ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

Identification of Lung Cancer and Stage prediction using Machine Learning Approach

Yash A. Patade[1], Keyur S. Karve[2], Saurabh P. Joshi[3], Neel S. Shevatkar[4], Navnath S. Bagal[5]

B. E, Dept. of Computer, SPPU/PVPIT, Pune, Maharashtra, India[1]

B. E, Dept. of Computer, SPPU/PVPIT, Pune, Maharashtra, India[2]

B. E, Dept. of Computer, SPPU/PVPIT, Pune, Maharashtra, India[3]

B. E, Dept. of Computer, SPPU/PVPIT, Pune, Maharashtra, India[4]

Professor, SPPU/PVPIT, Pune, Maharashtra, India[5]

ABSTRACT: Identification of lung disease is the most fascinating exploration zone of analyst's in early days. The proposed framework is intended to identify lung malignancy in untimely stage in two phases. The proposed framework comprises of numerous means, for example, image extraction, pre-processing, binary translation, thresholding, Division, highlight extraction, and neural system identification.

In our system we developed Lung Cancer detection system based on machine learning and neural network. It decreases the chances of getting harm to human by early detection of cancer. Presently a few frameworks are proposed and still a large number of them are theoretical plan. Convolutional Neural Network based Classification and location arrangement of lung tumor.

KEYWORDS: Convolutional Neural Network, Computed Tomography, Support Vector Machine, Small Cell Lung Cancer.

I. INTRODUCTION

Due to large prevalence of smoking and air pollution around the world, lung cancer has become one of the most common and deadly disease in recent decades. It often takes long time to develop and most people are diagnosed with the disease within the age bracket 55 to 65. Early identification and treatment isthe best available option for the infected people. Reliable identification and classification of lung cancer requires pathological test, namely, needle biopsy specimen and analysis by experienced pathologists. However, because it involves human judgment of several factors and a combination of experiences, a decision support system is desirable in this case. Recent developments in image processing, pattern recognition, dimensionality reduction and classification methods has paved the way for alternate identification and classification approaches for lung cancer.

Thus, we are trying to give a tremendous solution over this high rate of lung deaths due to lung cancer. We got conclusion that the main reason behind these deaths is late detection of lung cancer diseases. We have highly skilled doctors and digital and powerful treatment techniques even though deaths happened. Main scenario behind this implementation to give such noble implementation to health care sector to early detect lung cancer and prevent number of deaths in world.

In addition to machine learning methods is used to train our stages CT scan images dataset by using tensor flow python. After getting highly trained model of python we come on last stage of implementation. At last we test create our test model of lung cancer stages. The test model is used to find out accuracy rate of our trained model. We got nearly 90 to 98 % accuracy.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

II. LITERATURE SURVEY

Lilik Anifah, Haryanto,Rina Harimurti et al, [1] states research focuses on detection of lung cancer utilizing Artificial Neural Network Back-engendering based Gray Level Co-event Matrices (GLCM) highlight. The lung information utilized begins from the Cancerimagingarchive Database, information utilized comprised of 50 CT-pictures. CT-picture is gathered into 2 bunches, typical and lung disease. The means of this exploration are: picture preprocessing, locale of intrigue division, highlight extraction, and recognition of lung disease utilizing Neural Network Back-spread. The outcomes demonstrate framework can identify CT-picture of ordinary lung and lung malignancy with exactness of 80%. Early discovery of lung malignant growth will recoup the patient. Instrument used to recognize lung malignant growth is through CT Scan (Computed Tomography).

