WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

DETECTION OF LUNG CANCER USING MACHINE LEARING

P. Guna Sri Charan Dr. K. Aravind GMR

Institute of Technology, Rajam, India
20341a05d7@gmrit.edu.in

Abstract:

Machine learning (ML) is a type of artificial intelligence that allows software applications to become more accurate at predicting outcomes without being explicitly programmed. These algorithms can also be used in medical field to detect some of the dangerous diseases like “Lung Cancer ”. Lung cancer also referred as lung carcinoma, is a disease which is malignant tumor leading to the uncontrolled cell growth in the lung tissue. They have too many methods are developed in recent years to diagnose lung cancer, most of them utilizing CT scan images and some are using x-ray images. CT scan images are more suitable for accurate results. Image processing and machine learning approaches have shown a great potential for detection and classification of lung cancer. GLCM (Gray Level Co-occurrence Matrix) features play an important role in medical image analysis. Many Image segmentation methods such as Region Growing, Marker Controlled Watershed and Marker Controlled Watershed with Masking are used for image segmentation. Among them Marker Controlled Watershed with Masking obtained higher results. Machine learning algorithms like Support Vector Machine, Backpropagation Learning Algorithm and Long-Short Term Memory Network etc, are used to perform lung cancer prediction. Using this Machine Learning Model and Algorithms can detect the lung cancer at very starting stage which can save many people’s lives.

Keywords: Machine Learning, Lung cancer, Support Vector Machine, Marker-Controlled Watershed-Based Segmentation, K-Nearest-Neighbor Algorithm (KNN)

Introduction

Lung cancer is the second most common cancer, accounting for about one out of five malignancies in men and one out of nine in women. Unfortunately, over the past several years, while the incidence of lung cancer has gradually declined in men, it has been rising alarmingly in women . The cancer is called a malignant tumor, which is caused by an irregular division of any tissue or organ in cell. When the treatment is not possible , it causes serious discomfort or death . Small-cell lung carcinoma(SCLC) and non-small-cell lung carcinoma are the two main kinds of lung cancer . SCLC is nearly linked to smoking and is growing faster . The most dangerous about the disease is the progression of lung cancer without symptoms.
Although cancer cases can be diagnosed more quickly with the help of developing technology, it may cause death in cancer cases which can grow rapidly or metastasis to different regions . Cancer is the abnormal growth of cells. A tumor called nodule from the cells in the airways of the respiratory system causes lung cancer . Since the lung nodules using raw chest X-ray imaging cannot be detected quickly , the presence of nodule from a large 3D lung CT scan would have to be detected by the computer-aided diagnosis

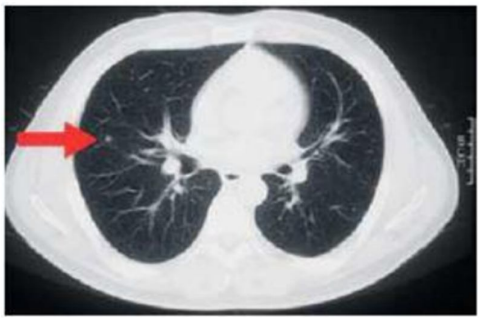

Fig.1: LUNG NODULE

Literature Survey

Dakhaz Mustafa Abdullah & Nawzat Sadiq Ahmed, 2021. "A Review of most Recent Lung Cancer Detection Techniques using Machine Learning," International Journal of Science and Business, IJSAB International, vol. 5(3), pages 159-173
The lung cancer might be detected with high precision. For segmentation threshold and marker-controlled watershed-based segmentation was used.

