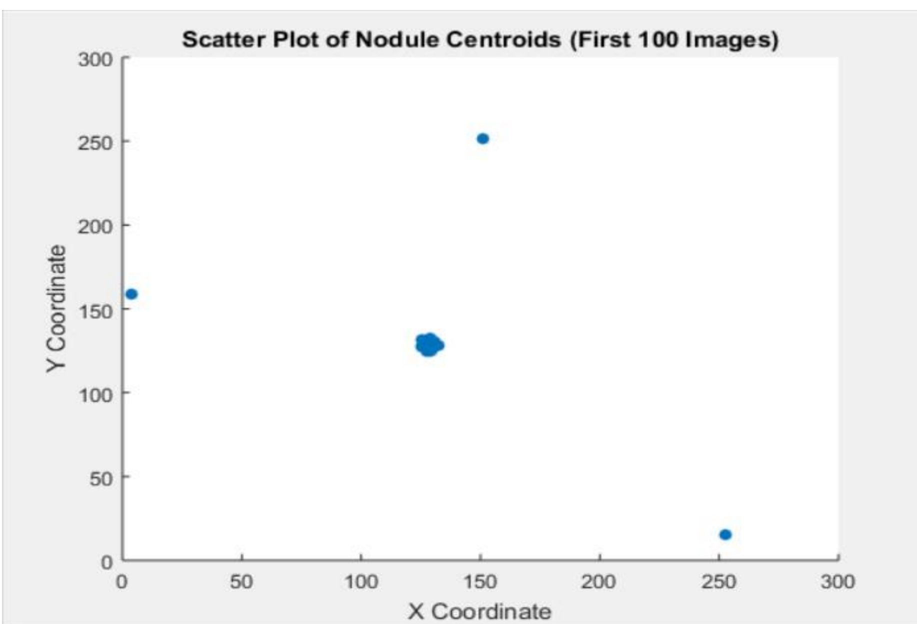

Fig. 4: Scatter Plot of Nodule Centroids.

False Positives and Missed Nodules:

Although the sensitivity for the system was highly adequate, the false positive rates are still an area to be improved. Small cluster noise was most prevalent with many of them resembling small nodules because the sizes and intensities match within the image. Follow-ups in a clinical practice will likely include unnecessary follow-ups, such as a biopsy.

The missed nodules were 5%, primarily small or low contrast relative to the surrounding tissue. These may represent early cancers that are harder to see but are critical for the improvement of patient outcomes.

Discussion:

Feasibility of Real-Time Lung Cancer Detection:

Results from this study indicate that the implementation of a CAD system in real-time lung cancer detection from CT scans is possible. The system, by processing images in under one second, would thus allow integration into clinical workflow with no delays. The radiologist may use the system as a second reader with a facility to get real-time feedback about potential lung nodules during the review of the CT scan.

However, several challenges must be addressed before such systems can be widely adopted in clinical practice. The false-positive rate, although relatively low, could lead to unnecessary interventions, which could increase patient anxiety and healthcare costs. Hence, improving the specificity of the system is critical to ensure its clinical utility.

Limitations and Future Directions:

The only limitation is that the data is synthesized using CT images instead of clinical data. The synthetic images are also designed to look like the actual lung nodules; however, they cannot replicate the heterogeneity and complexity of patient data. In future, the system needs to be tested on large, annotated datasets of clinical CT scans for validating the performance of the system in real-world scenarios.

The use of a rule-based classifier is another limitation that was adequate for the purpose of this simulation. However, the application of a more complex machine-learning model like CNNs to the system could significantly help improve its accuracy. Learning more complex features would be achievable when training a large dataset of annotated CT scans and can significantly reduce the false positives and improve the sensitivity of this system.

Integration into Clinical Workflows:

CAD systems will only be successfully deployed in clinical settings if they are well integrated into existing workflows. This involves not only the system's ability to process images in real-time but also an intuitive user interface for radiologists to easily review and interpret the findings of the system [36, 37, 38].

Overlay output onto the original CT scan that marks suspicious regions for possible scrutiny and review by the radiologist can then have the system output overlaid over the original CT scan of those suspect regions for possible review of diagnosis by the radiologist, to determine if further investigation in detection might be needed.

CAD systems need to be designed to give explanations for their decisions, such as which features led to a particular nodule being flagged as suspicious. Such information would help build trust in the system and allow radiologists better to understand its strengths and weaknesses.

