International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Lung Cancer Examination and Risk Severity Prediction using Data Mining Algorithms

Dr. P. Thamilselvan*[1]

Submitted: 10/09/2022 **Accepted**: 20/12/2022

Abstract: Lung cancer is one among the primary wellsprings of disease passing, worldwide and even more particularly in India. As the symptoms of the lung cancer can't show explicitly, early disclosure of lung tumor is enchant; the endurance pace of lung disease is more, if it is found early. To propel the early finding of lung tumor, the patient should insight screening quickly after the secondary effects are taking note. In this paper, a construction for risk seriousness expectation of lung cancer is proposed to further develop the forecast exactness of the lung tumor. To improve the nature of the image and to remove the commotion by proposed Profuse Grouping Algorithm for Image Denoising. After completion denoising stage, the denoised images are tested with Enhanced k-nearest neighbor method for detecting the cancer. To increase the segmentation process Advanced Classification and Regression Tree algorithm is used to segment the lung cancer properly. At last, Fuzzy logic method has been used to find the detection level of the lung cancer and to identify the risk severity of the lung.

Keywords: *Classification and Regression Tree, Detection, Fuzzy Logic, K-Nearest Neighbor, Particle Swarm Optimization, Profuse Clustering, Segmentation, Severity Prediction*

1. Introduction

In image processing concept pre-processing [1] stage is a significant stage in various functions such as classification strategies, image denoising techniques and etc. In recent years pre-processing and super pixel classification has improved among numerous scientists. The nature of the image is impacted by various antiquities like non-uniform power, varieties moving, shift what's more clamor. So the picture is handled by specific strategies like thresholding, histogram adjustment and so forth, to eliminate repetition present in the checked pictures without influencing the elements of the picture in the discovery of lung cancer [2]. The attractive reverberation imaging (X-ray) [3] result can be used to make images of any piece of the human body and it gives a useful and easiest way for identifying the different cancer. Computer aided design framework helps the doctors to take choices and goes about as a 'second per user' in the analysis of diseases. Of the current imaging modalities, Computed Tomography (CT) is a clear cut one dependent on exactness [6]. In the information mining, K Nearest Neighbour is one of the top most calculations for order. The methods like order and identification are utilized generally utilized in the clinical imaging, for the most part in the Lung cancer order and identification [5]. In the clinical field, the difficult task is the ID of the underlying stage lung disease [4]. For the underlying stage recognition of cancer patients, the mechanized cancer, finding framework depends on the AI is more significant than others.

[1] *Assistant Professor, Department of Computer Science, Bishop Heber College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli – 620017.*
thamilselvan1987@gmail.com
ORCID ID : 0000-3343-7165-777X

2. Literature Review

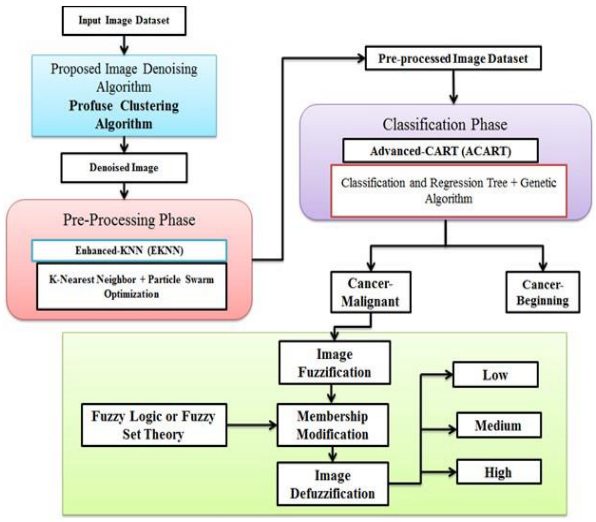
Firmino et. al. [16] proposed computer aided diagnosis method for finding cancers in CT lung images. Shankar et. al. [17] implemented apriori algorithm to find the severity of the algorithm for predicting risk factors in hepatitis disease. McRoberts et. al. [18] developed k-nearest neighbor method for calculating forest biomass for airborne laser scanning records.
Moffy et al. [19] implemented image processing technique for separting benign and malignant lung tissues. Xu et. al. [20] anticipated deep convolutional neural network for classifying and segmenting in histopathological images. Aniket et. al. [21] developed k-means algorithm for segmenting lung nodules in CT images. Zou et. al. [22] SLIC superpixel algorithm for preprocessing and removing the noise from SAR images. Drouyer et. al. developed Sparse stereo method for hierarchical image segmentation [23]. Angalaparameswari et. al. deliberated median filter with canny edge detection algorithm for removing the unwanted noise [24].
Venkatesan et. al. proposed LMS algorithm for an efficient noise removal method [25]. Portilla et. al. deliberated wavelet transform method to scale mixtures gaussian and image denoising [26]. Yao et. al. developed non-local sparse model with low rank matrix for denoising the image [27]. Cruz et. al. implemented convolutional neural network method for image denoising [28].