Prof. Anuradha Deshpande, Dhanesh Lokhande et al, [2] presentedto locate the beginning times of lung malignancy and increasingly exact outcome by utilizing distinctive strategies like combination, upgrade and division process. Already the vast majority of the malignant growth location strategies relies upon human experience by watching the picture of CT-filter. It will be a bogus discovery of lung malignant growth arrange. Utilizing Image Processing we can rapidly and precisely distinguish tumor of malignant growth. Utilizing Image Processing viable procedures, we gather data from complex restorative pictures. In combination procedure, the critical highlights of various unique pictures are consolidated together to acquire the required data in a Fused Image. In medicinal application there are different plans to enhance the substance of picture shape CT and MRI like CT picture examines the denser tissues and MRI filters the delicate tissues, so by joining pertinent data of the two pictures, we get proper data of melded picture. This procedure additionally enhances the nature of the melded picture.

Rachid Sammouda et al, [3] exhibits. In this paper an improved technique for Hopfield Artificial Neural Network Classifier show is proposed to section extricated lung locales from human chest Computer Tomography pictures. The pictures are procured utilizing Computer Tomography imaging methods from typical subjects and others as possibility for lung malignant growth determination. A blend of bit-planes of every pixel are utilized to upgrade edges' recognition of lung area flaps. Three indicative guidelines are confirmed too characterized channels of applicant malignant locales from the status of possibility to false or genuine positive status.

Abbas Khosravi, Amin Khatami et al, [4]examines In this paper, we propose a novel neural-arrange based calculation, which we allude to as entropy debasement technique (EDM), to recognize little cell lung malignant growth (SCLC) from processed tomography (CT) pictures. This exploration could encourage early identification of lung malignant growths. The preparation information and testing information are high-goals lung CT examines given by the National Cancer Institute. We chose 12 lung CT filters from the library, 6 of which are for sound lungs, and the rest of the 6 are examines from patients with SCLC. We arbitrarily take 5 filters from each gathering to prepare our model, and utilized the staying two outputs to test. Our calculations accomplish a precision of 77.8%.

Qing Wu, Wenbing Zhao et al,[5]has been findsIn this work, a profound auto encoder order system is proposed which initially adapts profound highlights and after that prepares a fake neural system with these educated highlights. Test results demonstrate the profound educated classifier beats every single other classifier when prepared with all properties and same preparing tests. It is likewise exhibited that the execution enhancement is factually huge.

Zhi-HuaZhouet al,[6]is proposed, which uses an artificial neural network mainly to detects lung cancer cells in the MRI of the specimens of needle biopsies gained from the different persons of the subjects to be diagnosed. The ensemble is made up of two-level ensemble architecture. The first-level ensemble is used to judge whether a cell is normal with high confidence where each individual network has only two outputs respectively normal cell or cancer cell. The predictions of those individual networks are combined by a novel method presented in this paper, i.e. full voting which judges a cell to be normal only when all the individual networks judge it is normal.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

Sameer Kumar et al. [7] has been introduced Lung cancer is one of the most dangerous health problems in the world. Lung Computer-Aided Diagnosis (CAD) is a powerful method to provide a range of computational tasks such as early cancer and disease identification, analysis of disease progression. Edge detection in Lung image is a main step in CAD system and many methods have used before edge detection in lung systems.

Nisar Ahmed Memonet al. [8] has beenstudied thatDivision is a significant advance in medicinal picture investigation and characterization for radiological assessment or PC supported finding. The CAD (Computer Aided Diagnosis) of lung CT by and large first section the region of intrigue (lung) and after that investigate the independently acquired territory for knob identification so as to conclusion the sickness. For ordinary lung, division can be performed by utilizing magnificent difference among air and encompassing tissues. Anyway this methodology falls flat when lung is influenced by high thickness pathology.

M.Gomathiet al. [9] has been examined that to create a fruitful Computer Aided Diagnosis framework, a few issues must be settled. Division is the main issue to be viewed as which helps in age of competitor area for identifying disease knobs. The second issue is ID of influenced knobs from all the hopeful knobs. At first, the essential picture preparing procedures, for example, Bit-Plane Slicing, Erosion, Median Filter, Dilation, Outlining, Lung Border Extraction and Flood-Fill calculations are connected to the CT check picture so as to distinguish the lung district.

