Volume: 1 Issue: 16 Pages: 121-124

International Journal of Research in Academic World

E-ISSN: 2583-1615
Impact Factor: 4.714

Received: 20/November/2022 Int. J Res. Acad. World. 2022; 1(16):121-124 Accepted: 23/December/2022

Machine Learning Algorithm for Predicting and Classification of Lung Cancer

*[1]Dr. M Deepa, [2]MP Venkat Vijay, [3]S Sri Ranjani, [4]V Sowmiya and [5]E Tamizhan

*[1]Associate Professor, Department of Computer Science & Engineering, Paavai College of Engineering, Namakkal, Tamil Nadu, India.

[2, 3, 4, 5]Student, Department of Computer Science & Engineering, Paavai College of Engineering, Namakkal, Tamil Nadu, India.

Abstract
Lung cancer has repeatedly shown itself to be one of the most fatal illnesses in the history of mankind. Additionally, it is one of the malignancies that causes the most fatalities and is the most prevalent. Lung illnesses are rising quickly. In India, there are roughly 70,000 instances per year. The illness is frequently asymptomatic in its early stages, making detection practically difficult. Because of this, early cancer identification is crucial to preserving lives. A patient may have a greater chance of recovery and cure with an early diagnosis. Effective cancer detection is greatly aided by technology. On the basis of their findings, several scholars have suggested various methodologies. The goal of this study is to list, debate, contrast, and analyse a number of strategies for classifying and detecting lung cancer in its early stages, as well as numerous methods for picture segmentation and extraction of features.

Keywords: Lung cancer, computer aided diagnosis, machine learning

1. Introduction

Lung cancer was blamed for an estimated 9.6 million fatalities in 2018 according to estimates. If one considers the many varieties and the prevalence of each, lung cancer comes out on top. Lung cancer is thought to affect 2.09 million people worldwide, with 1.76 million deaths, or around 84% of all fatalities [1]. Lung cancer has earned the distinction of being one of the deadliest illnesses as a result. Lung cancer tumors are created by the proliferation of aberrant cells. Due to the presence of lymph fluid and blood streams in lung tissue, cancer cells have a tendency to spread very quickly. Generally speaking, cancer cells commonly move to the centre of the chest as a result of regular lymphatic flow. Metastasis happens as cancer cells spread to different tissues. Because lung cancer symptoms only appear in the latter stages and because it is practically impossible to save a person's life in this stage, lung cancer is challenging to diagnose.

Imaging methods such as computed tomography (CT), positron emission tomography (PET), magnetic resonance imaging (MRI), and X-ray are used to obtain images of the lungs for assessment. The CT imaging technique is the most popular of the approaches discussed since it may provide a picture without overlapping structures. For doctors, interpreting and identifying cancer is challenging. Lung cancer may be accurately diagnosed using CT images. Deep learning and image processing techniques will be utilised to detect lung cancer. These strategies can increase accuracy. Finding a tumour and figuring out its shape, size, and location is a difficult undertaking. Time is greatly reduced with prompt detection. In this research, tumours will be categorised into one of two classes, namely Malignant and Benign, using pre-processing (removing any noise), post-processing (segmentation), and classification approaches.

A benign tumour is one that is not malignant and really doesn't extend to other body areas. Malignant cells can infect nearby tissues and divide abnormally uncontrollably. This paper's main objective will be to investigate several approaches to lung cancer diagnosis. A 3D picture of the chest may be created by using computed tomography to acquire images of the lungs in various dimensions. In this research, tumours will be categorised into one of two classes, namely Malignant and Benign, using pre-processing (removing any noise), post-processing (segmentation), and classification approaches. A benign tumour is one that is not malignant and really doesn't extend to other body areas. Malignant cells can infect nearby tissues and divide abnormally uncontrollably. This paper's main objective will be to investigate several approaches to lung cancer diagnosis. A 3D picture of the chest may be created by using positron emission tomography to acquire images of the lungs in various dimensions.

