UDC 616.24-006:004.932

LUNG CANCER DETECTION ON CT IMAGES USING IMAGE PROCESSING

Akbayeva, A. Shilikbay, A. Toleubekova, A. Zhumagaliyeva

Senior Robotics Engineering students at School of Engineering and Digital Sciences, Nazarbayev University, Astana, Kazakhstan

This proposal outlines the development of a Lung Cancer Detection System using advanced imaging and processing techniques. Lung cancer is deadly, and early detection is vital for improving survival rates. The system will assist medical professionals in identifying potential cases from chest CT scans, aiming to enhance diagnostic accuracy and efficiency, ultimately improving patient outcomes.

Keywords: Lung cancer, Medical imaging, Machine learning, Image analysis, Computer-aided diagnosis, Chest Computed Tomography (CT) scan.

Introduction.

Lung cancer is a rapidly growing tumor which can also spread to other organs within the human body. The approach to treating lung cancer depends on its type, but a prevalent method of its identification involves the utilization of computed tomography (CT) images. According to [1], computer-aided diagnosis (CAD) became an essential part of the diagnosis process in early detection. This project aims to address these issues by developing an advanced Lung Cancer Detection System that employs image processing algorithms to automate the analysis of medical images.

Related works.

Numerous studies have explored the use of image processing and machine learning in medical imaging, particularly for lung cancer detection. Ginneken [6] categorized lung cancer region extraction approaches into two types: rule-based and pixel classification-based methods. Most methods fall into the rule-based category [1-11], which applies steps, rules, and tests during extraction. Techniques like local thresholding, region growing, edge detection, ridge detection, and morphological operations are commonly used. In contrast, pixel classification assigns each CT image pixel to an anatomical class (lung, background, or other classes like heart, mediastinum, diaphragm) [12].

Image preprocessing is crucial in medical image analysis, especially for lung cancer detection. Most techniques, including those for lung cancer, rely on fundamental methods like median filtering, Gabor filters, smoothing, enhancement, binarization, and Fast Fourier Transform (FFT). Bit-slicing is widely used for its speed and independence from data and user input, making it superior to the threshold method [11]. Another technique involves adding salt-and-pepper noise to reduce distortion, followed by median filtering [8]. Research shows Gabor filters perform better than FFT for image segmentation, with Gabor providing superior image quality [1].

Thresholding and watershed transform are the primary image segmentation methods [1]. In lung cancer detection, thresholding isolates lung tissue from the chest cavity to identify suspicious nodules or lesions [7], while watershed transformation separates touching objects by using internal and external markers [5]. Watershed segmentation has been found to offer greater accuracy and quality than thresholding [1].

The CAD (Computer-Aided Diagnosis) system focuses on feature extraction to detect and isolate portions of lung images, helping to identify cancer nodules [9]. Features like average intensity, area, perimeter, and eccentricity aid in classifying regular versus irregular cell images [2]. Morphological characteristics like nucleus area, perimeter, and eccentricity are used in a module to classify lung cancer cells [4]. CAD is favored for its high sensitivity (90%) and low false positive rate (0.05 per image), aiding early detection without hindering the radiologist's diagnosis [10]. The system can detect early-stage lung nodules, improving early diagnosis and patient survival rates [3].

Proposed approach.

1. Obtain a CT scan of the lung.
2. Convert the image to grayscale.
3. Apply high-pass filtering to preserve edges.
4. Add salt-and-pepper noise, followed by median filtering to enhance the image.

Image Enhancement.

Techniques such as Gabor filtering and auto-enhancement are employed to improve image clarity. The Gabor function is a valuable tool for texture analysis, because of favorable localization characteristics in both the spatial and frequency domains and it can be clearly seen that it produces an image with the best quality.

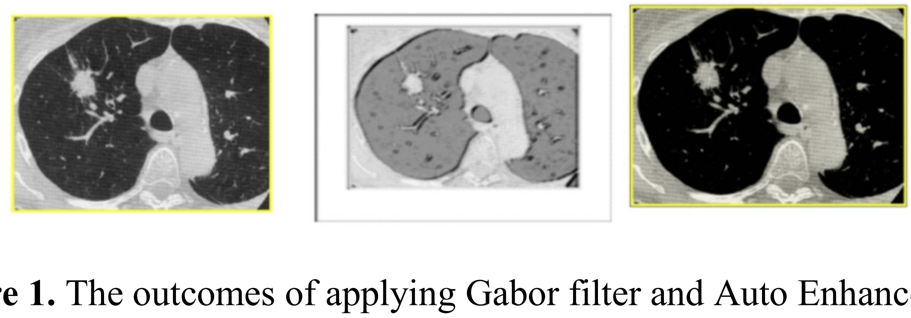

Figure 1. The outcomes of applying Gabor filter and Auto Enhancement methods: (a)Original Image (b)Enhanced Image (c)Auto Enhancement

