International Journal of Advancements in Technology

OPEN ACCESS Freely available online

Research Article

Lung Cancer Detection Using Machine Learning

Waykule Jyoti M*

Department of Electronics Engineering. Sanjay Ghodawat University Kolhapur, Maharastra, India

ABSTRACT

Pulmonary cancer, also known as lung carcinoma, is the primary cause of cancer death worldwide. Every year, early-stage cancer detection using Computed Tomography (CT) could save hundreds of thousands of lives. However, analysing hundreds of thousands of these scans is a huge burden for radiologists, and they frequently experience observer tiredness, which can harm their performance. As a result, there is a requirement to efficiently read, detect, and evaluate CT scans. As a result, there is a requirement to efficiently read, detect, and evaluate CT scans. Using the midpoint of the lung cancer provided in the dataset, the author cropped 2D cancer masks on its reference image and trained a model with various techniques. The proposed system consists of many steps such as image acquisition, pre-processing, binarization, thresholding, segmentation and feature extraction. In first stage, Binarization technique is used to convert binary image and then compare it with threshold value to detect lung cancer. In second stage, segmentation is performed to segment the lung CT image and a strong feature extraction method has been introduced to extract the some important feature of segmented images. Extracted features areused to train the neural network and finally the system is tested any cancerous and noncancerous images.

Keywords: Gabor filter; Thresholding Approach Marker-controlled Watershed Segmentation; Binarization

INTRODUCTION

Lung cancer is a disease in which abnormal cells develop and form tumours in the lungs. Cancer cells can be carried away from the lungs in the bloodstream or in the lymph fluid that surrounds lung tissue. Lymph is transported by lymphatic vessels to lymph nodes in the lungs and the chest centre when a cancer cell leaves the place where it started and travels through the bloodstream to a lymph node or another part of the body, this is called metastasis [1]. Primary lung cancer is cancer that begins in the lungs. Lung cancer comes in a variety of forms, including Carcinoma, Aden carcinoma and Squamous. According to the rank order of cancer cases among Jordanians in 2008, 365 cases of lung cancer were diagnosed, accounting for 7.7% of all newly diagnosed cancer cases. Lung cancer struck 297 (13.1%) men and 59 (2.5%) women, with a male-to-female ratio of 5:1 and lung cancer ranking second among men and tenth among women [2]. The first stage begins with a collection of CT images (normal and abnormal) from the IMBA home data base [3]. The second stage employs a number of image enhancement techniques in order to achieve the highest level of clarity. The third stage employs image segmentation algorithms, which play an important role in the image processing stage, and the fourth stage obtains general features from an enhanced segmentation algorithm, which provides an indicator of image normalcy or abnormally. According to so many stages of discovery of the cancer cell in the lungs, lung cancer is the most harmful and common cancer cell in the world, so the mechanism of early detection of the disease plays a very significant and vital role in avoiding severe advance stages and reducing its percentage of dissemination [4].

Literature review

The primary goal of this research is to evaluate various computer-aided techniques by analysing the current best practises technique, defining its limitations and drawbacks, and finally proposing a new model that enhance the drawback. As a result, a new system is proposed. The proposed system is used to locate a cancerous nodule in the body. Watershed segmentation was used to detect the nodule and SVM is used for classification .This model detect the cancer with 92% accuracy and classifier has accuracy of 86.6% by [5]. The time factor was taken into account to discover the abnormality issues in target images. An image improvement strategy is being developed for earlier disease identification and treatment stages; The core factors of this research are image quality and accuracy, as well as image quality assessment and enhancement stages, where low pre-processing techniques based on Gabor filter within Gaussian rule were used by [4]. The Proposed method is developed

Correspondence to: Waykule Jyoti M, Department of Electronics Engineering. Sanjay Ghodawat University Kolhapur, Maharastra, India; E-mail: jyoti.waykule@sanjayghodawatuniversity.ac.in

Received: June 14, 2021, **Accepted:** June 21, 2021, **Published:** June 30, 2021

Citation: Waykule JM (2021) Lung Cancer Detection Using Machine Learning. Int J Adv Technol 12:17.

