IJCSE International Journal of Computer Sciences and Engineering Open Access

Research Paper Vol.-7, Issue-5, May 2019 E-ISSN: 2347-2693

A Novel Automated CNN Based Lung Cancer Prediction Technique (CNNLCPT) for CT scan images

M. Muthuraman[1*], S. Ravichandran[2]

[1,2] PG and Research Department of Computer Science, H.H. The Rajah's College (A), Pudukkottai, Tamilnadu, India

*Corresponding Author: muthuraman70@gmail.com, Tel.: +91-7010751943

DOI: https://doi.org/10.26438/ijcse/v7i5.14401444 | Available online at: www.ijcseonline.org

Accepted: 26/May/2019, Published: 31/May/2019

Abstract— There are many incidents of Lung Cancer in the globe. This Cancer is curable if diagnosed early stage, where screening plays an important role in prevention of the disease. Computed Tomography (CT) scans can provide medical information, but its access is limited in rural areas. Computer-aided diagnosis (CAD) which can assist in screening of cancer from medical images can also provide help to doctors in remote areas. Previous studies have promoted and proposed CAD based systems for predicting lung cancer. Their findings have laid the foundation of promise lung cancer diagnosis using the deep learning approaches. This paper proposes and demonstrates a novel Automated CNN Based Lung Cancer Prediction Technique (CNNLCPT), a set of unique steps in image processing for predicting lung cancer from medical CT scans. The accuracy of predictions is also demonstrated in the paper.

Keywords— Automation, Deep Learning, Image processing, Lung cancer, Prediction.

I. INTRODUCTION

Cancer is a leading cause of death in humans with the disease claiming above 8 million deaths in 2015, where 20% of them were due to cancerous lungs [1].This type of cancer though fatal, can be treated properly when treated early. Low dose computed tomography (LDCT) help in screening lung cancer [2] which can be evaluated further with Computed Tomography (CT) scan images [3]. Rural populations use medical images which are readily available, but suffer from low quality. Thus, image processing techniques are used to enhance the quality of images, while diseases can be predicted from images using data mining techniques. These techniques can be implemented in CAD systems for better accuracy and efficiency in diagnosis of diseases like lung cancer. Radiomics is an approach that analyzes features like shape, size, location and texture of pulmonary nodules from radiological images. It helps in detection and diagnosis of cancerous parts and specifically lung nodules [4]. Figure 1 depicts Cancer Lung Nodules.

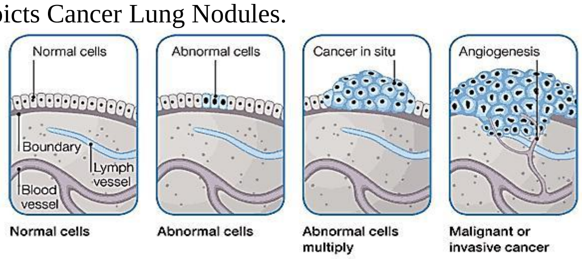

Figure 1. Cancer starting phase

Deep learning technique, Convolutional neural network (CNN), has been used effectively in many classification tasks and studies have detected outlier information in medical images using CNN [5]. Thoracic pathologies were also detected from medical images using CNNs in [6]. Thus, CNNs are emerging as an alternative approach in healthcare to analyze lung nodules and have been successfully applied in chest analysis [7, 8, 9]. However, Lung cancer predictions using deep learning techniques face one major issue namely the sample size which needs to be large. The techniques have to be adjusted to a specific and targeted image dataset. The availability of lung cancer images is an issue and the datasets are far too small for data mining and adjusted results have a problem of accuracy. The proposed work attempts to overcome this problem by proposing a set of unique steps. The proposed scheme called CNNLCPT (CNN Based Lung Cancer Prediction Technique) attempts to improve the performance of lung cancer prediction using CNNs. Figure 2 briefly describes the stages in lung cancer prediction using imaging techniques. The first step after image acquisition is image pre-processing which improves the quality of images followed by image segmentation. Features are then extracted before the disease is diagnosed using classification techniques. Figure 2 depicts the generic processing steps of medical image processing.