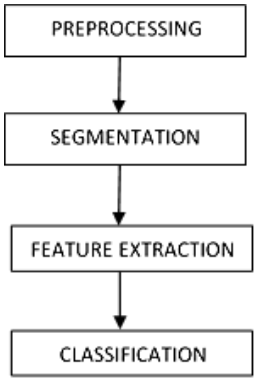

Fig 1: Several approaches to lung cancer diagnosis

***Corresponding Author:** Dr. M Deepa

2. Literature Survey

Moradi *et al*. (2019) [2] saw the comparison of many methods for separating lung cancer modules from non-modules. They developed the 3D Convolutional Neural Network Technique to lessen or completely remove the false positive predictions. There are various sizes of nodules, therefore utilizing only one CNN might lead to erroneous detections. Accordingly, they separated the nodules into four groups based on size. Additionally, four different sizes of 3D CNN were employed. To achieve superior results, they integrated all four classifiers. Each CNN is made up of a number of 3D CNNs of various sizes. A combination of all four classifiers was used to get superior results. They have decided to use a classifier for logistic regression that receives input from four CNNs and generates a better model. They used a gradient boosting model and analysis tools classifier to execute regression models. The whole model was trained using the LUNA16 dataset. The CT scans from the LIDC collection provide the foundation for LUNA16. They observed that the fused classifier's outcome was superior to each of the individual classifiers as a consequence [2].

To diagnose lung cancer, Mesut Togacar *et al*. (2020) [3] suggested a CNN-based method. A total of 100 pictures from 69 different patients have been collected, 50 of which are malignant. Because there weren't as many photos, augmentation was utilised to create a robust dataset. The study made use of VGG-16 CNNs, LeNet, and AlexNet. For AlexNet and VGG-16, Stochastic Gradient Descent was employed as an optimization technique to update the weights for each training set. Additionally, RMSProp and ADAM were employed as the optimization techniques (for LeNet). To extract the features, the mRMR method was employed. Following the CNN designs, certain conventional machine learning models are also employed, including DT, LR, LDA, SVM, and KNN. The Principal Component Analysis technique was used to enhance the performance. By selecting KNN with CNN & mRMR, 99.51 accuracy was attained [3].

A approach for creating an automated lung nodule detection system was put out by QINGHAI ZHANG *et al*. (2020). The public dataset LIDC-IRDI was utilised to test the suggested methodology. Multi-Scene Deep Learning Framework is the suggested methodology employed for this study, and it consists of multiple phases. The probability distribution of various grey levels is produced from the raw CT images using threshold segmentation, also known as histogram. The primary goal of the lung parenchyma segmentation procedure is to correct the smooth lung outlines. Identification of the nodules morphology is aided by the replacement of the lung's venous system [4].

Samaiya Dabeer *et al*. (2019) [4] suggested utilising a CNN-based technique to detect cancer in a histological picture. BreakHis, MITOS-ATYPIA14, and Original Data Set (UC Irvine Machine Learning Repository). There has been use of the BreakHis database network. The RGB colour model's 2480 benign and 5429 malignant samples were used to train the model. As a result, the suggested method shown in Fig. 2 uses an efficient classification model to categorise breast tissue as benign or cancerous. First, the deep net implementation is completed by processing the dataset's picture data. Data redundancy has to be eliminated since it complicates networks and is out of date. The accuracy is reported to be 90.55% for the benign and malignant classes and 94.66% respectively [5].

In their comparison research, Aicha Majda *et al*. (2019) [5] offered four distinct feature extraction techniques: CNN, PCA, Restricted Boltzmann Machines (RBM), and 2D-DFT. Three hidden layers of a neural network were utilised to further assess which strategy performed the best. This neural network was trained using the LIDC-IDRI dataset. In order to increase the volume of the data set, patches of the lung nodule areas are retrieved from the CT scan using a narrative file. In this investigation, CNN was shown to be producing superior outcomes compared to other approaches. Except for CNN 2D-DFT, which nearly matched the findings in terms of accuracy, it had a large variance and bias problem [6].

Anum Masood *et al*. (2018) [7] presented a way to use IoT and CNN based methodology to detect symptoms and lung cancer in the early stages. They have suggested a system based on the Internet of Things (IoT) that includes wearable smart devices and certain symptom charts that can be used to check if the patient is displaying any significant symptoms that might alert the doctor. These patients' CT pictures were then used as input for the CNN model. A pre-processing technique was the Gabor filter. In order to obtain the Region of Interest, threshold was employed. The primary classification model was DFCNet. On the LIDC-IDRI dataset, the suggested model demonstrated 86.02% accuracy, 83.91% sensitivity, and 80.59% specificity. Other datasets were used in this experiment as well including a real time dataset from hospital [7].

