Lung Cancer Detection through Image Processing

Abhigna B.S[1]
School of Computer Science and Engineering, Lingaya's University, Haryana, India,
abhigna17@gmail.com

Nishant Sai[2]
School of Computer Science and Engineering, Lingaya's University, Haryana, India,
nsai56@gmail.com

Prof. Tapas Kumar[3]
[3]School of Computer Science and Engineering, Lingaya's University, Haryana
kumartapas534@gmail.com

Abstract: Lung cancer has turned out to be the most widespread genre of cancer among both men and women. When it comes to improving the survival rate, the only conceivable solution is its early detection. CT images are useful in diagnosing the presence of lung cancer as the doctor analyses the CT image of lungs and predicts the presence of tumour. Now, as the chances of false detection are more in case of manual detection, we need a computerized technique for this purpose. Lung cancer detection system can be developed by using several image processing techniques, among which neural networks is used most commonly.The prediction of presence of the lung nodule by the machine involves three stages, namely - pre-processing stage, feature extraction stage and lung cancer cell identification.Neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. So, this helps in predicting the presence of tumour in the lung.

Keywords: neural networks; neuron; nodule; tumour; segmentation.

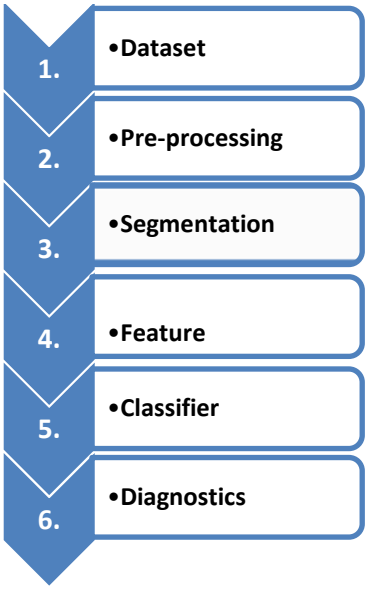
1. INTRODUCTION

It's a matter of regret that cancer is found in every nook and corner of the world causing numerous deaths every year. Lung cancer is the most lethal type of cancers. Around 158,080 people (**85,920** in men and **72,160** in women) die of lung cancer every year around the globe and the death toll seems to be increasing boundlessly year by year. While thistoll among men has reached upland, it's still rising among women.Lung cancer has been proven to be more fatal then the next three most common cancers combined (colon, breast, Pancreatic) and has been the topic-of-the-hour of almost every medical community.According to the American Cancer Society, In 2014, about 224210 new cases of lung cancer and an estimated 159,260 deaths from lung) were reported in the US.

Lung Cancer is basically the uncontrolled growth of abnormal cells which starts off in one or both lungs.It is diagnosed from the CT-Scan images of lungs. Normally a doctor analyses the CT images of the lungs and detect the presence of cancer in the same. But however, in this manual method of detection there are chances of false detection which may occur due to the presence of ribs and blood vessels, presence of air in bronchi, etc.

Hence, it is essential to develop a computerized method for detection of cancer. Image processing is a very handy concept for developing such a method. So, when CT image of lung is processed by certain image processing tools and techniques,the machine specifies whether a cancer nodule is present or not in the lung.

The schema of this system is as below –



This system intends to detect cancer nodules with minimum false negative rate. The proposed system consists of some steps such as: collect lung CT scan image dataset, pre-processing, extraction of the lung region using ROI, feature extraction and to train the classifier to classify the images as normal or abnormal. So, basically this paper is focused on building an efficient and accurate computerized method for lung cancer detection.

2. LITERATURE REVIEW

Almost every researcher in this field has aimed to develop a system which accurately predicts and detects the cancer in its early stages and at the same time, they tried to improve the accuracy of the Early Prediction and Detection system by pre-processing, segmentation feature extraction and classification techniques applied on the extracted database. Given below are a few major contributions of the research –

Hao Han etal.[1]- Theyproposed the developing of a novel system CADe (computer-aided detection) for fast as well as adaptive detection of the pulmonary nodules in the input CT-scan images through the hierarchical vector quantization method. The high level VQ gives more accurate segmentation of the lungs from chest volume and hence is used for segmenting lung region. The low level VQ proved effective for INCs detection and segmentation. False-positive reduction was obtained by using rule-based filtering operations in combination with a feature-based support vector machine classifier. This system was validated on 205 dataset from the publically available online LIDC (Lung Image Database Consortium) database. Finally, diagnostic indicator achieves sensitivity of CADe system 82.7% at specificity of 4 FPs/scan.

M. New Begin et al.[2] - Here, they pointed out the fact Lung cancer, being one of the most dangerous diseases across the globe, usually spreads internally due to the unusual cell growth of tissues in the lungs. It is interesting to note that on being detected early, the survivability of the patient of this cancer can be increased. This paper revolves around the vivid concepts of data mining that are used for prediction purposes of Lung cancer. Also, Ant Colony Optimization (ACO) technique has been explained briefly. This technique is supposed to increase/decrease the disease prediction value of the disease under analysis. So, this study focuses on assorting data mining and ACO techniques for rule generating and classification purposes of the tumour and also, provides the basic framework for simplifying the medical diagnosis.