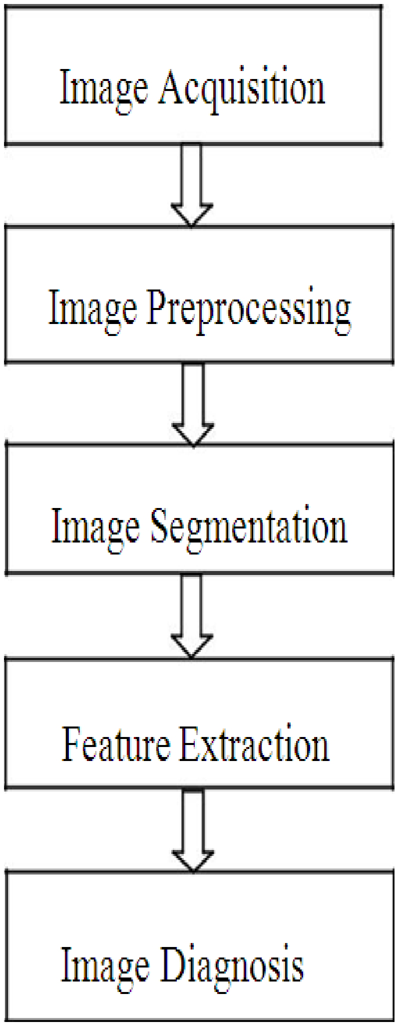

Figure 2. Generic medical image processing steps

II. CNNLCPT

The proposed scheme is a model for lung cancer prediction from CT scan images and can be implemented in CAD systems. The scheme is explained in this section. Image Acquisition is the first step in medical image processing. An existing dataset Lung Image Database Consortium (LIDC) has been used in this study. Since, lung CT images having low noise when compared to other images, CT images were taken to test the proposed scheme. The methodology followed in CNNLCPT for detecting lung abnormalities from CT scan images is pre-processing of images for image enhancements, filtering for removal of noises, segmentation for separating the lung area in CT scan images and finally prediction of abnormalities after extraction of features. The CNNLCPT scheme is pictured as Figure 3.

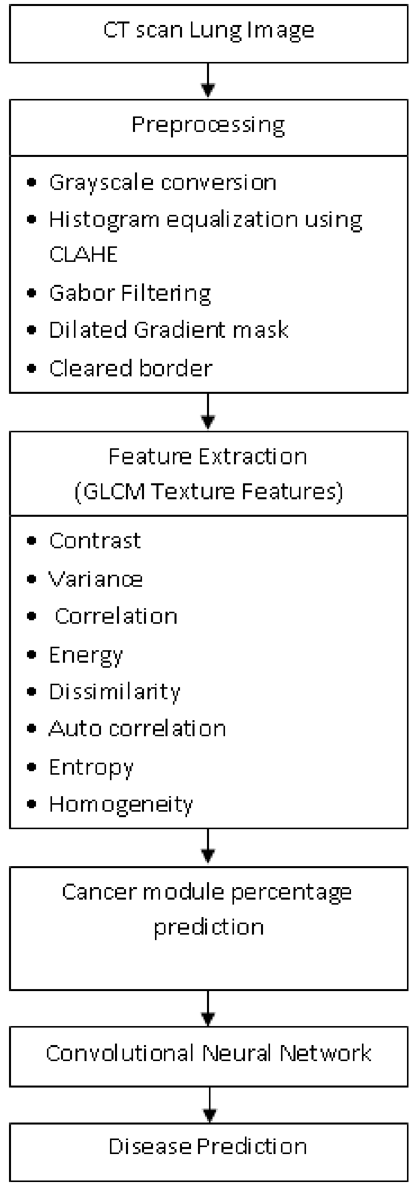

Figure 3. CNNLCPT scheme

III. CNNLCPT IMAGE PREPROCESSING

Image pre-processing is an attempt to make diagnostics using images more reasonable by improving the quality of images so that the consequential image is better than the original image.

