

PULMONARY NODULE DETECTION AND CLASSIFICATION IN CT IMAGES OVER EMPLOYING R-CAD SYSTEM

VIKUL J. PAWAR¹ and Dr. P. PREMCHAND²

¹Research Scholar, CSE Department, UCE, Osmania University Hyderabad, Telangana, India.

²Professor, CSE Department, UCE, Osmania University Hyderabad, Telangana, India

Abstract

The lung is important organ of human body, the functioning of lung is to exchange exchanges the gas in and out from human body. The lifestyle modernization is badly impacting on human body, one of the leading cancer disease is spreading worldwide nowadays in lung cancer disease. These days the Lung cancer patients are increased widely in the world, the major cause of this disease is smoking and increase of air pollution. Correspondingly the medical fields are upgrading the infrastructure to detect the cancerous nodules in pulmonary region. However, the computer science is providing enormous solution with aid of technological ground. The current trends in Image process and Machine learning techniques are offering enormous solution for medical professionals to identify distinct types of cancer disease. This article is contemplating and exploring by employing Robust Computer Assisted Diagnosis (R-CAD) system, the proposed R-CAD system frameworks shapes the various stages through implementation of the model. The major steps involved in implementing R-CAD system are Pre-processing, Segmentation, Feature Extraction, and Classification. This paper exhibits the enhance segmentation technique to segment the ROI from CT images; secondly the feature extraction technique extracts the learnable feature values through GLCM, GLRM and LBP methods. Lastly the modified CNN model classifies the detected pulmonary nodules are benign or malignant. The performance measurement of proposed model is evaluated through Accuracy, Specificity and Sensitivity, and this R-CAD model is delivering the accuracy of 92.9%.

Keywords: Lung Cancer, Pre-processing, Feature Extraction, Segmentation, Classification, Image Processing, Machine Learning.

1. INTRODUCTION

In the past few years the Lung Cancer became the major health diseases in human body. It is quite difficult to diagnose Lung Cancer in early stage of it, and later stage disease will increase risk of disease spreading in other organs of body. The early diagnosis of disease will help to doctors for offering correct treatment to the particular patients, there are numerous kinds of medical treatments are available such as surgery, Radiology, and Chemotherapy to control over spreading of cancer disease in human body. The early detection will not only control the disease it will save the life of patients, but the survival rate is also of 14% for the patients up to 5 years. Initially the cancer disease in lungs is developed in bronchial tree of lungs that is known as epithelium [1]. As per study cancer disease are detected after 55 years of human age but is also seen that there are few patients are also diagnosed with cancer disease after age of 45 years.

In current technological era image processing and machine learning techniques are providing robust solution for analyzing the stage of disease using CT Images and MRI. Correspondingly, the lung cancer disease is detected by employing the CAD system with the help computer based algorithm. The powerful tools of image processing are applied enhanced the quality of input image and using Segmentation algorithm it is possible to segment the suspicious region CT

images. The machine learning algorithm is applicable to train the system to detect the nodules in CT image automatically and classify that the detected nodule is cancerous or non-cancerous [2] [3].

The proposed R-CAD prototype is developed to provide robust solution for detection of pulmonary nodule and classifying the nodule as benign or malignant. This work improves the accuracy for classification nodules in CT images. The proposed R-CAD model is using CT images as input for pulmonary nodule detection in Lung Cancer [4] [5].

It is very challenging for medical professionals to detect the cancer nodules which are smaller in size in input CT images [5] [6]. The foremost objective of this study is to improve the accuracy in classification of Lung Cancer Diagnosis. The proposed model is employing the improved algorithm for preprocessing to improve the quality of input CT images, the second algorithm is employing the novel Segmentation algorithm to separate the Region of Interest (ROI), and segmented from ROI measurable feature are extracted, and those features are provided as input for the stage of classification to classify the nodules as benign or malignant.