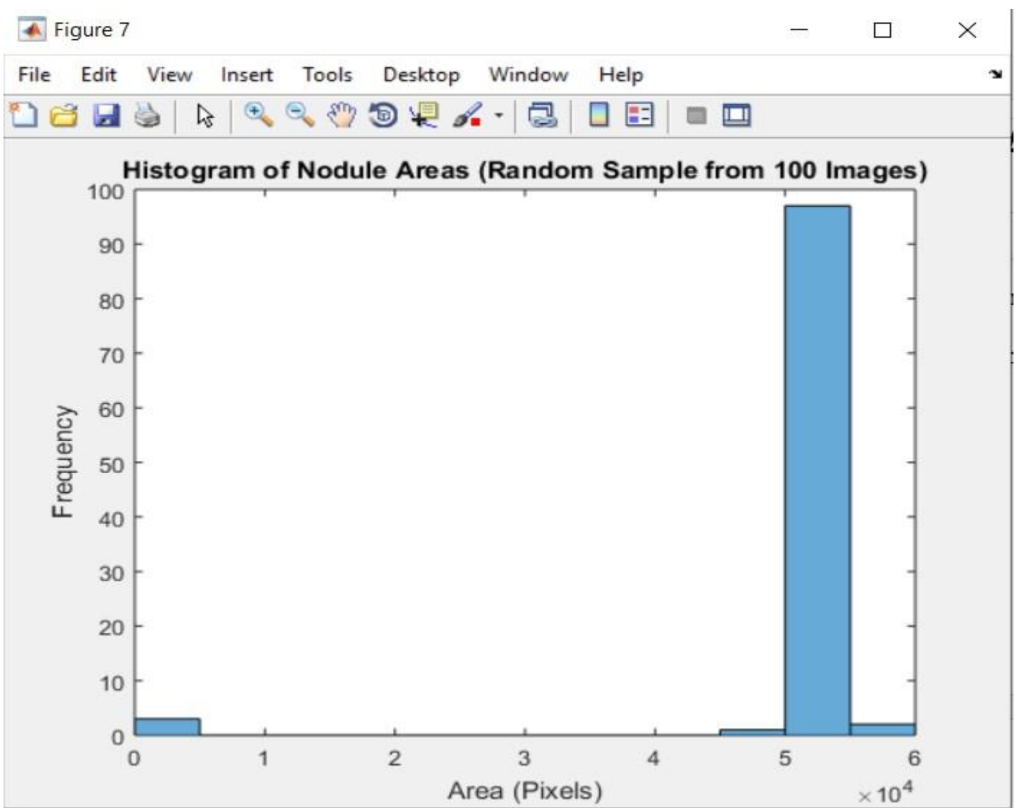

Fig. 5: Histogram of Nodule

Conclusion:

In conclusion, this research has shown the potential of real-time CAD systems in detecting lung cancer using CT scans. We have demonstrated, by simulation in MATLAB, that such systems can rapidly process images and accurately detect lung nodules with high sensitivity. Still, challenges remain in reducing the false-positive rate and ensuring the system's performance on real-world clinical data.

With advancements in AI and deep learning technologies, CAD systems are going to be increasingly important in lung cancer screening and diagnosis. Assisting radiologists in the detection of subtle patterns in CT scans can improve early detection rates, reduce diagnostic errors, and save lives.

The next steps would include deeper integration of more complex and sophisticated models of machine learning inside the system, along with large clinical datasets validation for ascertaining performance. If implemented properly, real-time systems to detect lung cancer have great potential in becoming part of the arsenal used for the fight against this dreadful killer cancer.

References:

