TECHNO INTERNATIONAL NEW TOWN

(Formerly known as Techno India College of Technology)

Block-DG, Action Area 1, New Town, Kolkata -700156, West Bengal, India

Department of Computer Science & Engineering

Sixth Semester Project-I Report (PROJ-CS601)

Lung Cancer Detection

Prepared by

Saikat Sheet (Roll No: - 18700120024)

Souvik Ghosh (Roll No: - 18700120031)

Rajdip Manna (Roll No: - 18700120042)

Sayak Pal (Roll No: -18701620010)

Academic Year: 2022-2023

Stream: - Computer Science & Engineering

Year of Study: 3rd

Affiliated to

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL

MAULANA ABUL KALAM AZAD UNIVERSITY OF TECHNOLOGY, WEST BENGAL

(FORMERLY KNOWN AS WEST BENGAL UNIVERSITY OF TECHNOLOGY)

CERTIFICATE

This is to certify that the project report entitled " Lung Cancer Detection " has been prepared by the following students of the Department of Computer Science & Engineering under my supervision in partial fulfillment for the degree of Bachelor of Technology (B.Tech.) in Computer Science & Engineering which is affiliated to Maulana Abul Kalam Azad University of Technology, West Bengal (Formerly known as West Bengal University of Technology) in the academic year 2021-2022.

It is to be understood that by this approval, the undersigned does not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn thereof, but approves the report only for the purpose for which it has been submitted.

Sl. No.	Name	Roll Number
1	Saikat Sheet	18700120024
2	Souvik Ghosh	18700120031
3	Rajdip Manna	18700120042
4	Sayak Pal	18701620010

…….
Prof. Bikash Sadhukhan,
HOD, Computer Science & Engineering,
Techno International New Town

Abstract

Lung cancer is the crucial disease, which is one of the most human beings and is considered one of the deadliest cancer in the world and it causes to millions of deaths around the globe. Therefore early detection and classification of lung cancers can save millions of lives. In recent times, machine learning has been a better technique for diagnosing lung cancer. In recent times, the need for the application of machine learning in health care has become crucial. Most healthcare administrators and researchers use various machine learning techniques to improve diagnosis, detection, and prediction of health related diseases to support experts in decision making. This study focused on the application of machine learning methods to diagnose lung cancer. The research identified the exiting prediction methods and tools used to detect, predict, compare lung cancer systematically. However, the conventional methods were failed to give result the better classification performance. We present an approach to detect lung cancer from CT scans using deep learning. We depict a pipeline of pre-processing techniques to highlight lung regions to cancer and extract features. Thus, this is implemented using deep learning convolutional Neural Network, Mobilenet and VGG-16 model for lung cancer detection and classification operations. Initially, pre-processing of Computed Tomography (CT) based lung images were performed using histogram equalization and thresholding segmentation, which effectively localizes the region of interest (ROI) of cancer.

CONTENTS

Introduction

Lung cancer is the leading cause of cancer-related deaths all around the world. The main reason behind death is that it is not recognized at earlier stage, and the rate of death of lung cancer patient is higher compared to other cancer types such as blood breast, and prostate cancer. In addition, the mortality rate of lung cancer must be kept increasing. One of the important steps in detecting early stage cancer is to find out whether there are any pulmonary nodules in the lungs which may grow to tumor in recent future. This work aims to determine the likelihood of a given CT scan images of lungs to be cancerous. The pre-processing processes have a direct influence on the ultimate CT scan preparation. Subsequently, more accurate segmentation procedures for lung CT scan are a fascinating problem deserving of investigation, having urgent realistic necessity and clinical value. Multiple lung divisions procedures have been investigated, and a fraction of the conventional approaches integrate threshold region-development techniques. Regardless, the results are not very promising, and the procedure is time-consuming and tedious. Deep learning is a fundamental image segmentation approach that is reliant on the area of the CT scan. It has the ability to quickly and efficiently separate the interstitial lung borders. While this approach is effective, it is time-consuming, and the developing model is sensitive to boundary conditions.

