A CLASSIFICATION IMAGE PROCESSING APPROACH TO BE USED FOR LUNG CANCER

Dr. T.SRINIVAS, Professor & Principal,
Department of ECE,
Mother Theressa College of Engineering & Technology, Peddapally, Telangana

ABSTRACT

When recognizing problems in target photos is time-sensitive, as is the case with tumors of various diseases (lung cancer, breast cancer, etc.), image processing techniques have recently been widely applied across a number of medical sectors to enhance images for early identification and treatment. This research focuses on using low-level pre-processing methods based on the Gabor filter and the Gaussian distribution to guarantee good quality and accuracy in the final image. Following segmentation rules allows us to zero in on the most relevant region of a picture for feature extraction. Using generalizable features allows for a normality comparison. In this research, we find that the proportion of pixels and mask-labeling are the most important features for reliable image comparison.

Keywords: Cancer Detection; Image processing; Feature extraction; EnhancementWatershed; Masking.

1. INTRODUCTION

Lung cancer is caused by the growth of a tumor composed of abnormal lung cells. Lung cancer cells can travel through the bloodstream or the lymph fluid that surrounds lung tissue. Lymph is transported through lymphatic veins to lymph nodes in the thoracic cavity and pulmonary system. Lung cancer frequently spreads to the middle of the chest due to the lymphatic system's natural drainage pattern. Metastasis occurs when a cancer cell travels via the circulation to a lymph node or another organ far from its original site. Lung-originating cancers are referred to as primary lung cancers. Non-small cell lung cancer is further split into three kinds (carcinoma, adenocarcinoma, and squamous cell carcinoma), while small cell lung cancer is divided into two categories (adenocarcinoma and squamous cell carcinoma).

In 2008, lung cancer accounted for 7.7 percent of all new cancer diagnoses in Jordan (male and female combined), with 356 cases reported. Lung cancer was the second greatest cause of death for males and the tenth leading cause of death for women, with a male to female ratio of 5 to 1. In Figure 1, we see the four main components of a system designed to detect lung cancer. The first step is to compile a set of normal and abnormal CT scans from the IMBA Home (VIA-ELCAP Public Access) Database. In the second stage, you'll use several image enhancement techniques to bring the level of sharpness and detail to its maximum potential. The third phase involves the use of picture segmentation techniques, which are essential for later image processing steps. The fourth stage involves obtaining the enhanced segmented image and then using its general features to establish whether the image is normal or abnormal.

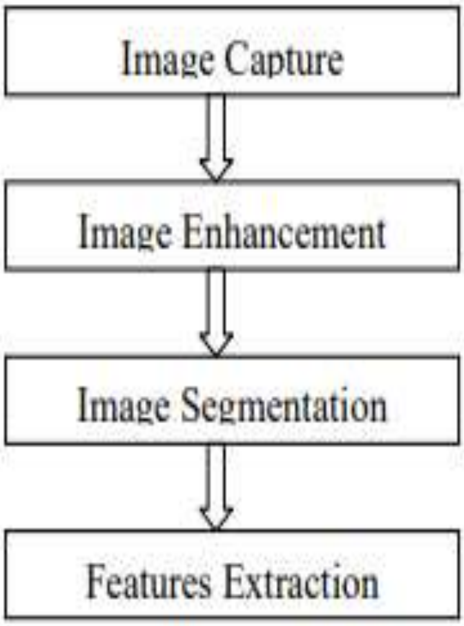


Figure 1. Lung cancer image processing stages

When broken down by diagnosis stage, lung cancer is both the most lethal and most frequent form of the disease. Because of this, early diagnosis is essential for preventing the disease's more advanced stages and decreasing the disease's prevalence. The goal of this research was to find a way to accurately compare images by employing identified information like pixel percentage and mask labeling.

2. MATERIAL AND METHOD

For more accurate results, we divided our work in this study into three stages:

Image Enhancement stage:

To enhance an image by doing away with background noise, lens distortion, or other flaws. Three methods are used for this purpose: the Gabor filter (best results), the Auto augmentation algorithm (excellent results), and the FFT Fast Fourier Transform (worst results).

Image Segmentation stage:

The enhanced images are divided and segmented according to the ROI of the image (only the two lungs) using the Thresholding method and the Marker-Controlled Watershed Segmentation method (which produces better results than Thresholding).