Ada et al. [3]-In this paper, they developed an automated diagnostic system for early detection and prediction of Lung cancer survival using neural network classifier to check the state of a patient in its early state, whether it is normal or abnormal. In the pre-processing stage, histogram equalization is used on images. Features are extracted via GLCM, and then binarization approach and PCA. The results have been shown on 909 CT images of different classifier by using WEKA data mining tool.

DasuVaman et al. [4] - In this paper, image quality and accuracy are the core factors of this research, image quality assessment as well as improvement are dependent on the enhancement stage where low pre-processing techniques is used, based on Gabor filter within Gaussian rules. Following the segmentation principles, an enhanced region of the object of interest that is used as a basic foundation of feature extraction is obtained. Relying on general features, a normality

comparison is made. In this research, the main detected features for accurate images comparison are pixels percentage and mask-labelling.

FatmaTaher et al. [5]- This paper described a Bayesian classification and a Hopfield Neural Network algorithm for extracting and segmenting the sputum cells for the purpose of lung cancer early diagnosis. The HNN segmentation algorithm outrun the Fuzzy C-Mean clustering and gave successful results after extraction of nuclei and cytoplasm regions. HNN algorithm outperforms better results after using morphological operations on the segmented area.

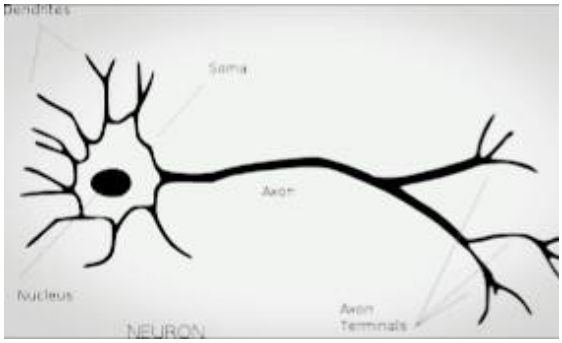
S Vishukumar et al. [6] - Here, authors mostly focus on significant improvement in contrast of masses along with the suppression of background tissues is obtained by tuning the parameters of the proposed transformation function in the specified range. The manual analysis of the sputum samples is time consuming, inaccurate and requires intensive trained person to avoid diagnostic errors. The segmentation results will be used as a base for a Computer Aided Diagnosis (CAD) system for early detection of cancer, which improves the chances of survival for the patient. In this paper, authors proposed Gabor filter for enhancement of medical images. It is a very good enhancement tool for medical images.

Disha Sharma et al. [7] - In their paper, developed an automatic CAD system for early detection of lung cancer by analysing LUNG CT images. First, extracting the lung regions from the CT image using several image processing techniques, including bit image slicing, erosion, and Weiner filter. To convert the CT image into a binary image, Bit plane slicing techniques is used in the extraction process. After extraction, the extracted lung regions are segmented using region growing segmentation algorithm. To classify cancer nodules rule based technique was used. From the extracted features, set of rules were generated and diagnostic indicator achieved accuracy of 80%.

3. ARTIFICIAL NEURAL NETWORK

In lung X-Ray, pulmonary nodule appears as a spherically shaped mass. It can be distorted by adjacent anatomical formation. There is no limit fixed on size or spreading in lung tissue. Pre-diagnosis approaches help to locate the risk of lung cancer disease in very early stage.

In the diagnosis of lung cancer, several approaches such as- genetic algorithms, artificial neural networks, supervised learning methods are used.



An Artificial Neural Network (ANN) is an information processing model that is highly inspired by the way biological nervous systems, such as the brain processes information. The key element of this paradigm is the novel structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in unison to solve specific problems. ANNs, like people, learn by example. An ANN is usually configured for a particular application, such as pattern recognition or data classification, through a learning process. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. This is true for ANNs as well.Such models have three simple sets of rules, namely - multiplication, summation and activation. [8]

At the entrance of every artificial neuron, every input value is multiplied by individual weight in the middle section of artificial neuron. The result would be the sum function that gives the total of all weighted inputs and bias, whereas, at the exit of artificial neurons the sum of previously weighted inputs and bias is passed through the activation function that is also referred to as the transfer function.

