



Application of K-MTh Algorithm for Accurate Lung Cancer Detection

Aouf, Mohamed*
Higher Technological Institute Tenth of
Ramadan City, Cairo, Egypt
maoufmedical@yahoo.com

Abstract. The primary goal of this paper is how to detect the lung cancer tumor based on the K-means clustering thresholding (K-MTh) segmentation technique, rate of death from lung cancer diseases is increased in the last years so the discovered of lung cancer early can protect a lot of people from died . The technique of image processing is utilized ,in image processing ,there are a lot of steps are done for improving the performance of medical diagnostic machine .The technique which used is considered very important to classify the degree of a tumor by improving the thresholding technique before using the classification methods such as support vector machine (SVM). Actually, we have applied Gabor Gaussian filtration method to improve and denoise the CT-image, then we applied the segmentation method (K-MTh) and SVM. Finally, the system has been achieved accuracy more than have been expected for classification method after applying K-MTh (more than 90%).

Keywords: — Lung cancer, Gabor filter; SVM; Thresholding; FFT; K-means; Clusters.

1 Introduction

Nowadays, Computer vision plays important role in medical image analysis. It is really significant, especially in the diagnosing process [1]. The most effective role of computer vision and image processing is how to increase the pace of healing by using early diagnosis. Tumors are considered to be one of the most popular diseases which, depend on image processing for diagnosis.

In early researches, Image processing improves the methods of classifications and detections of tumors to aid the doctors in there diagnosis. Researchers have tried in each research to improve the methods of extractions or classifications. In the following, we will show some of the last researches and how the authors try to improve methods to classify and detect the tumors such as Lung tumors or brain tumors [2-5].

Some of the researchers have focused on any image of cancer and how to detect the degree of tumor, such as have been shown in [6]. In [6] authors have described some of the techniques for segmentation and features exactions of any tumor by using a CAD system. They have attempted to extract features of a tumor and actually, they have achieved 100% in detections.

On the other hand, in [7] authors have applied Gabor wavelet to extract features of MRI –images. They also have compared results with statistical features methods. The comparing technique for the feature extraction method was founded on some classifiers. Authors have evaluated the methods and Gabor wavelet achieved 95.3% based on some of classification techniques. In [8], the authors have introduced the technique of using scanning electrochemical microscopy to detect the lung cancer cells and they reported and described how it was more practical in cancer diagnosis.

Nowadays, the hybrid technique plays important role in biomedical processing. In [9], the authors have achieved how to apply a hybrid method based on some of information and support vector machine (SVM) to classify the types of tumors. The accuracy of this method has achieved 90.3% for some specified types of cancer. Some of the researchers focus only on the methods of classifications and how to apply it such in [10]. In [10], the authors have presented the main studying of support vector machine (SVM) and how to apply it for diagnosing diseases and how to improve it.

Further, in [11], the authors have applied new model based on the latent support vector machine as a model for cancer prognosis. The operations of this model have based on microarray operations and some of the gene expression and they improved the algorithm and technique of this model.

There are some authors have invested new approaches for cancer features classifications such as in [12]. Actually, in [12] the authors have applied ANFIS (Neuro fuzzy) as classification method. They have used the fuzzy system to detect the severity of the lung nodules depend

on IF-Then rule method. They also have applied 150 images in computer added diagnosis (CAD) system. They have achieved sensitivity of 97.27%, specificity of 95% with an accuracy of 96.66%. Furthermore, in [13], the authors have applied the machine learning to detect Leukemia based on machine learning image processing. They have applied the analysis of Leukemia image and they extracted features and then classified by SVM. There were some authors have presented a survey about the most recent of image segmentation processing, especially for a medical images such as in [14]. In [14], the authors have introduced three methods of segmentation based on the features of images. They also have evaluated the advantages and disadvantages of the three methods. Furthermore, the authors have introduced the recent researches on the segmentation of the image and the best method of segmentation for medical images. On the other hand, in [15, 16] the authors have applied the detection of lung cancer based on the Computer Aided Detection (CAD) system. Actually, the thresholding and morphological is the main operations of the segmentation process in this system. Moreover, the system has achieved a sensitivity of 77.7%, specificity of 69.5 %and accuracy of 70.53% with an average 4.1Fps/scan.