Features Extraction stage:

The better segmented image can be roughly seen using a Binarization and Masking Approach.

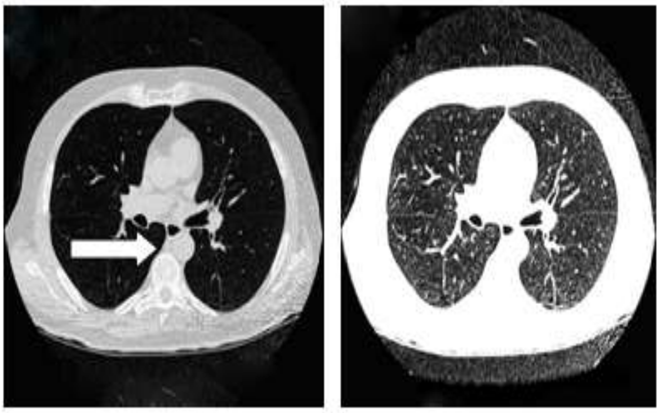
3. SYSTEM DESIGN

Image Enhancement

A Quick Survey To prepare an image for use as input by automated image processing software or for human comprehension, the first step in pre-processing is image enhancement. The two most common types of picture enhancement techniques are those that operate in the spatial domain and the frequency domain. There is no universally accepted theory that can be used to define what constitutes "good" picture enhancement, and individual opinions vary widely. Nice things are what they seem to be. In contrast, quantitative metrics can be utilized to zero in on the most productive picture enhancement strategies, laying the groundwork for more image processing. As part of this procedure, we applied a number of methods to enhance the photos, including the Gabor filter, Auto-enhancing, and Fast Fourier transform.

Gabor Filter

Gabor-based image display allows for simultaneous (and optimal) logon localization in the spatial and frequency domains, providing a local and multi-scale decomposition of logons. In order to determine the impulse response of a linear filter known as a Gabor filter, we multiply a harmonic function by a Gaussian function. The impulse response Fourier transform of a Gabor filter is the convolution of the Fourier transforms of the harmonic function and the Gaussian function, as stated by the multiplication-convolution property (Convolution theorem). Figure 2 shows both the unaltered (a) image and the improved (b) image that was created by applying the Gabor Filter.



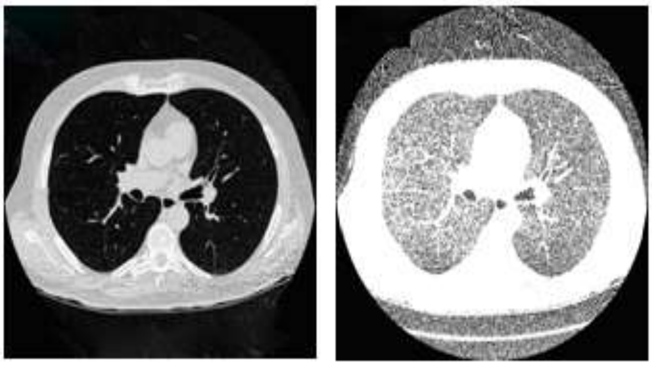
(a) Original Image (b) Enhanced by Gabor

Figure 2. The result of applying Gabor enhancement technique

The success of the auto-enhancement technique relies heavily on objective and subjective measurement of mean and variance. In this research, we calculated a rate of improvement of 38.025%.

Fast Fourier Transform

The Fast Fourier Transform processes an image by using the Fourier transform. The intensity values in a picture "I" change throughout a certain frequency range, which is represented by the values at each image position F in the frequency domain. Here, Fast Fourier Transform is used as a filter to enhance images. As can be seen in Figure 3, the FFT technique yields a 27.51% improvement over the original photos.



(a) Original Image (b) Enhanced by FFT

Figure 3. Auto enhancement technique using FFT

In Table 1, we can see how these three methods of improving images compare and contrast with one another. Table 1 shows that Gabor Enhancement is the best way to improve image quality. Images that have been processed in this way are improved in terms of clarity and contrast, but they also contain details that were not there before.

