Medicon Medical Sciences
Volume 2 Issue 3 March 2022
Research Article

A Novel Method for Mining Heterogeneous Lung Cancer from Computer Tomography (CT) Scan

Denny Dominic* and Krishnan Balachandran

School of Computer Science and Engineering, Christ (Deemed to be University), Bengaluru, India

***Corresponding Author:** Denny Dominic, School of Computer Science and Engineering, Christ (Deemed to be University), Bengaluru, India.

Received: February 20, 2022 ; **Published:** March 03, 2022

DOI: 10.55162/MCMS.02.023

Abstract

Early detection of any type of cancer, especially lung cancer, one of the world's worst diseases, can save a lot of lives. By using the early forecast, we will be able to increase life expectancy and decrease death. While there are other approaches for detecting lung cancer cells, such as X-rays and CT scans, CT pictures were found to be more preferred. CT scans, for example, use 2D images to get more accurate results. The proposed approach here will cover how to interpret CT pictures from a Computer Tomography (CT) Scan with the Confusion Matrix for Mining Heterogeneous Lung Cancer. This study will look at how image conversion may be performed using various image processing approaches to get better outcomes from CT images. The Confusion Matrix is a tool for calculating inequality in an image pattern. Following Confusion Matrix's examination of the processed images, a final accuracy of 93 percent is attained.

Keywords: Computer Tomography (CT); Mining; Confusion Matrix; Lung Cancer

Introduction

Early diagnosis of lung cancer disease is difficult due to insufficient prognostic techniques now accessible. If cancer is identified early and treated properly, the chances of survival are better [1]. The lymph flows into the lymph nodes in the lungs and the centre of the chest through lymph veins. A CT scan is more effective than an X-ray in detecting lung cancer. Lung cancer is considered to be caused primarily by smoking. As a result, a framework must be established in order to build an intelligent computer assisted diagnostic system.

This procedure must be rapid and trustworthy in order to improve CT image grading. A single characteristic would be insufficient to effectively classify all classes [1]. To diagnose lung cancer, a high-resolution chest computed tomography (CT) scan is often used. The little cell appears to be difficult to distinguish in the early stages. Predicting lung cancer in its early stages is a challenging undertaking in any case. Lung cancer is one of the most feared diseases.

Early diagnosis is widely advocated in order to save people's lives. Lung cancer can be identified using a variety of methods, including x-rays and CT imaging, though CT images are preferred. Medical images are always preferred to have superior results on the identical disease. In this work, the suggested approach will describe how Watershed segmentation can be utilized to pre-process CT images.

To improve the accuracy of picture scans, the data is put into deep learning algorithms. The medical examiner then uses the results to confirm the accuracy of the findings.

The majority of the pre-processing is done with the Median Filter, Gabor Filter, and Watershed segmentation.

In this study paper, we will look at how picture change can be done to improve CT scan results utilising various image processing techniques. The suggested method's design will include image smoothing with median filters, image enhancement, and picture segmentation.

Precision in pre-processing is critical if doctors are to develop a more effective method for detecting the types of lung nodules seen on CT images.

The primary sites where pulmonary nodules occur are the lung, airway, chest wall, artery, and pulmonary fissure, making it difficult for specialists to pinpoint the exact position of the malignant nodules.

The major goal of this study is to learn more about the disparities in Lung Cancer nodules portrayed in CT scans so that cancer cells can be predicted early utilising the Confusion Matrix to discover image disparities.

Literature Review

The first step in detecting lung cancer is noise filtering. Denoising and Weiner filtering are the two most common pre-processing techniques. White noise is one of the issues in image processing [3]. Watershed segmentation using topographic features and object markers has been implemented by Ilya Levner [2]. Pixel sorting and segmentation are important operations in image processing.

If objects belonging to the same specified class are close to each other, pixel grouping is required. The watershed algorithm may be used in an unsupervised setting to achieve a better result in a collection of non-overlapping areas. The concept of quantification has become a key worry in terms of similar nodule characteristics of resemblance, conjecture, and clarification [7-9].

As a result, the necessity for quantitative information derived from image segmentation has become critical. A pulmonary nodule can be found almost everywhere in the lungs. The final product is explained in the extraction of the attribute as if the processed image has some normality or abnormality [4], and image segmentation is the basis of classification.

Ginneken [4] divided lung image extract classification into two categories: rule-based and pixel-based. The majority of categories [5, 6] prefer a rule-based approach, with separate processes and the rule added to achieve a certain result.

Region growth, Thresholding, edge detection, ridge detection, fitting of geometric models and functions, morphological operations, and dynamic programming are the primary study topics in image scan.