A. CNNLCPT gray scale conversion

Captured raw images when converted into digital form vary in their intensities. These color images are converted to gray scale for further processing techniques which are dependent on them. The intensities range from 0 (black) to 255 (white) based on grayness levels. Figure 4 depicts a CNNLCPT converted gray scale image

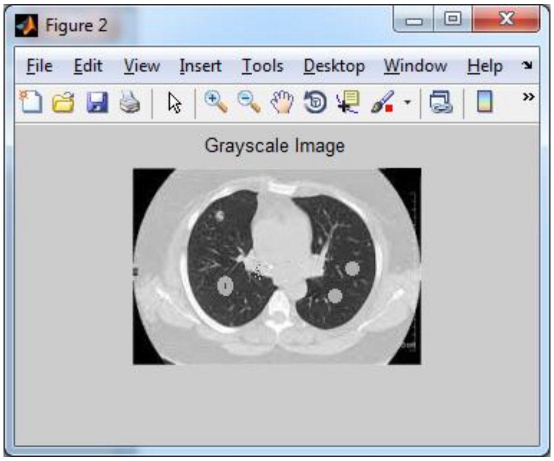

Figure 4. CNNLCPT gray scale image

B. CNNLCPT histogram equalization using CLAHE

Some scanned medical images may result in narrow ranges of gray scale and Contrast enhancement is used to increase features of interest areas for better visibility without distortion and thus increasing the overall image quality. Transformation functions are used for highlighting available structures [10]. Adaptive Histogram Equalization (AHE) is used to compute a pixel's histogram in a local window and then map. It can may result in the amplification of noise or artifacts at the edges. A simplification of AHE called Contrast limiting AHE (CLAHE) is used for its flexibility in selecting the histogram mapping function. It clips amplifications and subtracts the background, thus resulting in reduced artifacts at the boundaries. CNNLCPT uses CLAHE for the aforesaid reasons. Figure 5 depicts CNNLCPT Histogram equalization using CLAHE

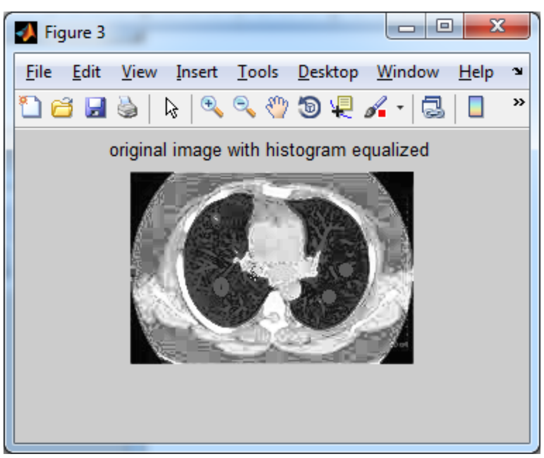

Figure 5. CNNLCPT histogram equalization output

C. CNNLCPT gabor filtering

Gabor filter is a linear filter that is popular for extracting spatially localized spectral features [11]. Figure 6 depicts CNNLCPT Gabor Filter Output

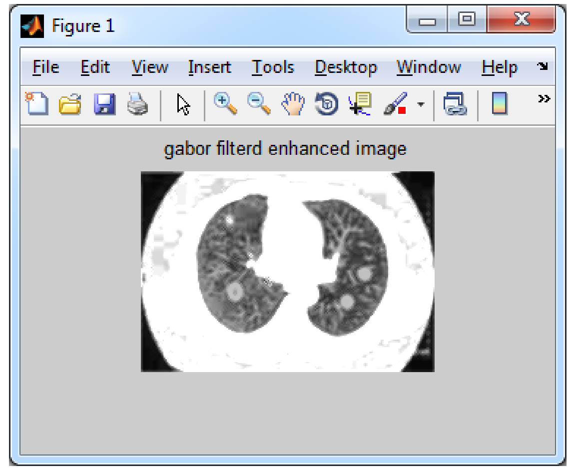

Figure 6. CNNLCPT gabor filter output

D. CNNLCPT dilated gradient mask .

Image objects can be identified with Gradient changes and Morphing an image helps in finding gradient in images. CNNLCPT uses Dilated Gradient mask for edge detections. CNNLCPT Dilated Gradient mask Output is depicted in Figure 7.