3. Proposed Framework for Risk Severity Prediction of Lung Cancer

In this paper, a framework is to be proposed to predict the risk severity of lung cancer by utilizing the proposed techniques in the

image denoising, pre-processing stage [7], order stage [8] and the dynamic stage. In the picture denoising stage, abundant grouping calculation is implemented to eliminate the commotion in the picture. In pre-processing stage, Enhanced K-Nearest Neighbor is anticipated by hybridizing the K-Nearest Neighbor and Particle Swarm Improvement. In the order stage, Advanced Classification and Regression Tree are proposed by hybridizing the CART and Genetic Algorithm. To predict the risk seriousness of lung cancer, Fuzzy Logic is utilized in this proposed architecture. The proposed framework of whole work is shown in below figure 1.

Figure 1. Proposed Framework for Risk Severity Prediction of Lung Cancer



4. Proposed Profuse Clustering Technique for Image Denoising

Image Denoising is a significant pre-processing task before further processing of the image like classification, clustering, segmentation, highlight extraction, surface examination. The motivation behind denoising is to eliminate the commotion while holding the edges and other itemized includes however much as could reasonably be expected. This commotion gets presented during procurement, transmission and gathering and capacity and recovery processes. To improve the function of image compression with highest classification accuracy, the each pixesl are supplanted with super pixels. The main issues in the image processing applications, the calculation cost is higher in the preprocessing stage. Without affecting the classification accuracy, the superpixels have been created by numerous superpixels calculation yet SLIC provides better result. Still, there is an enhancements required for superpixels calculation as per decrease the expense of calculation and limit a

This proposed calculation is made out of Simple Linear Iterative grouping, K-Means Clustering, Fusing Optimization calculation. This Profuse grouping calculation is on the superpixel grouping. In this commitment, an informative picture is given with white Gaussian commotion. What's more the underlying superpixel is by Simple Linear Iterative Clustering (SLIC) calculation [9]. K Means bunching strategy [10] is utilized to bunch superpixel into K groups. The superpixel is refined by intertwining advancement calculation [11]. The proposed Profuse Clustering Algorithm are briefly explained in the previous work [12].

5. Proposed Enhanced K-Nearest Neighbor Algorithm

The choice of best k-esteem with the less misclassification rate, this calculation is proposed. For the tasks like order and bunching, the one of the most normally utilized calculation is K-Nearest Neighbor (K-NN) calculation. In this approach, each examples are tested and it is well-organized by the nearest k example depending on the preparation information. The distance valuation measured by Euclidean distance calculation method. The distance of this approach is determined dependent on the test image utilizing Distance weighted recipe as. Distance(x)=

$$d(p,q) = \sqrt{\sum_{i=1}^{n}(qi - pi)2} \text{..... (1)}$$

Where x and y are images, n is the amount of highlights. Hence, the nameless test esteems are chosen generally pertinent to the class from KNN calculation and it is utilized to observe the genuine worth from nameless example esteems. In this proposed strategy, k esteem is gotten by PSO. This PSO method is instated with a gathering of arbitrary particles (arrangements). The molecule's wellness esteem is assessed on each emphasis. Assuming it is the best worth the molecule has accomplished, the molecule stores the area of that esteem as best (molecule best). The area of the best wellness esteems accomplished by any molecule during any emphasis is put away as best (worldwide best).