Methodology

Description about the data

The CT Scan is a more advanced version of the X-ray in which the gadget is connected to an X-ray equipment. The computer converts photos captured from various angles and distances into three-dimensional, cross-sectional (tomographic), and fragmented views. Bones, muscles, blood arteries, and organs can all be seen clearly. CT scanning is important for diagnosis, therapy, and medical advancement.

CT pictures are collected as real-time data from several medical colleges in this study, and they are used as a standard for Digital Imaging and Communication in Medicine (DICOM).

Information and associated data from medical imaging are shared and managed using the Digital Imaging and Medical Communication (DICOM) standard.

DICOM is widely used to consolidate and transport medical images by a variety of manufacturers, including scanners, servers, workstations, printers, network gear, and Picture Archiving and Communication Systems (PACS).

The patient ID in the file, for example, must be included in the chest X-ray image file to avoid the image being accidentally removed from that information. This is the same manner that image formats like JPEG incorporate tags in order to specify and read the file.

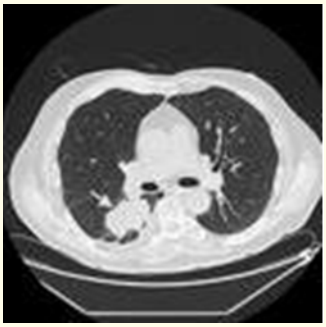

Figure 1: C.T Image.

Image Preprocessing

Preprocessed photos were gathered from numerous medical institutes in India as well as the LIDC (Lung image database collaboration). Because the size and shape of the photographs obtained vary, they are scaled to 128 × 128 pixels. The photos are re-converted to grayscale images. The pretreatment and processing processes are used to alter images.

Pre-processing is performed through picture smoothening, augmentation, and, lastly, segmentation. The poor contrast in lungs images makes it harder to see nodules.

As a result, we applied Contrast Limited Adaptive Histogram Equalization to improve the image (CLAHE).

This method divides images into blocks, then uses a fixed number of grey levels to produce the histogram values for each block. This method divides images into blocks, then uses a fixed number of grey levels to produce the histogram values for each block. This study used the LIDC (Lung Image Database Consortium), which comprised 221 lung CT pictures from cancer patients. This study used the LIDC (Lung Image Database Consortium), which comprised 221 lung CT pictures from cancer patients.

Image Segmentation

The image with different segments is used to make any picture easier to analyse and understand. They divide the image into the various parts and objects that make it up. Whole images are filled by different segments or a set of contours obtained from the image as a result of segmentation [5]. The principal application is for segmenting grey, white, and other types of materials.

The contrast between the background and the lung nodule is quite striking. As a result, utilising threshold segmentation to derive ROI (Receiver operating characteristic) of lung nodules is rather difficult. To reduce the interference, CNN is utilised.

To segment images, the CNN (Convolution Neural Network) algorithm is employed. A CNN contains many perceptron layers, such as convolutional and Re-Lu layers. Dropout, convolutional, and pooling are examples of normalised layers. Pooling and convolutional layers are often used in CNN architecture for picture combining. Pooling layers conduct two actions: mean pooling and maximum pooling. Max pooling calculates the largest number of feature points, whereas mean pooling calculates neighbouring pixels.

Mean pooling reduces size errors, but max pooling reduces estimated errors. CNN algorithms are used for image segmentation. In this way, it completes every CNN.

Image Processing

The preprocessed image is processed using various features, such as local processing of binary patterns, to improve the accuracy of the lung scan, which will accurately differentiate cancer cells. The variation in LBP distribution across different locations in a comparable surface image can be astounding.

Zero component analysis (ZCA) Whitening

Whitening is a data transformation in which the covariance matrix equals the identity matrix. As a result, whitening enhances some qualities. It is used as a preprocessing method. If you have *N* data points in *Rn*, then the matrix of covariance is *R n*n*n* this has estimation equation.1 as:

$$\Sigma\, jk^{(n)} = 1/(N+1)\ \Sigma(1=0)^{n}\ (x)^{*}ij - xj)(x^{*}ik - xk)\ ..\ 1$$

Where '*xj* denotes the j^{th} component of the samples x estimated mean. Any *W-Rn*n* matrix that satisfies the *WTW* = *C-1* condition whitens the results. ZCA whitening is the *W* = *M-(1/2)* alternative.

Local Binary Pattern(LBP Feature)

A close example of a picture pixel is identified by comparing its dim value to that of its neighbours, as set by the LBP system administrator. The LBP rotation invariant is used to create an LBP image. Every pixel in the image is evaluated. The LBP administrator labels the picture pixel as a close example, and its dim value is calculated by comparing it to that of its neighbours. The rotation invariant of LBP is used to construct the LBP picture.