by using image processing and machine learning algorithms such as gray scale conversion, noise reduction, and binarization to describe lung cancer and its phases. All this algorithms are used for the pre-processing of the given CT scan image. ROI (Area, Perimeter, and Eccentricity) extracted from CT scan image and this is useful for identifying lung cancer at an early stage.To focus on the tumor portion of the abnormal images, segmentation is used. Features derived from the images are used to classify them using SVM by used SVM and Image Processing techniques to develop an effective method for detecting lung cancer and its phases, with the aim of proto producing more accurate results [6-7]. Used Region Increasing, Marker Controlled Watershed, and Marker Controlled Watershed with Masking to introduce and analyse three image segmentation methods for analysing lung cancer used an active spline model to segment X-ray images of lung cancer [8]. For noise detection, a pre-processing median filter is used. Then, segmentation is used, followed by K-mean and fuzzy C-mean clustering to extract features. Thus, this paper focuses on image segmentation and detection of lung cancer by determining normality and abnormality and for classification it gives preference to SVM algorithm in MATLAB the proposed system is intended to detect lung cancer in its early stages in two stages [9]. The proposed system includes various stages such as image acquisition, pre-processing, binarization, threshold ingsegmentation, feature extraction, and neural network detection. First step is binary thresholding, and then feature extraction, and then using these features to train and evaluate the neural networkThis proposed system successfully detects the lung cancer from CT scan Images with overall. 97.67% success rate of the system has focused on ML algorithms such as Naive Bayes, Support Vector Machine (SVM), Logistic regression, and Artificial Neural Network (ANN) have been used in the healthcare field to analyse and predict lung cancer prognosis because of its precise outcomes [10].

MATERIALS AND METHODS

Image used for database

The presented work makes use of a collection of digital images that includes 20 images of small-cell lung cancer and 20 images of non-small-cell lung cancer, for a total of 40 images (samples) that are each 200 × 200 pixels in size. The photos were gathered from private hospitals and viewed on the internet in the JSRT (Japanese Society of Radiological Technology) public database. The digital images are saved in JPEG format with an 8-bit resolution per plane. All images are saved as raw 200 × 200 × 256 JPEG data (Figure 1).

Pre-processing

Image enhancement: Image enhancement is the first stage of image pre-processing; the goal of image enhancement is to improve the interpretability or perception of information included in the image for human viewers or to provide better input for other automated image processing techniques. During the image enhancement stage, we used three techniques: the Gabor filter, Auto-Enhancement, and Fast Fourier transform. The Gabor filter: Gabor filters' frequency and orientation representations are similar to those of the human visual system, and the Fourier transform of the harmonic is particularly suitable for texture representation and discrimination. Use the following three techniques in the image enhancement stage: Gabor filter, Auto-Enhancement, and Fast Fourier transforms. Gabor filters have similar frequency and orientation representations to the human visual system, and the Fourier transform of the harmonic is especially useful for texture representation and discrimination.

The method of segmenting a digital image into different segments, such as sets of pixels, also known as super-pixels, is known as image segmentation. The main goal of segmentation is to transform an image's representation into something that is easier to understand.