2. RELATED WORK

Ahmed Elnakib et al. (2019) [6], this authors of this article proposes a model nodules detection in pulmonary region this work uses (LDCT) CT images as input, the major stages in this models are employing pre-processing which helps to improve CT images, in second step applied transfer learning is utilize to extract features thru deep learning from low dose CT images, in third step the Genetic Algorithm is applied to categorized the most related cancerous nodules from training subset of extracted features dataset, at last is fourth step the classification is performed using Support Vector Machine (SVM) for detecting cancerous pulmonary nodules, this proposed system has achieved the 92.5% of Accuracy, 90% Sensitivity, 95% Specificity using online public lung image database of ELCAP project. In proposed system applying genetic algorithm by deep learning will create more complexity and processing time requires more, and during the stage of classification SVM algorithm incompatible with noisy dataset, target classes are overlapping, moreover raising the features in dataset will impact on performance of SVM.

Allison et al. (2017) [7], The proposed work is implemented using manifold preprocessing method and deep learning algorithm to detect the cancer nodules in CT images, the objective of this proposed work is to enhance the accuracy for diagnosis of disease [8-10]. This model is unique in comparison with other model; in this work researchers are using two networks to training purpose. Firstly the input images are goes under the preprocessing stage, for preprocessing Gaussian filter is applied to improve the quality of input data. These preprocessed input images are used as input for CNN classification model. The highest accuracy attained by this model is 97.5%, the limitation of this model is the size of input dataset, and consistency in classifying benign nodules is poor.

Hanan M. et al (2018) [11], the authors of this paper are proposing a CAD model to detect the pulmonary nodules from CT images, the this model implemented in four steps. The first steps are of preprocessing, which enhance the quality of input CT images under this process the image

contrast is improved and various noises are removed. The second steps applying morphological operation to separate the blood vessels and pulmonary nodules using double-fold thresholding techniques. The third steps is of feature extraction through fusion technique which extracts the HOG Histogram-of-oriented gradients feature, VH value Histogram features, First and second order of statistical features and GLCM Grey Level co-occurrence of Matrix centered on wavelength coefficient. The final step is of classification where three distinct type of classifier namely as (MFNN) Multi_layer_feed_forward neural-network, Radial_basis_function neural-network (RBNN) and SVM Support-Vector Machine. The performance of this proposed model is measures based on the evaluation metrics such as CAR Classification Accuracy, SP-Specificity, and SN-Sensitivity. The final obtained results are CAR=99.06%, SP=99.2%, SN=100%. This model uses 40 CT images for testing the robustness of this model which is comparably very less.

3. PROPOSED MODEL

The proposed model of robust computer assisted diagnosis system R-CAD has demonstrated in Figure 1, The key steps involved in R-CAD system are 1. Preprocessing, 2. Segmentation, 3. Feature Extraction, and 4. Classification. The contribution in proposed work is implementing the enhanced algorithm in distinct stage of this model.

a) Pre-processing

In the initial stage of preprocessing this model removes the noises from input CT images; the input CT images are applied from LIDC-IDRI dataset [12]. Image enhancement is applying various methods to remove the noise from CT images and improves the quality of images, the diverse types of spatial and frequency domain filters are used in image smoothing, preprocessing is applying filtering technique using median filter, the purpose of to use median filter to remove the noise from input CT images. Median filter are mainly applicable to remove the speckle noise similarly it remove the salt and pepper noise from input image, median filter are edge preserving [13].

The functioning of median filter is demonstrated as shown below: for processing the input image under preprocessing technique median filter is applied, sample the values of odd number are ranked. It is consider that the length signal is to be finite, the range of samples are $X(0)$ to $X(L-1)$.

