International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

The Detection of Lung Cancer in Computed Tomography (CT) Scans Using Image Analysis Techniques

Eesa Salaar Adnan

Student, Repton School Dubai

Abstract

This research paper aims to present and discuss the utilization of image analysis techniques for the detection of lung cancer through CT scans. To be specific, region growing, edge detection, and thresholding are explored and presented as image analysis techniques for the identification and determination of abnormal regions, which may indicate the presence of tumors. All image analysis techniques were implemented and tested using a dataset of lung images, and the overlapping regions are examined for increased accuracy in tumor identification. This research aims to prove that the utilization of multiple image processing techniques may increase the accuracy of tumor identification, providing a computational tool for radiologists for early tumor detection. Moreover, it aims to show that traditional image analysis techniques are effective and easily understandable for clinical decision support.

Keywords: Lung Cancer, CT Scan, Medical Imaging, Image Analysis, Region Growth, Edge Detection, Thresholding, Overlapping Regions (Intersection), Tumor Detection, Computer Algorithm

1. Introduction

Early detection is an essential aspect that can help in improving the treatment of cancer. In modern medical practices, diagnostic imaging is used to examine the internal parts of the body to detect abnormal tissue growth. In this regard, the Computed Tomography (CT) scan is commonly used to examine the internal parts of the body. Radiologists couple the use of CT scans and their own experience to detect abnormal tissue growth that could potentially occur within the body.

The interpretation of medical scan images may differ among medical practitioners based on their level of experience and their subjective judgement. In cases where the abnormality is not clear, it is challenging to ensure consistency in the interpretation of medical scan images. In the modern medical field, the number of such images scanned is rising, along with a significant interest in the development of a computational tool to assist medical practitioners in the interpretation of medical scan images.

The pixels in the image are arranged in grids, with each pixel carrying a numerical value that corresponds to physical properties of the tissue being scanned, e.g., density in the case of a CT scan. The numerical nature of the information contained in the pixels means that mathematical techniques can be employed to analyze the image, rather than purely visual means.

From a mathematical standpoint, the image itself is effectively represented as a matrix, where each element in the matrix corresponds to the value of the pixel at that location. This mathematical representation of the image allows various image-analysis techniques to be employed to identify structures within the image. Algorithms are capable of employing such techniques, which in turn can help identify regions which contain abnormal tissue growth.

The purpose and objectives of this research paper are to investigate the ways in which image processing techniques could be used in the analysis of the results obtained in the CT scan in order to detect the possible presence of cancerous regions in the body. In this context, the study aims to examine the ways in which techniques such as thresholding, region growth, and edge detection could be used to detect the presence of cancerous regions in the lung scan, and in the process, illustrate the ways in which mathematical techniques could be used in the objective analysis of the results obtained in the scan.

2. Background of the Study

Despite the many advances in medical science, cancer still ranks as one of the major causes of death around the globe, and thus early and accurate detection of the disease is of prime importance for increasing the chances of survival for those afflicted with it. Modern medical science relies heavily on medical imaging techniques such as CT scans, as they offer the advantage of viewing the internal structures of the body without resorting to any invasive procedures. CT scans offer detailed images of the internal structures of the body, and doctors can easily locate abnormal growths or nodules in the body.

The development of new and improved medical imaging technology has greatly enhanced the resolution and the overall volume of medical data generated in clinical environments. Though improved resolution scans offer many advantages in the accurate detection of small abnormalities, they also make the interpretation of the results more complex, as the radiologists need to study a large volume of medical data and make accurate distinctions between normal and abnormal structures, and those structures that may pose a threat to the body.

The numerical nature of the pixels in the digital image means that the image itself is treated as a mathematical object rather than a visual representation. Each image from the CT scan can be represented as a matrix, where the density of the tissue is represented by the pixels. This numerical nature means that mathematical algorithms can be used to examine the structure of the image.

