International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 06 | June - 2023 SJIF Rating: 8.176 ISSN: 2582-3930

LUNG CANCER DETECTION USING MACHINE LEARNING

Rohini Sarpale, Saniya Kumawat, Komal Pawar

ABSTRACT

Lung cancer is one of the most common diseases among humans and one of the major causes of growing mortality. Medical experts believe that diagnosing lung cancer in the early phase can reduce death with the illustration of lung nodule through computed tomography (CT) screening. Examining the vast amount of CT images can reduce the risk. However, the CT scan images incorporate a tremendous amount of information about nodules, and with an increasing number of images make their accurate assessment very challenging tasks for radiologists. Recently, various methods are evolved based on handcraft and learned approach to assist radiologists. In this paper, we reviewed different promising approaches developed in the computer-aided diagnosis (CAD) system to detect and classify the nodule through the analysis of CT images to provide radiologists' assistance and present the comprehensive analysis of different methods.

INTRODUCTION

Cancer is a group of diseases characterized by maturation, growth and/or disorganized proliferation of abnormal cell groups . According to the World Health Organization , cancer is a leading cause of death worldwide. In the U.S alone, 1,660,290 new cases and 580,350 deaths from the disease are estimated for the year of 2013. Lung cancer is one of the most common cancers , with estimated 228,190 new cases and 159,480 deaths in the U.S. alone in 2013 . One way to try to minimize this high mortality rate is through early detection and treatment. Recently, advances in computed tomography (CT) has allowed early diagnosis of the disease. According to Awai , the detection rate of lung cancer using CT is 2.6 to 10 times higher than by using analog radiography. However, the use of CT is directly impacting the workload of radiologists who need to analyze an increasing number of screening tests in a short time. result in errors in

detection (failure to detect) or misinterpretation (inability to properly diagnose a tumor).

Therefore, computational systems are needed to assist radiologists in the interpretation of images, nodule detection and determination of their characteristics are needed.This workload can result in errors in detection (failure to detect) or misinterpretation (inability to properly diagnose a tumor). Therefore, computational systems are needed to assist radiologists in the interpretation of images, nodule detection and determination of their characteristics are needed.

LITERATURE SURVEY

SR.NO.	Title	Author	Description
1.	Lung Cancer Disease Diagnosis Using Machine Learning Approach	Swati Mukherjee, Prof . S. U. Bohra	The analysis and study of lung diseases has been the most intriguing investigation zone of medical experts from early days to the present day. To address this concern, a diagnosis system like this can only help diminish the odds of getting risk to human live by early discovery of malignant growth.
2.	Classification of non-small cell lung cancer using one-dimensional convolutional neural network	Dipanjan Moitra , Rakesh Kr. Mandal	Non-Small Cell Lung Cancer (NSCLC) is a major lung cancer type. Proper diagnosis depends mainly on tumor staging and grading. Pathological prognosis often faces problems because of the limited availability of tissue samples. Machine learning methods may play a vital role in such cases. 2D or 3D Deep Neural Networks (DNNs) has been the predominant technology in this domain.

			Contemporary studies tried to classify NSCLC tumors as benign or malignant.
3.	Boosted neural network ensemble classification for lung cancer disease diagnosis	Jafar A. ALzubi a, Balasubramaniyan Bharathikannan b, Sudeep Tanwar c, Ramachandran Manikandan d, Ashish Khanna e, Chandrasekar Thaventhiran	Accurate diagnosis of Lung Cancer Disease (LCD) is an essential process to provide timely treatment to the lung cancer patients. Artificial Neural Networks (ANN) is a recently proposed Machine Learning (ML) algorithm which is used on both large-scale and small-size datasets. In this paper, an ensemble of Weight Optimized Neural Network with Maximum Likelihood Boosting (WONN-MLB) for LCD in big data is analyzed.

4.	A Comparative Study of Lung Cancer Detection using Machine Learning Algorithms	Radhika P R Rakhi.A.S.Nair	The growth of cancerous cells in lungs is called lung cancer. The mortality rate of both men and women has expanded due to the increasing rate of incidence of cancer. Lung cancer is a disease where cells in the lungs multiply uncontrollably. Lung cancer cannot be prevented but its risk can be reduced So detection of lung cancer at the earliest is crucial for the survival rate of patients. The number of chainsmokers is directly proportional to the number of people affected with lung cancer. The lung cancer prediction was analysed using classification algorithms such as Naive

			Bayes, SVM, Decision tree and Logistic Regression. The keyobjective of this paper is the early diagnosis of lung cancer by examining the performance of classification algorithms.
5.	Lung Cancer Detection and Classification usingDeep Learning	Ruchita Tekade1 , Prof. Dr. K. Rajeswari2	In recent years, so many Computer Aided Diagnosis (CAD) systems are designed for diagnosis of several diseases. Lung cancer detection at early stage has become very important and also very easy with image processing and deep learning techniques. In this study lung patient Computer Tomography (CT) scan images are used todetect and classify the lung nodules and to detect the malignancy level of that nodules.