Finally, after we have displayed the important ideas of the previous studies, we have chosen the Lung Cancer CT-image for our application. The research of our paper will focus on how to improve the thresholding segmentation technique using K-Means cluster (K-MTh) to achieve high accuracy in classifications. It is an unsupervised learning algorithm that classifies the input data points into multiple classes based on their inherent distance from each other and it's also referred to as hard clustering [17]. On the other hand, we will apply the Gabor filter (Gaussian kernel) before K-MTh to denoise CT-image. Further, we chose SVM for classification method to evaluate our new segmentation process. The organization of this paper is as follows; in section two we introduce our system methodology and concept of the work. In section three we discuss the results of the system proposed compared to other research work. At the end, we conclude with our conclusion and suggesting directions for future work.

2 Methodology

Our system consists of three main units (the preprocessing, the KMTh segmentation process and the classification process). Fig (1) shows the block diagram of the system and we will describe and identify each block briefly in the following sections.

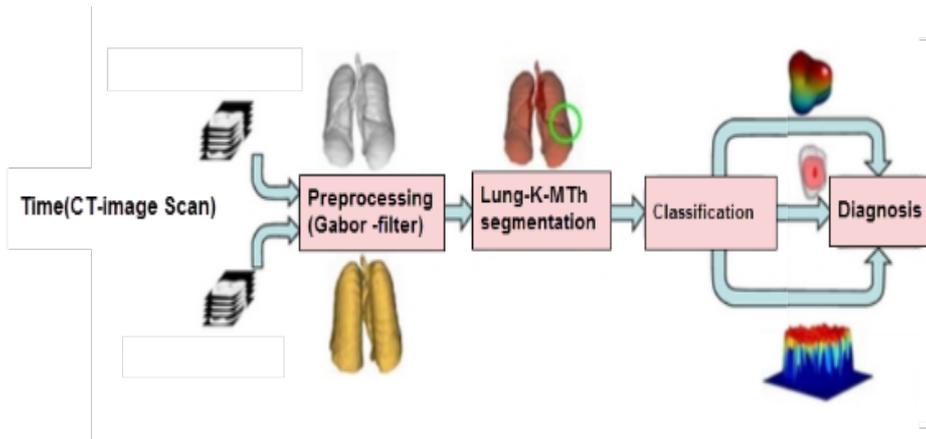


Figure (1) Block diagram of the system

2.1 Preprocessing

In this stage, CT-image will divide into two steps: image denoising and image enhancement

a) Image denoising

In this step, we have used Gabor filter based on a Gaussian kernel. Gabor filter in image processing, is a bandpass filter which is used for feature extraction, and texture analysis [18]. The impulse response of this filter is created by multiplying a Gaussian envelope function with a complex oscillation. Gabor [19] showed that the space (time)-uncertainty product is minimized by these elementary functions. It is possible to create filters which are selective for orientation, by extending these functions to two dimensions [20]. Under certain conditions the phase of the response of Gabor filters is approximately linear. It was claimed by some authors that simple cells in the visual cortex of mammalian brains can be modeled by Gabor functions [20- 22]. Thus, it is believed by some that image analysis with Gabor filters, to be similar to perception in the human visual system. The characteristic of the Gabor filter is being shown in equation (1) [23]. Fig (2) shows the Gabor filter denoising and enhancement of the CT scan image of our case.

$$g_e(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-x^2}{2\sigma^2} \cos(2\pi w_0 x)} \quad (1)$$

b) Image enhancement

The main purpose of image enhancement is to process a given image so that the result is suitable to the original image for a specific application.

- The enhancement doesn't increase the inherent information content of the data, but it increases the dynamic range of the chosen features so that they can be detected easily [12].



Figure (2) (a) after enhancement

2.2 K-MTh Segmentation

At this point, we have applied the thresholding based on the Otsu's method. After that, we have improved this method by applying K-means cluster technique. K-means is one of the most practical techniques in clustering methods. The K-means clustering is an unsupervised learning algorithm, and an iterative technique that is used to partition an image into K clusters, where K is a positive integer, according to some similarity feature like gray level intensity of pixels and distance of pixel intensities from centroid pixel intensity [17, 24]. Partitioning the objects into mutually exclusive clusters (K) is done with it in such a way that objects within each cluster remain as close as possible to each other but as far as possible from objects in other clusters. The main idea is to define the number of K centroids, one for each cluster. These centroids should be placed in cunning or a clever way because different location causes different result. Thus, the more beneficial choice is to set them far away from each other as much as possible [17].