Binsheng Zhao et al. [10] has been examined that progressively, computed tomography (CT) offers higher goals and quicker securing occasions. This has brought about the chance to recognize little lung knobs, which may speak to lung diseases at prior and possibly increasingly reparable stages. Notwithstanding, in the current clinical practice, many such thin-sectional CT pictures are created for every patient and are assessed by a radiologist in the conventional feeling of taking a gander at each picture in the pivotal mode. These outcomes in the possibility to miss little knobs and along these lines conceivably miss a malignant growth.

III. EXISTING SYSTEM APPROACH

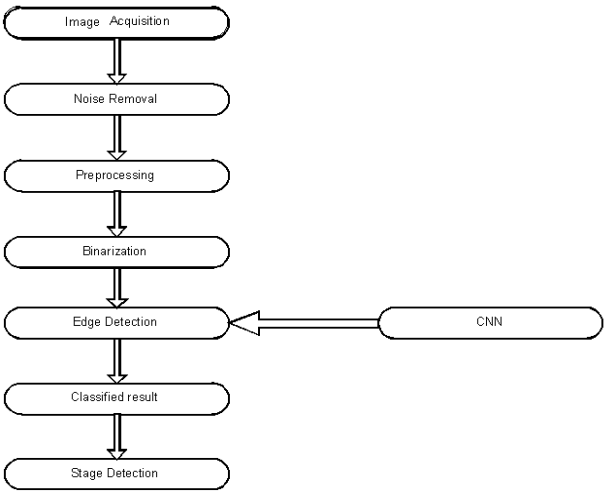


Fig.1Existing System Architecture

The existing system presented uses Support Vector Machine Classier and Watershed Segmentation of cancer detection. Also uses ANN back-propagation based on Gray Level Co-occurrence Matrix features. Which methods donot fulfilproject need, sowe have to implement our system to overcome all problems.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

Disadvantages

1. Due to large prevalence of smoking and air pollution lung cancer has become one of the deadliestdiseases in recent decades.
2. System needed by improving the prepossessing process, image segmentation, feature extraction, and learning process.

IV. METHODOLOGY USED IN PROPOSED SYSTEM

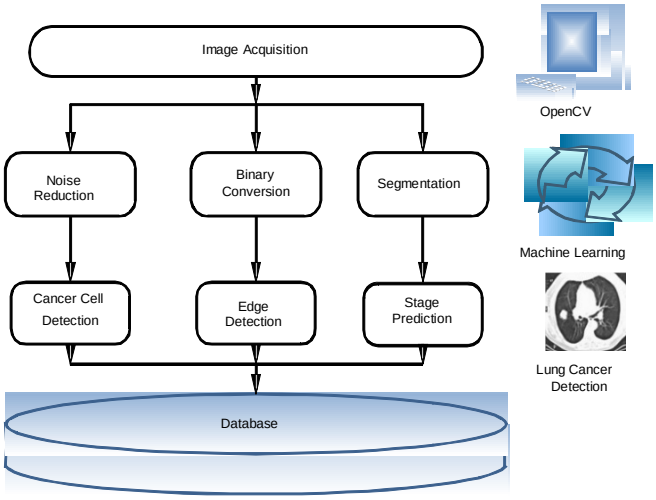


Fig.2 System Architecture

We propose a novel Lung detection and Stage prediction mechanism is proposed which first learns deep features and then trains an artificial neural network with these learned features. Experimental results show the deep learned classifier outperforms all other classifiers when trained with all attributes and same training samples. It is also demonstrated that the performance improvement is statistically significant. Classification of lung cancer using a low population, high dimensional data-set is challenging due to insufficient samples to learn an accurate mapping among features and class labels. Current literature usually handles this task through handcrafted feature creation and selection. Deep learning is found to be able to identify the underlying structure of data through the use of CNN and other techniques.

Advantages

1. This shows that application of machine learning has the potential to significantly detect and classify with almost high accuracy for the low population.
2. High dimensional lung cancer data-set without requiring any hand-crafted, case specific features.
3. High processing speed enhanced CNN classifier model.