International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

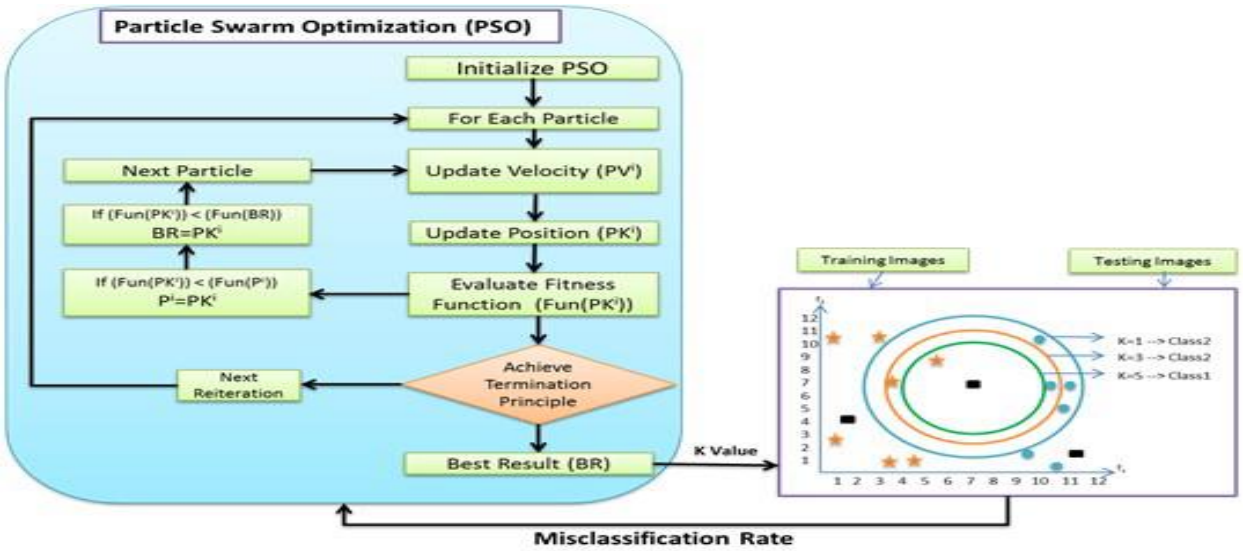


Figure 2. Conclude 'K' value using PSO for EKNN.

6. Advanced Classification and Regression Tree Algorithm

The accompanying figure 6 shows the proposed ACART. This ACART method is a non parametric number shuffling method and it convey staggered design of a tree. An average layout of portrayal and backslide tree yield is displayed in figure 3. This ACART start with one middle point', having the whole model, called a root center point. This procedure reviews each and every possible split likewise, picks the one from twofold social affairs that is fluctuating from one to another factor of part. The root center apportioned into two tyke center point considering picking free component. Gathering and backslide tree simply aspects root center point into two sub center points. This ACART computation gives information about the tumor tissue and obliging tissues and also hazardous tissues in the image data. In this ACART strategy, the boundaries and choice areas are determined by utilizing Genetic Calculation. This CART have the following phases:

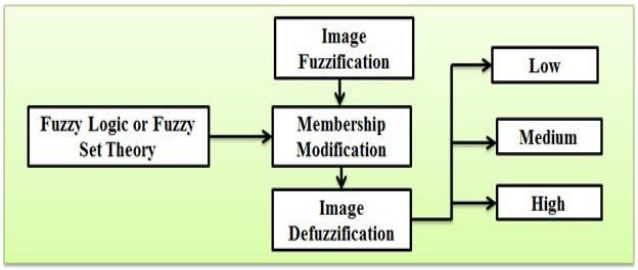


Figure 3. Framework for Proposed ACART method

7. Risk Severity Prediction by Fuzzy Logic

In the progression of edge identification, the fuzzy methodology can be used. Fuzzy image handling is the social affair of all techniques that cycle, comprehend, and address their highlights and portions as fuzzy logic. The handling and portrayal depend on the named fuzzy strategy to tackle the given issue. The three principle phases of fuzzy image handling are Image Fuzzification, Membership Values and Image Defuzzification. The accompanying figure 4 portrays the danger seriousness expectation of Lung Cancer by utilizing Fuzzy Logic.

