MACHINE LEARNING APPROACH FOR LUNG CANCER DETECTION AND STAGES PREDICTION

Swati Mukherjee*, Prof. Sneha Bohra**

* *Student, Department of Computer Science & Engineering, GHRU, Amravati, India*
** *Assistant Professor, Department of Computer Science & Engineering, GHRU, Amravati, India*

Abstract- Diagnosing the patients correctly and administering treatments are a major challenge for medical practitioners as critical decisions are made based on diagnosis. The proposed framework is intended to identify lung malignancy in untimely stage in two phases. The proposed framework comprises of numerous steps, for example, image extraction, pre-processing, binarization, thresholding, Division, feature extraction, and neural system identification.

In our system we will develop 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. In recent past, there has been a lot of progress on data mining and machine learning techniques to predict the various types of diseases.The proposed model is a tool which will take input as the CT scan images and it will predict the possibilities of the disease and its stages. Thus we will try to provide a direction to medical practitioners to make quick intelligent clinical decisions which can help in prophylaxis of the disease and thereby reduce any treatment costs.

Keywords- Convolutional Neural Network (CNN), Computed Tomography(CT), Support Vector Machine (SVM), Lung Cancer.

I. INTRODUCTION

Cancer is a condition in which some cells in the body grow deliriously, and in some case, spread and occupy organs in other parts of the body — a process called metastasis. This abnormal growth is detrimental because it doesn't just replace healthy cells in organs, but also causes changes in our body's biochemistry leading to weight loss and a compromised immune system thereby, leading to death.

There are over 200 different types of cancers but Lung Cancer is one of the heinous diseases that attack the patients abruptly and sometimes there is no proper diagnosis system and hence it is not treated accurately due to lack of time. The earlier the detection is, the higher the chances of successful treatment.

In 2018,[7] the lung cancer led to around 7 lakh deaths in India alone. The Indian Council of Medical Research (ICMR) estimates that the country is plausible to enroll over 17 lakh new cases and record over 8 lakh deaths by 2020.

So it is really necessary to track down this deadly disease on time but categorization of lung cancer requires careful pathological test, namely, needle biopsy specimen and analysis by expert doctors.

Thus it involves sensible human insight and their intellect. To curtail the human effort, an automated computerized system became the need.

Present-day progress in machine learning, image processing, pattern recognition, dimensionality reduction and classification methods has eventually lead to alternate classification method for lung cancer detection.
Prognosis based system, thus, widely uses algorithms such as Artificial Neural Network, Decision tree, Fuzzy Logic, SVM, autoencoders and other image processing methods.
Motivated by the success of deep learning in relevant fields, a deep learning based classification method will be investigated in this work.

II . LITERATURE SURVEY

1. **Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches.**

In this paper, CAD system(Computer Aided Diagnosis) was developed for characterizing of lung nodules in CT images. The system had two sections.[1] First section is based on supervised learning approach which uses a 3D CNN and MTL (Multi Task learning) for classification. Second section is based on unsupervised learning approach which uses clustering and a Proportion SVM to classify tumors. The paper described the list of different experiments performed using both supervised and unsupervised learning along with their evaluation sets. Compared to other methods, 3D CNN with multi task learning [1] approach showed accuracy of 91.26%.Thus they evaluated both supervised and unsupervised learning algorithms to find characterization of lung nodules and pancreatic cysts.

2. **Cancer lung detection on CT scan image using ANN back propagation based gray level co-occurrence matrix feature.**

This research focused on detection of lung cancer using ANN back propagation based GLCM features. The paper illustrated the two types of CT images that are categorized as normal lung CT image and lung cancer CT image[2]. Image processing techniques used for intensity improvement were grayscale, binary, median filters and adaptive histogram equations. This was done for image enhancement, noise removal and contrast, results of which are shown in the document. Second step was image segmentation which helps to find segment that has a potential as a tumour. The next statistical technique for texture feature extraction was GLCM(Gray Level Co-occurence Matrix).GLCM[2] features used were homogeneity, energy, contrast, correlation and variance. The GLCM features extrication results were then trained using ANN back propogation method which has 3 layers and it could distinguish between cancerous and normal images with accurateness of 80%.

3. **Lung cancer detection with fusion of CT and MRI image using image processing**

Using Image processing ,the tumor of malignant growth can be rapidly distinguished. Fusion of both CT and MRI images were examined. CT image scans[3] the denser tissues and MRI scans the softer tissues so by combination, more quality data is acquired from fused image. The process followed steps such as image capture, enhancement, segmentation, feature extraction and then cancer detection.

