# An Improved Computer Aided System for Lung Cancer Detection using Image Processing Techniques

Manoj Mhaske[1(✉)], Ramesh Manza[1], Pallavi Pradhan[1], and Kavita Khobragade[2]

[1] Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India
mhaskemanoj@gmail.com

[2] Department of Computer Science, Fergusson College, Pune, Maharashtra, India

**Abstract.** Early detection and prevention is the only way to treat lung cancer to avoid the loss of life. Where Computed Tomography (CT) screening is viewed as perhaps the best technique for discovering the early indications of lung malignant growth. The primary goal of this study is nodule detection and classification of collected CT scans images as benign or malignant. Sometimes some human errors can occur in the checking of a long series of CT slices of a single patient manually. This automated system (CAD-Computer Aided System) can help to radiologist or doctors to know the current stages and condition of the disease to diagnose correctly and quickly on a single click which will useful for radiologists and doctors to avoid the serious disease stage. The key four processes of our proposed system are input CT images, pre-processing, features extraction, and classification. In the proposed approach firstly we read all the CT image database (70 thoracic lung CT scans) having Dicom format then applied some pre-processing techniques of Matlab to enhance the image quality and obtained texture features. Using texture features, we extracted several features. At the end, we classified the dataset as benign or malignant using the K-means clustering method, and we achieved an accuracy of 92.8 percent.

**Keywords:** ROI · K-means clustering · CAD-computer aided system · Lung · Nodules

## 1 Introduction

The body text starts with a standard first-level heading like INTRODUCTION or any other heading suitable to the content and context. First level headings are in all caps. Copy the content and replace it for other first-level headings in remaining text. Reference citations should be within square bracket [1]. Headings should always be followed by text. Lung cancer is diagnosed in at least 12 million patients per year. In 2018, cancer will be the primary cause of about 9.6 million deaths. Lung cancer is the second most frequent cause of cancer in comparison to other types of lung cancers. Nearly 1.76 million of the 2.09 million cases of lung cancer reported by the World Health Organization


R. Manza et al. (Eds.): ACVAIT 2022, AISR 176, pp. 4–13, 2023.
https://doi.org/10.2991/978-94-6463-196-8_2

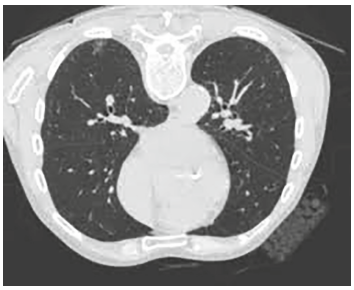


**Fig. 1.** Lung CT Image

are directly attributable to the disease [1]. The second most common lung disease is persistent malignancy, which affects one in five men and one in nine women. Coughing, coughing up blood, chest pain, weight loss, and shortness of breath are among the most typical lung cancer symptoms [2]. Cancerous cells are eliminated during the production cycle, and red blood cells degrade. Internally, they change the plasma membrane's shape and make-up, which results in a rise in RBCs and a shortening and ultimately rupture of the veins and arteries' walls. Nowadays, CT (computed tomography) is frequently employed in the clinical diagnosis of lung cancer. An unregulated proliferation of lung cells is known as a pulmonary nodule, which appears as a circular formation with a diameter of 3 to 30 mm on lung CT images. The formation of malignant lung tissue comes in two different forms. The first is small cell lung cancer (SCLC), while the second is non-small cell lung cancer (NSCLC).

Lung disease malignancies found around one out of five in men and one out of nine in women and it is the second most regular malignant growth. Malignant and benign are the two structures where tumors comes into. Benign tumors are not dangerous, in this manner they don't develop and spread to the degree of destructive tumors. Benign tumors are normally not perilous or life threatening. The Malignant or cancerous tumors are Harmful tumors, which growth can develop and spread to different parts or regions of the human body. Travelling of the disease cells from the underlying tumor site to different pieces of the body is called as metastasis [3] (Fig. 1).