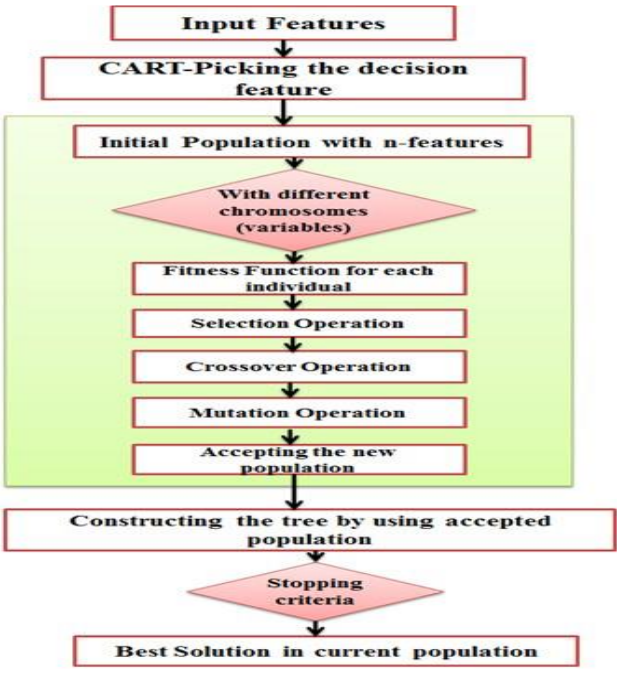


Figure 4. Frisk Severity Prediction of Lung Cancer by using Fuzzy Logic

In this strategy, it is needed to address the computerized image as

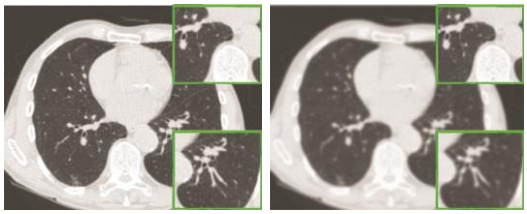
a fuzzy logic set and this process is called as fuzzification, at the same time image defuzzification is a process where the change the qualities. Back to fresh qualities in the spatial space. An image I of size M XN and L dim levels can be considered as a variety of fuzzy singletons. This enrollment process level of brilliance comparative with some splendor level l = 0, 1…. L-1. The enrollment work portrays a reasonable property of an image. As far as we might be concerned, in our technique, the property that is significant is Edginess. The way to decide find the pixel in on edge pixel gives us edge enrollment work. The Membership work isn't one of a kind, so there can be distinctive participation capacities, however they should meet the proper limitations. Bit by bit Procedure for Risk Severity Prediction by Fuzzy Logic.

7.1. Algorithm for Fuzzy Logic

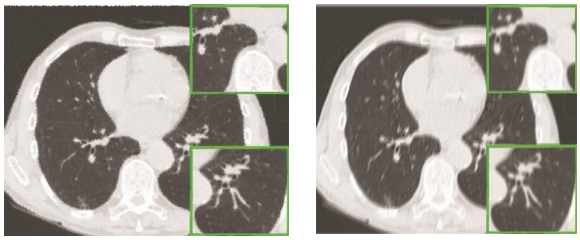
Step 1: Load malignant image as input and find the proportions M is equal to width, N is equal to Height.
Step 2: If it is a color image then it converted into grayscale image (i.e. black and white image). Iterating the pixel of image and set their intensities.
Step 3: Iterate all the pixels in the image and calculate the edge detection using μedge
Step 4: Every pixel should adjust by its membership rate wirh maximum value.
Step 5: For each pixel edges are created and to find out the value of grayscale

8. Result and Discussion of Proposed Clustering Algorithm

The proposed clustering technique is tested with lung cancer images with the noise level $\boldsymbol{\phi = 60}$ and the result is shown in Table1. Moreover the proposed method is compared with other denoising algorithm mentioned here K-SVD, BM3D and NLM. The implemented PCT technique removes the noise in the lung even in high level of noise. There are some statistics are used to measure the value of PSNR, SSIM, and FOM to approve the proposed profuse clustering technique. The FOM and PSNR give the greatest quality test. The noise level is set to 60 by proposed Profuse Clustering technique and the group number is set to 100, then, SSIM comes to their greatest worth. In an ideal arrangement, the noise level is set to 60. To calculate proposed method the profuse clustering calculation is contrasted and the following denoising calculations like NLM, K-SVD (K Means-Singular Vector Decomposition), BM3D and 3D Filtering).



a. Input Image b. Result image of K-SVD



c. BM3D d. NLM

d. Profuse Clustering Technique

Figure 5. Denoising evaluation in cancer image when level of noise is 60.

Table 1. Evaluation of the Proposed method with other methods

Methods used	$\boldsymbol{\phi = 60}$		
	PNSR (in dB)	**FOM**	**SSIM**
K-SVD	21.66	0.6874	0.4398
BM3D	22.57	0.5876	0.4234
NLM	18.20	0.4572	0.4321
Proposed PCT	23.73	0.6989	0.5671

Table 2. Noise removing using PSNR with different methods

Level of Noise (ϕ)	Methods used					
	K-SVD	**BM3D**	**NLM**	**Proposed PCT**		
5	36.54	37.50	32.65	**38.71**		
15	29.16			31.74	26.34	**31.78**
25	26.56			26.61	23.11	**29.71**
40	23.78			24.24	20.12	**26.32**
60	20.21			22.14	17.98	**24.78**

Table 3. Noise removal using SSIM with other methods

Level of Noise (ϕ)	Methods used			
	K-SVD	**BM3D**	**NLM**	**Proposed PCT**
5	0.7654	0.8652	0.8762	**0.9167**
15	0.6198	0.6621	0.5899	**0.8652**
25	0.5342	0.6142	0.4682	**0.7583**
40	0.6373	0.5782	0.4727	**0.7892**
60	0.5737	0.6838	0.3662	**0.8366**

