International Journal of Advance Research in Science and Engineering
Volume No. 11, Issue No. 04, April 2022
www.ijarse.com
IJARSE
ISSN 2319 - 8354

Lung cancer detection using ml

Atharva Khedekar[1],Onkar ekre[2] ,Mrs.Sushadevi Adagale[3]

[1] *Student,Computer Engineering, Trinity Academy of Engineering, Pune, India*

[2]*Student,Computer Engineering, Trinity Academy of Engineering, Pune, India*

[3]*Associate Professor,Computer Engineering, Trinity Academy of Engineering, Pune, India*

Abstract

Lung cancer is one of the most deadliest diseases in today's decades. It has become one of the causes of death in both man and woman. There are some reasons for which lung cancer occurs so classification of tumor and predicting it in the right stage is important part. In this paper we are focused on the numerous approaches has been derived for lung cancer detection from different literature survey to advance the ability to detect cancer. Image processing and data mining these are also important because for prediction either image dataset or statistical dataset is used so for pre-processing and the image dataset digital image processing is applied for statistical dataset data After the pre-processing, segmentation and feature extraction we can apply various machine learning algorithm for the prediction of the lung cancer. So first we have provided the sketch of the Machine learning and then various fields like in the image data or statistical data where machine learning has been used for classifiction. So once the classification is done confusion matrix is also generated for calculating accuracy, sensitivity, precision, these method is used to measure the rate of accuracy in the proposed model

I. INTRODUCTION

The rapid growth in machine learning is interesting for many people due to its numerous applications in the different areas like it can also be used for fraud detection, computer vision, bioinformatics, medical image diagnosis etc. This is also used in prediction in cancer based on the medical reports like Ct scan, X-Ray, and MRI etc, and has been proven that due to various machine learning technique it has become easier for the doctor so they can predict disease at right stage. Cancer is leading cause of death globally and by 2018 it has been estimated as 9.8 million deaths and the estimation has been provided by world health organization, and the most common cancer is lung cancer, and death rate due to lung cancer is more compared to other all type of cancer .

It not only affects timing of the people but also emergency vehicles such as ambulance and fire brigade vehicles are either stuck or delayed to the specific destination resulting in a heavy loss .The traffic density pattern also remains irregular till the end of day.

Though congestion can also be controlled by flow structure, utilizing various models like vehicle detectors and sensors but the noise pollution caused due to congestion is still a major problem in traffic

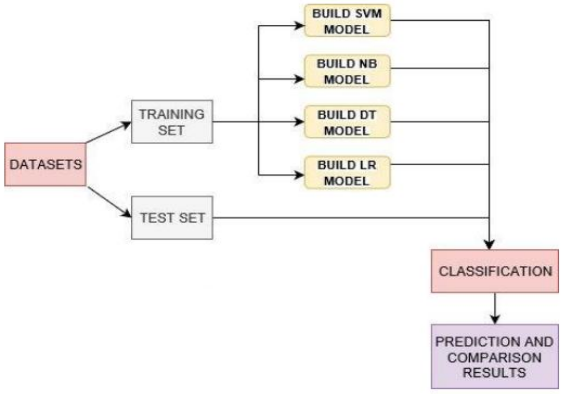

Fig. 1. Model of Traffic Management System

Lung cancer is one of the leading causes of cancer death in men and women . There are various reason for lung cancer such as smoking, explorer to radon gas etc but it is not necessary that the person who smoke will only suffer from lung cancer, it can also occur due to secondhand smoking. The treatment therapy monitoring and the lung nodule analysis by using the computed tomography (CT) medical images that are having useful strategies to diagnosis the lung cancer early and also to monitor the severity This paper consist of various machine learning techniques used for the prediction of cancer in both image data that is CT scan report through which we can predict the location of tumor or the size of tumor and CSV file which contain the data like age, gender smoking rate etc The Lung Cancer datasets used for this study are taken from UCI Machine Learning Repository and Data World. First, the given datasets are divided into training and test data by using k-fold cross validation technique. Then using classification algorithms such as SVM classification models are implemented using the given training data. The classification models are created using training data and the corresponding models are evaluated using test datato get the accuracy of the models. Finally, we can compared the accuracy rates of each and every classificationmodels that we implemented and arrived at a conclusion.