The main advantages of K-mean clustering algorithm are its simplicity (easy to implement) and low computational cost, which permit it to function efficiently on large data sets. The main drawback is that K the number of clusters must be determined beforehand [25], it does not yield the same result each time the algorithm is executed and the resulting clusters depend on the initial assignments of centroids. Actually, this approach works in the gray level and also can work in RGB. We have achieved high accuracy in both approaches when we have applied them in the classification stage [26-29]. The following algorithm of applying K-MTh before classifications:

Main algorithm of System block diagram

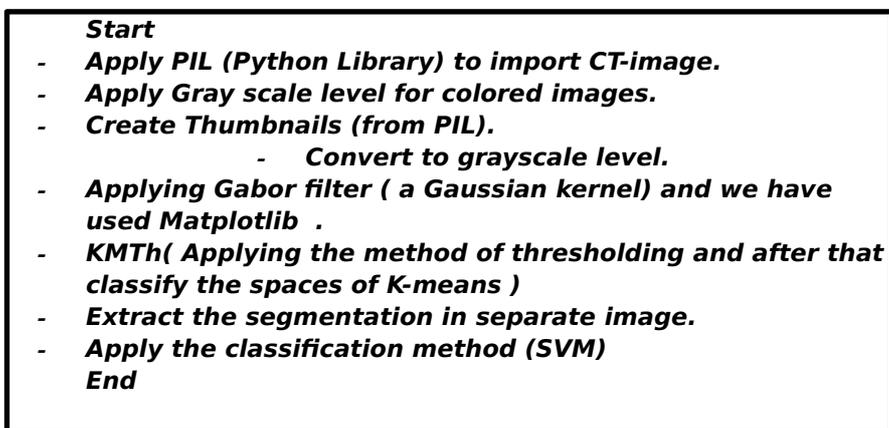


Figure (3) has showed the result of K-MTh segmentation. The general flow chart of our system has been shown in Fig (4)