Table 1. Sub and final averages for three techniques used for image enhancement stage

Subject	Auto Enhancement	Gabor Filter	FFT Filter
Sub1	37.95	80.975	27.075
Sub2	47.725	80	36.825
Sub3	36.825	79.5	25.625
Sub4	34.775	81.8	25.175
Sub5	32.85	81.4	22.85
Final Average	38.025	80.735	27.51

Image Segmentation

Segmenting images is the first step in every image analysis project. The results of segmentation are used in many contemporary methods for describing and recognizing images, for instance. Segmentation is the act of separating out the parts of a picture. Two-dimensional medical picture segmentation has several clinical uses, including tumor and polyp detection, tissue quantification and classification, and organ visibility and volume estimation.

Segmentation is a method used to modify an image in some way that makes further analysis more manageable. Objects and boundaries (lines, curves, etc.) inside a picture can be identified with the help of image segmentation. picture segmentation is the process of separating a picture into groups of similar-looking, tagged pixels. In order to build contours from an image, it must first be segmented into smaller, more manageable sections or its borders must be identified. All of the pixels in a given area have the same value for a given computed characteristic, such as color, contrast, or texture. Even in relatively close regions, there can be significant differences in the same trait.

Segmentation algorithms rely heavily on two features of intensity values: their discontinuity and similarity. In the first method, the image is broken up according to its edges and other distinct changes in brightness. The second type employs a fixed set of rules to segment the image into coherent areas. Histogram thresholding is one of the methods included.

Thresholding approach

Thresholding is a powerful method for dividing up photos into distinct parts. When compared to gray level images, which typically have 256 levels, the thresholded image provides advantages in terms of storage space, processing speed, and simplicity of manipulation. This is why threshold strategies have been so prevalent for the past 20 years. The thresholding process, which divides a grayscale image into two groups, above and below the threshold value, is a non-linear one. In this study, we employ Otsu's technique of utilizing the (gray thresh) function to determine the global picture threshold. Otsu makes decisions about thresholds using statistical methods. Otsu proposed calculating an appropriate threshold by summing the within-class variances of object and background pixels and weighting the results. In other words, the maximum of the variation between classes is comparable to the minimum of the variation within classes, as we know from experience. When working with bimodal histogram images, this approach performs admirably. In this technique, the threshold value will be between 0 and 1, resulting in a divided image at that moment. The thresholding approach has visible results, as seen in Figure 4.

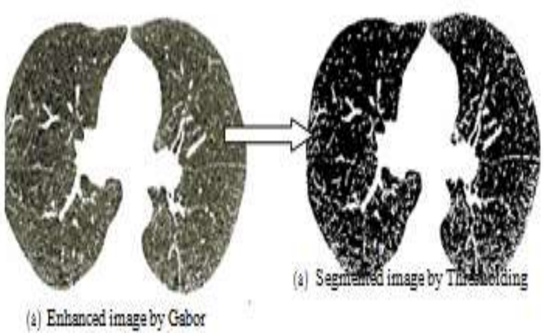


Figure 4. Normal enhanced image by Gabor filter and its segmentation using Thresholding approach

Marker-Controlled Watershed Segmentation Approach

The seeds that tell us whether a given area of an image is foreground or background can be extracted using a marker-driven watershed segmentation technique. The watershed technique is then used to set markers at regions with a local minimum on the topological surface (often the initial input image's gradient). The watershed transform is commonly used to solve the challenging challenge of separating touching items in an image. When looking at a watershed, the external marker-controlled method takes into account factors in the broader context, while the internal marker-controlled method narrows in on the study's foci.

If we can "mark" foreground items and background locations, the watershed transform for image segmentation will be able to locate "catchment basins" and "watershed ridge lines" by analyzing the image as a surface with bright and dark areas. Figure 5 depicts watershed-based picture segmentation.

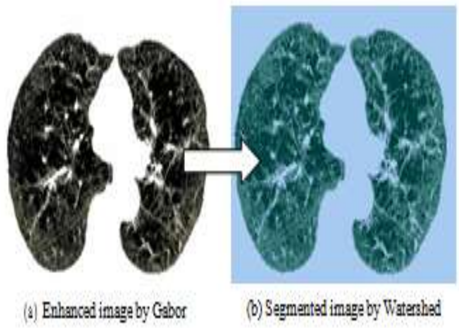


Figure 5. Normal Enhanced image by Gabor filter and its Segmentation using Marker-Controlled Watershed approach

Both the Thresholding and Marker-Controlled Watershed Segmentation techniques are displayed in Table 2 alongside the results of a subjective experimental evaluation of the segmentation step.

