Mehran University Research Journal of Engineering & Technology
Vol. 38, No. 2, 351-360 April 2019
p-ISSN: 0254-7821, e-ISSN: 2413-7219
DOI: 10.22581/muet1982.1902.10

Lung Cancer Detection Using Digital Image Processing Techniques: A Review

MUTIULLAH*, MEHWISH BARI**, ADEEL AHMED***, MUHAMMAD SABIR*, AND SAJID NAVEED*

RECEIVED ON 09.03.2018 ACCEPTED ON 25.05.2018

ABSTRACT

From last decade, lung cancer become sign of fear among the people all over the world. As a result, many countries generate funds and give invitation to many scholars to overcome on this disease. Many researchers proposed many solutions and challenges of different phases of computer aided system to detect the lung cancer in early stages and give the facts about the lung cancer. CV (Computer Vision) play vital role to prevent lung cancer. Since image processing is necessary for computer vision, further in medical image processing there are many technical steps which are necessary to improve the performance of medical diagnostic machines. Without such steps programmer is unable to achieve accuracy given by another author using specific algorithm or technique. In this paper we highlight such steps which are used by many author in pre-processing, segmentation and classification methods of lung cancer area detection. If pre-processing and segmentation process have some ambiguity than ultimately it effects on classification process. We discuss such factors briefly so that new researchers can easily understand the situation to work further in which direction.

Key Words: CT-Scan Images Enhancement, Feature Extraction, Watershed Segmentation Thresholding, Matching, Multi-Label Classification.

1. INTRODUCTION

Cancer comes from the uncontrolled division of abnormal cells in any part of the body and in some cases these cell also spread in other parts of the body and consequently become responsible for multiple diseases. Moreover, there are many types of cancer such that breast cancer, lung cancer, ovarian cancer, cervical cancer, brain cancer etc. Lung cancer is major type of cancer which is becoming the leading cause of deaths in the current generation of the United State. Primary lung cancer is Carcinoma. There are two types of lung cancer namely SCLC (Small-Cells Lung Carcinoma) which is related to smoking of any types and the other one is NSCLC (Non-Small Cell Lung Carcinoma), 10-15% lung cancer diagnostics as SCLC while 85% cases of lung cancer are NSCLC [1]. Among multiple symptoms of lung cancer, some common symptoms are cough that get worse, chest pain, weight loss, shortness of breath, coughing of blood and weariness etc. In order to overcome the death rate due to lung cancer, one should adopt screening, chest radiograph (X-Ray), MRI (Magnetic Resonance Imaging)

Authors E-Mail (mutiullah@cs.qau.edu.pk, mehwishbari@gmail.com, adeel.ahmed@cs.qau.edu.pk, xajidnaveed@gmail.com)
* Department of Computer Science, National College of Business Administration & Economics, Bahawalpur, Pakistan.
** Department of Mathematics, National College of Business Administration & Economics, Bahawalpur, Pakistan.
*** Department of Computer Science, Quaid-e-Azam University, Islamabad, Pakistan.

and CT (Computed Tomography) scan. The detection and diagnose of lung cancer can be processed on three basic stages which are pre-processing, segmentation and finally followed by post-processing. The CT scan image is pre-processed to remove Gaussian white noise using non-local mean filter technique [2]. This paper concentrated on various techniques utilized in pre-processing and segmentation process since 2011-2018 is given in Table 1. The given below third column techniques are useful to perform the task in fourth column applications.

TABLE 1. TECHNIQUES USED IN IMAGE PROCESSING FROM YEAR 2011-2018.