Figure (3) segmentation after KMTh from code result

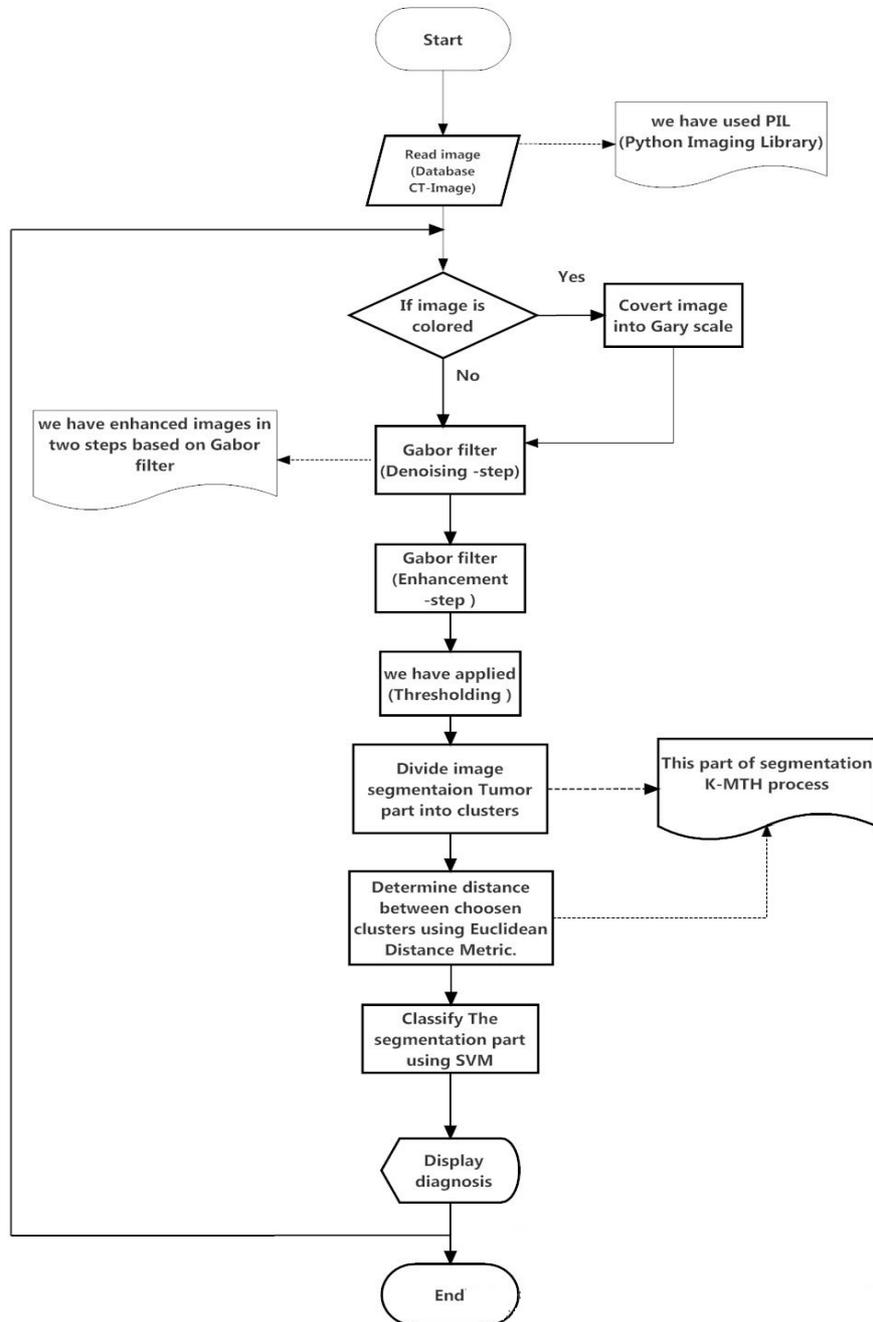


Figure (4) Flow chart of our system

2.3 Classification

The classification method is one of the most important stages to show the degree of tumor. In this stage, we have classified the degree of lung tumor based on support vector machine (SVM) [30, 31]. Actually, in this part, the accuracy of classification has been depending on the K-MTh segmentation process. The K-MTh has helped to classify the type of tumor Malignant or benign. The percentage of accuracy ranges from 85% to more than 91%.

3 RESULTS AND DISCUSSION

We have utilized our system based on open access databases Lung cancer images. From pervious section, we have applied Gabor filter Method to filter CT-images of lung cancer. The second step, we have applied K-MTh and extract features from the segmented image of the Lung cancer. On the hand, there was problem with k-means clusters faced us in the structure of K-means because the most cases of clusters based on RGB but actually when we have been applying it after thresholding method especially "Otsu's method", it has been achieved in high accuracy.

On the other hand, the classification methods can achieve high accuracy, but based on the good detection of the main features of Lung image. Further, In our case, the K-MTh have achieved good accuracy raised more than 91% in SVM compared to other research work such as in [9].

CONCLUSION

This paper aims to how to improve the segmentation process to classify the lung cancer from CT-image with high accuracy. We have proved that our method covered the learning technique in classification based on K-MTh segmentations. Ultimately, this work can improve in the future to apply the deep-learning technique in classification and detection.