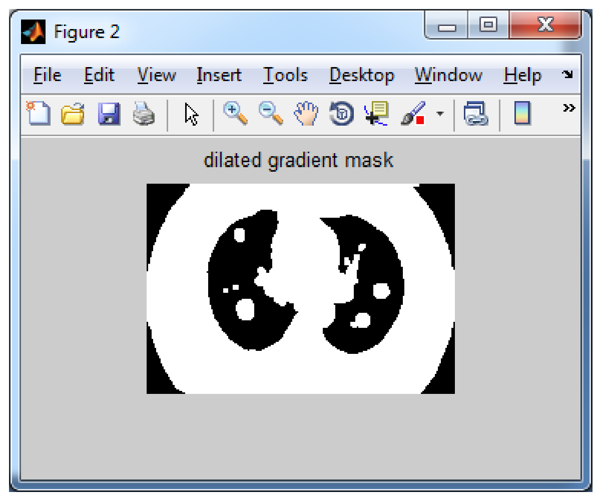

Figure 7. CNNLCPT dilated gradient mask output

E. CNNLCPT Cleared border

CNNLCPT uses Cleared border to identify high contrast objects and segment them in a CT scan image in its preprocessing step. Pixels Hamlets formed in the image are isolated and a segmented output is obtained. Figure 8 depicts CNNLCPT Cleared border Output

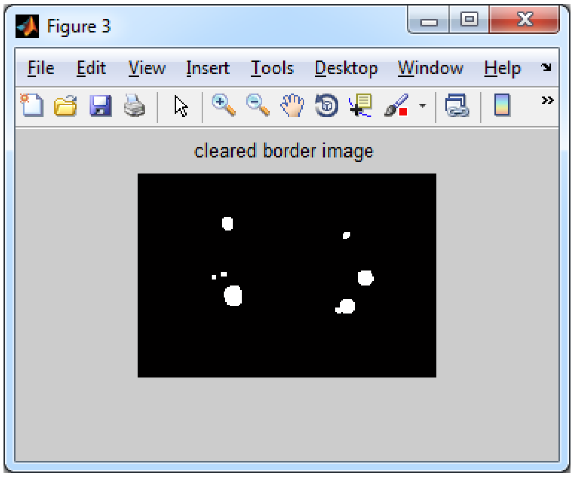

Figure 8. CNNLCPT cleared border output

IV. CNNLCPT Feature Extraction

CNNLCPT extracts relevant texture features from segmented CT scan images using gray-level co-occurrence matrix (GLCM) which are tabulated using equation (1)

$$P_{i,j} = \frac{V_{i,j}}{\sum_{i,j=0}^{N-1} V_{i,j}} \quad (1)$$

Where a row is i and J the column. A contrast process that minimizes intravenous contrasts on non-contrast CT scans in CNNLCPT. Variances in images are also addressed in feature extraction. Correlation assess the degree and relationship between pixels given by equation (2)

$$\frac{\sum_x \sum_y (Axy - A)(Bxy - B)}{\sqrt{(\sum_x \sum_y (Axy - A)^2)(\sum_m \sum_N (Bxy - B)^2))}} \quad (2)$$

Where Axy is the pixel intensity at point (x,y) and Bxy is the grayscale value at point (x,y). CNNLCPT uses energy and dissimilarity measures in images for improved accuracy and is given in equation (3)

$$d_{sum}(I_1, I_2) = \sum_{i,j} \Delta(x_{1i}, x_{2j}) \quad (3)$$

Where I1 and I2 are images x_i is the i^{th} region of interest and delta the dissimilarity between objects or regions of interest. CNNLCPT calculates entropy as the final step in its feature extractions. Entropy of a discrete random variable X with a probability distribution pX = (p1, . . . , pn) is defined in equation (4)

$$H(X) = H(\mathrm{p}_X) \triangleq \sum_{i=1}^{n} p_i \log(1/p_i), \quad (4)$$

Where 0 log ∞ =0 and the base of the logarithm determines the unit.