Table 2. Image Segmentation experimental result

Subject	Thresholding	Watershed Filter
Sub1	81.625	85.375
Sub2	82.2	85.25
Sub3	82.125	85.55
Sub4	81.725	84.75
Sub5	81.5	84.9
Final Average	81.835	85.165

Features Extraction

Superb visuals Specific characteristics of an image are isolated and removed during the extraction process, which makes use of algorithms and techniques. Binarization and masking both use data closely related to lung anatomy and information of lung CT imaging to make predictions about the existence of lung cancer.

Binarization Approach

To create a threshold, we started tallying the number of black pixels in both normal and abnormal images. If the number of black pixels in a new image is greater than the threshold, we classify it as normal; otherwise, we label it as abnormal. This is because in typical lung scans there are a lot more black pixels than white ones. We find a TAR of 92.86%, FAR of 7.14%, and FAR of 0.7% when using a cutoff of 17178.48. The binarization procedure and the binarization check are depicted in Figure 6 and a flowchart, respectively, below.

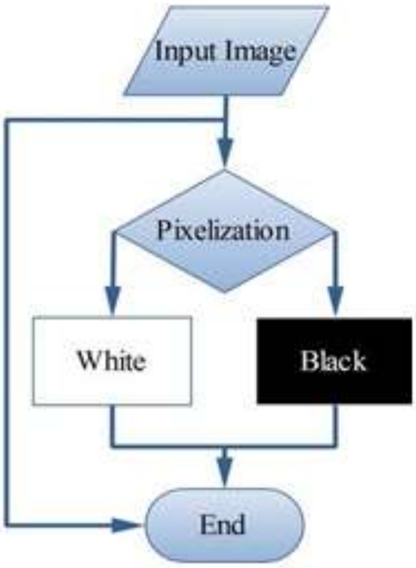


Figure 6. Binarization method procedure

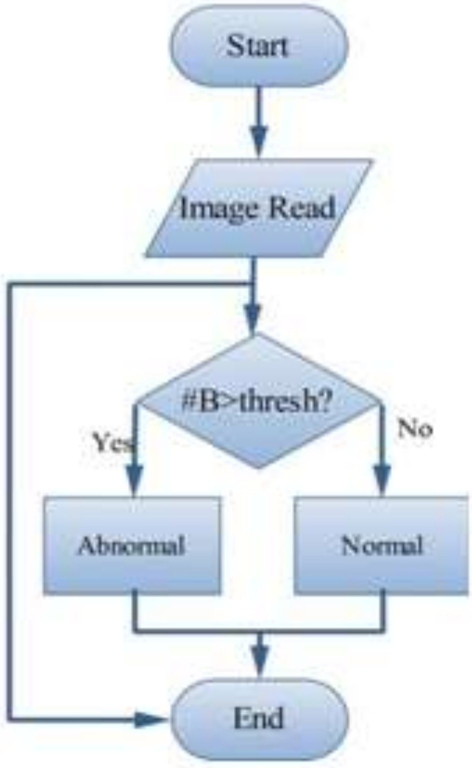
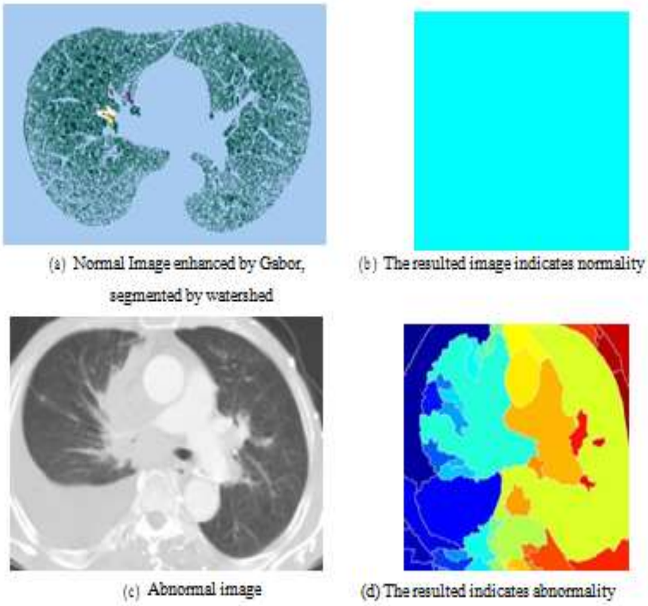


Figure 7. Binarization check method flowchart

Masking Approach

The masking method makes use of the fact that lung cancer tumors typically manifest as white, linked patches within the lung ROI. The TAR of this approach is 85.7%, while the FAR is only 14.3%, because a uniform blue color is indicative of health, and the emergence of RGB masses is diagnostic of malignancy. In Figure 8, we see both typical and abnormal examples of what can happen when the Masking tool in MATLAB is used.