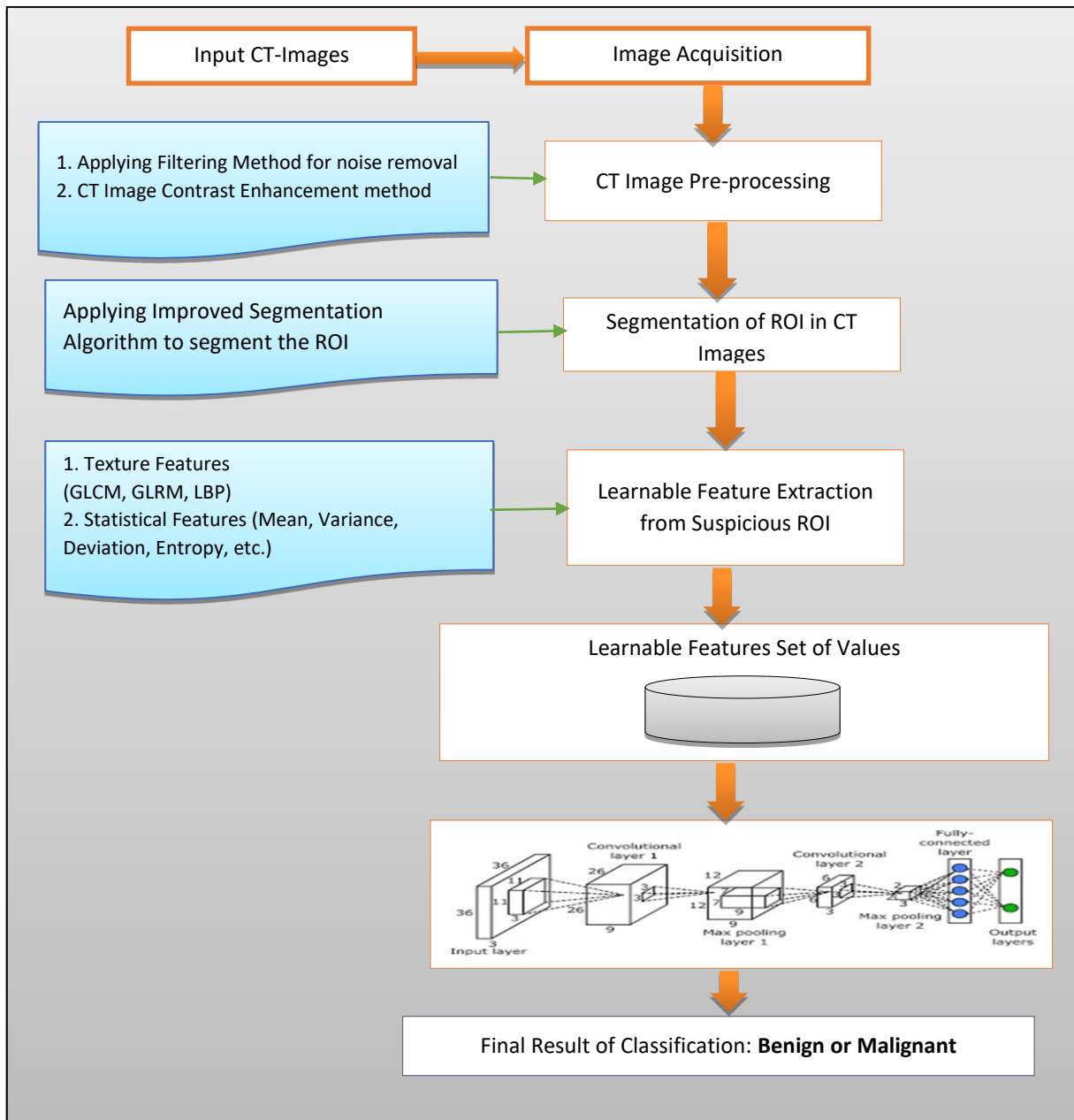
The window length for median filter is $N=2k+1$,

$$Y(n) = \text{med}[X(n-k) \dots X(n) \dots X(n+k)] \quad (1)$$

Histogram Equalization

Histogram equalization technique is computer-based image processing technique which is used to improve the contrast in image; histogram equalization is the technique which alters image intensities to improve the contrast. Pixel intensities ranging from 0 to $L-1$. L is the number of possible intensity values, often 256 [14].