II. LITURATURE SURVEY/BACKGROUND

In 2019 Nidhi S. Nadakarni et al. proposed an automated system for lung cancer detection at an early stage. CT images from the Cancer Image Archive Database were used in DICOM format. These images were then pre-processed using various image enhancement techniques such as Median Filtering, Smoothening, and Contrast Adjustment to remove noise and improve image quality. Further Morphological opening operations were performed after transforming the grayscale image into a binary image for image segmentation. In the feature extraction method features like area, perimeter, and eccentricity (roundness) are evaluated. Using these features classification of images is done into normal and abnormal using SVM supervised learning classifier. The proposed methodology as said by the authors detects cancer in the early stages accurately

In 2019, Sanjukta Rani Jena et al. proposed a method that focuses on texture analysis based on feature extraction of images and then classifying them. In image pre-processing, several filters are used to remove the unnecessary noise and stabilize the image. In the feature extraction part, shaped based FETs (Area, Perimeter, Median,

Mean, and Variance) and intensity-based FETs (Contrast, Uniformity, Homogeneity) are used. Then the local binary pattern (LBP) is used for texture matching. The performance of LBP is better than other available textual patterns. Then an SVM classifier is used for classification. A hyperplane is chosen such that it maximizes the margin

III. PROPOSED SYSTEM

System details: tumor called nodule from the cells in the airways of the respiratory system causeslung cancer. These cells in chest X-rays are always in direct contrast and take theform of a spherical object. If lung nodules can be reliably detected at an early stage,the patient survival rate can be dramatically improved. Since lung nodules usingraw chest Xray imaging cannot be detected quickly, however, interpreting thesediagnostic photographs has become repetitive and a very complicated task (Jothilak-shmi and SV, 2020). The presence of a small nodule from a large 3D lung CT scanwould have to be detected by the computer-aided diagnosis (CAD) system (Zeeba-ree et al., 2019; Zebari et al., 2020). Fig.1. gives an example of an early-stage lungcancer nodule seen in a 2D slice of a CT scan. Noise from the tissues around it, air,and bone fills the CT scan, so this noise would first have .Image segmentation is a process which divides the image into several segmentbased on the pixel, once the image segmentation is over the feature extraction can beapplied. Feature extraction is a type of dimensionality reductionwhere a set of rawdata is reduced to more manageable group image data for extracting the feature likeregion and texture. After extracting the feature different machine learning techniqueis used to classify the image.

Pre-Processing

• A CAD system cannot directly use CT images. They needto be well pre-processed before the actual use. Various Image pre-processing tech-niques are used to discard noise and to make images suitable for use. This helpsin the betterment of the performance of the whole system and hence the accuracy

B. Image segmentation

The method of partitioning an image into several segmentsis known as image segmentation. Segmentation of image is done majorly to findboundaries in the given image. The process of analyzing the image becomes easieras segmentation reduces the image complexity

C. Feature Extraction Feature Extraction is a method by which we aim at re-ducing the number of dimensions that our raw data contains so that it is easier toprocess and is in a form of manageable classes. Variables in a huge number requiring computational resources in order to process andproduce results is characteristicfor the massive amounts of data. Feature Extraction techniques deal with simplifying the data while at the same time ensuring that no data is lost. These techniquesare responsible for picking and merging the features to minimize the amount of data

International Journal of Advance Research in Science and Engineering
Volume No. 11, Issue No. 04, April 2022
www.ijarse.com
IJARSE
ISSN 2319 - 8354