Techniques	Applications	References
Gabor Filter	Optical character recognition	Al-Tarawneh [1]
Image Processing and Classification	Remove Gaussian white noise	Malik et. al. [2]
Weiner Filter	Noise reduction,Signal detection.	Sharma and Jindal [3]
Layer Separation	Used to separate layer of image	Rani [4]
Gray scale Image	Used to convert color in gray	Gajdhaneand Deshpande [5]
Enhancement	Used to sharpen the image	Patiland Jain [6]
Gabor filter	Feature extraction	Onizawa et. al. [7]
Gabor Filters, Discrete Wavelet Transform and Auto Enhancement Algorithm	Identify Cancerous Cells	Avinashet. al. [8]
Fast Fourier Transform	Image reconstruction	Gauthier et. al. [9]
Sparsity-based image modeling	Image Layer Separation	Gu et. al. [10]
Edge detection-based methods	Lane edge detection	Yan and Li [11]
	Canny algorithm	
Matching	Local matching	Nagao et. al. [12]
	3D Elastic matching	
Classification	Cellular dependency	Shao et. al. [13]
Support Vector Machine, Fuzzy C-Mean, Conventional Neural Networkand Computer Aided Design	Segmentation	Al-Zubaidi et. al. [14]
Wiener filter	Image Restoration	Zubair [15]
Gray conversion	Histogram equalization	Li et. al. [16]
Image segmentation	Labeling	Berahim et. al. [17]
Thresholding	Deep learning algorithms and convolutional networks	Talukdar et. al. [18]
Region-based segmentation	Region growing	Jain and Laxmi [19]
	Region splitting and merging	
Clustering techniques	Seed Point Selection Algorithm	Chowdhury et. al. [20]
Morphological segmentation	Watershed algorithm	Wang et. al. [21]
	Cell nuclei	
Weibull multiplicative model	Image Segmentation	Chouhan et. al. [22]
Marker-controller segmentation	Magnetic Resonance Imaging	Vesal et. at. [23]
	Watershed	
Classification	Support Vector Machine	Chandra et. al. [24]
Classification	Supervised and Unsupervsed Tumor Characterization	Hussein et. al. [25]
Classification	Multi-label Classification	Read et. al. [26]

2.1 Pre-Processing of Lung Cancer

Pre-processing is an improvement of the image data that suppresses unwanted distortions or enhances some image features for further processing.It is needed to minimizing the effects of distortion found in imaging device such as light fluctuation, to remove blueness and in the same time pre-processing is required to remove unwanted areas from the images and some time it is used for enhancing the image features like lines, boundaries and textures of image so that we can easily divide the contents of images in two parts, wanted and un-wanted contents of image. For removing noise from the image, many researchers use different filtering techniques which depends on type of noise. In medical imaging all types of filtering techniques may be used depending on noise present in image [1]. Detail is given below:

(a) **Gaussian Noise**: Outside the Normal distribution values, usually we cannot see in the image.

(b) **Salt and Paper Noise**: Tiny white and black points randomly appear in the image.

(c) **Poisson Noise**: In Poisson distribution, mean and variance are equal. Noise is present due to non-linear response of image detectors and recorders.

(d) **Impulse Noise:** Usually it appears in the result of electromagnetic interference, scratches on the recorded disks

(e) **Speckle Noise:** Appearance ofwaves which are found in many microscopic diffused reflections which create hurdles to understand the image components. This noise follow Gamma distribution found in ultrasound, SAR (Synthetic Aperture Radar) and CT scan images.

De-Noising techniques categorized in two parts [2-3].

(i) Spatial Domain Filtering

(a) Linear Filter i.e. Wiener Filter or Mean Filter

(b) Non-Linear Filter i.e. Median Filter

(ii) Transform Domain Filtering i.e. Wavelet Transform

2.2 Image Enhancement

In the process of features extraction from the image it is necessary that all properties of an object present in the image should be clear. So, for growing the digital objects we need image enhancement [3]. Image enhancement is divided into two categories. First one is spatial domain and other is frequency domain.

In spatial domain, operations are performed on the pixel values directly so it is easy to understand and analysis. While in frequency domain the methods is used to explain the analysis of signals and mathematical formulas with respect to frequency and function. Image enhancement is achieved when we are able to interpret and threshold the image into two parts, one is known as ROI (Region of Interest) and second is compliment of it.

2.3 Gabor Filer

Gabor filter is belonging to the linear filter mentioned in Section 2.1. It is used to analyze the texture patterns in terms of pixel values so it is very helpful to analyze the lung cancer texture within the image. Gabor filter is used in many texture analysis applications by many researchers. For example, face recognition and vehicle veriûcation [4-5].

2.4 Wiener Filter

Wiener filter is also linear filter belong to special domain. By the help of this filter we can get uncorrupted image with minimal error. In this technique we take 9 or sixteen neighborhood values and take mean and also replace central value with mean value and repeat this process to all over the image. In case of nine values, boundary of the image is not replaced by new value so it means if the tumor is present at the boundary of the image then we are unable to enhance that region unless we add dummy values with all boundary of the image(Fig. 1(a-c)) [6-7].