Figure 1: The proposed R-CAD model for pulmonary nodule detection



4. PREPARE YOUR PAPER BEFORE STYLING

b) Segmentation

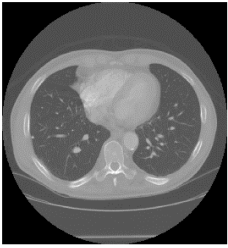

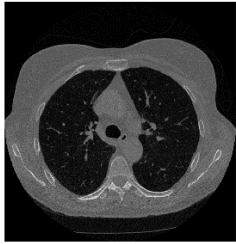

The segmentation is the most crucial phase in this work, though it is challenging to recognize the suspected region in the CT image, the suspected region indicates the presence of cancer

nodule in pulmonary region [15]. This research model pioneered the enhanced Level set segmentation method to segment the region of interest (ROI). The enhanced Level set algorithm identifies moving curve velocities and curvature of suspicious regions in CT image [16]. Initially the level is set as zero in all three directions in plane i.e $\phi(x, y, z) = 0$. The external energy is computed for curvature moving velocity in equation (2).

$$\xi_o(u, v) = \left(\left[(FM + KM) * 0.5 * (-5\chi) \right] \right) \quad (2)$$

The boundary of distinct region is separated with help of Heaviside function; the further stage will compute the length and area for suspected region using conventional statistical computation method. In this way the segmentation step will separate the suspicious region is shown in Figure (2).

Figure 2: Input CT images and Segmented ROI in input image

Input CT Image from LIDC-IDRI Dataset	Segmented Image after applying Enhanced Segmentation Method
	
	

c) Feature Extraction

The Feature Extraction is a step where the learnable input values are collected from ROI; these learnable values are applied to classification stage. The leading extracted features are Grey-level co-occurrence matrix (GLCM), Grey-Level Run Length Matrix and LBP segmentation is the most crucial phase in this work, though it is challenging to recognize the suspected region in the CT image, the suspected region indicates the presence of cancer nodule in pulmonary region.

This research model is extracting the GLCM, GLRM, and LBP features and these features are computed the using distinct equations of it.

- **GLCM**

The GLCM features are the texture features which calculate the features such as Contrast, Homogeneity, Energy, dissimilarity, Mean, ASM, GLCM_Std, GLCM_Max, Skewness, Entropy, and Kurtosis [17]. The mathematical equations of GLCM features are shown in Table 1.

- **GLRM**

The GLRM is measuring the pixels intensity value in certain defined angle from the given source of point. The GLRM measures the SREmpasis, LREmpasis, GreyLN, GreyLNN, RunLN, RunLNN, RunP, LowGLRE, HighGLRE [18] [19]. The mathematical equation of GLRM features are shown in Table 2.

Table 1: GLCM Features Equation

GLCM Features	Equation
Energy	$J = \sum_{i=1}^{G-1} \sum_{j=1}^{G-1} (px(i, j))^2$
Correlation	$\sum_{i,j=0}^{G-1} P(i, j)(i - \mu_i)(j - \mu_j) / \sigma_i \sigma_j$
Contrast	$\sum_{i,j=0}^{N-1} P(i - j)^2$
Homogeneity	$\sum_{i,j=0}^{N-1} \frac{P_{ij}}{1 + (i - j)^2}$
Entropy	$- \sum_{i=1}^{G-1} \sum_{j=1}^{G-1} (px(i, j)) \log(px(i, j))$
Mean	$\mu = \sum_{i,j=0}^{N-1} iP_{ij}$
Standard Variance	$\mu = \sum_{i,j=0}^{N-1} P_{ij} (i - \mu)^2$
Skewness	$\sum_{i,j=0}^{N-1} P_{ij} (i - \mu)^3$
Kurtosis	$\sum_{i,j=0}^{N-1} [P_{ij} (i - \mu)^4] - 3$

