Lung Cancer Detection using Machine Learning

Harpreet Singh[1], Er. Ravneet Kaur[2]

[1]Research Scholar, [2]Assistant Professor,

[1,2]Baba Banda Singh Bahadur Collage, Fatehgarh Sahib, Punjab, India

ABSTRACT

Modern three-dimensional (3-D) medical imaging offers the potential and promise for major advances in science and medicine as higher fidelity images are produced. Due to advances in computer aided diagnosis and continuous progress in the field of computerized medical image visualization, there is need to develop one of the most important fields within scientific imaging. From the early basis report on cancer patients it has been seen that a greater number of people die of lung cancer than from other cancers such as colon, breast and prostate cancers combined. Lung cancer are related to smoking (or secondhand smoke), or less often to exposure to radon or other environmental factors that's why this can be prevented. But still it is not yet clear if these cancers can be prevented or not. In this research work, approach of segmentation, feature extraction and Convolution Neural Network (CNN) will be applied for locating, characterizing cancer portion.

KEYWORDS: Lung Cancer, Image Processing, Machine Learning, K-means, Gray-Level Co-Occurrence Matrix (GLCM)

INTRODUCTION

The image processing is a technique which is used for the enhancement of unprocessed pictures or images captured from different cameras from different origins. With the help of image processing, the significant data can be retrieved efficiently. In the past decades, various methods have been evolved in image processing techniques for the extraction of complicated information in an effective manner. Image processing approach is widely utilized in army, clinical and investigational areas [1]. Some associations also use image processing approach for simplifying the manual workload and execution of positive actions. The image processing is applied inside numerous applications inclusively in order to improve the optical description of pictures. For the preparation of pictures, different calculations are implemented as well. Image processing also known as Digital Image Processing (DIP) comprises both visual and analog image processing which involves different methods. Image acquisition is also termed as imaging [2]. The visual and digital image processing can be performed with the help of imaging. This technique utilizes several domains like computer graphics for the generation of pictures [3]. This technique also provides assistance in the manipulation and modification of pictures. The picture or image is analyzed with the help of processor hallucination or computer vision. In lung cancer, anomalous cells multiply and grow in the form of a tumor. The lymph fluid which environs lung tissue carries the cancerous cells from lungs to blood. The lymph streams via lymphatic vessels. These lymph fluid drains into lymph nodules deployed in the lungs and in the middle region of chest area. The growth of lung tumor always carried out towards the middle area of chest due to the regular flow of lymph fluid towards the chest center. When a cancer cell leaves its origin area, metastasis happens [4]. This cancerous cell now goes towards a lymph nodule or to different body part with the help of blood flow. The prime lung tumor is a kind of cancer which originates from the lung. The compilation of lung pictures for the creation of data sample in the initial step. In Image Enhancement, image is processed and smoothened. This process enhances the picture quality and also eliminates noise from the picture. Thus, this process offers superior key for the digital image processing [5]. Image enhancement is an important pillar of image pre-processing. Image Segmentation allocates a digital picture into different sections like sets of pixels also recognized as super-pixels. The key objective of this process is the alteration of a picture demonstration in an easier investigative manner. Picture sectioning is utilized for identifying the location of objects, limits and borders in pictures. In this process, a label is assigned to each pixel in a picture and thus the pixels with the identical label share definite features [6]. In Feature Extraction feature plays an extremely significant character. Different image preprocessing approaches such as binarization, thresholding, normalization, masking approach etc. are implemented on the sampled picture before the attainment of features. Various classifiers are used for performing the classification on the basis of retrieved characteristics. SVM (Support Vector Machine) is a classification algorithm that is based on optimization theory. As it maximizes the margin it is also known as a binary classifier. All the data points of an individual class are separated by the best hyperplane, this can be identified through the classification provided by

support vector machine[7]. The main aim of Naïve Bayesclassifier is the implementation of a strategy where future objects are assigned to a group in the presence of a pattern of objects for every class. The applied variable vectors are demonstrated with the help of future entities. Decision Tree Classifier is considered as non-parametric supervised learning techniques and used for categorization and deterioration [8]. The main aim of this approach is the development of a model for the accurate prediction of an intended variable in accordance with several key variables. K-Nearest neighbor classifier depends on the learning by similarity. The n dimensional arithmetic qualities are utilized for the description of training sets.