2.5 Fast Fourier Transform

FFT (Fast Fourier Transform) algorithm takes the signals in specific time or space and divide it into its frequency forms. FFT lies in frequency domain and helpful in that modularity in which we get images from the response of signals like ultrasound images and MR (Magnetic Resonance) images. FFT is commonly used to enhance medical images [8].

2.6 Layer Separation

There are many imaging model like RGB (Red, Green, Blue) HSV (Herpes Simplex Virus), HIS (Humane Society International), YCbCr (Luminance; Chroma: Blue; Chroma: Red) and Lab model. In RGB model, RGB are also called color Channels. Sometime we take single channel which is consistent in different situation or distortion during the imaging process. The separation of a channel from other channels is called layer separation, which is key step to produce the satisfactory results [9].

3. GRAY CONVERSION

Medical image datasets are found in number of different formats. For the simplicity we divide images in three following types:

(i) Color image

(ii) Gray image

(iii) Binary image

Color image has three channels each channel has 256 variant intensity values ranging 0-255. In RGB if RGB have zero value then color is black while all channel values equal to 255 then color is white. Computer can easily distinguish each color shade while human vision is limited as compare to machine vision. In color image there are almost 16777216 different shade. In gray image there are 256 shades in the form of black and white where 0 represents black color and 256 represents white color (Fig. 2).

(a) INPUT IMAGE (b) OUTPUT OF GAUSSIAN FILTERED IMAGE (c) SEGMENTED LUNG

FIG. 1. STEPS INVOLVED IN THE SEGMENTATION OF SUSPECTED LUNG MODELS

In the other hand, binary image has only two shades. 0 is used for Black and 1 is used for white. Usually we convert color image to gray image for extracting image objects from the whole image by adjusting histogram of the image [10].

3.1 Image Segmentation

The process of subdividing the image into two parts one is wanted and other is unwanted parts. Wanted part is also called ROI. In the case of lung cancer our target is to identify the tumor present in the Lung image (Fig. 3(a-b)). So tumor is our ROI and other part is unwanted area [11]. It is crucial task for machine to automatically detect the tumor because of variant texture properties of Lung tissue in abnormal region such as reûective of tumors or level of a cancer malignancy [12]. For segment the image there are two strategies found in various literature such as edge based segmentation and region based segmentation. Region based segmentation is further divided as:

- Thresholding,
- Region Growing
- Clustering in Feature Space

3.2 Thresholding

A value which is able to segment the ROI is called threshold value. In lung cancer case we convert our image into gray image and then decide the threshold value on the basis of tumor properties. For example, we found that tumor has intensity values 115-255 then we can convert all values which are less than 115 into 0 and all other values between 115-255 is equal to 1 in binary image. In this way all unwanted area become black and wanted area become white region [13].

3.3 Region Growing

In region growing method we group the pixels or subparts of the image into the large region, means to say we analysis

FIG. 2. GRAY LEVELS OF GRAY IMAGE

FIG. 3(a). SHOWS ORIGINAL LUNG CT SCAN IMAGE

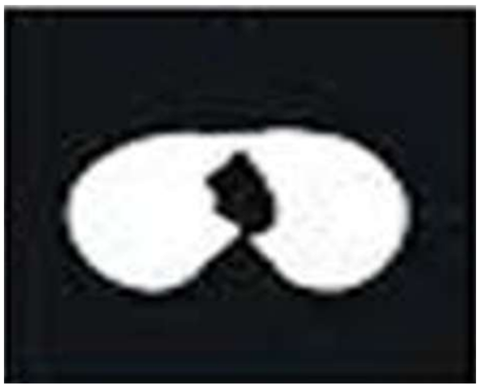

FIG. 3(b). SHOWS SEGMENTED CT SCAN IMAGE the pixel value and then we combine the various intensity values which are similar properties such as color, gray level values, texture and shape features. While in region splitting method we take large image and then split it into small region on the basis of homogeneity of that region [14].

3.4 Clustering Techniques

In clustering base segmentation, first of all find out the number of features like mean, median, standard deviation, quartiles, and texture features like homogeneity, entropy, correlation, and wavelet or regression parameters and then group the region depending on the minimum distance within the region. Region based clustering give better results in medical image processing and also detecting lung cancer tumor in the image [15-17]. Clustering techniques are considered as unsupervised classification methods in which k-means clustering algorithm is very famous in machine learning techniques where the output is not given in the features vector (Fig. 4) [18].