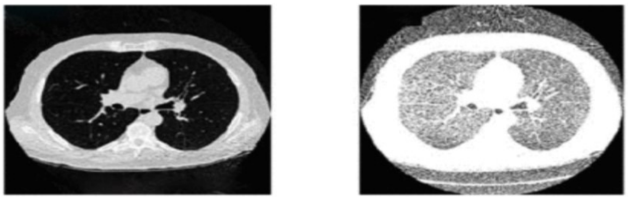

Figure 2. The outcome of applying FFT enhancement technique: (a)Original Image (b)Enhanced Image

Segmentation divides the image into regions based on attributes like intensity and shape. Thresholding and watershed segmentation are used to isolate lung tissue and detect nodules.

Thresholding: This method is used to differentiate between the foreground and the background in an image [5]. By selecting a suitable threshold value (T), a gray-level image can be converted into a binary image. The process involves classifying pixels based on their gray level values: below the threshold are black(0), and those above the threshold are white(1). The researchers utilized Otsu's method, a statistical approach, to determine an optimal threshold for binarization.

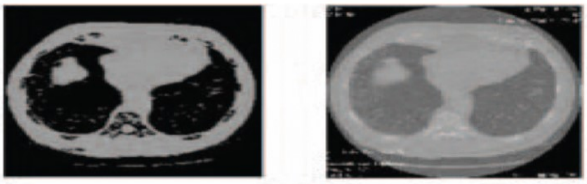

Figure 3. The outcomes of applying Thresholding Segmentation: (a)Original Image (b)Segmented Image

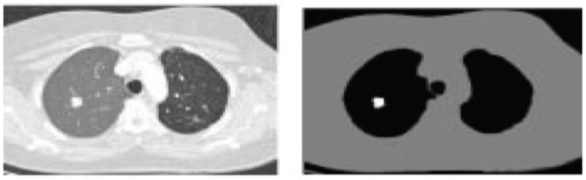

Figure 4. The outcomes of applying Watershed Segmentation: (a)Original Image (b)Segmented Image

Watershed segmentation: The watershed transform is often applied to separate touching objects on an image which include the internal markers, the markers of interest, and the external markers associated with the background [5]. The main advantage is that it is a fast, simple, intuitive method, and the image can be separated into regions even if the contrast is poor.

Feature Extraction: The Gray-Level Co-occurrence Matrix (GLCM) is used to extract features such as Contrast, Correlation, Energy, and Entropy from the images, which provide insights into the tumor characteristics [4].

Tumor Classification: Support Vector Machine (SVM) classifiers are used to distinguish between benign and malignant tumors based on the extracted features [4].

EXPERIMENTS & RESULTS

The primary objective is to demonstrate the system's reliability in distinguishing cancerous nodules, thereby enhancing diagnostic accuracy and aiding in treatment planning. The Lung Cancer Detection System involves the following steps, as illustrated in Figure 5.

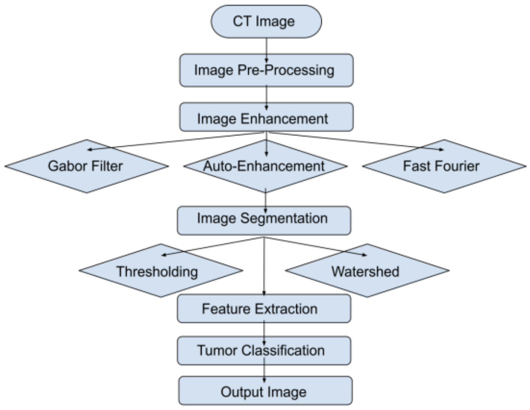

Figure 5. Graphical Abstract of the Project

A. CT Image Acquisition

CT scan images of lung cancer nodules were acquired from local hospitals. These images, captured as screenshots from lung scan computer topographies, were saved in JPEG format.

B. Image Pre-processing

The CT images were prepared for further
analysis through several image pre-processing techniques:

1. **Grayscaling**: Converting images to grayscale reduces potential errors compared to RGB.
2. **High-pass Filtering**: This technique suppresses low-frequency elements while highlighting crucial details.
3. **Salt and Pepper Noise**: Random noise was introduced into the image.
4. **Median Filtering**: This step preserves image edges while minimizing interference caused by noise.