## 2 Literature survey

Number of researchers carried out the research on development of lung cancer detection system. In [4], Sayani Nandy and Nikita Pandey put forth a plan for identifying cancer cells in lung CT scan pictures. This research offers a technique for extracting the majority of cancer cells from a CT scan image. In order to identify diseases, Prof. Samir Kumar B. created a system using computer-aided diagnosis (CAD) to extract edges from lung CT scan pictures [5]. Thresholding algorithm [6] offers filtering to notice the sputum cell from the raw image for early detection via Fatm Taher et al. M. Tan et.al [7] proposed CADe (Computer-Aided Detection) system in their work to classify nodules or non-nodules by means of genetic algorithms and Artificial Neural Network; with total of 360 nodules of 3–30 mm in diameter of 134 sufferers enrolled in LIDC society. This CADe

system had a sensitivity of 87.5 percent and four FPs (false positives) per scan. In [8], 420 CT scans of 420 individuals were randomly selected from the LIDC database, and the potential malignant nodules were identified by SVM classifier. These scans had 3–30 mm in diameter and 379 possible malignancies. This system achieved a segmentation stage accuracy of 97 percent, a CADe system sensitivity of 94.4 percent with 7.04 FP per scan, and a classification stage sensitivity of 93.9 percent with 7.21 FP each case.

A computerized intelligent method for nodule detection and lung cancer classification in CT images was created by Amjed et al. [9]. They applied morphological image processing methods and geometrical facets in their study. Watershed transform and image mapping were used in [10] by C. Panyindee and W. Chiracharit to identify lung nodules present in PET/CT data. Few studies combined multiple classification techniques with feature extraction from lung CT images using Haralick texture features [10, 11]. For texture quantification, Balaji et al. [11] employed a selective scale-based picture filtration. When classifying the tumors in CT images, Mir Rayat et al. employed a backward search algorithm and the Chi-square distance metric to identify important features [12]. A CAD machine was invented by Tidke et al. for the early detection of lung cancer nodules using chest computer tomography scans. The grey level co-occurrence matrix was used to extract textural components from the lung nodules [13]. With the use of the Sobel component identification methodology, Pandy et al. established a novel method of malignant cell detection from lung CT scan images [14] in those works. Sudha et al. segmented the lung area in their proposed system with use of thresholding and morphological operations [15]. Ada et al. estimated and detected lung cancer survival the usage of neural community classifiers [16]. For the pre-processing of the images, they employed histogram equalization in this. The early stages of the patient's condition are examined using feature extraction techniques and neural network classifiers to determine if it is normal or abnormal. Using a Gabor filter and smart system, Sankar et al. [17] enhanced the structure for lung cancer cell detection. A CAD method was created by Silva et al. [18] to find lung nodules. 33 exams were subjected to the application of SVM as a classification approach. Their suggested strategy has a 95.21 percent accuracy rate.

## 3 Methodology

The system described here is composed of four fundamental steps. The collection of lung CT scans from the database is the initial phase, which is followed by a quantitative analysis of the images. We used pre-processing methods in Matlab in the second step to improve the quality and clarity of the original lung CT images. These methods included binarizing all CT images, separating the lung parenchyma mask, and binarizing the nodule candidates again.

Finally, we finalized the region of interest to segment the image and obtain the lung nodule candidates. The third step includes feature extraction to produce numerical features. The fourth phase uses K-means clustering to categorize diseases as benign or malignant. Figure 2 shows schematic steps of proposed algorithm of this system.

The database for this study taken from The Cancer Imaging Archive (TCIA), which was sponsored by SPIE, NCI/NIH, University of AAPM, and Chicago. We employed the SPIE-AAPM CT challenge dataset for this study. A total of 70 CT image data from

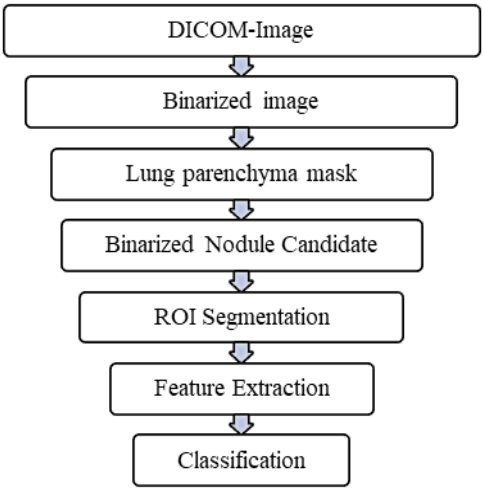



**Fig. 2.** Flow of Methods

patients are included in the data collection. Out of 70 CT scans of the chest, 10 are used for training and 60 are used for testing [19, 20]. These CT scans include DICOM images with 512 × 512 pixel size.