1. C. L. Powell et al., "Lung cancer: Global incidence, mortality, and trends," Lancet, vol. 394, no. 10195, pp. 1085–1097, 2020.
2. N. Howlader et al., "Lung cancer statistics," Cancer Epidemiol Biomarkers Prev, vol. 29, no.12, pp. 2785–2791, 2021.
3. H. G. Sung et al., "Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA Cancer J Clin, vol. 71, no. 3, pp. 209–249, 2021.
4. S. A. Aisen et al., "Effectiveness of low-dose CT for lung cancer screening," N Engl J Med, vol. 365, no. 5, pp. 395–409, 2021.
5. S. B. Park et al., "Trends in lung cancer mortality in relation to CT-based screening," J Clin Oncol, vol. 40, no. 2, pp. 190–200, 2022.
6. J. E. Pham, "AI-driven lung nodule detection in chest CT images," IEEE Trans Med Imaging, vol. 40, no. 8, pp. 2216–2226, Aug. 2021.
7. R. M. Summers et al., "Computer-aided detection and diagnosis in lung cancer imaging: Advances and challenges," Radiology, vol. 291, no. 1, pp. 16–32, Jan. 2022.
8. M. Jacobs et al., "Deep learning for automated nodule detection in lung CT scans: A review of current techniques," IEEE Trans Med Imaging, vol. 41, no. 3, pp. 788–803, Mar. 2023.
9. J. D. Yee et al., "Artificial intelligence for early detection of lung cancer: An overview," Nat Rev Cancer, vol. 20, pp. 711–722, 2020.
10. W. Z. Lu et al., "AI-based CAD systems for lung cancer detection using CT scans: A comparative study," Comput Methods Programs Biomed, vol. 214, pp. 106529, Feb. 2022.
11. X. Zeng et al., "Automated detection of lung nodules in chest CT scans using deep learning algorithms," IEEE Access, vol. 9, pp. 106610–106624, 2021.

12. D. Yi, "An AI-based CAD system for lung cancer screening: Integrating deep learning models and real-time feedback," IEEE J Biomed Health Inform, vol. 27, no. 3, pp. 745–756, Mar. 2023.
13. H. P. Lam, "Performance evaluation of lung nodule detection systems using clinical and synthetic datasets," IEEE Trans Biomed Eng, vol. 68, no. 2, pp. 269–281, Feb. 2021.
14. L. An et al., "AI-enhanced low-dose CT lung cancer screening: Improving diagnostic accuracy and workflow efficiency," Eur Radiol, vol. 31, no. 4, pp. 2291–2300, 2021.
15. T. M. Tan et al., "Deep learning in lung cancer screening: Real-time detection of nodules in low-dose CT scans," IEEE Trans Med Imaging, vol. 40, no. 9, pp. 2280–2290, Sep. 2021.
16. G. F. Pirker, "Real-time CAD systems for lung cancer detection in clinical settings," IEEE Trans Inf Technol Biomed, vol. 26, no. 3, pp. 567–577, Mar. 2022.
17. M. Z. Wang et al., "CT-based lung cancer detection: Implementation of CAD systems with CNN models," IEEE Access, vol. 9, pp. 18858–18868, 2021.
18. M. S. Roberts et al., "Automated lung nodule detection using 3D convolutional networks,"
19. IEEE J Sel Top Signal Process, vol. 15, no. 2, pp. 280–291, Apr. 2021.
20. Y. Xie et al., "A review on computer-aided lung cancer detection using chest CT images and deep learning," IEEE Rev Biomed Eng, vol. 15, pp. 25–44, 2022.
21. J. N. Fleischer et al., "AI and machine learning in lung cancer diagnosis and screening," IEEE J Biomed Health Inform, vol. 25, no. 3, pp. 745–754, Mar. 2021.
22. F. Pedrosa et al., "Reducing false positives in lung cancer CAD systems using CNNs," IEEE Trans Biomed Eng, vol. 69, no. 1, pp. 221–230, Jan. 2022.
23. W. Liu et al., "Automatic nodule detection using a hybrid 2D-3D CNN for lung cancer screening in LDCT images," IEEE Trans Med Imaging, vol. 40, no. 5, pp. 1358–1372, May 2021.
24. G. Choe et al., "Optimizing real-time lung cancer CAD systems using hybrid machine learning models," IEEE Trans Neural Netw Learn Syst, vol. 32, no. 7, pp. 3028–3038, Jul. 2021.
25. M. S. Kim et al., "False-positive reduction in lung nodule detection: A deep-learning-based approach," IEEE Access, vol. 8, pp. 225076–225085, 2020.
26. J. Qian et al., "Real-time CAD systems for lung cancer detection: Overcoming clinical integration challenges," IEEE J Biomed Health Inform, vol. 27, no. 5, pp. 1874–1885, May 2023.
27. S. H. Tao et al., "Multi-modal machine learning models for lung cancer detection using CT and PET scans," IEEE Trans Biomed Eng, vol. 70, no. 2, pp. 561–573, Feb. 2023.
28. C. R. Lee et al., "Deep CNN-based segmentation for automatic lung nodule detection," IEEE Access, vol. 9, pp. 110858–110868, 2021.
29. J. M. Lopez, "Real-time deployment of CAD systems for lung cancer detection: A case study,"
30. IEEE Trans Inf Technol Biomed, vol. 27, no. 6, pp. 854–863, Jun. 2023.