Every pixel in the image is scrutinised. This is how a picture's texture is represented. As seen in Figure 2, the technique is used to describe the texture of a scene and can also be used to estimate an image's local binary pattern distribution. The disadvantage is that there is a lack of local statistical texture data from the image. This method ignores differences between small regions and calculates pattern distribution over the full image, as seen in Figure 2.

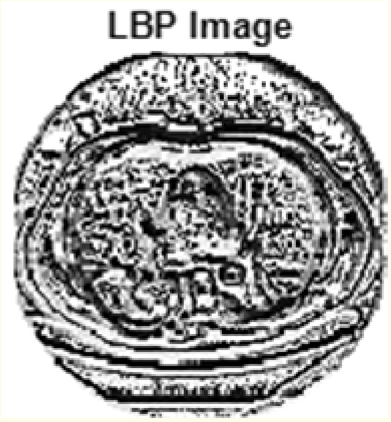

Figure 2: LBP Image.

Training with Deep Learning Methods

Convolutional neural networks (CNNs, or ConvNets) are powerful deep learning algorithms that excel in interpreting image data. CNNs, for example, could be used to classify images. To forecast continuous variables like angles and lengths, you can employ a regression layer at the network's conclusion. They combine the input data with the training data set's projected result. It provides "ground truth" data to their model, which is often generated by a human or semi-automated procedure.

During the machine learning testing phase, these data can be used to assess how well the model is trained and its attributes. During the testing phase, they use this model to train their model by mixing the inputs with the expected outputs.

Once trained, it builds a Convolutional neural network architecture that is used to predict the angles of rotated handwritten digits.

These predictions can help optical character recognition. Synthetic photographs of handwritten numerals are included in the data set, as are the corresponding rotation angles (in degrees) for each picture.

There are 5000 photos in each of the training and validation data sets.

Conclusion and Future Scope

This study works with segmented lungs images for classification with malignant nodule and without malignant nodules. By using classification and consistency, the outcomes are evaluated. This experiment was carried out Using MATLAB. This research presented inequality among the Lung images which is evaluated by confusion matrix. We measured an accuracy of 93.2735 percent with the aid of that confusion matrix. To evaluate a model for improved outcomes, the confusion matrix and Receiver Operating Characteristic (ROC) curve were utilized.

Confusion-matrix defines the performance and goal groups. The false positive rate was 3.1 percent, indicating that only a few of the specified parameters were above 50% because biomarker data was insufficient due to the fact that most hospitals did not do this test and hence could not obtain sufficient data. The percentage of false negatives was 100%, indicating that all of the specified features were within the parameter. The findings were derived using a Confusion matrix with a True positive value of 96.9% and a True negative value of 0%, indicating that all of the selected values were appropriate and no throwaway qualities existed. Finally, we have an accuracy of 98.3051 percent, which is better than the literature that was studied.

References

1. A Cruzroa, JC Caicedo and FA Gonzalez. “Visual pattern mining in histology image collections using bag of features”. Artificial Intelligence in Medicine 52.2 (2011): 91-106.
2. IlyaLevner and Hong Zhangm. “Classification driven Watershed segmentation”. IEEE transactions on image processing 16.5 (2007).
3. Anita chaudhary and Sonit Sukhraj Singh. “Lung Cancer Detection on CT Images Using Image Processing”, International transaction On Computing Sciences 4 (2012).
4. BV Ginneken, BM Romenyand and MA Viergever. “Computer-aided diagnosis in chest radiography: a survey”. IEEE, Transactions on medical imaging 20.12 (2001).
5. Disha Sharma and Gagandeep Jindal. “Identifying Lung Cancer Using Image Processing Techniques”. International Conference on Computational Techniques and Artificial Intelligence (ICCTAI’2011) 17 (2011): 872-880.
6. Nguyen HT., et al. “Water snakes: Energy-Driven Watershed Segmentation”. IEEE Transactions on Pattern Analysis and Machine Intelligence 25.3 (2003): 330-342.
7. Farag A, Ali A, Graham J, Elshazly S and Falk R. “Evaluation of geometric feature descriptors for detection and classification of lung nodules in low dose CT scans of the chest. Paper presented at the Biomedical Imaging: From Nano to Macro”. 2011 IEEE International Symposium, Chicago, IL, USA (2011).
8. Lin P-L, Huang P-W, Lee C-H and Wu M-T. “Automatic classification for solitary pulmonary nodule in CT image by fractal analysis based on fractional Brownian motion model”. Pattern Recognition 46.12 (2013): 3279-3287.
9. Farag A, Elhabian S, Graham J, Farag A and Falk R. “Toward precise pulmonary nodule descriptors for nodule type classification. Medical Image Computing and Computer-Assisted Intervention”. MICCAI 2010: Berlin, Germany: Springer (2010): 626-633.

Volume 2 Issue 3 March 2022