Pre-processing plays a vital role in medical imaging analysis, particularly for CT scans, where accuracy and precision are critical. Grayscaling simplifies the data, high-pass filtering enhances essential details, and noise filtering refines the image to improve diagnosis.

C. Image Enhancement (Gabor Filter, Fast Fourier Transform)

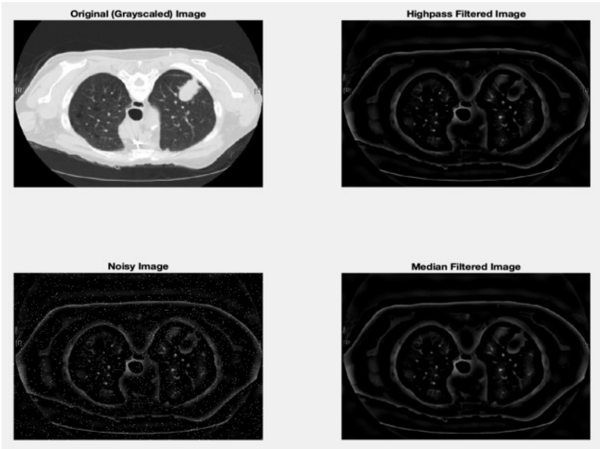

Figure 6. Image Pre-processing Stages

Next, image enhancement techniques were applied to improve image clarity and the visibility of potential lung nodules:

● **Gabor Filter**: This filter was employed to enhance image textures, aiding in the detection of tiny lung structures. By experimenting with different parameters like wavelength and orientation,

we fine-tuned the filter for optimal detection. The obtained results are demonstrated in Figure 7.

Figure 7. Applying Gabor Filter to Preprocessed Image

● **Fast Fourier Transform (FFT)**: The FFT was combined with a surface representation technique to highlight cancerous areas. White regions on a surface plot generated in MATLAB indicate potential cancer cells.

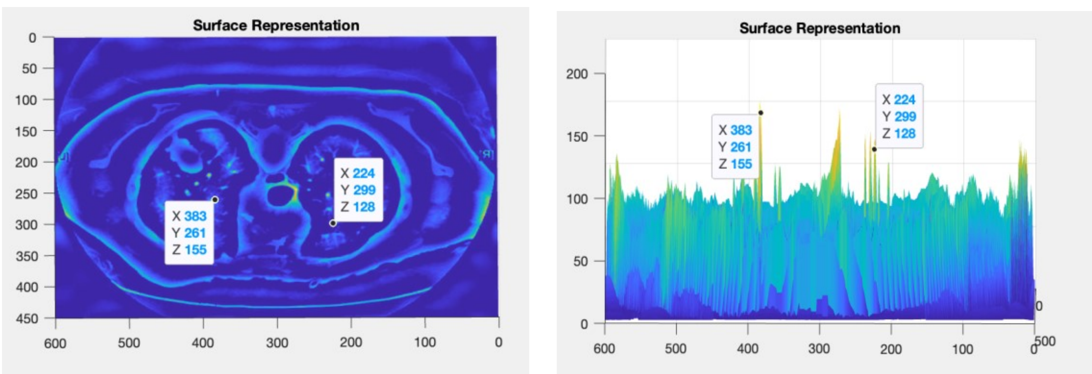

Figure 8 & 9. Surface Representation to detect Cancerous Cells using FFT

D. Image Segmentation (Thresholding, Proposed Algorithm)

● **Segmentation Algorithm**: The watershed algorithm was initially employed to segment lung structures and nodules in CT images. However, it proved insufficient for our task, prompting us to switch to Otsu thresholding combined with morphological operations like erosion and dilation. Figure 11 shows an example of watershed segmentation.

● **Otsu Thresholding**: This global thresholding method was used to binarize the image. Based on the variance between classes, a threshold value was calculated, ensuring optimal segmentation for cancer detection.

The outcome of this segmentation process is visually presented in Figures 11 and 12. Despite the segmentation's effectiveness, its performance relies on specific image characteristics, such as the tumor's relative size and its separability from lung tissues.