SVM binary classifier was used for classification purposes. The results showed that watershed with a highly precise and reliable masking technique was the best method for identifying major features. It has been shown that using a watershed and a masking technique produced better outcomes. Using a multiclass SVM classifier and adopted marker-controlled watershed-based segmentation for image segmentation to achieve the greatest accuracy result of nearly 97%. [1]

Abdullah, Dakhaz Mustafa, Adnan Mohsin Abdulazeez, and Amira Bibo Sallow(2021). "Lung cancer prediction and classification based on correlation selection method using machine learning techniques." *Qubahan Academic Journal* 1.2: 141-149.The accuracy ratios of three classifiers—the Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Convolutional Neural Network (CNN) are examined the informative indexes used in this investigation were derived from lung cancer patient databases at UCI.An picture from a CT scan that has been given to a classifier contains unrelated information. The most crucial features are selected using the Gray Level Cooccurrence Matrix (GLCM) (i.e., for removing features). This classification has an MLP accuracy of 98%, an SVM accuracy of 70,45%, and a KNN accuracy of 99,2%. According to the experimental findings, SVM achieves the greatest results (95.56%), followed by CNN (92.11%), and KNN (88.40%).[2]

Anita, C. S., Vasukidevi, G., Rajalakshmi, D., Selvi, K., & Ramesh, T. (2022). Lung cancer prediction model using machine learning techniques. International Journal of Health Sciences, 6(S2), 12533²12539.The GNB machine learning technique is employed to forecast lung cancer.UCI Machine Learning Repository is used to assess the effectiveness of the proposed GNB algorithm. In comparison to other machine learning techniques, the performance analysis reveals that the GNB prediction model achieves 98% accuracy. [3]

Bebas, Ewelina, Marta Borowska, Marcin Derlatka, Edward Oczeretko, Marcin Hładuński, Piotr Szumowski, and Małgorzata Mojsak. "Machine-learning-based classification of the histological subtype of non-small-cell lung cancer using MRI texture analysis." Biomedical Signal Processing and Control 66, pp: 102446, 2021.Model on machine-learning-based classification for early-stage prediction of lung cancer using Magnetic resonance image MRI texture analysis. In this paper they localize the metabolic active area in PET/MR of lung and select Region of interest and calculate the image texture parameters. Based on these parameters the classification process done by using different machine learning algorithms. These algorithms will classify the subtypes of lung cancer. Here in this study, the algorithm used for best outcome prediction is Support Vector Machine (SVM) This study was identification of the subtypes non-small cell lung cancer based on the MRI images. It can be helpful in defining patients with subtypes of lung cancer. It can help in better detection the indications for appropriate treatment and decrease risk of death.[4]

Nanglia, Pankaj, Sumit Kumar, Aparna N. Mahajan, Paramjit Singh, and Davinder Rathee. "A hybrid algorithm for lung cancer classification using SVM and Neural Networks." ICT Express 7, v.no. 3, pp: 335-341, 2021.In this paper the author proposed A hybrid algorithm for lung cancer classification using SVM and Neural Networks by using this we can diagnosis the lung cancer. In this paper the selection of the appropriate feature and anticipation with the classifier is the major challenge in most of the computer-Aided Design Module (CAD) models. In this algorithm initiated the selected features that have been extracted from SVM training. The same set of labels, which was supplied by the SVM, has been passed to Neural Network as well for the selected attribute set. This study describes utilization of SVM and neural network in an optimized way has been investigated in depth to create an efficient hybrid approach for lung cancer classification. The results of the classification are tested by varying number of neurons and the final selection is made on the basis of regression value. .[5]

Methodology

Five basic steps for detection of lung cancer in humans

Fig 2: Five step-based detection scheme

The dataset is used as the input for the methodology that is being proposed. To obtain the desired free noise images, these images are subsequently sent for pre-processing.

To learn the images and provide the appropriate results, machine learning models are used. Below is a detailed explanation of how machine learning models and data pre-processing work.

By extracting the valuable information from the images, image pre-processing reduces these undesirable noises and enhances classification performance.