Table 4. Noise removal of FOM with other methods

Level of Noise (ϕ)	Methods used			
	K-SVD	**BM3D**	**NLM**	**Proposed PCT**
5	0.8464	0.8464	0.84649	**0.9747**
15	0.8369	0.8363	0.7837	**0.9836**
25	0.7383	0.8263	0.6828	**0.8943**
40	0.7363	0.7373	0.5212	**0.8538**
60	0.5373	0.8363	0.3869	**0.9363**

From all the above tables 2, 3 and 4 address the upsides of PSR, FOM and SSIM in different noise levels. In table 1, the deliberated method gives the higher result. In the table 2, 3, 4 the implemented strategy gives the higher worth of PSNR, SSIM and FOM at different noise level. It is presumed that the presentation measurement gives the most extreme qualities just even in high noise level. The implemented EKNN strategy has been utilized in the pre-handling stage to channel the boisterous information what's more to work on the nature of the image.

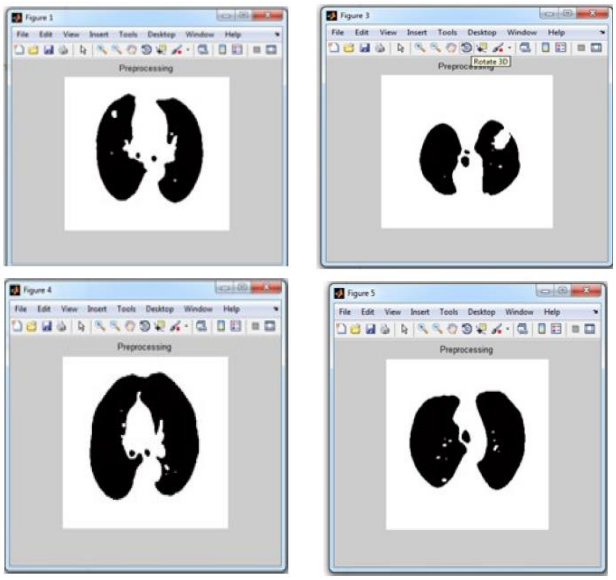


Figure 6. Result obtained by Particle Swarm Optimization, K-Nearest Neighbour and Enhanced K-NN

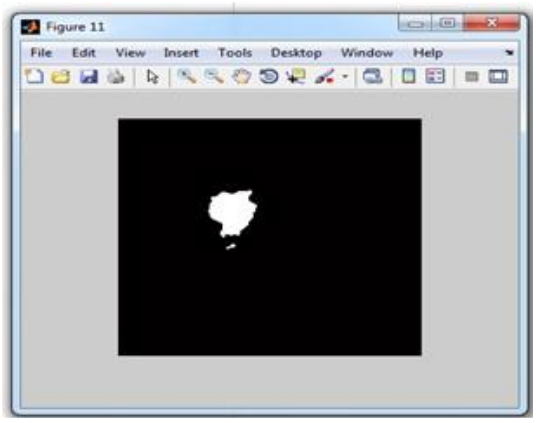
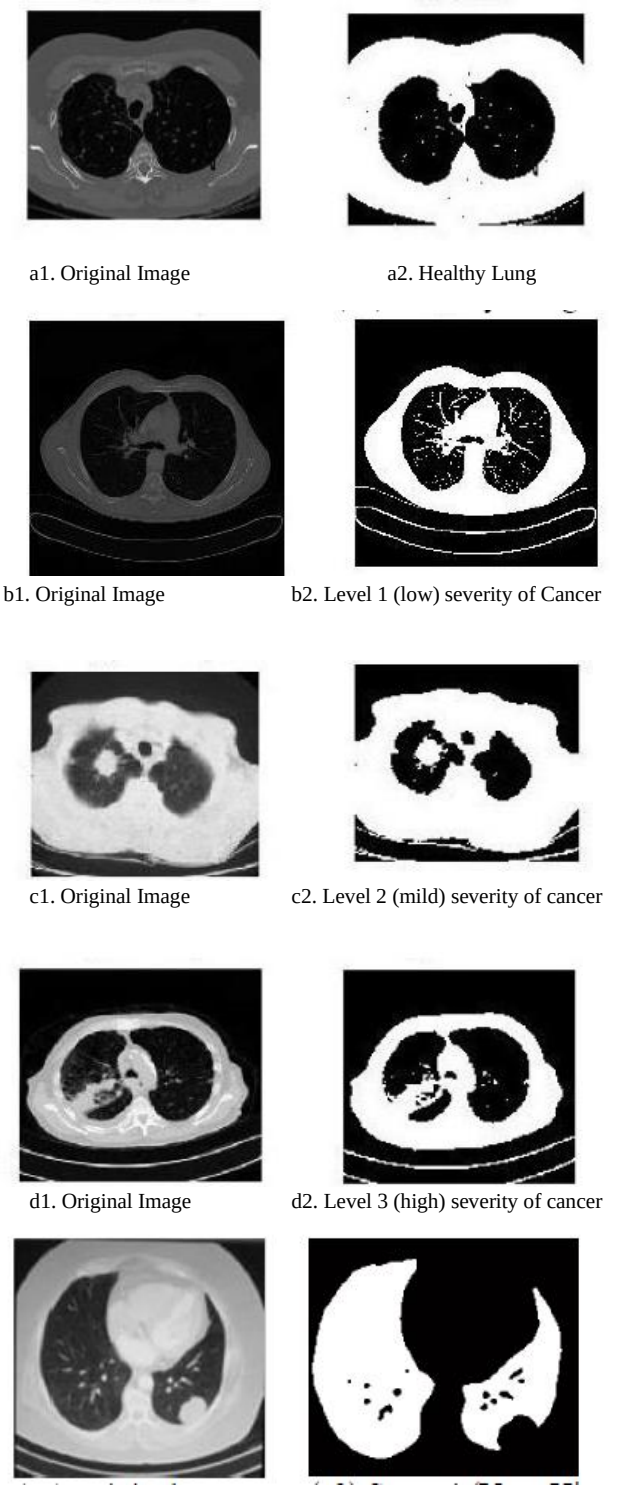