Table 2: GLRM Feature Equation

Grey Level Run-length Matrix	Equation
SRE	$SRE = \frac{\sum_{i=1}^{N_h} \sum_{j=1}^{N_k} \frac{P(i, j \theta)}{j^2}}{N_k(\theta)}$
LGE	$LGE = \frac{\sum_{i=1}^{N_h} \sum_{j=1}^{N_k} P(i, j \theta) j^2}{N_k(\theta)}$
GLN	$GLN = \frac{\sum_{i=1}^{N_h} \left(\sum_{j=1}^{N_k} P(i, j \theta) \right)^2}{N_k(\theta)}$
GLNN	$GLNN = \frac{\sum_{i=1}^{N_h} \left(\sum_{j=1}^{N_k} P(i, j \theta) \right)^2}{N_k(\theta)^2}$
RLN	$RLN = \frac{\sum_{i=1}^{N_k} \left(\sum_{j=1}^{N_h} P(i, j \theta) \right)^2}{N_k(\theta)}$
RLNN	$RLNN = \frac{\sum_{i=1}^{N_k} \left(\sum_{j=1}^{N_h} P(i, j \theta) \right)^2}{N_k(\theta)^2}$
RP	$RP = \frac{N_k(\theta)}{N_p}$

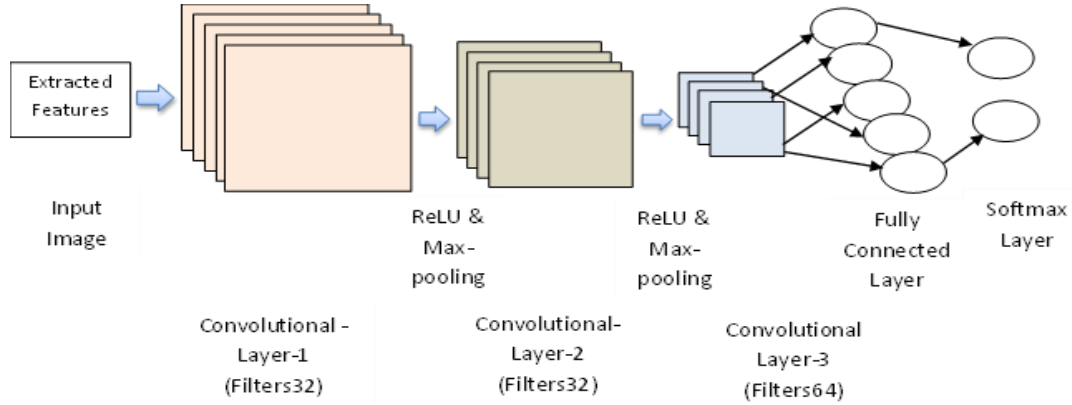
• **LBP**

The texture features are measured using LBP in image, these LBP features values are having two crucial properties, which are spatial patterns and contrast value, The computational procedure of LBP is straightforward, and it consist of better learnable values [20]. However, the pixel values in LBP are labeled in the form numerical values. For computing these labeled values the neighboring pixel values are used in computation. The middle pixel-value of image is most significant role in entire process of computation.

d) Classification

The proposed R-CAD system is employing modified M-CNN model, the two similar set of M-CNN modified layers are merged [21], and the proposed CNN model is overcome on the distinct challenges of it to deliver sustainable performance result in this work [22]. The kernel size of proposed CNN model is 3X3 with kernel size is 32, 32, and 64 are connected fully as demonstrated in Figure 3. In this model MAX-pooling and ReLU activation function are employed to stimulate the neurons in network, these activation functions are optimizing the feature value which are extracted from learnable feature function finally for concluding the classification result this model applies the soft-max function. The proposed R-CAD system trained the modified CNN using LIDC-IDRI dataset and delivers the final result as detected nodules are benign or malignant.