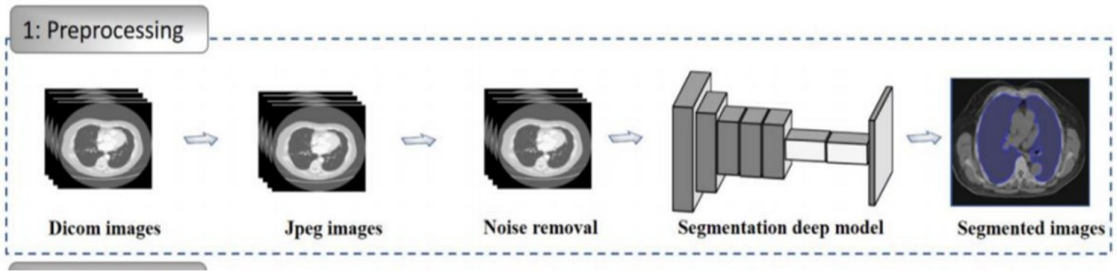

Fig3: Pre-Processing

Image Preprocessing:

Image processing is the process of transforming an image into a digital form and performing certain operations to get some useful information from it The images in the dataset are firstly preprocessed for image enhancement In this graphs , increasing in equivalation , mapping, interpolation and merging takes place It is done to remove noise

Marker-Controlled Watershed-Based Segmentation:

In this marker-controlled watershed-based segmentation was used.The images in the dataset are spreaded for the detection of nodules in the lung. Marker-driven watershed segmentation technique extracts seeds that indicate the presence of objects or background at specific image locations. Marker locations are then set to be regional minima within the topological surface (typically, the gradient of the original input image), and the watershed algorithm is applied [11]. Separating touching objects in an image is one of the most difficult image processing operations, where the watershed transform is often applied to such problem

Feature Extraction:

Image feature The extraction stage is a crucial step in which different desired areas or forms (features) of a given image are found and isolated using algorithms and procedures.Feature extraction the required area is extracted for the use of future prediction or classification.

SVM Algorithm:

One of the most widely used methods for Supervised Learning, Support Vector Machine (SVM), is used to solve Classification and Regression issues. In machine learning, it is generally employed to solve classification issues. SVM is a technique that creates a hyperplane dividing these classes with a maximum margin of separation.A hyperplane is created between the categorised set of classes using the SVM algorithm to predict the kind of lung cancer based on the type of NSCLC. . The goal of the SVM algorithm is to create the best line or decision boundary that can segregate *n*- dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine.

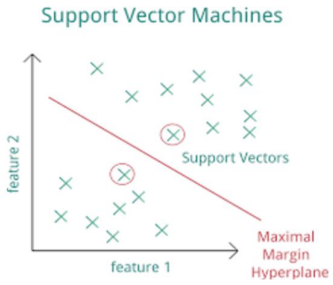

Fig4: Support vector machine

KNN Algorithm

KNN algorithm is used to classify the type of NSCLC and predict the type of lung cancer.K nearest neighbour classifier is the one of the simpler classifiers. KNN determines assigning a new point in the feature space to a specific class based on the distance from the current point to the k- nearest neighbour. K-NN is a **non-parametric algorithm,** which means it does not make any assumption on underlying data. It is also called a **lazy learner algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category

that is much similar to the new data.

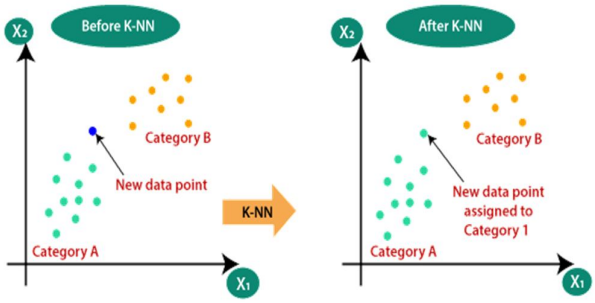

Fig5:K-NN

In this graph, we have to classify the new data point based on the number of nearest neighbours to that new point based on the number of the nearest point we have to classify the new data point.