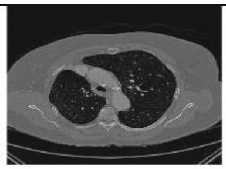
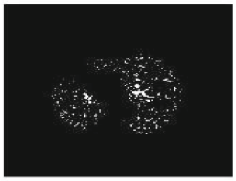
## 4 Result

Based on data from the database for the lung RAD system, the False Positive reduction is initialized as a consequence of the SPIE-AAPM lung CT challenge [19]. We read the original CT pictures using the given lung CT image database, binarized the input images, and then obtained the lung parenchyma mask. Then, lung ROI candidates are obtained by applying the nodule candidate's mask to the initial CT image (shown in Table 1).

This study focuses on the intensity and texture qualities of an image. Mean, variance, and skewness are the intensity feature parameters used in this work, whereas contrast, correlation, energy, and entropy are the texture feature parameters. In addition to these metrics, smoothness and kurtosis properties are also extracted [21, 22]. The segmented lung nodule is used for feature extraction when segmentation is finished. A feature could be a fact that is extracted from image and improves our understanding of it. In this work, the highlights, such as geometric and statistical features based on intensity, are retrieved. Physical dimensions known as shape measurements are used to describe how an object appears. All of these features are produced using the Co-occurrence matrix, which shows when different features occur together [23, 24].

Haralick et al. [25] suggested 14 texture quantities generated from the GLCM matrices for the texture evaluation. These measurements speak to the different dark level types that are connected to images smoothness, consistency, heterogeneity, and distinction. We suggest using ten of the 14 features in total, shown in Table 2, and independent features tests are shown in Table 3.

**Table 1.** Preprocessing & Image Segmentation

| Sr. No. | Image Processing Used | Output Image |
|---|---|---|
| 1. | Original Lung CT Image |  |
| 2. | Binarized Image using Initial Threshold |  |
| 3. | Lung Parenchyma Mask |  |
| 4. | Lung Nodule Candidates |  |
| 5. | Segmented ROIs |  |

The prosed system uses the K-means clustering method to divide a larger dataset into smaller groups. Unsupervised learning is a method utilized in the K-means algorithm, which is used for classification. It is known as unsupervised categorization since the system automatically categorizes items based on user-defined criteria. We used the K-means clustering technique for image segmentation, followed by morphological filtering for lung nodule detection from lung CT scans [26].

**Table 2.** Features Extraction

| Sr. No. | Type of Feature | Equation |
|---|---|---|
| 1. | **Contrast:** Measures the local fluctuations in the GLCM, in contrast It determines the difference in intensity between an image element and its neighbour. | $\sum_{i,j} \lvert i-j\rvert^2 p(i,j)$ |
| 2. | **Correlation:** It calculates the correlation between the joint probabilities of the required image element pair occurrences. | $\sum_{i,j} \frac{(\mathrm{i}-\mu_i)(\mathrm{j}-\mu_j)\mathrm{p(i,j)}}{\sigma_1\sigma_2}$ |
| 3. | **Energy:** The GLCM's total square components are provided by energy. Additionally, it is known as homogeneity or the angular moment. | $\sum_{i,j} \mathrm{p}(i-j)^2$ |
| 4. | **Mean:** For a random variable vector A made up of N scalar observations, it is used to find the average or mean value of the array. | $\sum_{i=0}^{G-1} ip(i)$ |
| 5. | **Entropy:** Entropy shows the dissimilarity in the image or ROI. Entropy and energy are inversely related because of this. | $\sum_{i} p(i) log_2(p(i))$ |
| 6. | **Variance:** It gives the variance of the A items along the first array dimension for which the size is not 1. | $\sum_{i=0}^{G-1} (i-\mu)^2 p(i)$ |
| 7. | **Smoothness:** Smoothness is a measure of relative smoothness of intensity in a region. | $R = -1\frac{1}{1+\sigma_2}$ |
| 8. | **Kurtosis:** Kurtosis measures how prone a distribution is to outliers. | $\sigma^{-4} \sum_{i=0}^{G-1} (i-\mu)^4 p(i) - 3$ |
| 9. | **Skewness:** Simply put, skewness is a measurement of how asymmetric the given data is in relation to the sample mean. | $\sigma^{-3} \sum_{i=0}^{G-1} (i-\mu)^3 p(i)$ |
| 10. | **IDM**: his is the homogeneity at the local level. It increases when the inverse grey level co-occurrence matrix is large and the local grey level is uniform. | $\sum_{i,j} \frac{1}{1+(i-j)^2} p(i,j)$ |

The steps below have been used for K-means.

1. Assign the k value to the number of clusters and select the k-cluster centers at random.
2. Determine the cluster's mean or center.
3. Next, determine the separation between each pixel and the center of each cluster.
4. If the cluster is close to the center, move there; otherwise, go on to the next cluster.
5. Re-evaluate the center and continue until it stops moving (Fig. 3 and Table 4).