CNNLCPT's feature selection is a crucial component in machine learning techniques [12]. This is done to reduce the dimensionality and define region for analysis which are then extracted. Moreover, CNNLCPT speeds up its execution and improves its predictive accuracy with feature extractions. Table 1 lists features extracted by CNNLCPT from CT scan images of the lungs. The extracted features are then used by CNNLCPT for classification of images. Table 1- CNNLCPT extracted features from CT scan images of the lungs

Table 1. Feature extraction

Feature Set	0°	45°	90°	135°
Contrast	0.194609	0.136549	0.240547	0.119334
Variance	1.966507	1.954332	1.966507	1.954332
Correlation	1.559461	1.58283	1.548966	1.59456
Energy	1.457335	1.471407	1.454274	1.475278
Entropy	1.815941	1.739614	1.840625	1.721473
Homogeneity	1.552643	1.564193	1.550544	1.567797
Dissimilarity	1.4565	1.1298	1.4609	1.0002
Auto Correlation	1.9949	2.0209	1.9959	2.0373

V. CNNLCPT CLASSIFICATION

CNNLCPT uses CNNs for its classification as it is the most popular neural network model for image classifications. As CNNLCPT reduces dimensionality with its feature extractions it greatly improves the time it takes to learn as well as reduce the amount of data required to train the model. The CNN is given enough weights to look at a small patch of the image for accuracy. Figure 9 depicts a Convnet.

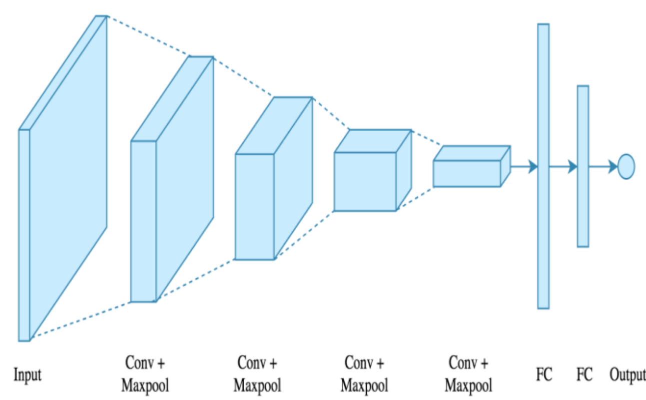

Figure 9. ConvNet

The convolution is a weighted sum of the values in the image within a sliding window, thus producing another image of a weighted matrix. Every CNN convolution layer generates many alternate convolutions. CNN is used by CNNLCPT since the number of parameters is independent of the size of the original image. Thus, the output identifies a diseased image. Figure 10 and 11 depict CNNLCPT's convolution output and its final prediction.

