RESEARCH ARTICLE OPEN ACCESS

A Survey Paper on "Computer-Aided Diagnosis: An Approach for Lung Cancer Nodule Detection"

Vidya Nikam[1], Pradnya Joshi[2], Pragya Sharma[3], Rupali Chaturvedi[4], Rasika Nimkarde[5]
[1](Assistant Professor at Department of Computer Engineering, Pune/D.Y.Patil College Of Engineering Akurdi
[2,3,4,5](Department of Computer Engineering, Pune/D.Y.Patil College Of Engineering Akurdi

Abstract:

Cancer is a group of diseases involving abnormal cell growth with the potential to invade within other parts of the body too. It is caused by an abnormal growth of tumour cells (a swelling of a part of the body, generally without inflammation). This contrast with benign tumour, which do not spread to other parts of the body.Lung cancer is the most powerful cancer which can often grow for long time before they are found. In order to treat the early lung cancer more effectively, it is important to develop an effective clinical marker or prediction model to more accurately predict cancer prognosis after cancer surgery. In this paper the focus given on diagnosing at ahead of schedule and critical stages with keen computational procedures with deferent noise elimination by segmentation strategies and calculations which is the root idea of digital image processing. Location of CT pictures received from cancer various research organizations like ACS (American Cancer Society) and NCI (National Cancer Institute). And pulse coded neural network is designed to effectively analyse the cancer. This improves the accuracy of the prediction and process at low cost compared with the Naive Bayesian network.

Keywords — **American Cancer society; Noise Removal; Segmentation; Morality Rate; CT scanning.**

1. INTRODUCTION

I. . BASIC CONCEPT

The purpose of this project is to present a system that will be used to detect the lung cancer nodule with its area of influence. The idea is intended for both the stakeholders and the developers of the system. This system will be designed to maximize the productivity by providing sophisticated way to detect the lung cancer nodules easily. This system will effectively used for cancer detection and produces reliable solution for cancer detection problems in its early stages.The terminologies and algorithms used in this project are easy to implement and improves efficiency, reliability of the existing systems.

2. RELATED WORK

USER CLASSES AND CHARACTERISTICS

[1] EDUCATIONAL LEVEL:

Users should have appropriate knowledge about CT scanning, various medical terms related to the cancer and database manipulation.

[2] EXPERIENCE:

Users should have prior information regarding the lung cancer symptoms and detection.

[3] SKILLS:

Users should have basic knowledge and should be comfortable using general purpose applications on computers.

3. THE OPERATING ENVIRONMENT:

Front-end:

[1]JAVA:

Java is a programming language expressly designed for use in the distributed environment. Java is used to create complete applications that may run on a single computer or be distributed among servers and clients in a network. It can also be used to build a small application module or applet for use as part of a Web page. Applets make it possible for a Web page user to interact with the page. It is used to build application module for use as a part of Web page and provides interaction between the pages.

Back-end:

[1]OPEN CV:

Open CV - Python is a library of Python bindings designed to solve computer vision problems.

Python is a general purpose programming language that became very popular because of its simplicity and code readability. It enables the programmer to express ideas in fewer lines of code and hence improves the readability.

Python can be easily extended with C/C++, which allows us to write computationally intensive code in C/C++ and create Python wrappers that can be used as Python modules.

This gives us two advantages: first, the code is as fast as the original C/C++ code and second, it easier to code in Python than C/C++. OpenCV-Python is a Python wrapper for the original OpenCV C++ implementation.

OpenCV-Python makes use of Numpy, which is a highly optimized library for numerical operations with a MATLAB-style syntax.

4. LITERATURE REVIEW:

In [1] surveyed paper American Cancer Society has estimated the numbers of new cancer cases and deaths that will occur in the United States in the year 2015 and compiles the most recent data on cancer incidence, mortality, and survival. Incidence data were collected by the National Cancer Institute (Surveillance, Epidemiology, and End Results [SEER] Program), the Centre's for Disease Control and Prevention (National Program of Cancer Registries), and the North American Association of Central Cancer Registries. Mortality data were collected by the National Centre for Health Statistics. A total of 1,658,370 new cancer cases and 589,430 cancer deaths are projected to occur in the United States in 2015.