Image processing techniques have become an important area of research in the field of medical imaging. This is because the techniques provide the ability to enhance or detect structures within the image using mathematical algorithms. Segmentation algorithms, or our image analysis techniques, for instance, attempt to separate the various components of the image based on common characteristics. This could potentially help to draw attention to possible abnormalities, which could include tumors.

3. Statement of the Problem

Though the usage of CT scans is common in the diagnosis of cancer, the identification of abnormal tissues from the scans is a challenging task. The human body contains many structures that show similar characteristics when viewed in an image. This makes it hard to distinguish between normal tissues and cancerous tissues in the early stages of development. The identification of small nodules is also a challenging task, especially when the nodules occur in complex areas of the lung. Furthermore, it is difficult to highlight the difference in densities of such nodules through visual analysis of CT scanned images.

The interpretation of medical images is also based on the judgment of the individuals involved. The individuals involved in the interpretation of the images may concentrate on different aspects of the image or make slight differences in the patterns of the image. This brings the possibility of inconsistent results, especially when the abnormal tissues are hard to identify from the images. At the same time, there is a large amount of medical image data coming out of the healthcare system. Careful examination of this data is time-consuming and requires a lot of expertise, which is a problem in the field of radiology and increases the risk of overlooking certain details. The problem has led to the idea of using a quicker, more efficient method to help analyze the medical scan data. Using mathematical image processing techniques on the data obtained in a CT scan, it is possible to highlight suspicious regions in the image and help detect potential cancerous structures.

4. Purpose and Objectives of the Research

The purpose of this research is to investigate the possibilities of using mathematical image processing techniques to analyze the images obtained by the CT scan and detect the potential cancerous regions in the lung tissue.

The objectives of this research are as follows:

- To investigate the possibilities of representing the digital medical images and analyzing them as numerical data.
- To apply the techniques of segmenting the medical images, which include thresholding, region growing, and edge detection, on the lung scan images.
- To investigate the possibilities of using different parameters to enhance the visibility of the suspicious regions in the images.
- To analyze the intersection of the techniques used in the segmentation of the images in order to investigate the possibilities of using a combination of techniques in the detection of the potential cancerous regions.

5. Literature Review

The initial methods for medical image analysis were based on manual inspection by trained medical experts. In medical image analysis, radiologists inspect medical imaging data and try to find patterns that might be related to diseases, based on their knowledge. Though this is an important aspect of medical imaging, it has some limitations, especially because of human factors and the amount of data generated by medical imaging systems.

In order to deal with these limitations, researchers have been working on developing computational methods for image analysis. Initially, researchers have been working on developing simple image processing operations such as filtering, thresholding, and edge detection. These methods involve mathematical operations on matrices that represent images.

Segmentation techniques also proved to be of vital significance in the field of medical image research. They are a set of algorithms used in image processing with the primary goal of segmenting the image based on the similarities in the pixel values or structures. Techniques used in segmentation include thresholding and region growing, which help in segmenting a medical image and identifying structures that could potentially be a tumor or nodule.

Recent research has also tried to combine these image processing techniques with advanced computer systems with the primary goal of assisting in the diagnosis of medical conditions. Though advanced techniques in artificial intelligence have been used in the field of image analysis, it is essential to first know the basic principles of image processing in the creation of a reliable system.

The present research is based on the findings of the earlier research in the field of medical image analysis, with the aim of determining the ways in which segmentation methods could be used to highlight suspicious areas in the CT scan images obtained from the lung tissue. This study aims to evaluate the ways in which mathematical image processing could help in identifying the areas that could potentially be affected by cancer.

6. Methodology

6.1 Dataset

Utilized in the study were ten CT scan images of lungs from a public dataset. Out of the ten images, five were confirmed to have cancerous nodules, while the remaining five did not have cancer. Although the study had a small database, it allowed for conducting an initial evaluation of the segmentation techniques by comparing the algorithms for the cancerous and non-cancerous images.

6.2 Image Representation

Each CT scan was represented as a matrix of pixel values, with each pixel value representing the density of tissue at that particular location. Higher intensity pixel values indicate denser tissue, while lower intensity pixel values indicate areas that are not as dense, such as the lungs.