Figure 7. Detection Cancer using ACART

The detection level founds the highest accuracy level in lung cancer tissue detection by using ACART method. The result can be named following stages strong (healthy), Stage 1 (Low level cancer), Stage 2 (Medium level of cancer), Stage 3 (high level of cancer) and Stage 4 (Very High level of cancer).

Table 5. Output Table for Risk Severity Prediction of Lung Cancer

Output Field	Range	Fuzzy Sets
Severity	2cm - 3 cm	Stage 1 (Low)
	3.1cm – 5 cm	Stage 2 (Medium)
	5 cm - 8cm	Stage 3 (High)
	> 8cm	Stage 4 (Very High)



a1. Original Image a2. Healthy Lung

b1. Original Image b2. Level 1 (low) severity of Cancer

c1. Original Image c2. Level 2 (mild) severity of cancer

d1. Original Image d2. Level 3 (high) severity of cancer

e1. Original Image e2. Level 4 (very high) severity of cancer

Figure 8. Risk Severity Prediction of Lung Cancer by using Fuzzy Logic

Table 6. Evaluation of the Fuzzy Logic in the detection of Lung Cancer in the given lung cancer images

Image	Area	Stage Detected
a1 Image	187.78 sq.mm	Healthy
b1 Image	477.112 sq.mm	Stage 1: Cancer is in initial stage, small tumor is detected.
c1 Image	925.632 sq.mm	Stage 2: Cancer is confined to the Lung
d1 Image	3652.78 sq.mm	Stage 3: Cancer has been confined to the chest.
e1 Image	6542.98 sq.mm	Stage 4: Cancer has been spread from chest to other parts of the body

To analyze the efficiency of implemented methods EKNN and ACART are compared with LSRE [13] method, ANN [14] and SVM [15]. Based on the results and comparison the proposed EKNN method shows 97% of accuracy in detecting the cancer and ACART method yields 98.3% of accuracy in segmenting cancer tissues

Table 7. Comparative performance suing Proposed Approach

Methodology	Classification Accuracy	Processing Time	Misclassification Rate
Enhanced K-NN	97%	3 sec	3%
Advanced CART	98 %	3.5 sec	1.73%

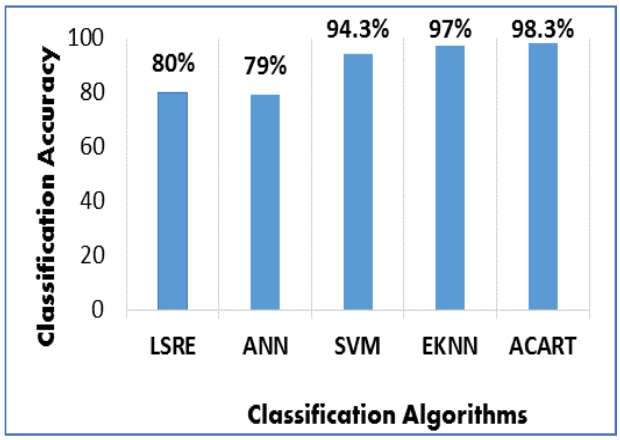


Figure 9. Comparison result of proposed method with other data mining algorithms

9. Conclusion

The various error levels are used in this work to identify the performance of proposed PCT method. This concept is used to remove the noise in the lung image by removing noise. The proposed PCT method is giving best result in denoising even at highest noise level of 60. From above mentioned figures and tables, it is proved that proposed PCT method gives most extreme result in different noise concept such as FOM, PNSR and SSIM. This strategy is utilized in the pre-processing level to eliminate the noise in the lung cancer image. This noise expulsion settles on the decision making simpler.