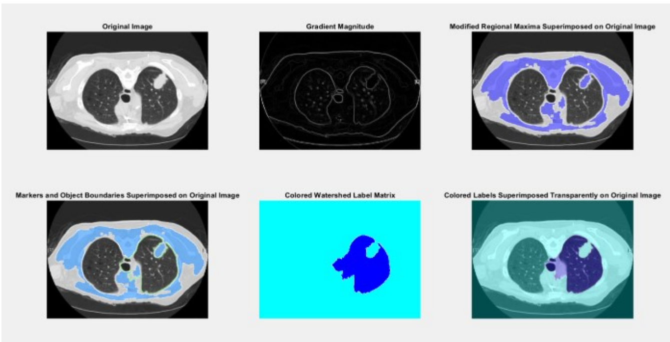

Figure 10. Watershed Segmentation

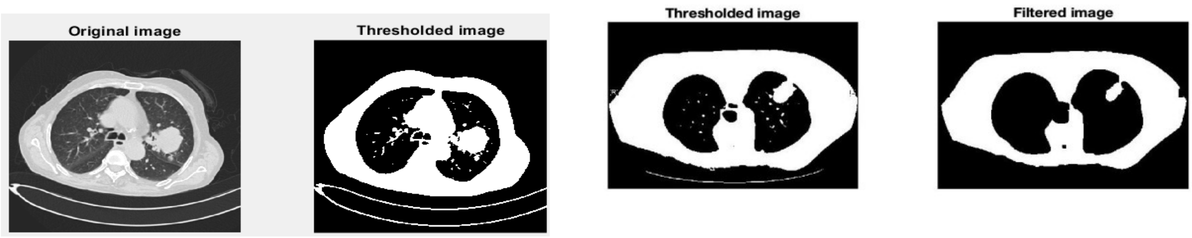

Figure 11 & 12. Otsu thresholding results & Application of Erosion and Median Filter

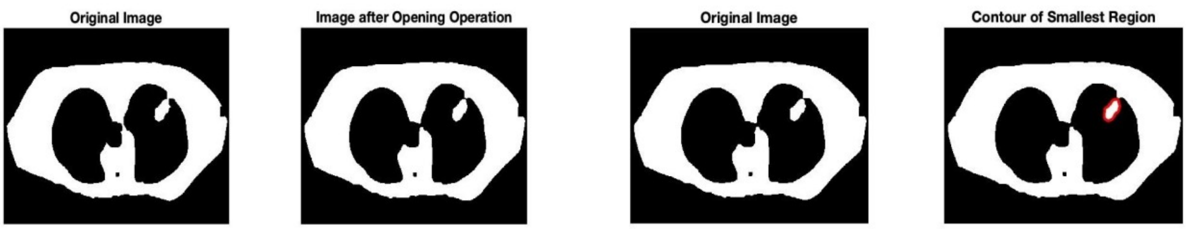

Figure 13 & 14. Otsu thresholding results & Application of Erosion and Median Filter

E. Feature Extraction

After segmentation, key features of the cancerous regions were extracted for analysis:

• **Area and Perimeter**: The area was calculated by counting the number of white pixels representing cancer cells, and the perimeter was determined by identifying boundary pixels.

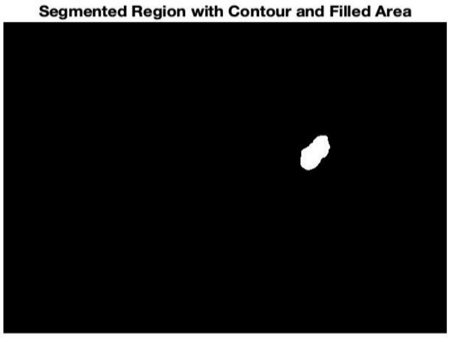

Figure 15. Segmented Region of Tumor

$$A = n\{1\} \quad (1)$$

$$P = |s_n \times s_1| + \sum_{i=1}^{n-1} |s_i * s_{i+1}| \quad (2)$$

• **Eccentricity**: This value helped characterize the shape of the cancerous region.

$$Eccentricity = \sqrt{1 - \frac{4 * \pi * Area}{PerimeterLength^2}} \quad (3)$$

• **Entropy, Contrast, Correlation, Energy, Homogeneity**: Texture features were extracted using the Gray Level Co-occurrence Matrix (GLCM) for further analysis.

$$Entropy = -\sum\sum p(i,j)log(p(i,j)) \quad (4)$$

$$Contrast = \sum\sum (i - j)^2 log(p(i,j)) \quad (5)$$

$$Correlation = \frac{\sum\sum(i - mi)(j - mj)p(i,j)}{stdev(i)stdev(j)} \quad (6)$$

$$Energy = \sum\sum (p(i,j))^2 \quad (7)$$

$$Homogeneity = \sum \frac{p(i,j)}{1 + |i - j|} \quad (8)$$

F. Classification. The classification algorithm, based on feature extraction, aimed to assign cancer stages to images. By comparing extracted features like area, perimeter, and entropy against reference values, the system successfully classified images into stages of lung cancer. A simple threshold-based method was employed, and the results showed promise for future refinements. Tables 1 and 2 present the classification results.