This representation enables image processing techniques to analyze each pixel in a systematic way to identify patterns of interest.

6.3 Segmentation Techniques

Three techniques were applied to each image: thresholding, region growth, and edge detection.

6.3.1 Thresholding

Thresholding is an image analysis technique which is used to separate the pixels in the image based on their intensity. It compares the intensity or brightness of the pixels with a given threshold value. Results will only show regions where the intensity of the pixel is greater than the intensity of the threshold.

6.3.2 Region Growth

Region growth constructs an image region starting with a chosen pixel, known as the seed pixel, and then continues to add neighboring pixels to the region that are similar in intensity to the initial pixel.

The algorithm searches the neighboring pixels to see whether the difference in intensity with the initial pixel is small enough; if it is, the pixel is added to the new region.

6.3.3 Edge Detection

Edge detection measures the rate at which the pixel intensity is changing. If the rate is high enough, then it is considered an edge. This creates an image where only the edges of the shapes are visible, with everything inside the shapes being eliminated.

6.3.4 Intersection of Thresholding and Region Growing

The intersection was determined by selecting only those pixels identified as belonging to a region by both thresholding and region growing techniques.

This approach sought to minimize false positives and emphasize regions identified by both techniques. This was done to question how accurately these techniques can be used together.

6.4 Parameter Selection

Parameters were chosen based on testing different values on the ten images:

Technique	Parameter	Reason for Choice
Thresholding	T = 160	Balanced highlighting dense areas while avoiding normal tissue.
Region Growing	sim = 10	Produced compact regions without including surrounding normal tissue.
Edge Detection	50–150	Highlighted structural boundaries without excessive noise.

7. Results

The image analysis techniques above were applied to all ten images. Observations are summarized below:

- **Cancer images (5):** The intersection of thresholding and region growing did not highlight any regions, indicating that the selected parameters did not isolate the tumor areas.
- **Non-cancer images (5):** Most non-cancer images produced highlighted regions, likely corresponding to normal lung structures such as vessels or dense tissue.
- **Anomaly:** One non-cancer image produced little or no highlighted region, showing variation in algorithm behavior.

7.1 Intersection Results

Image	**Cancer Present**	**Intersection Highlighted?**	**Notes**
1	Yes	No	No regions detected
2	Yes	No	No regions detected
3	Yes	No	No regions detected
4	Yes	No	No regions detected
5	Yes	No	No regions detected
6	No	Yes	Normal tissue highlighted
7	No	Yes	Normal tissue highlighted
8	No	Yes	Normal tissue highlighted
9	No	Yes	Normal tissue highlighted
10	No	No	Anomalous; no regions

Overall, the intersection method reduced false positives compared to thresholding alone but did not detect regions in the cancerous images.