Literature Review

Amir R. et.al (2019) reviewed the development of inclusive molecular description of tumor lump [9]. A fundamental role was played by the ailment biomarkers for the early detection and indulgent of tumor analysis. This work summarized the speedy development of biosensor equipments for lung tumor biomarkers discovery. More expansion in nanobiotechniques in association with nanobiocomposite and miniaturization approaches would considerably improve existing biodiagnostic capability for sensing tumor biomarkers in genuine organic models with sufficient compassion, accuracy, sturdiness and price efficiency.

Guobin Z., et.al (2019) presented a serious evaluation of the CADe scheme for automated lung cancer recognition with the help of CT descriptions for summarizing the existing developments [11]. These mechanisms included information attainment, preprocessing, lung image segmentation, nodule recognition and false positive diminution. A brief summary of superior nodule detection methods and classifiers was also provided on the basis of understanding, false positive value and other constrained data. After different studies it was evaluated that Computer aided diagnosis (CAD) scheme was essential for timely lung malignancy recognition.

JingS. et.al (2019) proposed a novel approach of microscopic hyper spectral imaging for the identification of ALK affected lung tumor [11]. In this approach, a household microscopic hyper spectral imaging scheme was utilized for capturing the pictures of five classes of lung tissues. In results, Group ALK obtained more relative proportion of cytoplasm of 77.3% than Group ALK-positive. The investigational outcomes related to quantitative scrutiny and ethereal curves demonstrated that the treatment of ALK affected lung tumor implemented with low concentrated medicines would be developed towards the ALK non-affected lung tumor.

Moritz S., et.al (2018) estimated the usefulness of machine learning for lung tumor recognition in FDG-PET imaging in the scenario of ultralow amount PET scan [12]. In the absence of pulmonary tumor, the recital of artificial neural network on selective lung cancer patients was examined. The sensitivity rate of 95.9% and 91.5% was attained by the artificial neural system for lung cancer detection. The deep learning approach for detecting lung cancer provided AUC of .989 for standard dose images, 0.983 for reduced dose images, and 0.970 for PET3.3% rebuilding. It was also suggested that more advancements in this technique could enhance the accurateness of lung tumor testing approaches.

Madhura J, et.al (2017) presented a review of noise reduction approaches for lung cancer diagnosis [13]. It was stated that lung cancer was a solemn ailment which caused due to the abnormal growth of cells in the lung tissues. Amongst all the other kinds of tumors, the lung tumor was identified as the most incident cancer. Therefore, this cancer became the reason of several cancer patients' deaths. This review work also described the different kinds of noises present in the pictures, techniques for the attainment of apparent pictures and noise elimination methods. A brief review on the existing noise elimination methods was also provided in this paper.

Suren M., et.al (2017) stated that CT images could be used for the lung tumor recognition. The major objective of this study was the evaluation of different automated technologies, investigation of existing finest method, recognition of its restrictions and disadvantages and the projection of a decisive system with several advancements [14]. For this purpose, the lung tumor recognition approaches were classified on the basis of their lung cancer analyzing accurateness. In every stage, these lung cancer recognition methods were examined and their restrictions and disadvantages were considered. It was identified that different lung cancer detection techniques showed different precision. Some techniques showed least precision rate while some techniques showed good precision rate for lung cancer detection but no technique showed 100% precise lung cancer detection.