Pre-processing had steps like Noise reduction using median filter and image enhancement by Gabor filter. Watershed segmentation[3] was performed to find distinct areas in the image and GUI was created using MATLAB. Determination of nodule from different segmented areas was done by SVM algorithm. Feature extraction was achieved based on parameters such as area, perimeter and eccentricity of the pictures.

4. Segmentation and analysis of CT chest images for early lung cancer detection.

In this paper, an improved technique of Hopfield ANN classifier was proposed to section extricated lung locales from human chest CT pictures. It was the CADetection [4] system to detect lesions from medical CT images. Classical image processing methods were employed in which bit planes representation of raw 3D CT pictures produces 2D slices. Some of the steps applied were Median Filter, Outlining , Lung Border Extraction and Floodfill Algorithm. After extraction, the retrieved lung patches were segmented using unsupervised Modified Hopfield NN classifier [4].

After segmentation, diagnosis rules (consist of 3 main filters) are applied to cancer nodules in lung. This CADe System implemented in MATLAB gave promising results when tested with a database of images given by National Cancer Centre of Japan.

5. Small cell lung cancer detection using a supervised ML algorithm

This paper has referred to small cell EDM calculation (EDM: Entropy Degradation Method) [5] with histogram which distinguished SCLC(small cell lung cancer) from CT images for early malignancy identification. It was demonstrated that EDM had great precision and calculation can be utilized in clinical setting. It was observed that precision was 77.8%, when tested with 12 lung CT filters[5]. Training data and testing data were based on lung CT scan and was provided by National cancer Institute.

6. Lung cancer classification using deep learning features on low population dataset

In this work, Deep Learning was used to classify lung cancer. Classification was initially difficult using High dimensional dataset but in this approach, classification was possible using autoencoders and deep learning [6].Input and labels were fed into the stacked classifier with deep features. This stacked classifier consisted of autoencoders and FNN (feed forward neural network) and then output is obtained. This research gave details on the accuracy levels of Deep NN compared to other neural networks.

III. EXISTING SYSTEM APPROACH

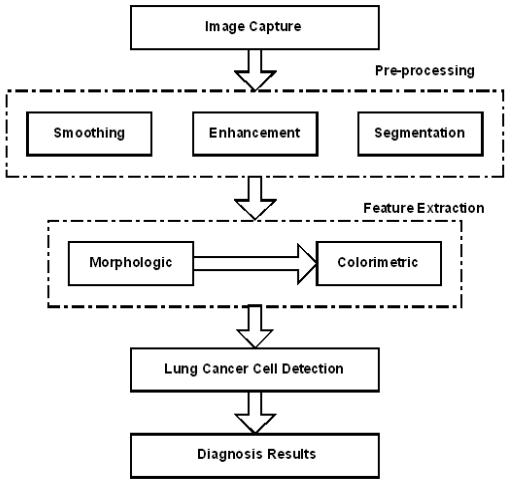

Fig.1 Block Diagram of Existing System

The existing system presented based on three main processes used in which the report; Pre-processing, feature extraction and finally the classification process. Implementation was done using MATLAB. The processes followed by existing lung cancer detection system are described in above Figure 1. Support Vector Machines was used as classifier to model data. They used generalization control with fusion of technique to solve curse of the dimensionality. The basic features examined for extraction are area, perimeter and eccentricity.

Disadvantages

1. Due to low dimensional dataset it needs more work for data cleaning and feature extraction.
2. A system is thus needed by improving the prepossessing process, image segmentation, feature extraction, and learning process.

IV. CONCLUTION

We are going to achieve high precision over existing work done on lung cancer detection using low dimensional datasets. We are going to utilize deep neural network for machine model creation which will show remarkably accurate results.

V. FUTURE WORK

For future work, we can impose this technique on real time large dataset images. Increasing the number of images and using combination like CT, PET,MRI etc images for the process, can preserve the accuracy of our system.

ACKNOWLEDGMENT

This work is supported in Lung Cancer Detection System for medical research and development of any state in India. Authors are thankful to Faculty of Engineering and Technology, Raisoni University, Amravati for providing the facility to carry out the research work.

REFERENCES

[1] Sarfaraz Hussein, Pujan Kandel, Candice W. Bolan, Michael B. Wallace, and Ulas Bagci, "Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches" , 2019.2894349, IEEE

[2] 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

[3] 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

[4] 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

[5] Qing Wu, Wenbing Zhao, "Small-cell lung cancer detection using a supervised ML algorithms." International Symposium on Computer Science and Intelligent Controls 978-1-5386-2941-3/17 2017 IEEE

[6] Syed Moshfeq Salaken, 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

[7] www.icmr.nic.in.