References

1. DEVI, M. Anousouya, et al. Classification of cervical cancer using artificial neural networks. *Procedia Computer Science*, 89: 465-472(2016).
2. ADALI, Terin; ŞEKEROĞLU, Boran. Analysis of micrnas by neural network for early detection of cancer. *Procedia Technology*, 1: 449-452(2012).
3. CHARRON, Odelin, et al. Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network. *Computers in Biology and Medicine*(2018).
4. YANG, Xin, et al. Co-trained convolutional neural networks for automated detection of prostate cancer in multi-parametric MRI. *Medical image analysis*, 42: 212-227(2017).
5. MAKAJU, Suren, et al. Lung Cancer Detection using CT Scan Images. *Procedia Computer Science*, 125: 107-114(2018).
6. LEE, Howard; CHEN, Yi-Ping Phoebe. Image based computer aided diagnosis system for cancer detection. *Expert Systems with Applications*, 42.12: 5356-5365 (2015).
7. NABIZADEH, Nooshin; KUBAT, Miroslav. Brain tumors detection and segmentation in MR images: Gabor wavelet vs. statistical features. *Computers & Electrical Engineering*, 45: 286-301 (2015).
8. NING, Xin, et al. Simultaneous detection of tumor markers in lung cancer using scanning electrochemical microscopy. *Journal of Electroanalytical Chemistry* (2018).
9. GAO, Lingyun, et al. Hybrid Method Based on Information Gain and Support Vector Machine for Gene Selection in Cancer Classification. *Genomics, proteomics & bioinformatics*, 15.6: 389-395 (2017).
10. SWEILAM, Nasser H.; THARWAT, A. A.; MONIEM, NK Abdel. Support vector machine for diagnosis cancer disease: a comparative study. *Egyptian Informatics Journal*, 11.2: 81-92(2010).
11. Ford, William, and Walker Land. "A Latent Space Support Vector Machine (LSSVM) Model for Cancer Prognosis." *Procedia Computer Science* 36 : 470-475(2014).
12. KHARE, Rakesh Kumar; SINHA, G. R.; KUMAR, Sushil. Cancer Detection Using Neuro Fuzzy Classifier in CT Images (2017).
13. Jagadev, Preeti. automated detection and classification of leukemia using image processing and machine learning.
14. Salem, M. A. M., Atef, A., Salah, A., & Shams, M. Recent survey on medical image segmentation. *Computer Vision: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*, 129 (2018).
15. El-Regaily, S. A., Salem, M. A., Abdel Aziz, M. H., & Roushdy, M. I. Lung Nodule segmentaion and Detection in Computed Tomography: In: 8th IEEE International Conference on intelligent computing and information systems (ICICs 2017), (2017).
16. El-Regaily, S. A., Salem, M. A., Abdel Aziz, M. H., & Roushdy, M. I. Survey of Computer Aided Detection Systems for Lung Cancer in Computed Tomography. *Current Medical Imaging Reviews*, 14(1), 3-18 (2018).
17. Bharati R.Jipkate, and V.V.Gohokar. A Comparative Analysis of Fuzzy C-Means Clustering and K Means Clustering Algorithms. *IJCER*, 2.3: 737-739 (2012).
18. Aach, T. Kaup, A., Mester, R. On texture analysis: Local energy transforms versus quadrature filters. In *Signal Processing*, 45:173-181(1995).
19. Gabor, D. Theory of communication. In *J. IEE*, 93: 429-457 (1946).
20. Daugman, J. G. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *J. Optical Society of America A*, 2.7:1160-1169 (1985).
21. Andrews, B. A., Pollen, D. A. Relationship between spatial frequency selectivity and receptive field profile of simple cells. In *Journal of Physiology*, 287:163-176 (1979).
22. Marčelja, S. Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America*. 70.11: 297–1300 (1980).
23. ADAK, Chandranath. Gabor filter and rough clustering based edge detection. In: *Human Computer Interactions (ICHCI), 2013 International Conference on. IEEE*, 1-5(2013).
24. V. K. Dehariya, S. K. Shrivastava, R. C. Jain. Clustering of Image Data Set Using K-Means and Fuzzy K-Means Algorithms. *International conference on CICN*, :386- 391(2010).

25. Soumi Ghosh, and Sanjay Kumar Dubey. Comparative Analysis of K-Means and Fuzzy C-Means. IJACSA, 4.4: 35-39 (2013).
26. Shafique, Sarmad; Tehsin, Samabia. Computer-Aided Diagnosis of Acute Lymphoblastic Leukemia.
27. M. M. D. Joshi, P. A.H.Karode and P.S.R.Suralkar, "Detection of Acute Leukemia Using White Blood Cells Segmentation Based on Blood Samples," International Journal of Electronics and Communication Engineering & Technology (IJECET), vol. 4, no. 3, pp. 148-153 (2013).
28. Bhattacharjee, Romel and L. M. Saini, "Detection of Acute Lymphoblastic Leukemia using watershed transformation technique," Signal Processing, Computing and Control (ISPC), 2015 International Conference on. IEEE, pp. 383-386 (2015).
29. L. M. Al-Hayali, D. M. Morsy and M. M. Abdul-Razak, "Microscopic Image Processing Of Automated Detection And Classification For Human Cancer," International Journal of Scientific & Technology Research, vol. 5, no. 9, pp. 231-236 (2016).
30. Ramesh K. Kulkarni and Niket Amoda, "Image Segmentation and Detection using Watershed Transform and Region Based Image Retrieval," International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), vol. 2, no. 2, pp. 89-94, April (2013).
31. G. Singh, P. Swarnalatha, B.K. Tripathy and S. Kakani, "Convex Hull Based WBC Coputation For Leukemia Detection," International Journal of Advanced Research in Electrical, Electronics and Instrumentation.