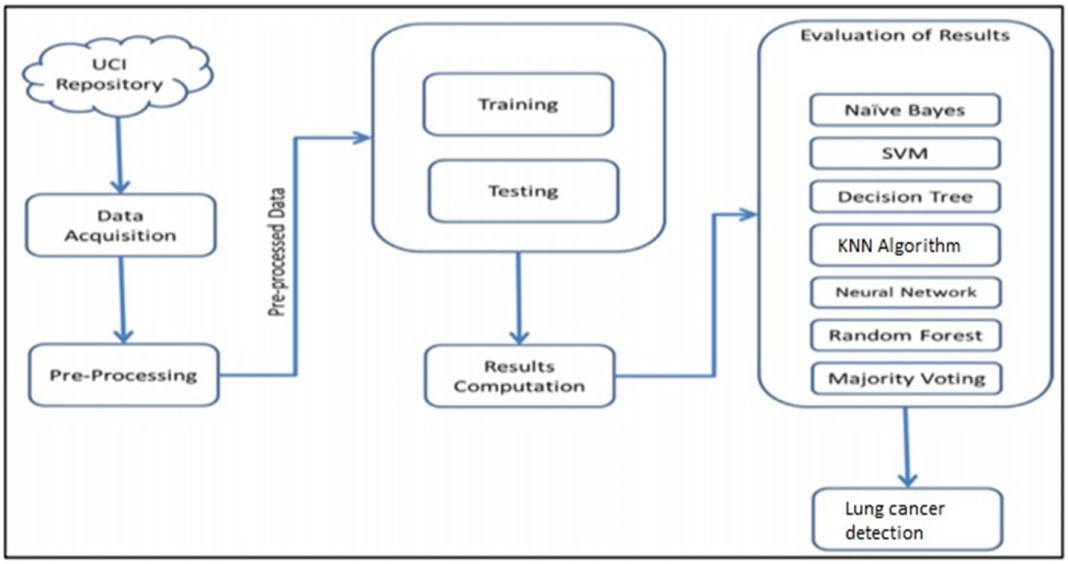

Fig 6:Architecture

4 .Discussion and Results

For preprocessing CT scan images segmentation method marker-controlled watershed-based segmentation has given best result.
In this KNN and SVM classification algorithms are used for the detection of the lung cancer in human lungs
Using multi class SVM classifier as well as adopting marker-controlled watershed-based segmentation for image segmentation. It is clear that highest accuracy result was about 97% obtained.

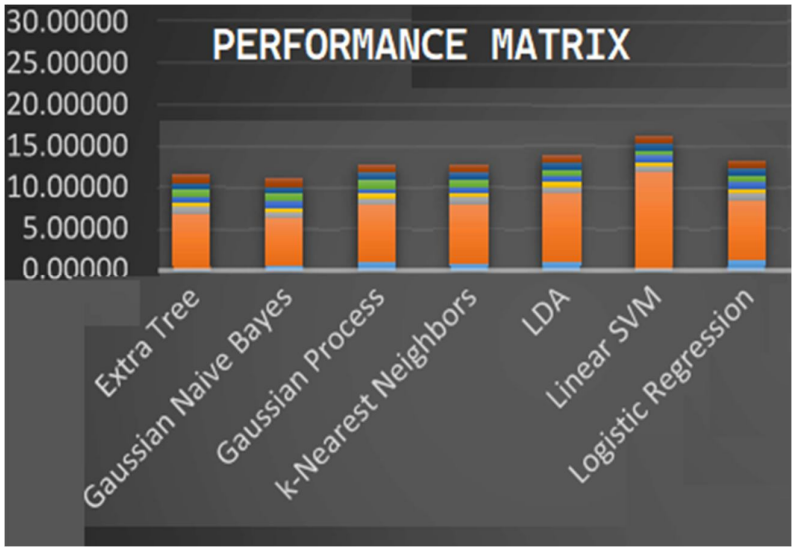

Fig7: Performance graph

Conclusion:

In Early detection of lung cancer is advantageous since treatment may start right away to stop the disease from having negative effects. Therefore, This document compiles a thorough study on numerous machine learning classification methods using CT scan pictures to detect lung cancer. There are many existing algorithms have worked on this detection but SVM gave best performance when compared with those earlier mechanisms. So, finally it concluded that SVM gives a good accuracy. And using marker-controlled watershed-based segmentation gives the best results.