3.5 Edge Based Segmentation Methods

In edge based segmentation we try to find edges of digital objects found in an image on the basis of abruptly change found in intensity values of a pixel (Fig. 5). Many digital objects have cloudy and broken edges. To remove such effects from the image we usually performs morphological operations which are discussed in next section and linear filters to give more distance and remove distance which is closed to the digital object [16,19]. Because tumor has no regular shape and intensity range so it is challenging task for CAD (Computer Aided Design).

3.6 Morphological base Segmentation

Morphological base segmentation is a process that combine morphological operations such as extended minima and morphological gradient with watershed flooding algorithm to segment grayscale image of any type of image. For example, gray scale with single band (8-bit image) and color image with three bands (24-bit image). In morphological operation, popular operations are Dilation and Erosion. Erosion operation is performed when two edges are merged with each another whereas Dilation operation is performed when there is broken edges are found in the image. This operation tries to fill the gap between two broken edges [20-21].

3.7 Template Matching

In template matching method we take a sample or template and match it with all possible sub-image within the whole image. This process is continue until the template is exactly matched in term of intensity values with any portion of an image [22]. There are two types of matching Local Matching and 3D Elastic Matching. The Local Matching method based on GGVF (Generalized Gradient Vector Flaw). One of the GGVF's features is to extend the influence of gradient also to position away from the edge and a homogeneous region while leaving desirable properties of vector field on the edge map. The 3D Elastic matching method based on smoothing shift vector. This method is minimization of energy function by sequentially updating the shift vector on the alignment points from initial vector. This initial vector is taken from shift vector obtained from local matching.

3.8 Marker-Controller Segmentation

To reduce the drawback of over segmentation in watershed algorithm, many researcher use marker controller segmentations in which there are two types of markers, one is internal marker and other is external marker. In internal marker we give the range of gray level values of required region while in external marker we give the range of gray values related to back ground of the image [23].The marker selection typically consists of two steps: one is preprocessing and other is definition of a set of criteria that markers must satisfy. The preprocessing scheme is to filter an image with a smoothing filter. This step can minimize the effect of small spatial detail, in other words, this step is to reduce the large number of irrelevant detail, which is the reason of over-segmentation. Many researchers used marker controller segmentation in lung cancer detection in medical imaging.