A. **Image Processing: -**

An image is made up of RGB colours. Pre-processing unit consists of noise removal, grey scale conversion,binary conversion of images followed by feature extraction. In future extraction five steps followed in which fingertips searches by eccentricity. Next elongations of images aremeasured by considering pixel segmentation as well as rotation of images.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

B. **Feature Extraction: -**

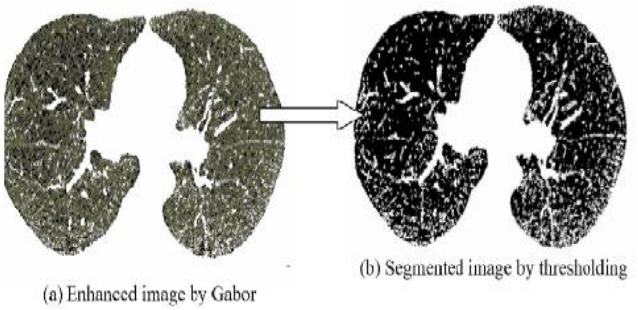
In feature extraction, algorithmic study used to find the feature vectors of systematic results combines K curvature and convex hull algorithms. In present work "K convex hull" algorithm which is used to detect fingertip with greater accuracy.In our system, Artificial Neural Network (ANN) is used for future recognition in which we having the input unit of training data set of images.

C. **Thresholding segmentation: -**

Thresholding is the simplest segmentation method. The pixels are partitioned depending on their intensity value converts a gray-scale image into a binary image Global thresholding, using an appropriate threshold T:

$g(x, y) = \quad 1, \text{if } f(x, y) > T$

$0, \text{if } f(x, y) \leq T$



(a) Enhanced image by Gabor (b) Segmented image by thresholding

D. **Edge Detection: -**

Edge defines the boundaries between regions in an image which helps in object detection. There are many edge detection operators and algorithms available. Edge Detection Operators and Algorithms used in our research like Convex hull method.

E. **Feature Recognition: -**

Brain-inspired systems used to replicate how humans learn. Consist of input, hidden and output layers that transform the input into something that the output layer can use. Excellent for finding patterns which is complex tohumanfor extract and teach the machine to recognize. ANN gathers their knowledge by detecting the patterns and relationships in data and learns (or is trained) through experience, not from programming.

V. RESULTS AND DISCUSSION

In Lung cancer detection and stage prediction system we have been implemented highly trained model that can accurately recognize lung cancer images. In this system we used Gaussian blur for grey scale conversion,Otsu's method for binary conversion of images after that we used convex hull for edge detection.

F. **Stages of Lung Cancer**

- *STAGE I:* Stage 1 is a part of number staging system and means your cancer is small. it hasn't spread to your lymph nodes and other distinct organ. they are of two types:

1. Type A: - Means that the cancer is 2 cm or smaller.
2. Type B: -Means that the cancer is between 2 to 3 cm.

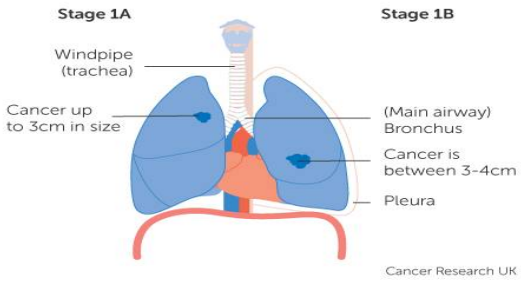
ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

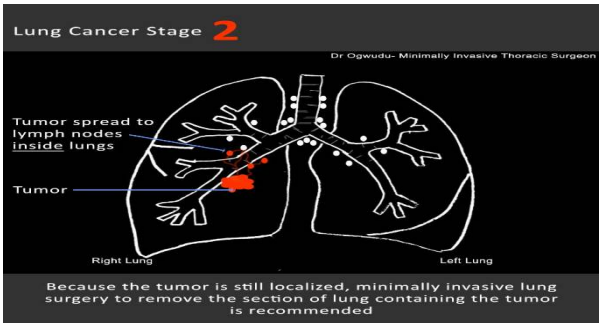
Website: www.ijircce.com

Vol. 7, Issue 4, April 2019



- *STAGE II:* These tumors are larger than the stage 1 tumors and start to spread to nearby lymph nodes or other structures.