To calculate the exact number of invasive cancer cases, a generalized linear mixed model was used for estimation of complete counts for each county from 2000 through 2014 using high-quality incidence data from 48 states. The effect of screening reductions on the incidence of advanced disease is being monitored closely. Cancer survival has improved since the mid-1970s for all of the most common cancers except those of the uterine cervix and uterine corpus,50 although increased survival for some cancer types are difficult to interpret because of changes in detection practice.

The [2] researched paper deals Computer Aided Diagnosing (CAD) system. This system was proposed in this paper for detection of lung cancer form the analysis of computed tomography images. Several problems were there for Computer Aided Diagnosis. This system first segments the area of interest (lung) and then analyses the separately obtained area for nodule detection in order to diagnosis the disease.

The initial step for detecting the lung cancer is done using noise removal and Weiner Filter. Wiener filter is used to filter out noise that has corrupted a signal based on statistical analysis. To implement the Wiener filter in practice we have to estimate the power spectra of the original image and the additive noise. The next phase is segmentation which is used is for the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. The result of image segmentation is a set of segments that collectively cover the entire image for further processing.

Sobel edge detection method is used to convolve the image g(r, c) to get smooth Image. After the segmentation is performed on lung region, the features can be obtained from it like:

• Area of the interest
•Calcification
• Shape and
• Size of nodule
• Contrast Enhancement

In the end the diagnostic indicators are applied so that the maximum of regions which does not considered as cancerous nodules are eliminated. In the end they have achieved an automatic CAD system for early detection of lung cancer by analysing LUNG CT images using several steps.

[3]rd paper illustrates that excision repair cross complementing group 1 gene product (ERCC1) and the regulatory subunit of ribo-nucleotide reductase (RRM1) were reported as being prognostic of outcome and predictive of therapeutic efficacy in patients with non–small cell lung cancer. This studies mainly focus on identifying more effective genomic biomarkers, demographic factors, and other clinical variables. In this study, they investigated a new quantitative image feature analysis approach using chest CT images and demonstrated two new study results. First result inhibits an image feature based classifier yields in significantly higher performance than two popular genomic biomarkers in predicting cancer recurrence risk and the second result states that image features and genomic biomarkers are not highly correlated and provide supplementary information. Resulting in the fusion of these two types of features and biomarkers further improved prediction performance. They also performed several additional experiments and observed a number of interesting results.

1. Optimal feature selection is crucial factor in developing a CAD-based quantitative image feature analysis scheme.
2. When using the original dataset of 79 cases to train two classifiers, the image feature based classifier yielded in high AUC.

Tested the performance by fusion of the normalized ERCC1 scores (from 0 to 1) and the prediction.

[4]th paper states that though surgery is the only potentially curative treatment for early-stage non-small cell lung cancer. The aim of customized adjuvant systemic treatment is to optimize the toxicity/benefit ratio for including those with stage I disease.

First, in the ERCC1 protein expression test, the cells showedstained yellow to brown colour is defined as positive cell to detect the coloured cells and estimates the percentage of positive cells within the examined specimen using a 4-scale grading system. As a result, visuallyestimated percentage of stained coloured cells with an assigned ERCC1 protein expression score for the case from 0 to 3, with the percentage of (1) < 10%, (2) 10% to 25%, (3) 26% to 50%, and (4) >50% were obtained.

Second, each case is gradedinto two categories represented by the technologists as eitherhigh (≥50%) or low (<50%) in RRM1 cell expression. They have used used a semi-automated computer-aided detection (CAD) scheme to segment lung tumors for segmentation.
Third, Quantitative tumor image features is done using WEKA data mining software package tool. The following 5 image features are extracted from it are:-

1. Standard deviation of tumor pixel value (CT number)
2. The maximum tumor diameter computed from the targeted CT slice
3. A tumor shape factor
4. The maximum CT number within the tumor volume
5. Gray-Level Non uniformity etc.

Fourth, they used AUC as an assessment index toanalyze and compare the performance of each feature to predict or classify the test cases of our dataset into two classes namely, positive (+) for cases with cancer recurrence and negative (-) for cases withoutcancer recurrence within 3 years.
Fifth, they built and tested a simple Multilayer perceptron based classifier to combine the two genomic biomarkers.
Last, they performed technology development study to find the clinical relevance or whether and how to optimally apply this new CAD-based quantitative image feature analysis scheme to assist decision-making of the clinicians.