(a) Normal Image enhanced by Gabor, segmented by watershed

(b) The resulted image indicates normality

(c) Abnormal image

(d) The resulted indicates abnormality

Figure 8. Normal and abnormal images using Masking approach

Using the assumptions from both Binarization and Masking, we can determine if the current state of affairs is abnormal. Everything is good if the image has more black than white pixels; otherwise, there is an issue.

4. SYSTEM ANALYSIS

RANDOM FORESTS

Random forests are quite effective in statistical learning. Bagging (in which we build a number forest of decision trees using bootstrapped training samples) is superseded by this method due to the fact that the trees are de-correlated. Construct a number of decision trees using bootstrapped training samples. We randomly select a subset of p predictors (often m = p) to act as split candidates when selecting how to split a tree into two branches. Let's imagine there's a reliable predictor available. The predictive power of the remaining components varies. The first-split strong predictor is present in the vast majority of bagged trees. Each packed tree will look and perform similarly to the others. Averaging heavily connected data is inefficient when seeking to reduce variation. To remove the correlation between bagged trees, a random forest is used.

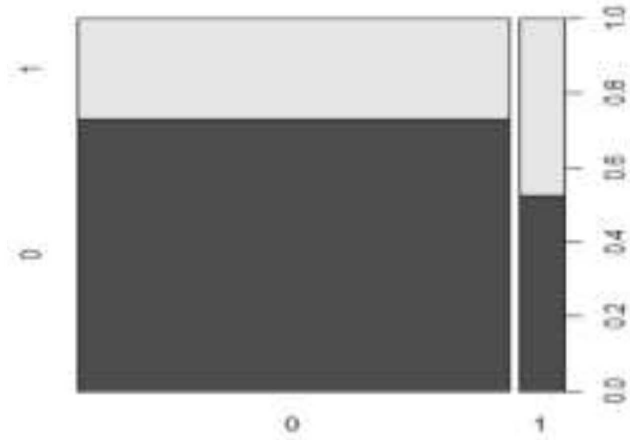


Fig.9. Classification using random forest

SUPPORT VECTOR MACHINE (SVM)

The accuracy of SVM was 71.71 percent before we tweaked the cost and gamma parameters. Next, we tuned the two parameters until we found that cost=1 and gamma=1 gave the best results. Our accuracy improved to 72.22 percent as a result. Figure 10 shows how SVM has been used to make predictions based on area and perimeter.

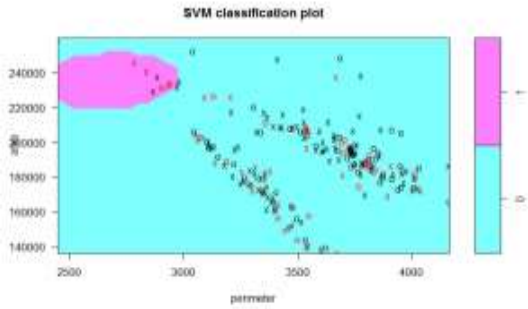


Fig.10. Classification using support vector machine

K-MEANS CLUSTERING

When applied to training data, k-means clustering achieved an accuracy of 52.97 percent when utilizing all predictors, while it achieved an accuracy of 54.67 percent when using just three. An accuracy of 47.47% was achieved by using all predictors on the test data, whereas an accuracy of 55.05% was achieved by using just three. Figure 11 shows k-means grouping based on area and perimeter.

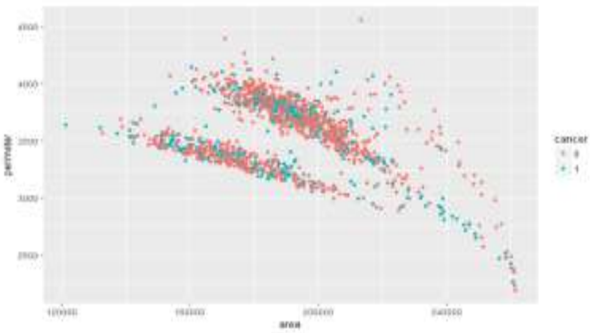


Fig. 11. Classification using k-means clustering

Figure 12 displays the overall accuracy. Total accuracy (blue) and accuracy with only three predictors (orange). In contrast, SVM employs color to denote the two alternative cost and gamma parameters. A closer inspection of the graph reveals that SVM achieves higher accuracy than QDA, classification tree, and random forest.