About 30% of the lungs cancer are diagnosed at these stage, and the treatment success rate is higher than the later stage.



- *STAGE III:* In this stage the tumors more than 3 cm wide and is spread in the lymph nodes and other parts outside the lungs. in this stage the affected lymph nodes are restricted to the same side of the body as the tumor.
- Stage IV: The cancer is metastasized or spread beyond the lungs into other areas of the body.

Primary Tumor(T)	Criteria
T1	< 3cm in diam; T 1a <= 2cm; T 1b > 2cm <= 3cm
T2	> 3cm <= 5cm; T 2a > 3cm <= 4cm; T 2b > 4cm <= 5cm
T3	>5cm<=7cm
T4	Any Size greater than above

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 4, April 2019

VI. CONCLUSION

This shows that application of machine learning has the potential to significantly detect and classify with almost accuracy for the low population, high dimensional lung cancer dataset without requiring any hand-crafted, case specific features.

For future work, we can implement this technique on some more images. Increasing the number of images used for the process, can improve the accuracy.

REFERENCES

[1] Lilik Anifah, Haryanto,Rina Harimurti, "Cancer lung detection on CT scan image using ANN backpropagation based gray level co-occurrence matrix feature." 978-1-5386-3172-0/17/ 2017 IEEE
[2] Prof. Anuradha Deshpande, Dhanesh Lokhande, "Lung cancer detection with fusion of CT and MRI image using image processing." (IJARCET) Volume 4 Issue 3, March 2015
[3] Rachid Sammouda, "Segmentation and analysis of CT chest images for early lung cancer detection." Global Summit on Computer & Information Technology 978-1-5090-2659-3/17 2017 IEEE
[4] Qing Wu, Wenbing Zhao, "Small-cell lung cancer detection using a supervised ML algorithm." International Symposium on Computer Science and Intelligent Controls 978-1-5386-2941-3/17 2017 IEEE
[5] Abbas Khosravi, Amin Khatami, "Lung cancer classification using deep learned features on low population dataset." Canadian Conference on Electrical and Computer Engineering (CCECE) 978-1-5090-5538-8/17 2017 IEEE
[6] Zhi-Hua Zhou, Yuan Jiang, Yu-Bin Yang, Shi-Fu Chen, "Lung Cancer Cell Identification Based on Artificial Neural Network Ensembles", Artificial Ingelligence in Medicine, 2002, vol.24, no.1, pp.25-36.
[7] Samir Kumar Bandyopadhyay, "Edge Detection from CT Images of Lung", International Journal of Engineering Science & Advanced Technology Volume - 2, Issue - 1, pg: 34 – 37, Jan-Feb 2012.
[8] Nisar Ahmed Memon, Anwar Majid Mirza, and S.A.M. Gilani, "Segmentation of Lungs from CT Scan Images for Early Diagnosis of Lung Cancer", World Academy of Science, Engineering and Technology 20 2006
[9] M.Gomathi, Dr.P.Thangaraj, "A Computer Aided Diagnosis System For Detection Of Lung Cancer Nodules Using Extreme Learning Machine", International Journal of Engineering Science and Technology Vol. 2(10), 2010
[10] Binsheng Zhao, Gordon Gamsu, Michelle S. Ginsberg, Li Jiang,Lawrence H. Schwartz, "Automatic detection of small lung nodules on CT utilizing a local density maximum algorithm", Journal of Applied Clinical Medical Physics, Volume 4, NUMBER 3, Summer 2003