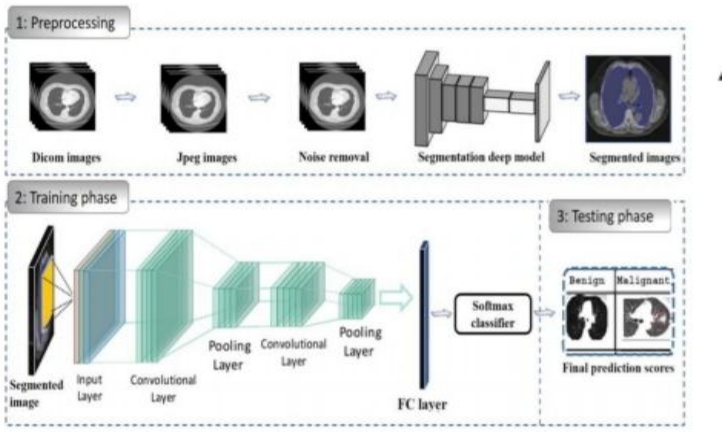

Fig. 2. Model of Traffic Management System

D. Image Classification

Classification of images is a basic task that seeks to interpret a picture as a whole. By assigning it to a particular label, the purpose is toidentify the image. Image Classification usually refers to images where only one ob-ject appears and is examined. Object identification, on the other hand, requires bothclassification and localization tasks and is used to examine more practical instancesin which an image may have several objects. Here the task is to classify lung nodulesas malignant or benign

E. Algorithm

A. Support Vector Machine is supervised learning method which fall under classification technique. Here object or the labeled data which are plot in n- dimensional space and then they are also separated by hyper plane which are also called as decision boundary, and the objects are called as support vec-tors

For a data are arranged in linear manner then in that case it can be separated by usinghyper plane and if the data are arranged nonlinearly then for partitioning two classeswe have to use kernel function. The type of kernel function are linear, nonlinear,polynomial, radial basis function and sigmoid. Diagram given in shows the work of SVM by separating two classes that is Class A is circle and Class B istriangle by the use of hyper plan

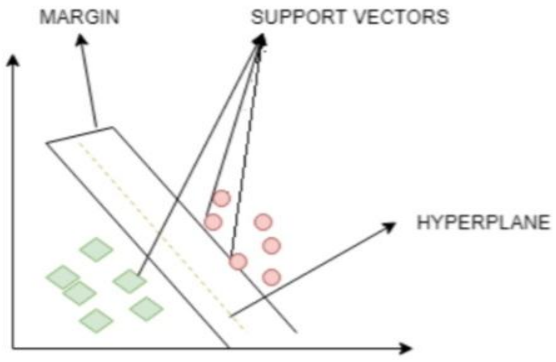

Fig. 3. Model of Traffic Management System

IV. CONCLUSION

In earlier times, the doctor have to do multiple tests in order to detect whether a given patient has lung cancer or not . But this is a very time consuming process.In a diagnosis sometimes a patient has to go through unnecessary check-ups or differ-ent tests to identify the disease of lung cancer. To minimize the process time and for unnecessary check-ups there needs to be a preliminary test in which both the patient and the doctor will be notified with the possibilities of lung cancer. The machine learning algorithms plays an important role in the prediction and classification in medical data. Logistic Regression, SVM, decision tree and Naive Bayes are the machine learning algorithms used in this comparative study. A comparative analysis of accuracy rates of each classifier are presented. The predictive performance of classifiers that are compared quantitatively. In the performance chart, different results are produced in each classifier on the lung cancer dataset. Looking at the correct classification (CA) and other metrics so the best result is given by the supportvector machine algorithm. SVM algorithm used high dimension to classify the observation so it's performance is the best. More accurate lung cancer detection can also be done using this technique. Therefore, there are less mistakes. Finally, by adding extra pre-processing the accuracy rate can be enhanced

REFERECNES

[1] Machine Learning Techniques for Prediction of Lung Cancer; Nikita Banerjee, Subhalaxmi Das

[2] A Comparative Study of Lung Cancer Detection using Machine Learning Algorithms ,Radhika P R Rakhi.A.S.Nair

[3] Prediction Lung Cancer– In Machine Learning Perspective,Nikita Banerjee, Subhalaxmi Das

[4] LUNG CANCER DETECTIONUSING MACHINE LEARNINGAPPROACH,Smita raut , shraddha patil

[5] A Review of most Recent Lung Cancer Detection Techniques using Machine Learning, Dakhaz Mustafa Abdullah Nawzat Sadiq Ahmed

[6] A Survey on machine learning techniques for the diagnosis of liver disease,harshpreet kaur mandeep kaur

[7] Lung Cancer Prediction using Machine Learning: A Comprehensive Approach,Syed Saba Raoof,M A. Jabbar

[8] Prediction and Classification of Lung Cancer Using Machine Learning Techniques Pragya Chaturvedi 1 , Anuj Jhamb 1 , Meet Vanani 1 and Varsha Nemade 1