The trial outcomes and discourse offer the Sensitivity, Specificity, and Accuracy rates attained in the suggested structure after the efficient characterization [27]. The degree of precisely identified negatives is measured by specificity. Sensitivity is often referred to as the true positive rate or the recall rate in various professions. It determines the degree to which real positives are correctly recognized. Accuracy used to represent the ratio of

**Table 3.** Extracted Sample Features

| Feature Extracted | Sample Values | | | | |
|---|---|---|---|---|---|
| | **Img1** | **Img2** | **Img3** | **Img4** | **Img5** |
| Contrast | 0.5042 | 0.3042 | 0.3679 | 0.4883 | 0.4405 |
| Correlation | 0.1032 | 0.1581 | 0.1231 | 0.1329 | 0.1704 |
| Energy | 0.8839 | 0.8015 | 0.8318 | 0.8870 | 0.8835 |
| Mean | 0.0059 | 0.0041 | 0.0041 | 0.0058 | 0.0063 |
| Entropy | 1.8331 | 2.9567 | 2.4440 | 1.3889 | 1.5578 |
| Variance | 0.0081 | 0.0080 | 0.0081 | 0.0081 | 0.0081 |
| Smoothness | 0.9565 | 0.9385 | 0.9390 | 0.9559 | 0.9589 |
| Kurtosis | 47.337 | 12.393 | 27.279 | 47.378 | 39.148 |
| Skewness | 4.4665 | 1.1712 | 2.2718 | 4.3207 | 3.3964 |
| IDM | 6.5730 | 0.6491 | 1.4005 | 3.7505 | 3.6540 |

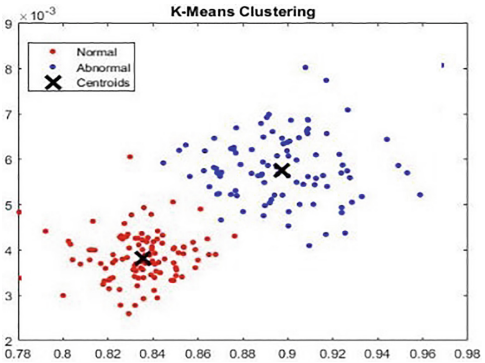



**Fig. 3.** K-Means Clustering

**Table 4.** Proposed System Result

| True Positive | True Negative | False Positive | False Negative |
|---|---|---|---|
| 32 | 33 | 3 | 2 |

the CT images that are classified correctly. We got overall 92.8% accuracy in proposed system.

## 5 Conclusion

This suggested CADe framework can speed up the analysis of lung cancerous growth and reduce human error when segmenting and categorizing lung nodules in CT images. This paper's main goal is to categorize the lung CT pictures as normal or abnormal. Radiologists and specialists can avoid the real illness with the help of the proposed technique. For this work, we used the TCIA lung CT scan image database. The texture features are retrieved after pre-purposing. K-means clustering is utilized to classify CT scans as normal or abnormal, and the proposed framework has a great accuracy of 92.8 percent.

**Acknowledgments.** Thanks to TCIA for providing the lung CT image database and also thanks to Dr. R. R. Manza for supervision and Department of CS & IT Dr. Babasaheb Ambedkar Marathwada University, Aurangabad (MS) India for providing the support.

**Authors' Contributions.** The main Conceptualization is held by R.M. and M.M., Methodology and framework processed by M.M. and P.P., Database collection, analysis and investigation done by K.K., P.P. and M.M., R.M. did the supervision, Writing-original draft, writing-review and editing is done by M.M.