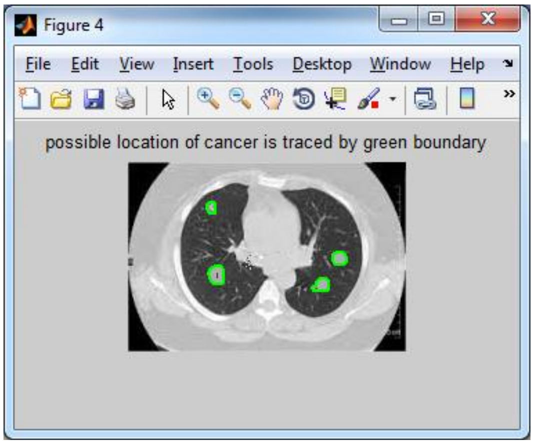

Figure 10. CNNLCPT convolution output

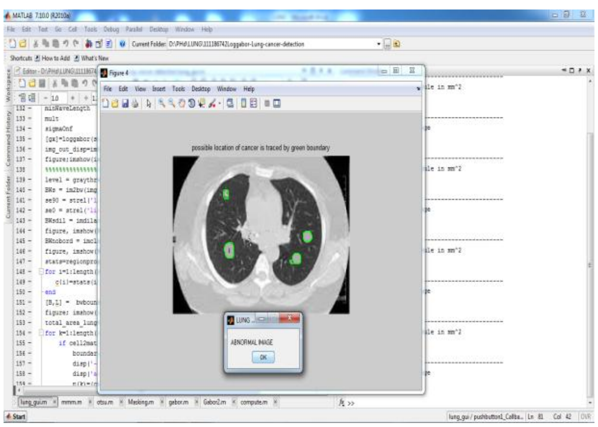

Figure 11. CNNLCPT Lung disease prediction output

VI. CONCLUSION AND FUTURE SCOPE

This work has proposed and demonstrated a novel technique, CNNLCPT, which is can be implemented in computer aided diagnosis (CAD) to detect abnormal lungs in CT scan images. This technique comprises of pre-processing, feature extraction, feature selection and final classification using a CNNs. It can be concluded that CNNLCPT is a viable scheme for detecting lung diseases from images using CAD. Further scope of this model can be in detecting lung diseases from other medical imaging modality images.

REFERENCES

[1] W. H. Organization, "Cancer," 2018. [Online]. Available: http: //www.who.int/mediacentre/factsheets/fs297/en/

[2] U. Kauczor, and H. Hatabu, "Standard-dose vs. low-dose ct protocols in the evaluation of localized lung lesions: Capability for lesion characterizationilead study," European Journal of Radiology Open, vol. 3, pp. 67 – 73, 2016.

[3] Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, Forster K, Aerts HJ, Dekker A, Fenstermacher D and Goldgof DB: Radiomics: the process and the challenges. Magnetic resonance imaging, 30(9), pp.1234–1248. (2012)

[4] M. T. Islam, M. A. Aowal, A. T. Minhaz, and K. Ashraf, "Abnormality detection and localization in chest x-rays using deep convolutional neural networks," vol. abs/1705.09850, 2017

[5] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2017.

[6] Hussein S, Gillies R, Cao K, Song Q and Bagci U: Tumornet: Lung nodule characterization using multi-view convolutional neural network with gaussian process. In Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on, pp. 1007–1010. IEEE (2017)

[7] Song Q, Zhao L, Luo X and Dou X: Using Deep Learning for Classification of Lung Nodules on Computed Tomography Images. Journal of healthcare engineering (2017) Proc Int Jt Conf Neural Netw. Author manuscript; available in PMC 2018 November 13.

[8] Jiang H, Ma H, Qian W, Wei G, Zhao X and Gao M: A novel pixel value space statistics map of the pulmonary nodule for classification in computerized tomography images. In Engineering in Medicine and Biology

[9] A. K. Jain, Fundamentals of Digital Image Processing. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[10] Z. Zhou, Y. Jiang, Y. Yang, S.Chen "Lung Cancer Cell Identification Based on Artificial Neural Network Ensembles", Artificial Intelligence in Medicine, 2002, vol.24, no.1, pp.25-36. @Elsevier

[11] A. Kalousis, J. Prados, M. Hilario, "Stability of Feature Selection Algorithms: a study on high dimensional spaces," Knowledge and information System, vol. 12, no. 1, pp. 95-116, 2007.

[12] Bartlett EC, Walsh SL, Hardavella G, et al. Interobserver Variation in Characterisation of Incidentally-Detected Pulmonary Nodules: An International, Multicenter Study. Available online: http://4wcti.org/2017/SS5-3.cgi

[13] Zinovev D, Feigenbaum J, Furst J, et al. Probabilistic lung nodule classification with belief decision trees. Conf Proc IEEE Eng Med Biol Soc 2011;2011:4493-8.

[14] Ciompi F, Chung K, van Riel SJ, et al. Towards automatic pulmonary nodule management in lung cancer screening with deep learning. Sci Rep 2017;7:46479.

[15] Aerts HJ, Velazquez ER, Leijenaar RT, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.