Block Diagram

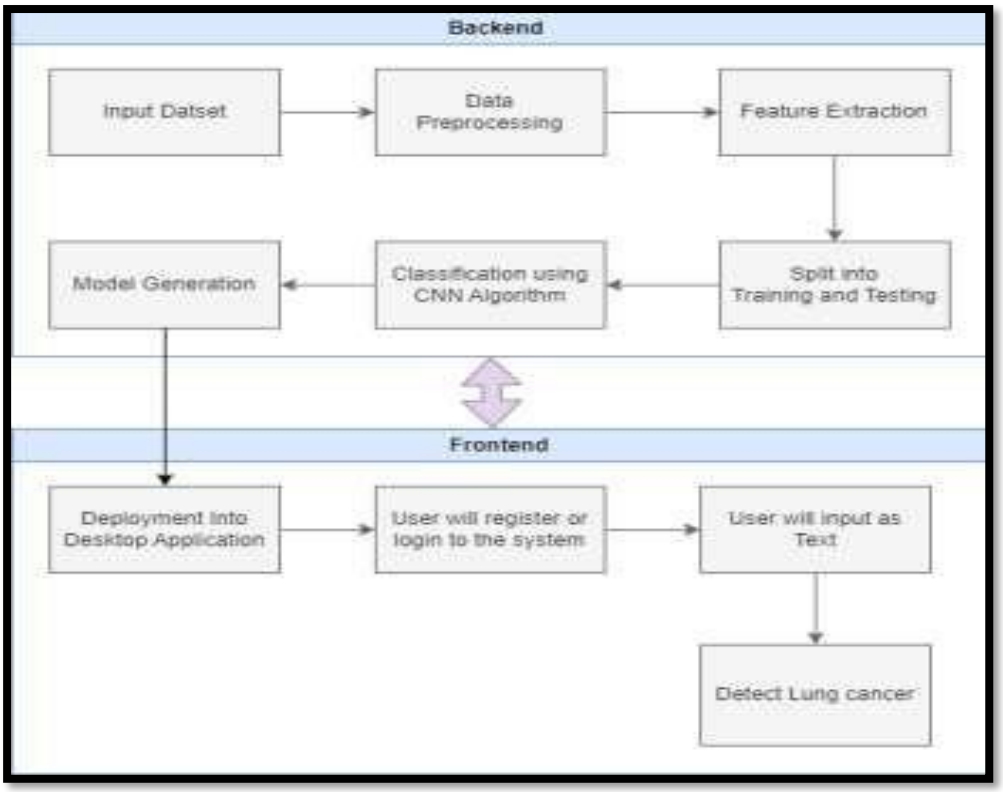

Figure 1.1. Block Diagram

WORKING

In our approach, we pre-process the data and select the most important features. Only those features are selected and the rest are eliminated. The model is trained and tested using various models like CNN, Random forest and hybrid neural network model. The efficiency of variousalgorithms are compared using different performance parameters.

• Dataset – Provide dataset (This means that the data collected should be made uniform and understandable for a machine that doesn't see data the same way as humans do.)

• Pre-processing –A real-world data generally contains noises, missing values, and maybe in anunusable format which cannot be directly used for machine learning models. Data pre-processingis required tasks for cleaning the data and making it suitable for a machine learning model which also increases the accuracy and efficiency of a machine learning model.

• Feature Extraction – Feature Extraction aims to reduce the number of features in a dataset by creating new features from the existing ones (and then discarding the original features). These new reduced set of features should then be able to summarize most of the information containedin the original set of features.

• Classification - The Classification algorithm is a Supervised Learning technique that is used to identify the category of new observations on the basis of training data.In Classification, a program learns from the given dataset or observations and then classifies new observation into a number of classes or groups

RESULT

• For effective diagnosis, early identification of lung cancer is key. Due to the fact that dermoscopic images provide the most viable alternative for utilizing modern image processing and machine learning algorithms. In consideration of the essential steps involved in lesion segmentation, feature segmentation, feature creation and classification.

• In this project, different phases of image processing were applied on input image. From these different image processing techniques, the fuzzy filter will provide the efficient denoising. Segmentation done by marker based watershed algorithm, gives various region of image. GLCM is used to extract the different features of image and which takes less time for generating the result. This results are passed through CNN Classifier, which classifies the nodules as benign or malignant. CNN classifier provides 92.5 percentage accuracy.

CONCLUSION

• We have Propose a system for Heart Disease Predication applied machine learning algorithms like CNN for diagnosis of heart disease. The goal of this study is to predict lung cancer disease based on image . The project is set up in such a way that the system takes the image as input and outputs a prediction of lung cancer disease.

REFERENCES

• Li Q: Recent progress in computer-aided diagnosis of lung nodules on thin-section CT. Comput Med Imaging Graph 2007, 4–5: 248–57.

• Kazuo A, Kohei M, Akio O, Masanori K, Haruo H, Shinichi H, Yasumasa N: Pulmonary nodules at chest ct: effect of computer-aided diagnosis on radiologists' detection performance. Radiology 2004, 230: 347–352.10.1148/radiol.2302030049

• Gomathi M, Thangaraj P: Computer aided medical diagnosis system for detection of lung cancer nodules: a survey. Int J Comput Intell Res 2009, 5: 453–467.

• El-Baz A, Beache GM, Gimel'farb G, Suzuki K, Okada K, Elnakib A, Soliman A, Abdollahi B: "computer-aided diagnosis systems for lung cancer: Challenges and methodologies". Int J Biomed Imaging 2013, 2013: 1–46.

• van Ginneken B, Schaefer-Prokop CM, Prokop M: Computer-aided diagnosis: How to move from the laboratory to the clinic. Radiology 2011,261(3):719–732. 10.1148/radiol.11091710

• Jeon K. N, Goo JM, Lee CH, Lee Y, Choo JY, Lee NK, Shim M-S, Lee IS, Kim KG, Gierada DS, Bae KT: Computer-aided nodule detection and volumetry to reduce variability between radiologists in the interpretation of lung nodules at low-dose screening computed tomography. Investigative Radiol 2012, 47: 457–461. 10.1097/RLI.0b013e318250a5aa

• Lee SLA, Kouzani AZ, Hu EJ: Automated detection of lung nodules in computed tomography images: a review. Mach Vision Appl 2012,23(1):151–163. 10.1007/s00138-010-0271-2

• Suzuki K: A review of computer-aided diagnosis in thoracic and colonic imaging. Quant Imaging Med Surg 2012,2(3):163–176.