Figure 3: Proposed Modified M-CNN classification model



The equation (3) exemplifies the feature mapping function proposed modified M-CNN classification model in R-CAD system

$$R_{p,r,q}^s = W_k^{sT} J_{p,r}^s + B_k^s \quad (3)$$

The B_k^s is define as biased in CNN and W_k^s is defined as weight in CNN, sequentially the weight optimization in M-CNN model is established using novel CSO method for tuning the weight in CNN classification model during the training the proposed CNN model.

4. PERFORMANCE ANALYSIS

The performance measurement of proposed model is analyzed based on standard measurement metrics such as Accuracy, Specificity, and Sensitivity. The results of proposed R-CAD system are computed using performance metrics and compared with existing classification model such as DBN [23], SVM [24] and conventional CNN [25] model.

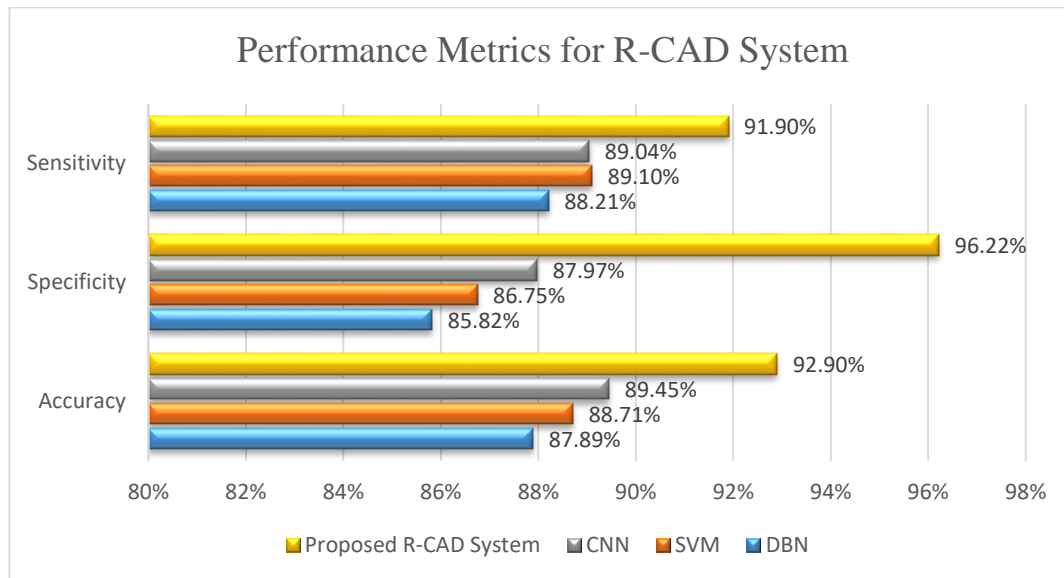
The Table 3 demonstrates the comparative result analysis of proposed R-CAD model with DBN, SVM, and CNN classification model. From the obtained performance measurement analysis it clearly understands that the proposed R-CAD system is superior in terms of all measuring metrics such as Accuracy, Specificity, and Sensitivity. The highest accuracy achieved of proposed R-CAD system is 92.90%.

Table 3: Performance Measurement analysis by evaluating Accuracy, Specificity, and Sensitivity

Classification Model	Measurement Metrics		
	Accuracy	Specificity	Sensitivity
DBN	87.89%	85.82%	88.21%
SVM	88.71%	86.75%	89.10%
CNN	89.45%	87.97%	89.04%
Proposed R-CAD System	92.90%	96.22%	91.9%

The result analysis of proposed model is shown in graphical representation of Figure 4, the result inference presumes that the proposed model attains the far superior accuracy of 91.9% for detecting the nodules in pulmonary region.