Then the resulted denoised images are applied in improved k nearest neighbor method to find the lung cancer tissues. The general classification accuracy is accomplished is 97% in after applying denoising concept. After that the resulted images are applied in ACART method for segment the cancer tissues in lung Images. After the completion of segmentation process the resulted images are established with fuzzy logic method to find the detection level and severity of cancer. The field of medical assurance restorative pictures goes up against a couple mechanical, logical and cultural troubles. The inventive types of progress in data picture mining have achieved an improved imaging overhaul. The implementation of the proposed estimation is differentiated and four estimations specifically Locality-obliged Sub-group Representation Outfit (LSRE), EKNN, counterfeit neural organization and backing vector mechanism. The proposed all the four method such as PCT (Denoising), EKNN (Finding cancer), ACART (Segmentation of cancer) and Fuzzy logic (find severity of the cancer) methods performs well all together in lung cancer examination.

References

[1] Apoorva Mahale, Chetan Rawool, Dinesh Tolani, Deepesh Bathija and Prof. Kajal Jewani, “A Survey on Lung Cancer Detection using Image Data Analysis and Machine Learning”, International Journal of Innovative Research in Computer and Communication Engineering, Volume 5, Issue 1, pp. 1066-1075, January 2017.

[2] P. Thamilselvan, Dr. J.G.R. Sathiaseelan, “Segmentation of Lung Malignant Cancer Tissues using PCA and ACART Method in MR Images for Huge”, Journal of Theoretical and Applied Information Technology, Volume 95, Number 6, pp. 1309-1317, 2017.

[3] Xinli Zhang, Zheng Fu, Guanzhong Gong, Hong Wei, Jinghao Duan, Zhaoqiu Chen, Xiangming Chen, Ruozheng Wang, and Yong Yin, “Implementation of Diffusion-Weighted Magnetic Resonance Imaging in Target Delineation of Central Lung Cancer Accompanied with Atelectas in Precision Radiotherapy”, Oncology Letters, Volume 14, Issue 3, pp. 2677-2682, 2017.

[4] C.Bhuvaneswari, P.Aruna, D.Loganathan, “Classification of Lung Diseases by Image Processing Techniques Using Computed Tomography Images”, International Journal of Advanced Computer Research, Volume-4 Number-1 Issue-14, pp. 87-93, 2014.

[5] M. Premchander, Dr. M. Venkateshwara Rao, Dr. T. V. Rajinikanth, “Detection of Lung Cancer in Medical Images using Image Processing Techniques”, InternationalJournal of Emerging Technology and Advanced Engineering, Volume 7, Issue 7, pp. 87-92, July 2017.

[6] Atsushi Teramoto, Tetsuya Tsukamoto, Yuka Kiriyama and Hiroshi Fujita, “Automated Classification of Lung Cancer Types from Cytological Images using Deep Convolutional Neural Networks”, BioMed Research International Hindawi Publications, pp.1-6, 2017.

[7] G. Niranjana and M. Ponnvaikko, “A Review on Image Processing Methods in Detecting Lung Cancer Using CT Images”, IEEE International Conference on Technical Advancements in Computers and Communications (ICTACC), pp. 18-25, 2017.

[8] Moffy Crispin Vas, Amita Dessai, “Classification of benign and malignant lung nodules using image processing techniques”, International Research Journal of Engineering

and Technology (IRJET), Volume: 04 Issue: 04, pp.3407-3411, April 2017.

[9] Xu, Jun, et al. “A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images”, Neurocomputing, pp. 214- 223, 2016.

[10] Aniket Bombale, C.G. Patil, “Segmentation of Lung Nodule in CT Data using KMeans Clustering”, International Journal Of Electrical, Electronics And Data Communication, Volume-5, Issue-2, pp. 36-39, February 2017.

[11] P. Thamilselvan, J. G. R.Sathiaseelan, “A Novel Profuse Clustering Technique for Image Denoising”, Procedia Computer Science, vol. 125, pp. 132–142, 2018.