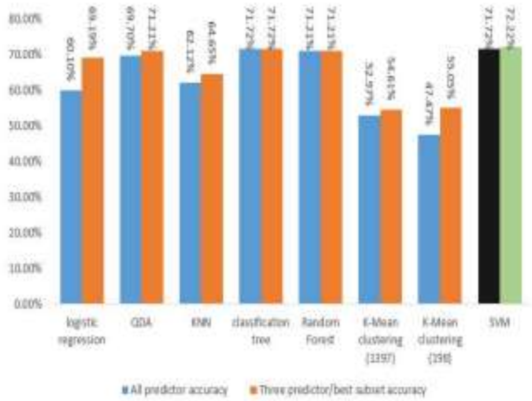


Fig. 12. Bar graphs comparing the accuracy level

5. CONCLUSIONS

As part of a novel image enhancement strategy for earlier disease identification and treatment, time was taken into account to find anomalous concerns in target images. This research aims to find cost-effective pre-processing methods based on the Gabor filter and the use of the Gaussian distribution that can guarantee high levels of image quality and precision. The suggested approach uses segmentation concepts to effectively generate features based on regions of interest. The results of the proposed method are very promising when compared to those of other methods. Using generalizable features allows for a normality comparison. Accurate picture comparison utilizing pixel percentage and mask-labeling is characterized by its high accuracy and reliable operation, two of its recently discovered qualities.

REFERENCES

1. Non-Small Cell Lung Cancer, Available at: http://www.katemacintyrefoundation.org/ pdf/non-small-cell.pdf, Adapted from National Cancer Institute (NCI) and Patients Living with Cancer (PLWC), 2007, (accessed July 2011).
2. Tarawneh M., Nimri O., Arqoub K., Zaghal M., Cancer Incidence in Jordan 2008, Available at: http://www.moh.gov.jo/MOH/Files/Publication/Jordan%20Cancer%20Registry_2008%20Report_1.pdf, 2008, (accessed July 2011).
3. Lung Cancer Database, Available at: https://eddie.via.cornell.edu/cgi-bin/datac/signon.cgi, (accessed July 2011).
4. Gonzalez R.C., Woods R.E., Digital Image Processing, Upper Saddle River, NJ PrenticeHall, 2008.
5. https://www.kaggle.com/c/data-science-bowl-2017
6. ilya levner, hong zhangm ,classification driven watershed segmentation , ieee transactions on image processing vol. 16, no. 5, may 2007
7. http://in.mathworks.com/help/images/examples/markercontrolled watershed-segmentation.html
8. Sivakumar, S. and Chandrasekar, C., 2013. Lung nodule detection using fuzzy clustering and support vector machines. International Journal of Engineering and Technology, 5(1), pp.179-185.
9. Jayaraj, D. and Sathiamoorthy, S., 2019, November. Random Forest based Classification Model for Lung Cancer Prediction on Computer Tomography Images. In 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 100-104). IEEE.
10. Neelima, S. and Asuntha, 2016. Image

Processing Used for Lung Cancer Detection in Medical Imaging. Journal of Chemical and Pharmaceutical Research, 4(8), pp.1044-1049.

11.]Manju, B.R., Athira, V. and Rajendran, A., 2021, January. Efficient multi-level lung cancer prediction model using support vector machine classifier. In IOP Conference Series: Materials Science and Engineering (Vol. 1012, No. 1, p. 012034). IOP Publishing.
12. Witschi, H. A short history of lungs cancer, Toxicol Sci, 2001. 64(1): p. 4-6.
13. Furat, et al., Machine-Learning Techniques for the Segmentations of Tomographic Images Data of Functional Materials, Frontiers in Materials, 2019. 6(145). https://doi.org/10.3389/fmats.2019.00145
14. E.R.M., The roles of computed tomography in the diagnosis and management of lungs cancer: an overview, The American College of Chest Physicians. Published by Elsevier Inc., 1986. 89(4): p.