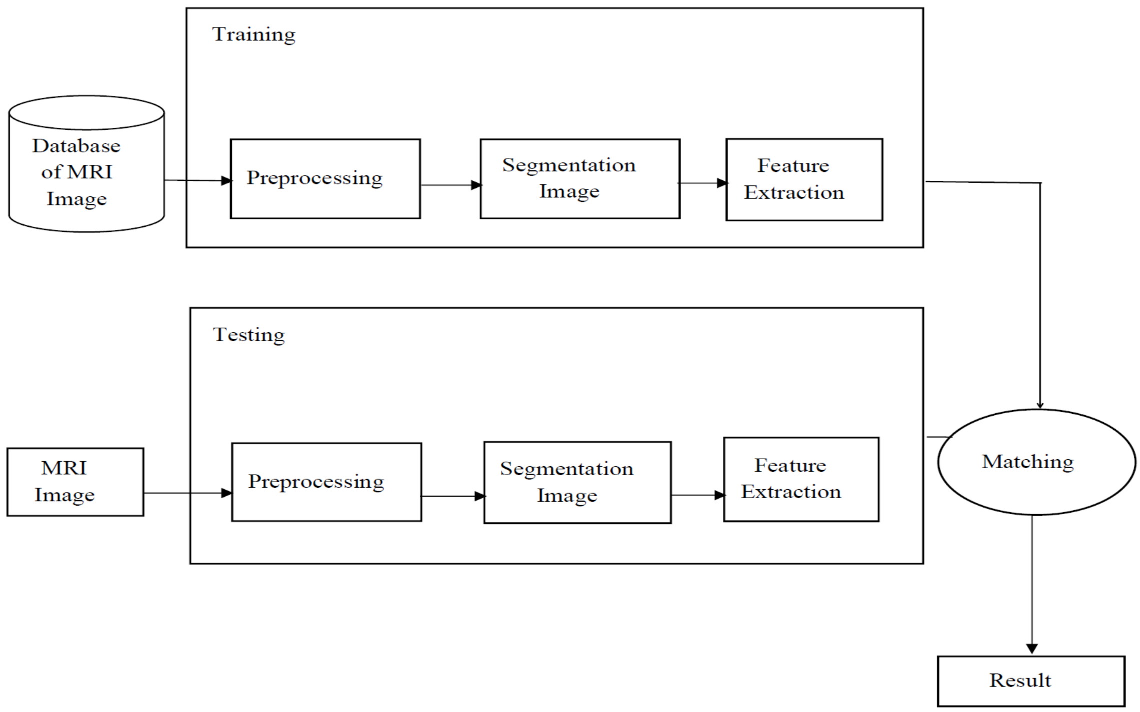

Figure 1: Image used for database.

Objects, margins, and boundaries in images are located using image segmentation. The process of assigning a label to every pixel in an image, so that pixels with the same label share certain characteristics, is known as image segmentation. A collection of segments that collectively cover the entire image, or a set of edges and boundaries extracted from the image, is the result of image segmentation. Some pixels in a given region share a common characteristic or property, such as colour, intensity, or texture. When it comes to the same characteristic, adjacent regions vary. The thresholding and watershed masking approaches are two methods used for image segmentation. Thresholding Methodology One of the most important methods for image segmentation is thresholding. As compared to grey level images, which typically contain 256 levels, the segmented image obtained by thresholding has the advantages of smaller storage space, faster processing speed, and ease of manipulation [3]. We have a greyscale image provided for a thresholding process, which transforms the RGB image into a binary image, i.e. a black and white image with only two colours, black and white, which denote levels 0 and 1. Since there are only two levels, the threshold value for this would be between 0 and 1. After reaching the threshold value, the image will be segmented accordingly. Marker-controlled Watershed Segmentation: The presence of artefacts or context at particular image locations is indicated using the marker watershed segmentation technique. There are two types of marker-controlled watershed segmentation approaches: external and internal. The external approach is associated with the context, while the internal approach is associated with the object of interest. If we can locate or "mark" foreground objects and background position, image segmentation using the Watershed transform works well, by treating an image as a surface with high light pixels and low dark pixels, "catchment basins" and "watershed ridge lines" can be found [11]. The Marker-Controlled Watershed Segmentation approach has more precision and consistency than the Thresholding approach, according to the experimental subjective evaluation during the segmentation stage. Machine learning is made up of two concepts: training and testing. In the context of supervised learning, training and testing are more easily explained:

Feature extraction: In the field of image processing, features are extremely significant. Various image pre-processing techniques, such as binarization, thresholding, normalisation, and masking approach, are applied to the sampled image before having features. And after that, feature extraction techniques are used to obtain features that can be used in image classification and recognition. Binarization relies on the assumption that if the number of black pixels in a digital x-ray image is significantly greater than the number of white pixels in the image, we infer that the x-ray report is normal lung image; otherwise, it is abnormal. As a result, we began counting the number of black pixels in an image for both normal and abnormal images in order to obtain an average that could be used as a threshold. If the number of black pixels in a new image is greater than the threshold that we determined, the image is normal; otherwise, if the number of black pixels is less than the defined value of a threshold, the image is abnormal.