Research Methodology

This research work is related to lung cancer detection from the CT (Computed Tomography) scan image using image processing techniques. The proposed methodology has the four phases for the lung cancer localization and characterization.
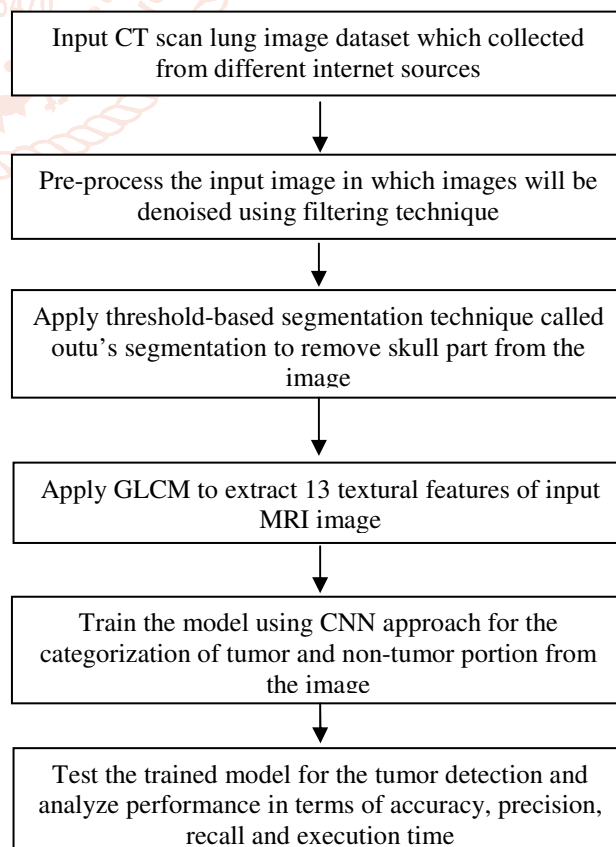
Input CT scan lung image dataset which collected from different internet sources

↓

Pre-process the input image in which images will be denoised using filtering technique

↓

Apply threshold-based segmentation technique called outu's segmentation to remove skull part from the image

↓

Apply GLCM to extract 13 textural features of input MRI image

↓

Train the model using CNN approach for the categorization of tumor and non-tumor portion from the image

↓

Test the trained model for the tumor detection and analyze performance in terms of accuracy, precision, recall and execution time

Figure 1: Proposed Flowchart

Following are the various phases of the lung cancer detection: -

1. **Pre-processing: -**
The pre-processing is the first phase in which CT scan image is taken as input. The technique of image de-noising will be applied which will remove noise from the input image. The output of this stage is an enhanced image. This is one of the most crucial stages in lung cancer detection.

2. **Segmentation:**
In the second phase, the approach of region-based segmentation will be applied which will segment the similar and dissimilar regions from the CT scan image. The otsu's segmentation technique is applied for the segmentation. The sectioned picture attained from thresholding comprises several benefits like lesser storage space, speedy dispensation velocity and easiness in exploitation in comparison with gray level picture that generally includes 256 steps. In the presented work, a gray scale picture is utilized for thresholding process. In this process, rgb picture is converted into binary picture. The obtained picture is in the form of black and white.

3. **Feature Extraction: -**
The feature extraction is the third phase, in which GLCM algorithm will be applied for the feature extraction of the CT scan image. In this step, the GLCM algorithm is applied for the feature extraction. The GLCM algorithm will extract the textural features of the input image. The GLCM algorithm extracts 13 features of the image for the tumor detection

$$Energy = \sqrt{\sum_{i,j=0}^{N-1} p_{i,j}^2}$$

$$Entropy = \sum_i p_i \log_x i$$

$$Contrast = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$$

4. **Classification: -**
In the last phase, the approach of CNN will be applied which can categorize and localize the cancer part .All the data points of an individual class are separated by the best hyper plane, this can be identified through the classification provided by CNN. In the CNN the largest the best hyper plane is described by the largest margin between the two classes. There are no interior data points when there is maximum width between the slabs parallel to the hyper plane which is also known as margin. The maximum margin in hyper plane is separated by the CNN algorithm.