Lung cancer detection using a 3D CNN-based method was suggested by Wafaa Alakwaa *et al*. (2017). Both the LUNA16 dataset and the Kaggle Data Science Bowl were used. As lung nodules were not tagged in the Kaggle dataset, LUNA16 One involved training the U-Net model to recognise lung nodules. During the pre-processing stage of the picture, segmentation, downsampling, normalisation, and zero centering were carried out. To segment the CT images, thresholding was applied after the pixel values were initially converted to Hounsfield units. Following segmentation, a 3D picture normalisation process was used to map values between 0 and 1. In all three dimensions, downsampling of 0.5 units has been carried out. The mean value of the pictures was then subtracted from the training dataset to accomplish zero-centering. Instead of feeding the segmented pictures straight into the classifier, a U-Net was trained using the LUNA16 dataset to identify the precise location of nodules. The results showed that the accuracy, false-positive rate, misclassification rate, and completely bogus rate were, respectively, 86.6%, 11.9%, 13.4%, and 14.7% [8].

A technique to automatically diagnose different types of lung cancer from cytological images using Deep CNN was proposed by Atsushi Teramoto *et al*. (2017) [9]. The image collection in question comprises 76 instances of cancer cells. Data augmentation is performed on pictures that were captured using a microscope and feature varying and direction-invariant cell sharpness. Convolutional edge enhancement filters and the Gaussian filter are both used for filtering. The stride of each layer and other variables, such as the filter size, are provided. Three layers make up the architecture: a convolutional layer, three pooling layers, and two fully linked layers. Using DCNN, 70% of the categorization is done accurately [9].

K. Punithavathy *et al*. (2015) [10] explained lung cancer detection based on texture features and Fuzzy C means. The paper mainly concentrates on the image pre-processing parts using different techniques to get better results and a clustering method to generate the outcome. In the pre-processing part, to increase the contrast present in the Computed Tomography Images (CT images), Contrast Limited Adaptive Histogram

Equalization (CLAHE) was applied. Instead of applying this technique to the whole image, it is applied to small regions of the images known as tiles. Bilinear interpolation is used to combine the different enhanced parts/regions of the image. Wiener filters are used to reduce the noise by a significant amount. The extraction of the required region is crucial to obtaining it. To get the required region, i.e., the region with lung lobes and leaving behind the blood arteries, bronchi, and all other internal components, morphological operations such as closure were performed. The closure procedure made advantage of the structural component of the disc form. Since intensity value is the incorrect parameter to extract features, texture-based features were focused during the feature extraction procedure [10].

Two approaches were suggested by P. Mohamed Shakeel *et al*. (2015) for the identification of lung cancer using CT images. The Cancer Imaging Archive (CIA) dataset was used in this investigation. This study makes use of an improved profuse clustering algorithm and a deep learning trained neural network. CT image pre-processing is done to get rid of the noise and low-quality pictures that are present in CT scans. Image histogram approaches are employed to enhance the quality of the photos since they are a very effective tool for a variety of images. A better CT picture created with IPCT is used to segment the areas impacted by cancer. The enhanced profuse clustering approach is used to separate the portions of the enhanced lung CT picture affected by cancer. Two processes of enhanced profuse approaches are used to examine the picture pixel and group comparable superpixels together in order to discover inconsistencies in the image pixels. During the segmentation process, when the pixels are continually analysed, the similitude of data is predicted using the pixel eigenvalue. Standard deviation, third-moment skewness, mean, and fourth-moment kurtosis are some of the spectral characteristics that are obtained from the segmented region and sent to the feature extraction stage because they are excellent indicators of lung cancer since they have connections between them. The system guarantees 98.42% correctness with a 0.038 minimal classifier. In figure 2, the Deep Learning network is displayed [11].

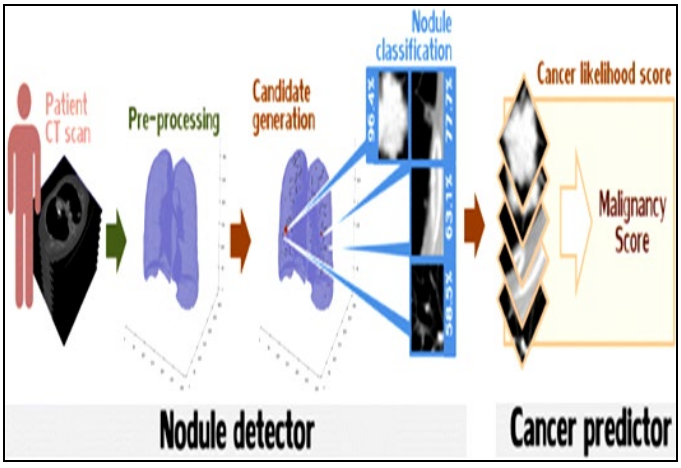

Fig 2: Deep Learning network

A technique that focuses on texture analysis and is based on feature extraction from pictures and classification after that was proposed by Sanjukta Rani Jena *et al*. (2015). Several filters are applied during picture pre-processing to eliminate extra noise and stabilise the image. Shape-based FETs (Area, Perimeter, Median, Mean, and Variance) and intensity-based FETs (Contrast, Uniformity, Homogeneity) are both employed in the feature extraction phase. In order to match textures, the local binary pattern (LBP) is employed. LBP outperforms other known textual patterns in terms of performance. After that, categorization is done using an SVM classifier. The margin is increased by selecting a hyperplane that maximises it (the distance between a few close points and the hyperplane) [12].