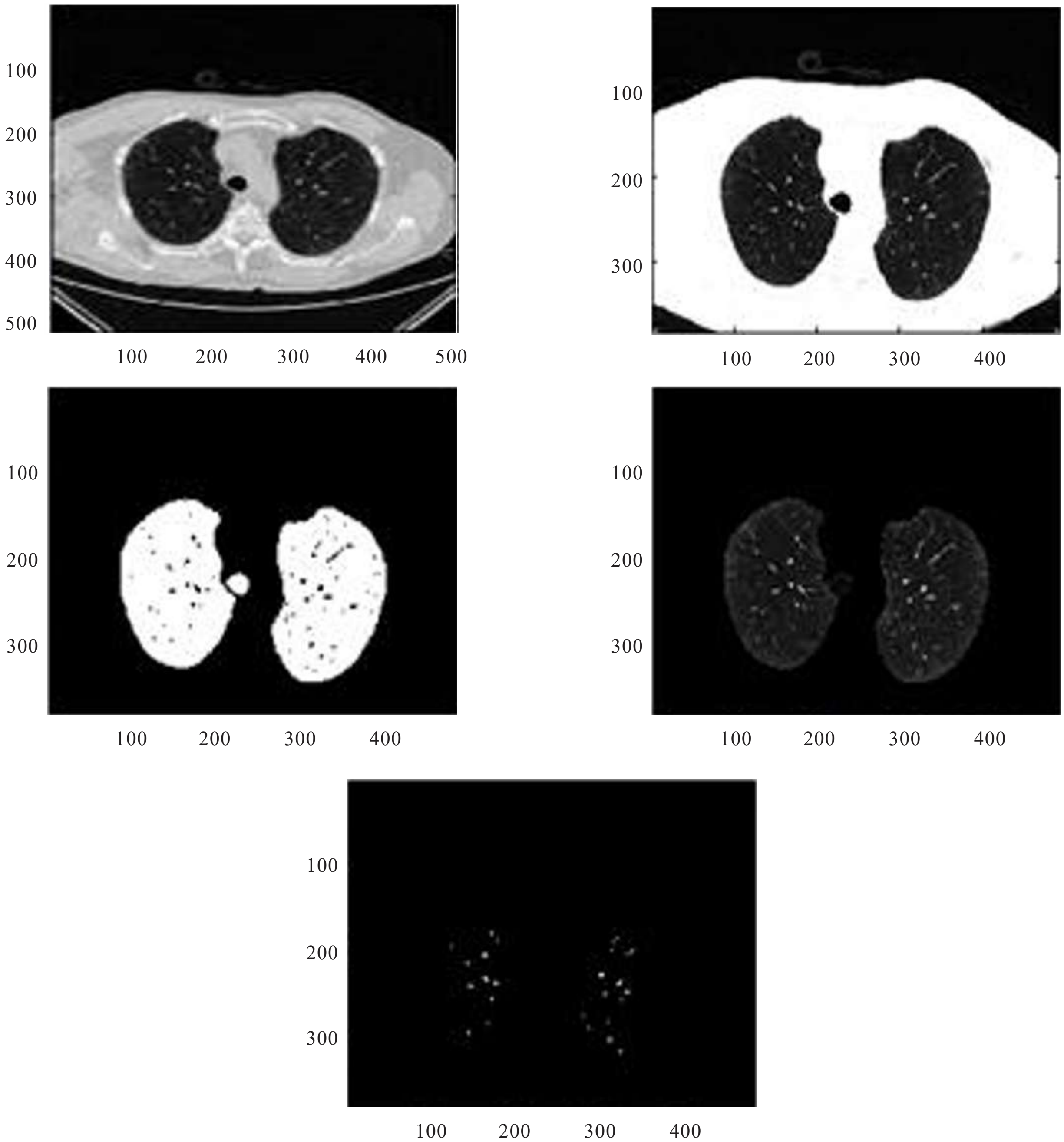

FIG. 4. MATLAB BASED SEGMENTED LUNG IMAGES BY DIFFERENT THRESHOLD VALUES

3.9 Feature Extraction

Each well define object has some features or properties which helps us to identify that object easily. To automatically classify the objects, we need some features. Collection of these features are called features vector. For the identification of lung cancer, we need to be find such features from the image. In image processing there are three types of features, first is structural features, second is statistical texture features and third is spectral features. Structural features are also called binary features like area, centroid, perimeter, orientation, projection, aspect ratio and Euler number etc. Statistical features also divide two classes' first order and second order statistical texture features. First order features are extracted directly from the gray level histogram while in second order texture features first we find co-occurrence matrix and then we find mean, entropy, and co-variance. In spectral features Gaber and wavelet features are very popular features. Many authors take different set of features to detect the lung cancer with different methodology with different available dataset but still there is no any method which can automatically segment and classify the malignant and benign area in the Lung cancer image. However, there are two approaches are used to extract the features vector first Banalization approach related to structural features which are calculated from binary image and other is masking approach in which we usually extract statistical texture and spectral features.

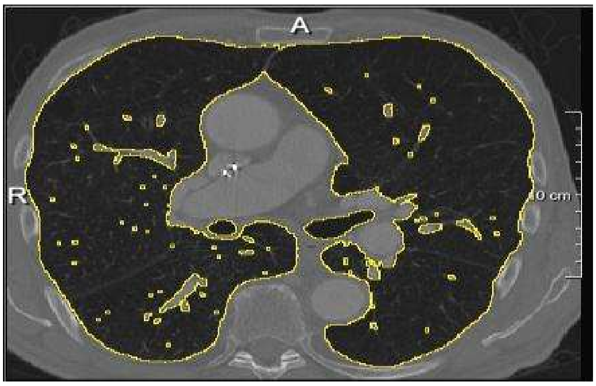

FIG. 5. EDGE BASED LUNG SEGMENTATION

3.10 Banalization

The banalization algorithm is also known as the threshold algorithm. The purpose of this algorithm is to find the suitable threshold [19]. By use of this threshold value the targeted image is divided into two parts foreground and background image. Foreground image is representing by white and background is representing by black color. This approach is based on the number of black pixels which are greater than white pixels if not then it means image is not normal.

3.11 Masking

Masking is also called filtering the word "filtering" has been borrowed from the frequency domain. Filters are classified as:

Low-Pass (i.e. preserve low frequencies)

High-Pass (i.e. preserve high frequencies)

Band-Pass (i.e. preserve frequencies within a band)

Band-Reject (i.e. reject frequencies within a band) [20].