Table 1. Reference Values for Each Stage of Lung Cancer

Feature	Stage 1	Stage 2	Stage 3	Stage 4
Area	386 – 2000 mm^2	2001 – 4000 mm^2	4001 – 6000 mm^2	6001 – 6560 mm^2
Perimeter	223 – 500 mm	501 – 650 mm	651 – 800 mm	801 – 994 mm
Eccentricity	0.92278 – 0.937	0.938 – 0.945	0.946 – 0.953	0.954 – 0.963
Entropy	0.010826 – 0.06	0.061 – 0.1	0.101 – 0.15	0.151 – 0.16
Contrast	$< 6e^{-5}$	$6.01e^{-5} – 9e^{-5}$	$9.01e^{-5} – 1.3e^{-4}$	$1.4e^{-4} – 1.8e^{-4}$
Correlation	< 0.94	0.941 – 0.96	0.961 – 0.975	0.976 – 0.979
Energy	> 0.986	0.976 – 0.985	0.966 – 0.975	0.947 – 0.965
Homogeneity	> 0.999	0.997 – 0.998	0.995 – 0.996	0.9932 – 0.994

Table 2. Classification Results

Image	Predicted Stage
Result1.jpg	Stage 2
Result2.jpg	Stage 2
Result3.jpg	Stage 1
Result4.jpg	Stage 4
Result5.jpg	Stage 3
Result6.jpg	Stage 2
Result7.jpg	Stage 2
Result8.jpg	Stage 1
Result9.jpg	Stage 1
Result10.jpg	Stage 2
Result11.jpg	Stage 1
Result12.jpg	Stage 1
Result13.jpg	Stage 3
Result14.jpg	Stage 2
Result15.jpg	Stage 3
Result16.jpg	Stage 2
Result17.jpg	Stage 1

Discussion.

A. Image Pre-processing. The preprocessing steps included grayscaling the CT images to minimize errors, followed by applying a high-pass filter to enhance critical details. "Salt and pepper" noise was added to simulate real-world noise, and median filtering was used to reduce it while preserving edges. These steps improved the image quality and highlighted important features for better tumor identification.

B. Image Segmentation. The watershed algorithm was ineffective for lung tumor CT scans, so Otsu thresholding combined with morphological operations (erosion, dilation) was used instead. However, this method is limited by the need for the tumor to be large and clearly separable from surrounding tissues.

Morphological operations may also reduce the tumor size, making early-stage detection challenging. Although this approach works well in certain conditions, CNNs or other deep learning methods could improve performance by detecting smaller or less distinct tumors.

C. Feature Extraction and Classification. Despite some uncertainty due to the quality of converted CT scans, we developed a method to classify tumor stages using feature extraction. Early-stage tumors were small, and texture-based features such as entropy and contrast were lower, while energy and homogeneity were higher.

Conclusion.

This paper provided a detailed plan for creating an advanced medical imaging and image processing system for the purpose of developing a lung cancer detection system. The suggested approach is divided into a number of stages, including image pre-processing, picture enhancement, thresholding and segmentation, feature extraction, and tumor classification. Through the examination of lung CT scans, the system seeks to support medical practitioners in the identification and diagnosis of lung cancer cases and contribute to the early detection of lung cancers. We can achieve that by increasing lung cancer diagnosis efficiency and accuracy. The study covered relevant literature, the suggested methodology, the trials, and findings, as well as the system’s disadvantages and limitations. This paper also made several recommendations for further improvement referring to deep learning applications, working with medical specialists, and utilizing a bigger and more varied data set in order to validate and enhance the accuracy of the outcomes. The study illustrated the potential of image processing techniques for medical imaging analysis, especially with regard to lung cancer identification.

Acknowledgements

We extend our heartfelt gratitude to our professor Zhanat Kappassov for his invaluable guidance throughout the course project and to the Nazarbayev University community for their support on this project.