Figure 4: Graphical Analysis of proposed R-CAD system with DBN, SVM, and CNN



However, the specificity of proposed model is extremely outstanding as compared to DBN, SVM, and CNN which is 96.22%. And ultimately the sensitivity of comparably more prominent than prevailing model.

5. CONCLUSION

The Lung Cancer is one of the leading causes of death in worldwide; the accessible medical examining techniques are striving hard for detecting the cancer disease in early stage of it in the patients. Providing constructive technical solution with aid of computer based system the previous model are implemented in Nodule detection in pulmonary region, however this proposed R-CAD system affording one of the best solution to detect cancerous nodules applying Image Processing and Machine Learning techniques using CT images. In this proposed model R-CAD system employed in major four stages, firstly preprocessing which improve the quality of input CT images by applying improved contrast enhancement techniques. Secondly, the second step is Segmentation which segmenting the suspected region, which is known as ROI, in third stage of Feature Extraction learnable feature value are extracted using distinct texture feature (GLCM, GLRM, and LBP). The final step in this work is of classification in which Modified M-CNN using improved weight optimization algorithm is employed to detect the pulmonary nodule in CT images of LIDC-IDRI dataset. The proposed R-CAD model achieved the superior Accuracy of 92.9%, Sensitivity of 91.9% and Specificity of 96.22% which are comparably improved results in contrast with prevailing DBN, SVM, and CNN classification model.

6. ACKNOWLEDGMENT

The authors of this paper are thankful to the Head and Professor of CSE department and Dean Board of Studies CSE, University College of Engineering, Osmania University Hyderabad, Telangana, for their continuous support and guidance.

REFERENCES

1. Asuntha A, and Srinivasan A 2020 Deep learning for lung Cancer detection and classification. *Multimedia Tools and Applications* pp 1-32.
2. Ozdemir O, Russell RL and Berlin AA 2019 A 3D Probabilistic Deep Learning System for Detection and Diagnosis of Lung Cancer Using Low-Dose CT Scans *IEEE Transactions on Medical Imaging* 1419-29.
3. Huang PW, Lin PL, Lee CH and Kuo CH 2013 A classification system of lung nodules in CT images based on fractional brownian motion model *Int. Conf. on system science and engineering (ICSSE)* pp. 37-40.
4. Šarić M, Russo M, Stella M and Sikora M 2019 CNN-based method for lung cancer detection in whole slide histopathology images *Int. Conf. on Smart and Sustainable Technologies (SpliTech)* pp. 1-4.
5. Alam J, Alam S and Hossan A 2018 Multi-stage lung cancer detection and prediction using multi-class svm classifier *Int. Conf. on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)* pp. 1-4.
6. Ahmed Elnakib, Hanan M Amer, Fatama E.Z. Abou-Chadi, "Computer aided detection system for early cancerous pulmonary nodules by optimizing deep learning features", 2019 Association for Computing Machinery ACM, ISBN 978-1-4503-6105-7/19/04 <https://doi.org/10.1145/3328833.3328856>
7. Zhou, A.M.R.W. Deep Learning for Categorization of Lung Cancer CT Images. in 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). 2017. Philadelphia, PA, USA: IEEE.
8. R. Fakoor, F.L., A. Nazi, and M. Huber, Using deep learning to enhance cancer diagnosis and classification, in *International Conference on Machine Learning*, 2013. 2013, MLR:W&CP: Atlanta, Georgia, USA.
9. H. Yang, H.Y., and G. Wang, Deep learning for the classification of lung nodules. 2016.
10. F. Ciompi, K.C., S. J. van Riel, A. A. A. Setio, P. K. Gerke, C. Jacobs, E. T. Scholten, C. Schaefer-Prokop, M. M. Wille, A. Marchiano, Towards automatic pulmonary nodule management in lung cancer screening with deep learning. 2016.
11. Hanan M. Amer, Fatma E.Z. Abou-Chadi, Sherif S. Kishk, Marwa I. Obayya, "A Computer-Aided Early Detection System of Pulmonary Nodules in CT Scan Images", *ICSIE '18*, May 2–4, 2018, Cairo, Egypt, 2018 ACM. ACM ISBN 978-1-4503-6469-0/18/05 DOI: 10.1145/3220267.3220291
12. Armato, I., Samuel McLennan, G., McNitt-Gray, F. R., Michael, Charles, Reeves, Anthony P., Clarke, Laurenc, Data From LIDC-IDRI. *The Cancer Imaging*. 2015.
13. Kim H, Nakashima T, Itai Y, Maeda S, kooi Tan J, Ishikawa S. Automatic detection of ground glass
 - a. opacity from the thoracic MDCT images by using density features. 2007 international conference on
 - b. control, automation and systems. *IEEE*; 2007, October. p. 1274–7.
14. R Gonzalez & R Woods "Digital Image Processing", 2nd Edition, Prentice Hall.
15. hen W, Liu C, Basu A, Pan B (2021) A hybrid level set model for image segmentation. *PLoS ONE* 16(6): e0251914. <https://doi.org/10.1371/journal.pone.0251914>