RESULTS AND DISCUSSION

Figures show the results of lung cancer detection from lung

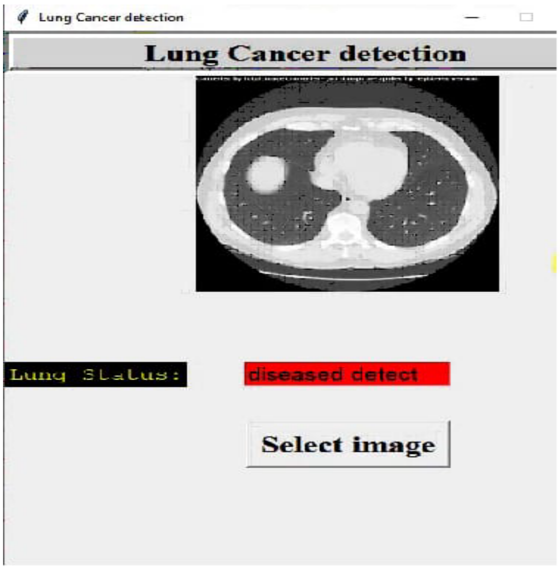

Figure 2: Shows the output image affected by lung Cancer disease. (Red color).

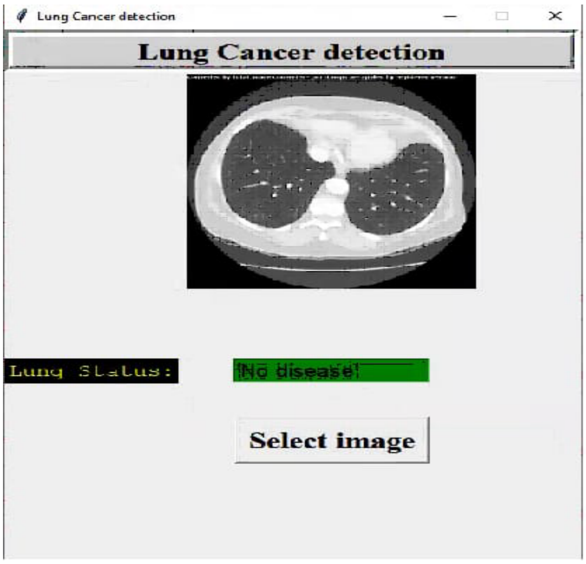

Figure 3: Shows the output as not affected by lung Cancer, indicating by green colour.

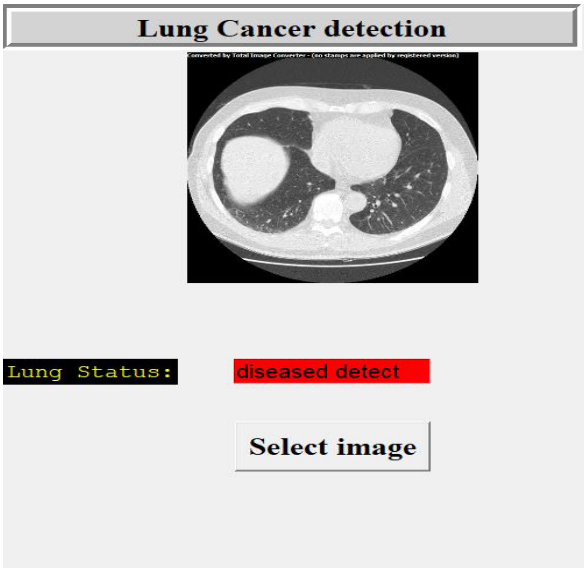

Figure 4: Shows the output image affected by lung Cancer. (Red color)

CT image, in (Figure 2 and Figure 4) shows the output of lung cancer disease detection for images that have been affected by lung cancer and (Figure 3 and Figure 5) shows the output as No disease, indicating that the picture is unaffected by lung cancer.