REFERENCES

1. M. S. Al-Tarawneh, "Lung Cancer Detection Using Image Processing Techniques," in Leonardo Electronic Journal of Practices and Technologies, 2012.
2. A. Chaudhary and S. S. Singh, "Lung Cancer Detection on CT Images using Image Processing," in International Conference on Computing Sciences, 2012.
3. P.P.Gawade and R.P.Chauhan, "Detection of Lung Cancer Cells using Image Processing Techniques," in 1st IEEE International Conference on Power Electronics Intelligent Control and Energy Systems, 2016.
4. K. Gopi and J. Selvakumar, "Lung Tumor Area Recognition and Classification using EK-Mean Clustering and SVM," in International Conference on Nextgen Electronic Technologies, 2017.
5. B. G. Patil and S. N. Jain, "Cancer Cells Detection Using Digital Image Processing Methods," in International Journal of Latest Trends in Engineering and Technology (IJLTET), 2014.
6. B. V. Ginneken, B. M. Romeny, and M. A. Viergever, "Computer-aided diagnosis in chest radiography: A survey," in IEEE Transactions on Medical Imaging, vol. 20, no. 12, 2001.
7. N.Hadavi,M.J.Nordin,andB.V.Ginneken, B.M.Romeny, and M.A. Viergever, "Computer-aided diagnosis in chest radiography: A survey," in IEEE Transactions on Medical Imaging, vol. 20, no. 12, 2001 A. Shojaeipour, "Lung Cancer Diagnosis using CT-scan Images based on Cellular Learning Automata," in International Conference on Computer and Information Sciences (ICCOINS), 2014.
8. P.R.Katre and A.Thakare, "Detection of Lung Cancer Stages using Image Processing and Data Classification Techniques," in 2nd International Conference for Convergence in Technology (I2CT), 2017.
9. S.Kalaivani, P.Chatterjee, S.Juyal, and R.Gupta, "Lung Cancer Detection Using Digital Image Processing and Artificial Neural Network," in International Conference on Electronics Communication and Aerospace Technology (ICECA), 2017.
10. S. Makaju et al., "Lung Cancer Detection using CT Scan Images," in Procedia Computer Science, 2018.
11. K. Punithavathy, M.M. Ramya and Sumathi Poobal, "Analysis of Statistical Texture Features for Automatic Lung Cancer Detection in PET/CT Images", International Conference on Robotics Automation Control and Embedded Systems (RACE), 18–20 February 2015
12. N.D. Thombare, "Detection of Lung Cancer Using Marker-Controlled Watershed Transform", International Conference on Pervasive Computing(ICPC), 2015.

ОБНАРУЖЕНИЕ РАКА ЛЕГКОГО НА ТОМОГРАФИЧЕСКИХ ИЗОБРАЖЕНИЯХ С ПОМОЩЬЮ ОБРАБОТКИ ИЗОБРАЖЕНИЙ

А. Акбаева, А. Шиликбай, А. Толеубекова, А. Жумагалиева

В данной статье описывается разработка системы обнаружения рака легких с использованием передовых методов визуализации и обработки. Рак легких смертельно опасен, и его раннее обнаружение жизненно важно для повышения уровня выживаемости. Система поможет медицинским работникам выявлять потенциальные случаи заболевания по снимкам КТ грудной клетки, что позволит повысить точность и эффективность диагностики и в конечном итоге улучшить состояние пациентов.

Ключевые слова: Рак легких, медицинская визуализация, машинное обучение, анализ изображений, автоматизированная диагностика, компьютерная томография (КТ) грудной клетки.

БЕЙНЕЛЕУ АРҚЫЛЫ ТОМОГРАФИЯЛЫҚ СУРЕТТЕРДЕ ӨКПЕ ОБЫРЫН АНЫҚТАУ

А. Ақбаева, А. Шілікбай, А. Төлеубекова, А. Жұмағалиева

Берілген мақалада бейнелеу мен өңдеудің озық әдістерін қолдана отырып, өкпенің қатерлі ісігін анықтау жүйесінің дамуы сипатталған. Өкпенің қатерлі ісігі өлімге әкеледі және оны ерте анықтау өмір сүру деңгейін жақсарту үшін өте маңызды. Жүйе медицина қызметкерлеріне кеуде қуысының КТ суреттері арқылы аурудың ықтимал жағдайларын анықтауға көмектеседі, бұл диагностиканың дәлдігі мен тиімділігін арттыруға және сайып келгенде пациенттердің жағдайын жақсартуға мүмкіндік береді.

Кілт сөздері: Өкпенің қатерлі ісігі, медициналық бейнелеу, машиналық оқыту, кескінді талдау, автоматтандырылған диагностика, кеуде қуысының компьютерлік томографиясы (КТ).