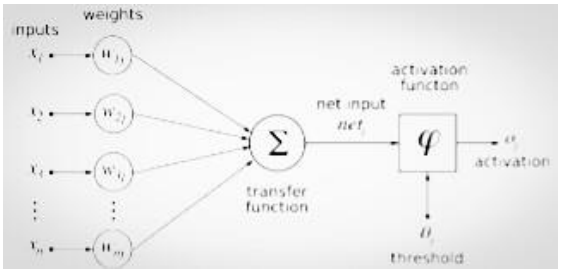


Figure-2 : Architecture of simplest form of Neural Network

Now the question arises why do we need Artificial Neuron Networks? The reason is that neural networks, with their remarkable ability to derive meaning from complicated or imprecise data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. A trained neural network can be thought of as an "expert" in the category of information it has been given to analyse. This expert can then be used to provide projections given new situations of interest and answer "what if" questions. Other advantages include:

1. Adaptive learning: An ability to learn how to do tasks based on the data given for training or initial experience.
2. Self-Organisation: An ANN can create its own organisation or representation of the information it receives during learning time.
3. Real Time Operation: ANN computations may be carried out in parallel, and special hardware devices are being designed and manufactured which take advantage of this capability.

But what makes ANNs different from the conventional computers is the kind of approach neural networks take to problem solving. Conventional computers use an algorithmic approach i.e. the computer follows a set of instructions in order to solve a problem. Unless the specific steps that the computer needs to follow are known the computer cannot solve the problem. That restricts the problem solving capability of conventional computers to problems that we already understand and know how to solve. But computers would be so much more useful if they could do things that we don't exactly know how to do.

Neural networks process information in a similar way the human brain does. The network is composed of a large number of highly interconnected processing elements(neurones) working in parallel to solve a specific problem. Neural networks learn by example. They cannot be programmed to perform a specific task. The examples must be selected carefully otherwise useful time is wasted or even worse the network might be functioning incorrectly. The disadvantage is that because the network finds out how to solve the problem by itself, its operation can be unpredictable.

On the other hand, conventional computers use a cognitive approach to problem solving; the way the problem is to solved must be known and stated in small unambiguous instructions. These instructions are then converted to a high level language program and then into machine code that the computer can understand. These machines are totally predictable; if anything goes wrong is due to a software or hardware fault.

However, neural networks and conventional algorithmic computers are not in competition, in fact they complement each other. There are tasks that are more suited to an algorithmic approach like arithmetic operations and tasks that are more suited to neural networks. Even more, a large number of tasks, require systems that use a combination of the two approaches (normally a conventional computer is used to supervise the neural network) in order to perform at maximum efficiency.

Neural networks do not perform miracles. But if used sensibly they can produce stupendous results.[10]

4. CONCLUSION

Lung cancer is the deadliest of all the cancers. Most of the tumours can be easily detected, provided the condition that the process parameters are selected carefully. However, the only silver ling of the cloud when it comes to the cure of this disease is its early detection. Indeed, it's a matter of fact that early detection can save the victims of this fatal disease from death. So, technologies that contribute in early detection are

a must in abundance but the issue in hand now is that these systems should be capable of detecting and quantifying cancer with both- high accuracy as well as less computational time, which would be favourable for the doctors to ease the treatment. Artificial Neural Network modelling has been a very attractive field in Image Processing area when it comes to medical analysis lately. Very few studies have been conducted on CT- images yet so the aspiration of the above study is to contribute in the development of an approach to detect tumours using CT images.

5. REFERENCES

[1] Hao Han, Lihong Li, Fangfang Han, Bowen Song, William Moorre and Zhengrong Liang "Fast and Adaptive Detection of Pulmonary Nodules In Thoracic CT Images using a Hierarchical Vector Quantization Scheme", IEEE journal of Biomedical and Health Informatics (2014).

[2] T. Sowmiya, M. Gopi, M. New Begin L.Thomas Robinson "Optimization of Lung Cancer using Modern data mining techniques." International Journal of Engineering Research (2014).

[3] Ada, Rajneetkaur, "Early Detection and Prediction of Lung Cancer Survival Using Neural Network Classifier" International Journal of Application Volume 2, Issue 6 (June 2013).

[4] DasuVaman Ravi Prasad, "Lung cancer detection using image processing techniques", International journal of latest trends in engineering and technology (2013).

[5] FatmaTaher, NaoufelWerghi, Hussain Al-Ahmad , RachidSammouda " Lung Cancer Detection Using Artificial Neural Network and Fuzzy Clustering Methods", American Journal Of Biomedical Engineering (2012).

[6] S. Vishukumar, K. Patela and PavanShrivastavab, "Lung Cancer Classification Using Image Processing", International Journal of Engineering and Innovative Technology Volume 2, Issue 3, (September 2012).

[7] Disha Sharma, Gagandeep Jindal "Identifying Lung Cancer Using Image Processing Techniques", International Conference on Computational Techniques and Artificial Intelligence (ICCCT AI '2011).

[8] FatmaTaher, NaoufelWerghi, Hussain Al-Ahmad, RachidSammouda :"Lung Cancer Detection using artificial neural network and clustering methods", American Journal of Biometric Engineering (2012).

[9] F. Taher and R. Sammouda," Identification of Lung Cancer based on shape and Color", Proceeding of the 4th International Conference on Innovation in Information Technology, pp.481-485, Dubai, UAE, Nov. 2007.

[10] M. G. Forero, F. Sroubek and G. Cristobal, "Identification of Tuberculosis Based on Shape and Color," Journal of Real time imaging, 2004.