The development of an automated method for early lung cancer detection was suggested by Nidhi S. Nadakarni *et al*. (2015). The Cancer Image Archive Database included CT scans in DICOM format. These photos were then pre-processed to reduce noise and boost image quality using a variety of image enhancement techniques, including median filtering, smoothening, and contrast adjustment. After converting the grayscale picture into a binary image for image segmentation, additional morphological opening procedures were carried out. Features including area, perimeter, and eccentricity (roundness) are assessed in the feature extraction approach. SVM supervised learning classifiers are used to classify pictures into normal and pathological utilising these attributes. According to the authors, the suggested approach accurately diagnoses cancer in its early stages [13].

Conclusion

Lung cancer is among the most lethal illnesses to have ever existed. Unfortunately, once the disease has spread or progressed significantly, it is very difficult to treat. One of the rapidly developing technologies, computer-aided detection (CAD), aids in the early detection of cancer by taking into account a variety of patient-related inputs, including scans like CT, X-ray, and MRI scans, patients' unique symptoms, biomarkers, etc. Several techniques are utilised to increase accuracy and facilitate the process, including SVM, CNN, ANN, Watershed Segmentation, Image Enhancement, and Image Processing. The most often used datasets for training are LUNA16, Super Bowl Dataset 2016, and LIDC-IDRI. We hope to highlight all the significant studies that have been conducted in recent years that can be improved upon to provide better findings through the use of this landmark study.

References

1. World Health Organisation‘s Official websitehttps://www.who.int/newsroom/factsheets/detail/cancer#:~:text=The%20most%20common%20causes%20of,Lung%20(1.76%20million%20deaths).
2. Moradi P, Jamzad M. Detecting Lung Cancer Lesions in CT Images using 3D Convolutional Neural Networks 4th Int. Conf. on Pattern Recognition and Image Analysis (IPRIA), 2019, 114-118.
3. Toğaçar M, Ergen B, Cömert Z. Detection of lung cancer on chest CT images using minimum redundancy maximum relevance feature selection method with convolutional neural networks Biocybernetics and Biomedical Engineering. 2020; 40(1):23-39.
4. Dabeer S, Khan MM, Islam S. Cancer diagnosis in histopathological image: CNN based approach Informatics in Medicine Unlocked. 2019; 16:100231.
5. Skourt BA, Nikolov NS and Majda A. Feature-Extraction Methods for Lung-Nodule Detection: A Comparative Deep Learning Study Int. Conf. on Intelligent Systems and Advanced Computing Sciences (ISACS), 2019, 1-6.
6. Liu Z, Yao C, Yu H, Wu T. Deep reinforcement learning with its application for lung cancer detection in medical Internet of Things Future Generation Computer Systems, 2019, 1-9.

7. Masood A, Sheng B, Li P, Hou X, Wei X, Qin J, Feng D. Computer-assisted decision support system in pulmonary cancer detection and stage classification on CT images Journal of biomedical informatics. 2018; 79:117-28.
8. Alakwaa W, Nassef M, Badr A. Lung cancer detection and classification with 3D convolutional neural network (3D-CNN) Lung Cancer. 2017; 8(8):409.
9. Teramoto A, Tsukamoto T, Kiriyama Y, Fujita H. Automated classification of lung cancer types from cytological images using deep convolutional neural networks BioMed research International, 2017.
10. Punithavathy K, Ramya MM, Poobal S. Analysis of statistical texture features for automatic lung cancer detection in PET/CT images Int. Conf. on Robotics, Automation, Control and Embedded Systems (RACE), 2015, 1-5.
11. Shakeel PM, Burhanuddin MA, Desa MI. Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks Measurement. 2019; 145:702-12.
12. Jena SR, George T and Ponraj N 2019 Texture Analysis Based Feature Extraction and Classification of Lung Cancer IEEE Int. Conf. on Electrical, Computer and Communication Technologies (ICECCT) pp. 1-5.
13. Nadkarni NS and Borkar S. Detection of Lung Cancer in CT Images using Image Processing 3rd Int. Conf. on Trends in Electronics and Informatics (ICOEI), 2019, 863-866.