Despite the limitations, this preliminary study provides avalid foundation to continue working in this new andpromising CAD field. It also helps to develop and optimize highlyperformed and robust risk prediction schemes that may have potential to assist clinicians in more accuratelyidentifying the patients with a higher risk of lung cancerrecurrence after surgery. Therefore, for these high riskpatients, the post-surgery chemotherapy is required to prevent or minimize the risk of cancer recurrence and thus increasetheir disease-free survival and overall survival time.
In [5] paper, the National Lung Screening Trial has estimated convincing evidence of a substantial mortality benefit of lung cancer screening with low-dose computed tomography (CT) for current and former smokers at high risk. The United States Preventive Services Task Force has recommended screening, triggering coverage of low-dose CT by private health insurers under provisions of the Affordable Care Act. The centres for Medicare & Medicaid Services (CMS) are currently evaluating coverage of lung cancer screening for Medicare beneficiaries. Since 70% of lung cancer occurs in patients 65 years or older, CMS should cover low-dose CT, thus avoiding the situation of at-risk patients being screened up to age 64 through private insurers and then abruptly ceasing screening at exactly the ages when their risk for developing lung cancer is increasing.

[6]th paper brings the advances in imaging technology have ushered in a new era for lung cancer screening in high-risk individuals using computed tomographic (CT) scans. Although most published studies are nonrandomized observational cohorts of volunteers, the ability of CT scans to detect early stage lung cancer is undisputable. What is unresolved is the ability of spiral CT screening to affect lung cancer-related mortality. A large randomized trial sponsored by the National Cancer Institute to address this question is currently under way. Genomic and proteomic approaches promise to complement

the ability of spiral CT to detect early lung cancer in the next few years.

5. EXISTING SYSTEM:-

There are various systems available that predicts whether the image contains cancer nodule or not.

One of them is "A Neural-NetworkBased Approach to Lung Nodule Detection" which is described below.

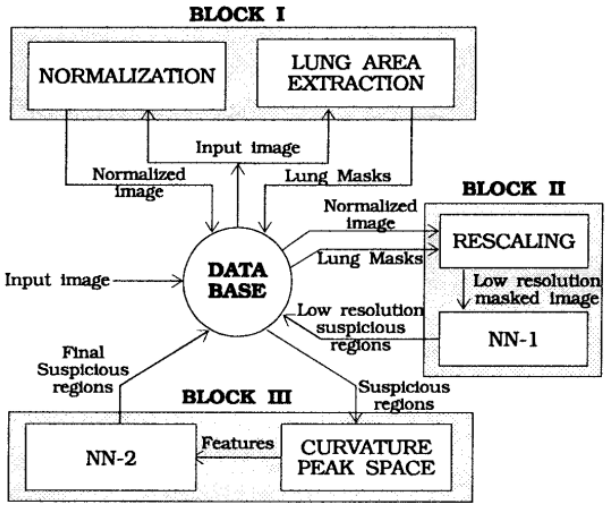
Working:-

It consist of a database storing the input image set.This system contains 3 blocks:-

1st block deals with the normalization of image and lung area extraction from it.

2nd block performs various operations on normalized image and low resolution suspicious regions are detected.

3rd block get the input as the suspicious regions and by extracting various features from it, it will finally predict the suspicious regions.



6. PROPOSED SYSTEM:-

Algorithm:-

The algorithm comprises of several phases as follows:-

1. In the very beginning we will be initiating the system by loading the web page and logging into the system.
2. In the second phase we will obtain the CT scanned image of logged in user. This image will be further used for processing so we will store it in the database.

3. The image obtained from the user is a CT image which has intensity levels lying between 0-1. To extract the features from an image we will convert it into digital image.