In masking approach, we take usually neighborhood pixels of CT scan images of lung cancer with effected areas and then calculate the statistical and structural features which are used to classify Lung cancer images in early stage and normal images as well.

3.12 Classification

Human brain is a great example of ability to learn and classify the different objects and things without any error. For example, distinctive mankind's face and pictures. At the same time, it still challengeable task for the human to get the ability via machines. However, there are lot of techniques have found in literature for data classiûcation. But no single classiûcation technique has been found with best accuracy for all kinds of dataset [24]. Two main approaches are used to make a machine capable to understand and classify the CT scan images are supervised and unsupervised classification. If the feature vectors are given with output label to classifier than it is called supervised classifier. KNN (K-Nearest Nabors), ANN (Artificial Neural Network) and SVM (Support Vector Machine) are the examples of supervised learning approach used in many medical image classifications. In other hand if feature vector is given to the classifier without output label then such classifier is called unsupervised classifier. K means, mixture models, hierarchical clustering, Principal Component Analysis, Singular Value Decomposition are the example of unsupervised learning methods [25].

3.13 Multi-Label Classification

Multi-label classification in concerned with the classiûcation about information for different class labels. MLC (Multi-Label Classification) is a major research area in the machine learning community and ûnds application in several domains. MLC problem by adopting a group-sparsity regularize to select common subsets of relevant features from different cellular compartments. In addition, we also add a cell structural correlation regularized Laplacian term, which utilizes the prior biological structural information to capture the intrinsic dependency among different cellular compartments [26].

4. CONCLUSION

An effort to diagnose the cancer in the lung using digital image processing techniques. Initially the CT scan/MRI captured the image and processed then cancer or tumor region is detected and diagnosis from the original image. In pre-processing phase, number of techniques are being used for filter the distortion of the images, for example Weiner Filter, Fast Fourier Transformation etc. After pre-processing, image segmentation phase should be done through Marker-Controlled Watershed and Watershed Segmentation. The next phase is features extraction is done by Area, Perimeter, and Eccentricity etc. These features help to detect the lung cancer. In classification, SVM (Support Vector Machine) and CNN (Convolutional Neural Network) are used for classification of normal and abnormal area in early diagnoses.

5. FUTURE WORK

In future researchers will try to build Image Processing Techniques on X-Ray, PET images for more accuracy. For the classification KNN or fuzzy, logic k/c-mean and clustering may be used. Furthermore, contrast between X-Ray and CT scan images for better result for early lung cancer diagnose.

ACKNOWLEDGEMENT

This work is supported by National College of Business Administration & Economics, Sub-Campus, Bahawalpur, Pakistan.

REFERENCES

[1] Al-Tarawneh, M.S., "Lung Cancer Detection Using Image Processing Techniques", Leonardo Electronic Journal of Practices and Technologies, Volume 11, No. 21, pp. 147-58, 2012.

[2] Malik, B., Singh, J.P., Singh, V.B.P., and Naresh, P., "Lung Cancer Detection at Initial Stage by Using Image Processing and Classification Techniques", Lung Cancer, Volume 3, No. 11, 2016.

[3] Sharma, D., and Jindal, G., "Identifying Lung Cancer Using Image Processing Techniques", International Conference on Computational Techniques and Artificial Intelligence, pp. 115-120, 2011.

[4] Rani, J., "Noise Removal in Medical Images Using Filters", International Journal of Engineering Research & Technology, Volume 2, pp. 1013-1016, March, 2013.

[5] Gajdhane, M.V.A., and Deshpande, L., "Detection of Lung Cancer Stages on CT scan Images by Using Various Image Processing Techniques", International Organization of Scientific Research Journal of Computer Engineering, pp. 2278-0661, 2014.

[6] Patil, B.G., and Jain, S.N., "Cancer Cells Detection using Digital Image Processing Methods", International Journal of Latest Trends in Engineering & Technology, Volume 3, 2014.

[7] Onizawa, N., Katagiri, D., Matsumiya, K., Gross, W.J., and Hanyu, T., "Gabor Filter Based on Stochastic Computation", IEEE Signal Processing Letters, Volume 22, No. 9, pp. 1224-1228, 2015.

[8] Avinash, S., Manjunath, K., and Senthilkumar, S., "Analysis and Comparison of Image Enhancement Techniques for the Prediction of Lung Cancer", 2ndIEEE International Conference on Recent Trends in Electronics, Information & Communication Technology, pp. 1535-1539, 2017.