Problem Definition

In this project we have focused on the detection of three types of lung cancer (adenocarcinoma, large cell carcinoma, squamous cell carcinoma).

These three different types of lung cancer has their characteristics and patterns.

Adenocarcinoma: Cancer that forms in the glandular tissue, which lines certain internal organs and makes and releases substances in the body, such as mucus, digestive juices, and other fluids.

Large cell carcinoma: Lung cancer in which the cells are large and look abnormal when viewed under a microscope.

Squamous cell carcinoma: Squamous cell lung carcinoma begins in the top layer of cells, called squamous cells, that line the large airways of the lung. It usually grows in the bronchi that branch off of the main left or right bronchus in the center of the chest.

Our first motive is to detect that the lung is cancerous or non-cancerous. We have four types of lung cancer data that is adenocarcinoma, large cell carcinoma, squamous cell carcinoma and normal (non-cancerous). Firstly histogram equalization is applied to each data to get clear images and then thresholding segmentation to get segmented images with proper outlines. Then there will be two types of data one is cancerous data and another is non-cancerous data. The data of adenocarcinoma, squamous cell carcinoma and large cell carcinoma is belongs to the

cancerous data and the data of normal lung is belongs to the non-cancerous data. Then several convolutional neural network CNN, MOBILENET and VGG-16 is used for training and testing in which it will show which type of conventional neural network is detecting more accurately based on that scores will be given.

Related Work

Lung cancer detection has earlier been studied using image processing techniques. With the advent of neural networks and deep learning techniques, these have recently been used in the medical imaging domain. Various researchers have tried to classify, detect lung cancer using machine learning and neural networks. Not many deep learning techniques have been applied to detect lung cancer. This is because of the lack of a large dataset for medical images especially lung cancer.

Kavitha and Prabakaran (2019) have presented lung nodules detection system using pre-processing approached with filtering techniques. Using segmentation, the uploaded image is divided into maximum of 6 strips. For optimization and grouping of the divided images Particle Swarm Optimization (PSO) with Fuzzy C-means approach is used. Using optimization approach, appropriated centre selection in cluster formation has been performed. The presented model has provided 95 % of detection accuracy rate with 5 number strips.

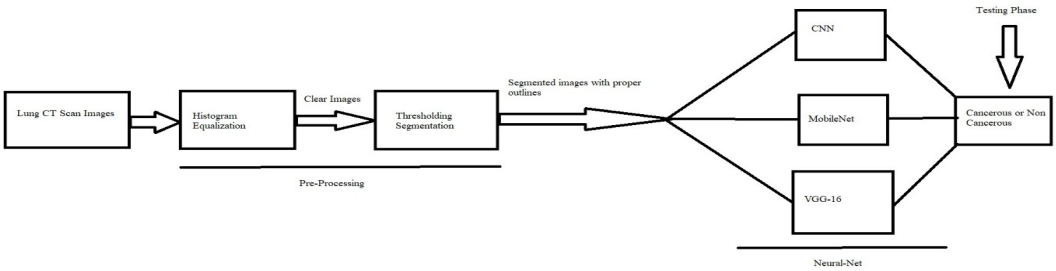
Prabukumar et al. (2019) have proposed automatic diagnosis system for lung cancer, which was integrated with fog computing technology. As we know that fog computing is used to save high quality CT images with better privacy. Here, for segmentation Region of interest is found, which was later integrated with Fuzzy c means approach. Then, for optimization of extracted features Cuckoo Search (CS) is used as an optimization algorithm. The classification of benign or malignant cancer is performed using Support Vector Machine (SVM) approach. The performance of the designed system has been observed in terms of sensitivity, specificity, and accuracy.

Perumal and Velmurugan (2018) have presented detection system for lung cancer using CT scan images. CT images helps to enhance the detection rate with small processing time. Here, the removal of noise in the CT images has been performed using Gabor filter. Otsu's threshold method has been used to find RoI of image, and to increase the processing speed and accuracy HAAR wavelet transform has been used. For the detection of cancer nodes in lungs an enhanced Artificial Bee Colony (ABC) is applied on the test CT image. The entire system helps radiologist to detect lung cancer and provide results with an accuracy of 92.4 %.