7.2 Images

Image 1 – Cancer

Figure 1: Original Image 1

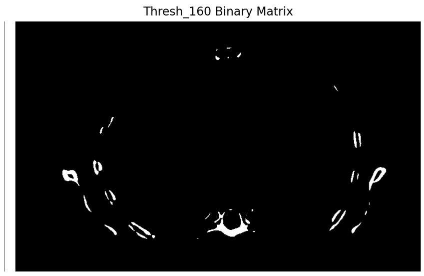

Figure 1.1: Image 1 Threshold

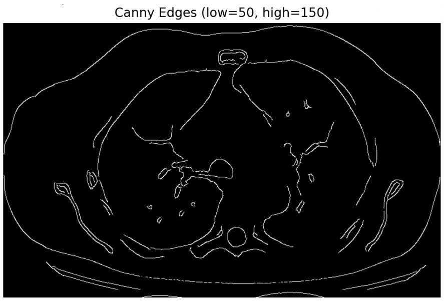

Figure 1.2: Image 1 Edge Detection

Figure 1.3: Image 1 Region Growth

Image 6 - No Cancer

Figure 2: Original Image 6

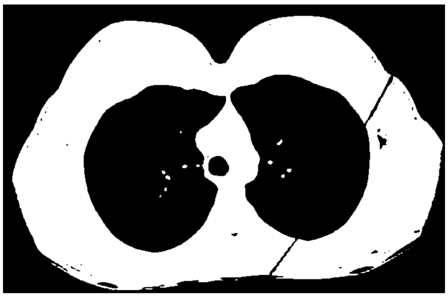

Figure 2.1: Image 6 Threshold

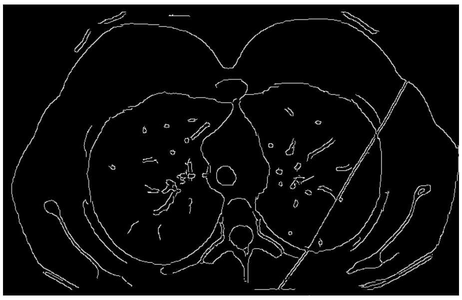

Figure 2.2: Image 6 Edge Detection

Figure 2.3: Image 6 Region Growth

Figure 2.4: Image 6 Intersection

8. Discussion

From the results, it is evident that the intersection of the thresholding and the region growing did not effectively highlight the cancerous areas in the images that contained tumors. The results show that several images that did not contain cancer had the areas that correspond to the normal lung structures highlighted. There are several possible reasons for the failure of the intersection of the algorithms to highlight the regions in the cancerous images. Firstly, the intensity of the tumor could have been similar to the intensity of the tissues. In the thresholding method, pixels that have an intensity value above the specified value are included. However, if the tumor does not have an intensity value significantly above the other tissues, the tumor may be excluded in this step. Secondly, the similarity parameter of the region growing method could have been too strict. In the region growing method, the new region grows to neighboring pixels that have an intensity value within the specified similarity range. If the similarity value is too strict, the new region may not grow enough to cover the entire tumor region.

These findings indicate that simple fixed parameter analysis might not be appropriate for reliable tumor detection, and that more sophisticated techniques, such as adaptive algorithms or machine learning, might be required.

9. Limitations

1. Small dataset: Only ten images have been analyzed.
2. Parameter sensitivity: Algorithm response is heavily reliant on the values of the threshold, similarity, and edge detection.
3. Verification: It is unknown if the areas that have been highlighted correspond to the actual tumors.
4. Simplicity: Only basic segmentation algorithms have been used, without the aid of advanced AI.

10. Conclusion

The study investigated thresholding, region growing, and edge detection techniques for highlighting abnormal regions that might be cancerous. Although the intersection between thresholding and region growing reduced false positives, it failed to highlight tumor regions in images containing cancer, instead highlighting normal regions.

The results show that simple segmentation techniques, despite their parameters, cannot be relied upon for cancer detection, though they offer a controlled means to investigate structural patterns in lung images. The study, despite its limitations, proves that image processing techniques can be systematically applied to medical imaging.

11. References

1. Dataset: Kabil, A. (2023). *LungCancer4Types* [Dataset]. Hugging Face. https://huggingface.co/datasets/Kabil007/LungCancer4Types
2. Harris, C. R., et al. (2020). *Array programming with NumPy*. Nature, 585, 357–362.
3. Bradski, G. (2000). *The OpenCV Library*. Dr. Dobb's Journal of Software Tools.
4. World Health Organization. (2023). *Cancer*. https://www.who.int/news-room/fact-sheets/detail/cancer
5. Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1), 62–66. https://doi.org/10.1109/TSMC.1979.4310076
6. Adams, R., & Bischof, L. (1994). Seeded region growing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(6), 641–647. https://doi.org/10.1109/34.295913
7. Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6), 679–698. https://doi.org/10.1109/TPAMI.1986.4767851
8. Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., & Yu, T. (2014). scikit-image: Image processing in Python. *PeerJ*, 2, e453. https://doi.org/10.7717/peerj.453
9. Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*. https://opencv.org/about/

10. Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Fernández del Río, J., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585, 357–362. https://doi.org/10.1038/s41586-020-2649-2