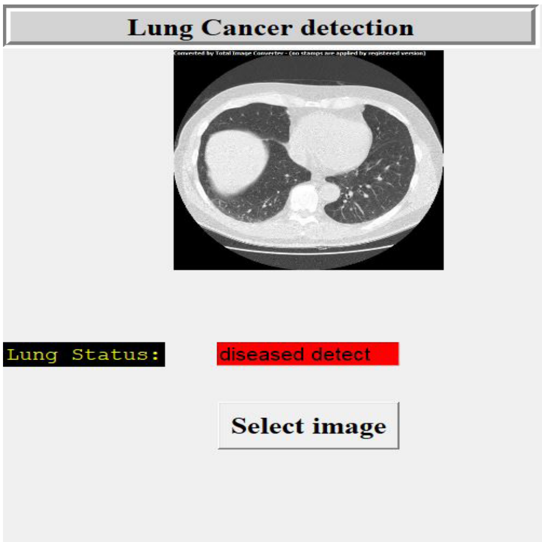

Figure 4: Shows the output image affected by lung Cancer. (Red color)

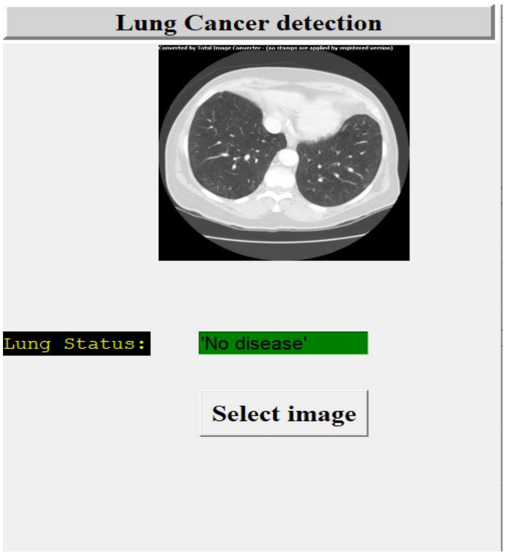

Figure 5: Shows the output as not affected by lung Cancer, indicating by green colour.

CONCLUSION

An automated lung cancer detection technique from a lung CT image was introduced in this paper. This developed algorithm successfully detects lung cancer in a CT scan of the lungs. This algorithm was tested on a large number of images and found to be effective in detecting lung cancer.

REFERENCES

1. Non-small cell lung cancer adapted from National Cancer Institute (NCI) and Patients Living with Cancer (PLWC). 2007. Available at: http://www.katemacintyrefoundation.org/ pdf/ non-small-cell.pdf
2. Tarawneh M, Nimri O, Arqoub K, Zaghal M, and Cancer incidence in Jordan. 2008. Available at: http://www.moh.gov.jo/MOH/ Files/Publication/Jordan%20Cancer%2Registry_2008%20 Report_1.pdf
3. Lung Cancer Database. 2011. Available at: https://eddie.via. cornell.edu/cgibin/datac/signon.cgi.
4. AL-Tarawneh M. Lunge cancer detection using image processing technique. 2012
5. Makajua S, Prasad PWC, Alsadoona A, Singh AK, Elchouemic A. “lung cancer detection using CT scan image. 2017.
6. Jondhale S. Lung cancer detection using image processing and machine learning Health care”.pages. 2018.
7. Abdillah B, Bustamam A, Sarwinda D. Image processing based detection of lung cancer on CT scan image. 2016;893:1.7.
8. Joon P, Bajaj SB, Jatain A. Segmentation and detection of lung cancer using image processing and clustering techniques. 2019: 1-6.
9. Alam Miah MB. Detection of lung cancer from CT image using image processing neural network. 2015:1-6.
10. Raoof S, Jabbar MA , Fathima SA.. lung cancer prediction using machine learning: a comprehensive approach.2020.
11. Levner I, Zhang H, Classification-driven watershed segmentation, IEEE Transactionson Image Processing.2007;16:1437-45.