References

1. Dakhaz Mustafa Abdullah & Nawzat Sadiq Ahmed, 2021. "A Review of most Recent Lung Cancer Detection Techniques using Machine Learning," International Journal of Science and Business, IJSAB International, vol. 5(3), pages 159-173.
2. Abdullah, Dakhaz Mustafa, Adnan Mohsin Abdulazeez, and Amira Bibo Sallow (2021). "Lung cancer prediction and classification based on correlation selection method using machine learning techniques." *Qubahan Academic Journal* 1.2: 141-149.
3. Anita, C. S., Vasukidevi, G., Rajalakshmi, D., Selvi, K., & Ramesh, T. (2022). Lung cancer prediction model using machine learning techniques. International Journal of Health Sciences, 6(S2), 12533²12539.
4. Bebas, Ewelina, Marta Borowska, Marcin Derlatka, Edward Oczeretko, Marcin Hładuński, Piotr Szumowski, and Małgorzata Mojsak. "Machine-learning-based classification of the histological subtype of non-small-cell lung cancer using MRI texture analysis." Biomedical Signal Processing and Control 66, pp: 102446, 2021.
5. Nanglia, Pankaj, Sumit Kumar, Aparna N. Mahajan, Paramjit Singh, and Davinder Rathee. "A hybrid algorithm for lung cancer classification using SVM and Neural Networks." ICT Express 7, v.no. 3, pp: 335-341, 2021.
6. Zhou, Yuxin, Yinan Lu, and Zhili Pei. "Accurate diagnosis of early lung cancer based on the convolutional neural network model of the embedded medical system." Microprocessors and Microsystems 81, pp: 103754, 2021.
7. Mechete, Imene, Lejla Alic, Maysam Abbod, Abbes Amira, and Jim Ji. "MR image-based attenuation correction of brain PET imaging: review of literature on machine learning approaches for segmentation." Journal of Digital Imaging, pp: 1-18, 2020.
8. Kavitha, M. S., J. Shanthini, and R. Sabitha. "ECM-CSD: an efficient classification model for cancer stage diagnosis in CT lung images using FCM and SVM techniques." Journal of medical systems 43, v.no. 3, pp: 73, 2019
9. Dwivedi, S. A., Borse, R. P., & Yametkar, A. M. (2014). Lung cancer detection and classification by using machine learning & multinomial Bayesian. *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)*, *9*(1), 69-75.
10. Xie, Y., Meng, W. Y., Li, R. Z., Wang, Y. W., Qian, X., Chan, C., ... & Leung, E. L. H. (2021). Early lung cancer diagnostic biomarker discovery by machine learning methods. *Translational oncology*, *14*(1), 100907.
11. Evangeline, P. P., & Batri, K. Detection of Lung Cancer by Machine Learning.
12. Tripathi, P., Tyagi, S., & Nath, M. (2019). A comparative analysis of segmentation techniques for lung cancer detection. *Pattern Recognition and Image Analysis*, *29*(1), 167-173.
13. Joshua, E. S. N., Chakkravarthy, M., & Bhattacharyya, D. (2020). An Extensive Review on Lung Cancer Detection Using Machine Learning Techniques: A Systematic Study. *Rev. d'Intelligence Artif.*, *34*(3), 351-359.
14. Tuncal, K., Sekeroglu, B., & Ozkan, C. (2020). Lung cancer incidence prediction using machine learning algorithms. *Journal of Advances in Information Technology Vol*, *11*(2).
15. Singh, G. A. P., & Gupta, P. K. (2019). Performance analysis of various machine learning-based approaches for detection and classification of lung cancer in humans. *Neural Computing and Applications*, *31*(10), 6863-6877.