Experimental Results
The proposed research is implemented in MATLAB and the results are evaluated by comparing proposed and existing techniques in terms of various performance parameters.
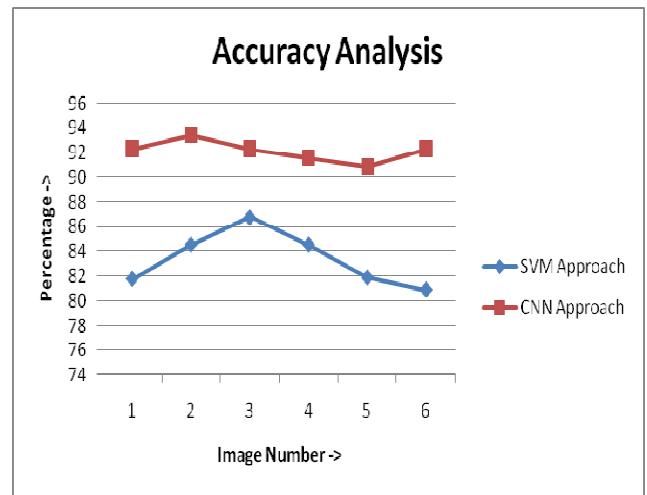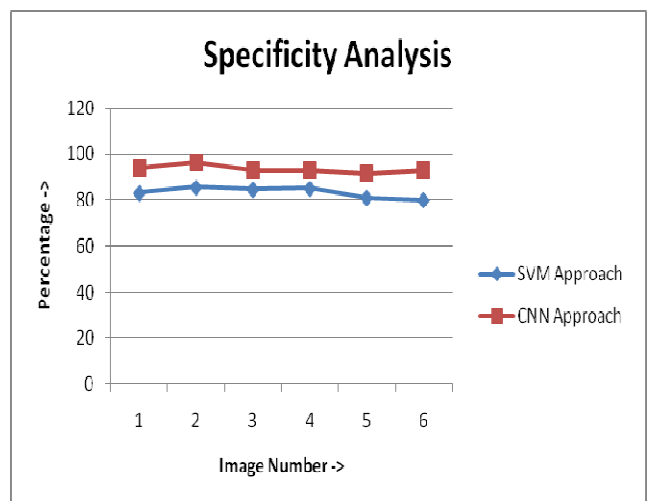

Figure 2: Accuracy Analysis

As shown in figure 2, the accuracy of the existing system which is SVM approach is compared with the proposed approach which is CNN approach. The system is tested on different number of images and it is analyzed that CNN gives better results as compared to SVM approach.
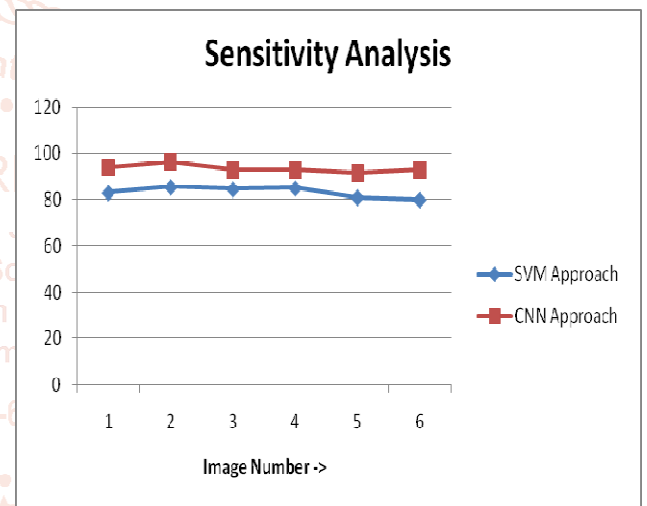
Figure 3: Sensitivity Analysis

As shown in figure 3, the sensitivity of the existing system which is SVM approach is compared with the proposed approach which is CNN approach. The system is tested on different number of images and it is analyzed that CNN give best results as compared to SVM approach.

Fig 4: Specificity Analysis

As shown in figure 4, the specificity of the existing system which is SVM approach is compared with the proposed approach which is CNN approach. The system is tested on different number of images and it is analyzed that CNN give better results as compared to SVM approach.

Conclusion

For lung cancer detection image processing is used. There are three steps for the detection of cancer nodule. To detect the presence of cancer nodule CT scan images are used. Further the pre-processing composed of two processes. Image enhancement and image segmentation are that two processes. The image segmentation process aims to partition the image into meaningful format and identify the object or relevant information from the digital image. The output from the segmentation process is applied to the feature extraction stage. Features such as area, perimeter and irregularity are found out in feature extraction. On the basis of the extracted features the abnormality in lung are found out by the cancer cell identification module. The approach of GLCM and CNN are implemented in this work for localizing and classifying cancer part from the CT scan image. The proposed approach is implemented in MATLAB and results are analyzed in terms of accuracy. It is analyzed that the proposed approach gives optimized results up to 8 percent.