## References

1. Adjei, A. A. (2019). Lung cancer worldwide. Journal of Thoracic Oncology, 14(6), 956.
2. Collins, L. G., Haines, C., Perkel, R., & Enck, R. E. Lung cancer: diagnosis and management. American family physician, 75(1), 56-63, 2007.
3. Elmar Rendon-Gonzalez, Volodymyr Ponomaryov, Automatic Lung Nodule Segmentation and Classification in CT Images Based on SVM, IEEE 978-1-5090-2267, 24 June 2016.
4. Sayani Nandy, Nikita Pandey, A Novel Approach of Cancerous Cells Detection from Lungs CT Scan Images', International Journal of Advanced Research in Computer Science and Software Engineering Volume 2, Issue 8, August 2012.
5. Prof. Samir Kumar Bandyopadhyay, Edge Detection from Ct Images of Lung, International Journal of Engineering Science & Advanced Technology, Volume - 2, Issue - 1, 34 – 37, 2012.
6. FatmTaher, Naoufel Werghi and Hussain Al-Ahmad, Extraction of Sputum Cells using Thresholding Techniques for Lung Cancer Detection, International Conference on Innovations in Information Technology, 2012.
7. M. Tan, R. Deklerck, B. Jansen, M. Bister, J. Cornilis, A novel computer aided lung nodule detection system for CT images, Medical Physics, October 2011, doi: https://doi.org/10.1118/1.3633941 [PubMed Open Access].
8. 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, Journal of Negative Results, 2016.
9. Amjed S. Al-Fehoum, Eslam B. Jaber, Mohammed A. Al-Jarrah, Automated detection of lung cancer using statistical and morphological image processing techniques, Journal of Biomedical Graphics and Computing, 4(2), pp. 33 - 42, 2014.
10. C. Panyindee and W. Chiracharit, Detection of Lung Tumors in Chest PET/CT Images by using Watershed Transform and Image Mapping, The 3rd International Symposium on Biomedical Engineering (ISBME 2008), p. 9–12, 2008.

11. Balaji Ganeshan, Sandra Abaleke, Rupert C.D. Young, Christopher R Chatwin and Kenneth A. Miles, Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage, Cancer Imaging, 10(1), pp. 137-143, 2010.
12. Mir Rayat Imtiaz Hossain, Imran Ahmed, Md. Hasanul Kabir, Automatic Lung Cancer Detection using GLCM features, Asian Conference on Computer Vision ACCV14, Singapore, 2014.
13. Ms. Swati P. Tidke, Prof. Vrishali A. Chakkarwar, Classification of Lung Tumor Using SVM, International Journal of Computational Engineering Research (ijceronline.com) Vol. 2 Issue. 5 September 2012.
14. Nikita Pandey, Sayani Nandy, A Novel Approach of Cancerous Cells Detection from Lungs CT scan Images, International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 8, August 2012.
15. Sudha. V, Jayashree. P, Lung Nodule Detection in CT Images Using Thresholding and Morphological Operations, International Journal of Emerging Science and Engineering (IJESE) ISSN: 2319–6378, Volume-1, Issue-2, December 2012.
16. Ada, Rajneet Kaur, Early Detection and Prediction of Lung Cancer Survival using Neural Network Classifier, IJAIEM Volume 2, Issue 6, June 2013.
17. K. Sankar, Dr. M. Prabhakaran, An Improved Architecture for Lung Cancer Cell Identification Using Gabor Filter and Intelligence System, International Journal of Engineering and Science (IJES) 2013.
18. Nunes, I. and Da Silva, H.S., Artificial neural networks: a practical course, Springer, 2018.
19. 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, Comput. Biol. Med., vol. 42, no. 11, pp. 1098–1109, 2012.
20. Public Database Open Access: https://public.cancerimaingachive.net/ncia/searcMain.jsf;jssionid=128369143ADF4D948378665461898B03, Accessed on 16 Apr 2019.
21. Ciurte and S. Nedevschi, Texture analysis within contrast enhanced abdominal CT images, 2009 IEEE 5th International Conference on Intelligent Computer Communication and Processing, Cluj-Napoca, 2009, pp. 73-78.
22. Manoj M. Mhaske, Ramesh R. Manza and Pallavi K. Pradhan, Detection of Lung Cancer Using SVM with Lung Nodule Segmentation, PENSEE International Journal, ISSN: 0031-4773, VOLUME 51, ISSUE 7 2021.
23. Sawakare, S., & Chaudhari, D., Classification of brain tumor using discrete wavelet transform, principal component analysis and probabilistic neural network, Int J Res Emerg Sci Technol, 1(6), 13-9, 2014.
24. Gonzalez, R. C., Woods, R. E., & Eddins, S. L., Digital image processing using MATLAB, Pearson Education India, 2004.
25. Haralick, R. M., Shanmugam, K., & Dinstein, I. H., Textural features for image classification, IEEE Transactions on systems, man, and cybernetics, (6), 610-621, 1973.
26. Mandwe, A. A., & Anjum, A., Detection of brain tumor using k-means clustering, IJSR, 5(6), 420-423, 2016.
27. MM Mhaske, RR Manza, YM Rajput, Detection of Retinal Images as Normal or Abnormal using Texture Feature Analysis to identify the Diabetic Retinopathy, ISSN No: 2309-4893 International Journal of Advanced Engineering and Global Technology Vol-03, Issue-03, March 2015.