[12] P. Thamilselvan, “Improving Medical Image Preprocessing using Denoising Technique”, ICTACT Journal on Image and Video Processing, Volume: 12, Issue: 03, pp. 2650-2654, 2022

[13] Yang Song, Weidog Cai, Heng Huang, Yun Zhou, Yue Wang and David Dagan Feng, “Locality-constrained subcluster representation ensemble for lung image classification”, Medical Image Analysis, 2015, pp. 1-14, 2015.

[14] Ling Ma, Xiabi Liu, Li Song, Chunwu Zhou, Xinming Zhao and Yanfeng Zhao, “A New Classifier Fusion Method based on Historical and On-line Classification Reliability for Recognizing Common CT Imaging Signs of Lung Diseases”, Computerized Medical Imaging and Graphics, pp. 1-20, 2014.

[15] Bruno Rodrigues Froz, Antonio Oseas de Carvalho Filho, Aristófanes Corrêa Silva, Anselmo Cardoso de Paiva, Rodolfo Acatauassú Nunes and Marcelo Gattass, “Lung nodule classification using artificial crawlers, directional texture and support vector machine”, Expert Systems with Applications, 2016, http://dx.doi.org/10.1016/j.eswa.2016.10.039.

[16] Firmino, Macedo, et al. "Computer-aided detection system for lung cancer in computed tomography scans: review and future prospects." Biomedical engineering online, Vol. 13(1), 2014.

[17] Shankar, K. "Prediction of Most Risk Factors in Hepatitis Disease using Apriori Algorithm." research journal of pharmaceutical biological and chemical sciences, Vol. 8(5), pp. 477-484, 2017.

[18] McRoberts, Ronald E., Erik Næsset, and Terje Gobakken. "Optimizing the k-Nearest Neighbors technique for estimating forest aboveground biomass using airborne laser scanning data." Remote Sensing of Environment, vol. 163, pp. 13-22, 2015.

[19] Moffy Crispin Vas, Amita Dessai, “Classification of benign and malignant lung nodules using image processing techniques”, International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 04, pp.3407-3411, April 2017.

[20] Xu, Jun, et al. “A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images”, Neurocomputing, pp. 214- 223, 2016.

[21] Aniket Bombale, C.G. Patil, “Segmentation of Lung Nodule in CT Data using KMeans Clustering”, International Journal Of Electrical, Electronics And Data Communication, Volume-5, Issue-2, pp. 36-39, February 2017.

[22] Zou, Huanxin, et al. "A PDF-based SLIC superpixel algorithm for SAR images." Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International. IEEE, 2016.

[23] Drouyer, Sébastien, et al. "Sparse Stereo Disparity Map Densification using Hierarchical Image Segmentation." International Symposium on Mathematical Morphology and Its Applications to Signal and Image Processing. Springer, Cham, 2017.

[24] Angalaparameswari Rajasekaran, Senthilkumar. P, “Image Denoising Using Median Filter with Edge Detection Using Canny Operator”, International Journal of Science and Research (IJSR), Volume 3 Issue 2, February 2014, pp.30-34.

[25] C. Venkatesan and P. Karthigaikumar, “An Efficient Noise Removal Technique using Modified Error Normalized LMS Algorithm”, National Academy Science Letters, Vol. 41, No. 3, pp. 155-159, 2018.

[26] J. Portilla, V. Strela, M.J. Wainwright and E.P. Simoncelli, “Image Denoising using Scale Mixtures of Gaussians in the Wavelet Domain”, IEEE Transactions on Image Processing, Vol. 12, No. 11, pp. 1338-1351, 2003.

[27] X.B. Yao, “Image Denoising Research based on Non-Local Sparse Models with Low-Rank Matrix Decomposition”, Ph.D. Dissertation, Department of Computer Science and Engineering, Xidian University, pp. 1-187, 2014.

[28] C. Cruz, A. Foi, V. Katkovnik and K. Egiazarian, “Nonlocality-Reinforced Convolutional Neural Networks for Image Denoising”, IEEE Signal Processing Letters, Vol. 25, No. 8, pp. 1216-1220, 2018.

Author Details

Dr. P. Thamilselvan is currently working as Assistant Professor, Department of Computer Science, Bishop Heber College (Affiliated to Bharathidasan University) Tiruchirappalli, Tamilnadu, India. He has been working as Assistant Professor since 2018. He has published more than 17 research articles in national and international journals which are published in IEEE, Elsevier, and Scopus. His area of research interests includes Artificial Neural Networks, Data Mining and Image Mining. E-mail: thamilselvan1987@gmail.com.