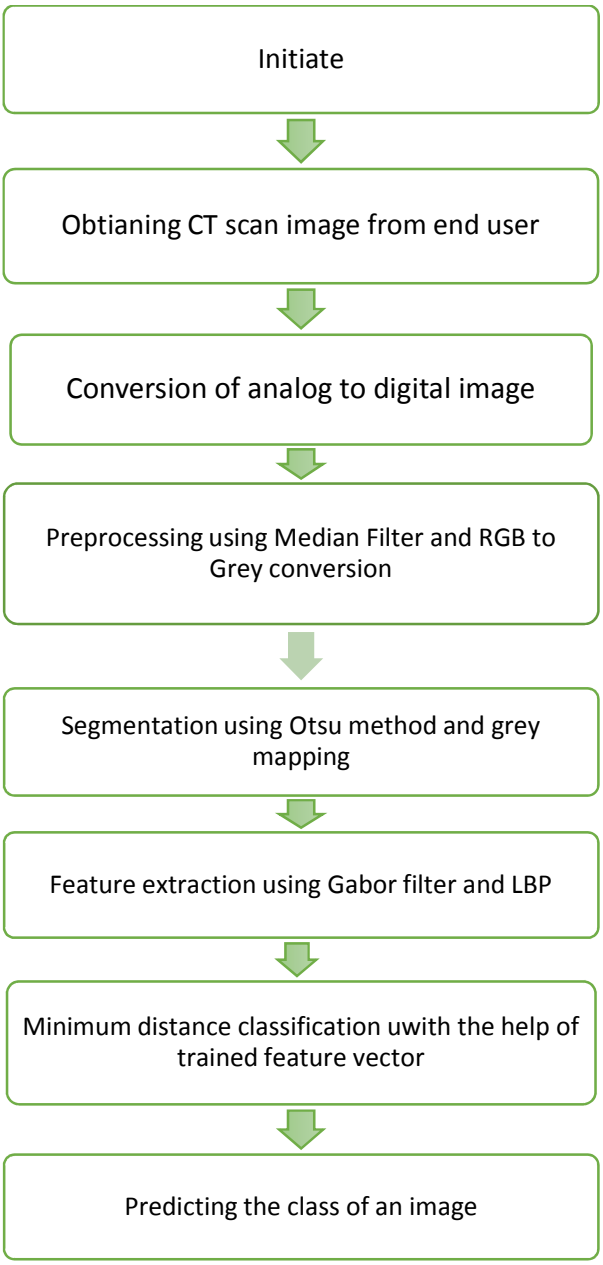
4. Since the converted image may conduct the blur edges so, to sharpen them external noise is added and sharpened images are obtained from this step.

5. Smoothening of image includes various image processing algorithms. That will help to gain features from an image.

6. In the end we will be using a classifier that will classify the cancerous and non-cancerous images. Such classifiers are termed as Binary classifiers.

7. At the very end we will display the result to the user. If the image entered contains cancer nodules then that part will be highlighted else "Null" reports will be given to the user.

FLOWCHART:



Advantages:

1) Portable system.
2) Easy To Use.
3) Faster Data Processing: All the tasks are controlled by single system which leads to fast data processing.

CONCLUSIONS:-

The main aim is to develop a new quantitative image feature analysis scheme and investigate its role using a Gabor filter.Gabor Filter is a linear filter whose impulse response is definedby a harmonic function multiplied by a Gaussian function. Because of the multiplication-convolution property (Convolutiontheorem), the Fourier transform of a Gabor filter's impulse response is the convolution of the Fourier transform of the harmonic function and the Fourier transform of the Gaussian function.

REFERENCES:-

[1] R. Siegel, *et al.*, "Cancer statistics, 2015," *CA Cancer J Clin.,* vol. 65, pp.5-29, Jan/Feb., 2015.

[2]Identifying Lung Cancer Using Image Processing Techniques Disha Sharma, Gagandeep Jindal [International Conference on Computational Techniques and Artificial Intelligence (ICCTAI'2011)]

[3]G. Bepler, *et al.*, "ERCC1 and RRM1 in the international adjuvant lung trial by automated quantitative in situ analysis,"

[4] Fusion of Quantitative Image and Genomic Biomarkers to Improve Prognosis Assessment of Early Stage Lung Cancer patients [DOI10.1109/TBME.2015.2477688, IEEE Transactions on Biomedical Engineering]

[5] L. Humphrey, *et al.*, "Screening for lung cancer with low-dose computed tomography: a systematic review to update the U.S. Preventive Services Task Force Recommendation", *Ann. Intern. Med.,* vol. 159, no. 6, pp. 411-420, Sep. 2013.

[6] American Cancer Society, (2014, Nov.). *Non-small cell lung cancer survival rates by stage.*