16. Chencheng Huang, Li Zeng, Level set evolution model for image segmentation based on variable exponent p-Laplace equation, *Applied Mathematical Modelling*, Volume 40, Issues 17–18, 2016, Pages 7739-7750, ISSN 0307-904X,
 - a. <https://doi.org/10.1016/j.apm.2016.03.039>.
17. Priyanka, Dharmender Kumar, "Feature Extraction and Selection of kidney Ultrasound Images Using GLCM and PCA",
 - a. *Procedia Computer Science*, vol. 167, pp. 1722-1731, 16 April 2020(Cover date: 2020).
18. Manavalan Radhakrishnan and ThangavelKuttiannan, "Comparative Analysis of Feature Extraction Methods for the Classification of ProstateCancer from TRUS Medical Images", *IJCSI International Journal of Computer Science Issues*, Vol. 9, no.1, January 2012.
19. Jong Kook Kim, Jeong Mi Park, KounSik Song and HyunWook Park "Texture Analysis and Artificial Neural Network
 - a. for Detection of Clustered Microcalcifications onMammograms" *IEEE*, pp.199 – 206, 1997.
20. T. Ojala, M. Pietikainen and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, July 2002
21. CHENHUI MA, XIAODONG MU, AND DEXUAN SHA, "Multi-Layers Feature Fusion of Convolutional
 - a. Neural Network for Scene Classification of Remote Sensing", *IEEE*, Digital Object Identifier 10.1109/ACCESS-2019-2936215.
22. Laith Alzubaidi1, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma,J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie and Laith Farhan, "Review of deep learning: concepts, CNNarchitectures, challenges, applications, future directions", *Alzubaidi et al. J Big Data (2021) 8:53*<https://doi.org/10.1186/s40537-021-00444-8>.
23. Shanthi; V, Sridevi; S, Charanya; R,andJosphin Mary "DEEP BELIEF NETWORK (DBN)CLASSIFICATION FOR LUNG CANCER PREDICTION USING KNN CLASSIFIER," *European Journal of Molecular & Clinical Medicine*ISSN 2515-8260 Volume 07, Issue 09, 2020.
24. Asuntha; A,Brindha; A,Indirani; S,Srinivasan; A,"Lung cancer detection using SVM algorithm and optimization techniques," *Journal of Chemical and Pharmaceutical SciencesJCPS Vol-9, Iss-4,ISSN: 0974-2115, December-2016*.
25. Pradhan; A, Sarma; B, Dey; B. K, "Lung Cancer Detection using 3D ConvolutionalNeural Networks," *International Conference on Computational Performance Evaluation,Shillong, Meghalaya, India. 978-1-7281-6644-5/20/\$31.00 ©2020 IEEE, 2-4-Jul-2020*.