Architecture & Detailed Working

At a very early stage, cancer is not being found because of the small size of the nodules and location of the glands. Machine Learning makes the diagnosis process easier and deterministic. This work has introduced one automatic lung cancer detection method to increase the accuracy

and yield and decrease the diagnosis time. The main objective of this work is to detect the cancerous lung nodules from a given input lung image and to predict the lung cancer using Deep Learning technique more efficiently. For deep learning technique several conventional neural networks CNN, MOBILENET, VGG-16 is used for detecting the cancerous lung and based on the accuracy of detection scores will be given. Better score means the better accuracy of the detection.



Methodology

At first CT scan images are taken as dataset and the dataset contains three types of lung cancer adenocarcinoma, large cell carcinoma, squamous cell carcinoma. For image pre-processing histogram equalization operation is performed on the dataset to get the clear images and thresholding segmentation operation is performed to get segmented images with proper outline.

Histogram Equalization: Histogram Equalization is a computer image processing technique used to improve contrast in images. It accomplishes this by effectively spreading out the most frequent intensity values, i.e. stretching out the intensity range of the image. This method usually increases the global contrast of images when its usable data is represented by close contrast values. This allows for areas of lower local contrast to gain a higher contrast.

Thresholding Segmentation: Thresholding is a type of image segmentation, where we change the pixels of an image to make the image easier to analyze. In thresholding, we convert an image from colour or grayscale into a binary image, i.e., one that is simply black and white. Most frequently, we use thresholding as a way to select areas of interest of an image, while ignoring the parts we are not concerned with.

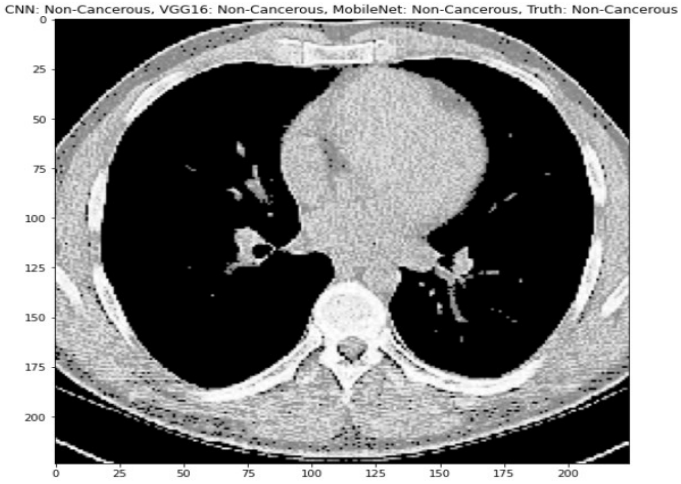
After pre-processing three convolutional neural network architecture is used for the recognition – **CNN**, **MOBILENET**, **VGG-16**.

CNN: A Convolutional Neural Network (CNN) is a type of deep learning algorithm that is particularly well-suited for image recognition and processing tasks. It is made up of multiple layers, including convolutional layers, pooling layers, and fully connected layers. The convolutional layers are the key component of a CNN, where filters are applied to the input image to extract features such as edges, textures, and shapes.

MOBILENET: Mobilenet is a model which does the same convolution as done by CNN to filter images. It uses the idea of Depth convolution and point convolution which is different from the normal convolution as done by normal CNNs. This increases the efficiency of CNN to predict images and hence they can be able to compete in the mobile systems as well. Since these ways of convolution reduce the comparison and recognition time a lot, so it provides a better response in a very short time and hence we are using them as our image recognition model.