[9] Gauthier, J.-P., Stephant, N., Rondeau, B., Cody, J.A., and Fritsch, E., "Aluminium Diboride-Type Structure in Ethiopian Opal-CT Revealed by Fast Fourier Transform", Journal of Applied Crystallography, Volume 51, No. 1, pp. 27-34, 2018.

[10] Gu, S., Meng, D., Zuo, W., and Zhang. L., "Joint Convolutional Analysis and Synthesis Sparse Representation for Single Image Layer Separation", IEEE International Conference on Computer Vision, pp. 1717-1725, 2017.

[11] Yan, X., and Li, Y., "A Method of Lane Edge Detection Based on Canny Algorithm", IEEE, pp. 2120-2124, 2017.

[12] Nagao, M., Miyake, N. Yoshino, Y., Lu, H., Tan, J.K., Kim, H., Murakami, S., Aoki, T., Hirano, Y., and Kido, S., "Detection of Abnormal Candidate Regions on Temporal Subtraction Images Based on DCNN", Proceedings of 17th International Conference on Control, Automation and Systems, pp. 1444-1448, 2017.

[13] Shao, W., Liu, M., Xu, Y.-Y., Shen, H.B., and Zhang, D., "An Organelle Correlation-Guided Feature Selection Approach for Classifying Multi-Label Subcellular Bio-images", IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2017.

[14] Al-Zubaidi, A.K., Sideseq, F.B. Faeq, A., and Basil, M., "Computer Aided Diagnosis in Digital Pathology Application: Review and Perspective Approach in Lung Cancer Classification", Annual Coference on New Trends in Information & Communication Technology, Application, pp. 219-224, 2017.

[15] Zubair, A.R., "Frequency Domain Image Restoration", International Journal of Computer and Information Technology, Volume 7, No. 2, pp. 64-73, 2018.

[16] Li, Y., Qian, M., Liu, P., Cai, Q., Li, X., Guo, J., Yan, H., Yu, F., Yuan, K., and Yu, J., "The Recognition of Rice Images by UAV Based on Capsule network", Cluster Computing, pp. 1-10, 2018.

[17] Berahim, M., Samsudin, N.A., and Nathan, S.S., "A Review: Image Analysis Techniques to Improve Labeling Accuracy of Medical Image Classification", Recent Advances on Soft Computing and Data Mining, pp. 298-307, 2018.

[18] Talukdar, J., and Sarma, D.P., "A Survey on Lung Cancer Detection in CT scans Images Using Image Processing Techniques", International Journal of Current Trends in Science and Technology, Volume 8, No. 3, pp. 20181-20186, 2018.

[19] Jain, S., and Laxmi, V., "Color Image Segmentation Techniques: A Survey", Proceedings of International Conference on Microelectronics, Computing & Communication Systems, pp. 189-197, 2018.

[20] Chowdhury, K., Chaudhuri, D., and Pal, A.K., "Seed Point Selection Algorithm in Clustering of Image Data", Progress in Intelligent Computing Techniques: Theory, Practice, and Applications, pp. 119-126, Springer, 2018.

[21] Wang, P., Xu, S., Li, Y., Wang, L., and Song, Q., "Feature-Based Analysis of Cell Nuclei Structure for Classification of Histopathological Images", Digital Signal Processing, 2018.

[22] Chouhan, S.S., Kaul, A., and Singh, U.P., "Image Segmentation Using Computational Intelligence Techniques", Archives of Computational Methods in Engineering, pp. 1-64, February, 2018.

[23] Vesal, S., Ravikumar, N., Ellman, S., and Maier, A., "Comparative Analysis of Unsupervised Algorithms for Breast MRI Lesion Segmentation", Bildverarbeitung Für Die Medizin, pp. 257-262, 2018.

[24] Chandra, M.A., and Bedi, S., "Survey on SVM and Their Application in Image Classification", International Journal of Information Technology, pp. 1-11, 2018.

[25] Hussein, S., Chuquicusma, M.M., Kandel, P., Bolan, C.W., Wallace, M.B., and Bagci, U. "Supervised and Unsupervised Tumor Characterization in the Deep Learning Era", arXiv preprint arXiv:1801.03230, 2018.

[26] Read, J., Pfahringer, B., Holmes, G., and Frank, E., "Classifier Chains for Multi-Label Classification", Machine Learning, Volume 85, pp. 333, 2011.