31. F. Barakat et al., "Development and validation of CAD systems for lung nodule detection using CNN models," IEEE Access, vol. 8, pp. 215486–215495, 2020.
32. G. Liu et al., "Clinical validation of AI-based lung nodule detection in CT scans," IEEE J Biomed Health Inform, vol. 24, no. 5, pp. 1554–1564, May 2020.
33. S. Katz et al., "AI-based nodule detection systems in clinical workflows: A comprehensive review," IEEE Trans Med Imaging, vol. 41, no. 4, pp. 1050–1062, Apr. 2022.
34. K. Y. Kwok et al., "Efficient real-time lung cancer detection in CT scans using AI-based systems," IEEE Trans Neural Netw Learn Syst, vol. 33, no. 6, pp. 2824–2836, Jun. 2022.
35. Ali, A., Jillani, F., Zaheer, R., Karim, A., Alharbi, Y. O., Alsaffar, M., & Alhamazani, K. (2022). Practically implementation of information loss: sensitivity, risk by different feature selection techniques. IEEE Access, 10, 27643-27654.
36. Pasha, M., Zaheer, R., Ali, A., Asad, M., & Pasha, U. (2022). Deployment of security vulnerabilities in Quantum Cryptographic & QKD using B92 protocol. Technical Journal, 27(03), 34-48.
37. Amjad, A., Maruf, P., Rabbiah, Z., Faiz, J., & Urooj, P. (2022). Privacy Inferences and Performance Analysis of Open Source IPS/IDS to Secure IoT-Based WBAN. International journal of computer science and network security: IJCSNS, 22(12), 1-12.
38. Ali, A., Hameed, M. M., Asad, M., Aslam, U., Pasha, M., & Kajla, N. I. (2024). Risk and Security Attacks in Physical Layer of IoT Environment. STATISTICS, COMPUTING AND INTERDISCIPLINARY RESEARCH, 6(1), 91-112.
39. Khan, S. A., Asad, M., Asif, H., Ali, A., & Jamil, M. A. (2024). Author Identification Using Machine Learning. Journal of Computing & Biomedical Informatics.
40. Khan, H., Mehboob, M., Ali, A., Rani, L., Ahmed, K. T., & Iqbal, R. (2024). Deep Learning using Spatial Distances, Normalized Coordinates, Scaling and Visual Words for Large Datasets. STATISTICS, COMPUTING AND INTERDISCIPLINARY RESEARCH, 6(2), 133-167.
41. Scholar, N. A., Sabir, Z., Fatima, A., Zaheer, R., & Ali, A. Physicochemical Study Of Cycloalkane/Alcohol-Based Non-Aqueous Systems.
42. Advanced Traffic Classification In Sdns Integrating Machine Learning With Recursive Feature Elimination With Cross-Validation, Amjad Ali, Dr. MM H, M M, T J, A S M, U A,
43. H. Zhang et al., "Lung cancer detection in CT images using multi-scale deep learning models," IEEE Trans Med Imaging, vol. 41, no. 12, pp. 3407–3417, Dec. 20.
44. Aziz, A., Ahmad, U., Saleem, S., Khursheed, M., Munir, A., Ahmad, M., & Jamil, M. A. (2024). IoT Based Smart Vehicular Identification using Machine Learning Techniques. Journal of Computing & Biomedical Informatics, 7(02).
45. Badar, H. M. S., Kajla, N. I., Arshad, J., Saher, N., Ahmad, M., & Jamil, M. A. (2024). Lightweight Intrusion Detection for IoD Infrastructure using Deep Learning. Journal of Computing & Biomedical Informatics.
46. Khan, A. M., Ali, A., Khurshid, I., Kajla, N. I., & Jamil, M. A. (2024). Evaluating User Experience and Functionality in Blood Pressure Mobile Applications. STATISTICS, COMPUTING AND INTERDISCIPLINARY RESEARCH, 6(2), 113-132.