VGG-16: A convolutional neural network is also known as a ConvNet, which is a kind of artificial neural network. A convolutional neural network has an input layer, an output layer, and various hidden layers. VGG16 is a type of CNN (Convolutional Neural Network) that is considered to be one of the best computer vision models to date. The creators of this model evaluated the networks and increased the depth using an architecture with very small (3 × 3) convolution filters, which showed a significant improvement on the prior-art configurations. They pushed the depth to 16–19 weight layers making it approx — 138 trainable parameters.

Result:



as we can see from result that CNN, MOBILENET and VGG-16 are able to detect that the lung is cancerous or non-cancerous.

But if we check the scoreline then it is clear that MOBILENET and VGG-16 is better accurate in terms of detection and their score is more higher than normal CNN.

	Accuracy	Precision	Recall	AUC	F1
CNN	0.867257	0.939394	0.704545	0.946476	0.740987
VGG16	0.982301	0.956522	1.000000	1.000000	0.979871
MobileNet	0.991150	0.977778	1.000000	1.000000	0.990000

Conclusion

This report has presented an approach used for early detection of lung cancer from CT images. From above it has been analyzed that Cancer is a serious health issue. The mortality rate of cancer victims associated with lung cancer is very high because even if they are found, their survival rate is very limited. The most widespread and scariest disease is lung cancer. Early analysis and identification can minimize the mortality of lung cancer patients. It is obvious that several different techniques are used for enhancement of detection accuracy. But the steps performed are same like other processes, like data collection, noise removal (pre-processing), segmentation, feature extraction, optimization, and classification. In future, we are planning to implement optimization approach with classification approach for early stage detection of lung cancer using CT images.

Future Scope of Project

Artificial intelligence (AI) simulates intelligent behavior as well as critical thinking comparable to a human being and can be used to analyze and interpret complex medical data. The application of AI in imaging diagnostics reduces the burden of radiologists and increases the sensitivity of lung cancer detection so that the morbidity and mortality associated with lung cancer can be decreased. In this report, we have tried to evaluate the role of artificial intelligence in lung cancer detection, as well as the future potential and efficiency of AI in the classification of nodules. Techniques such as deep learning and machine learning allow automatic characterization and classification of nodules with high precision and promise an advanced lung cancer detection method in the future. Even though several combination models with high performance have been proposed, an effectively validated model for routine use still needs to be improvised. Combining the performance of artificial intelligence with a radiologist's expertise offers a successful outcome with higher accuracy. Thus, we can conclude that higher sensitivity,

specificity, and accuracy of lung cancer detection and classification of nodules is possible through the integration of artificial intelligence and radiology.

References

- Tekade, Ruchita, and K. Rajeswari. "Lung cancer detection and classification using deep learning." 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). IEEE, 2018.
- Bhatia, Siddharth, Yash Sinha, and Lavika Goel. "Lung cancer detection: a deep learning approach." Soft Computing for Problem Solving. Springer, Singapore, 2019. 699-705.
- Radhika, P. R., Rakhi AS Nair, and G. Veena. "A comparative study of lung cancer detection using machine learning algorithms." 2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT). IEEE, 2019.
- Shakeel, P. Mohamed, Mohd Aboobaider Burhanuddin, and Mohamad Ishak Desa. "Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks." Measurement 145 (2019): 702-712.
- Asuntha, A., and Andy Srinivasan. "Deep learning for lung Cancer detection and classification." Multimedia Tools and Applications 79.11 (2020): 7731-7762.
- Joshua, Eali Stephen Neal, Midhun Chakkravarthy, and Debnath Bhattacharyya. "An Extensive Review on Lung Cancer Detection Using Machine Learning Techniques: A Systematic Study." Rev. d'Intelligence Artif. 34.3 (2020): 351-359.
- Alsinglawi, Belal, et al. "An explainable machine learning framework for lung cancer hospital length of stay prediction." Scientific reports 12.1 (2022): 1-10.
- Yan, Sha, et al. "Computed Tomography Images under Deep Learning Algorithm in the Diagnosis of Perioperative Rehabilitation Nursing for Patients with Lung Cancer." Scientific Programming 2022 (2022).