References

[1] Anjali Kulkarni, Anagha Panditrao, "Classification of Lung Cancer Stages on CT Scan Images Using Image Processing", 2014 IEEE International Conference on Advanced Connnunication Control and Computing Technologies (lCACCCT)

[2] Anam Tariq, M. Usman Akram and M. Younus Javed, "Lung Nodule Detection in CT Images using Neuro Fuzzy Classifier", 2013 Fourth International Workshop on Computational Intelligence in Medical Imaging (CIMI), Pages: 49 – 53

[3] Christian Donner, Naoufel Werghi, Fatma Taher, Hussain Al-Ahmad, "Cell Extraction from Sputum Images for Early lung Cancer Detection", 2012 16th IEEE Mediterranean Electrotechnical Conference, Pages: 485 – 488

[4] Fatma Taher, Naoufel Werghi and Hussain Al-Ahmad, "Comparison of Hopfield Neural Network and Mean Shift algorithm in Segmenting Sputum Color Images for Lung Cancer Diagnosis", 2013 IEEE 20th International Conference on Electronics, Circuits, and Systems (ICECS), Pages: 649 – 652

[5] Janee Alam, Sabrina Alam, Alamgir Hossan, "Multi-Stage Lung Cancer Detection and Prediction Using Multi-class SVM Classifier", 2018, International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)

[6] Moffy Vas, Amita Dessai, "Lung cancer detection system using lung CT image processing",2017, International Conference on Computing, Communication, Control and Automation (ICCUBEA)

[7] N. Werghi, C. Donner, F. Taher, H. Alahmad, "Segmentation of Sputum Cell Image for Early Lung Cancer Detection", IET Conference on Image Processing (IPR 2012), 2012, Pages: 1 – 6

[8] Shanhui Sun, Christian Bauer, and Reinhard Beichel, "Automated 3-D Segmentation of Lungs with Lung Cancer in CT Data Using a Novel Robust Active Shape Model Approach", IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 31, NO. 2, FEBRUARY 2012

[9] Amir Roointan, Tanveer Ahmad Mir, Shadil Ibrahim Wani, Mati-ur-Rehman, Khalil Khadim Hussain, Bilal Ahmed, Shugufta Abrahim, Amir Savardashtaki, Ghazaal Gandomani, Molood Gandomani, Raja Chinnappan, Mahmood H Akhtar, "Early detection of lung cancer biomarkers through biosensor technology: a review",2019, PBA 12266

[10] Guobin Zhang, Shan Jiang, Zhiyong Yang, Li Gong, Xiaodong Ma, Zeyang Zhou, Chao Bao, Qi Liu, "Automatic nodule detection for lung cancer in CT images: A review", 2018, CBM 3128

[11] Jing Songa, Menghan Hua, Jiansheng Wanga, Mei Zhoua,b, Li Suna, Song Qiua, Qingli Lia, Zhen Sun, Yiting Wanga, "ALK positive lung cancer identification and targeted drugs evaluation using microscopic hyperspectral imaging technique",2019, Infrared Physics & Technology

[12] Moritz Schwyzer, Daniela A. Ferraro, Urs J. Muehlematter, Alessandra Curioni-Fontecedro, Martin W. Huellner, Gustav K. von Schulthess, Philipp A. Kaufmann, Irene A. Burger, Michael Messerli, "Automated Detection of Lung Cancer at Ultralow dose PET/CT by Deep Neural Networks - Initial results", 2018, LUNG 5827

[13] Madhura J, Dr. Ramesh Babu D R, "A Survey on Noise Reduction Techniques for Lung Cancer Detection", 2017, International Conference on Innovative Mechanisms for Industry Applications

[14] Suren Makajua, P. W. C. Prasad, Abeer Alsadoona, A. K. Singhb, A. Elchouemi, "Lung Cancer Detection using CT Scan Images", 2017, 6th International Conference on Smart Computing and Communications, ICSCC