



In-Depth Examination of Machine Learning Algorithms for Lung Disease Prediction

Dr.P.A.Selvaraj(Associate professor)
Department of computer applications
Kongu Engineering
College,Perundurai-
638060,Erode,Tamil Nadu, India

Dr.M.Jagadeeshan(Associate
professor) Department of computer
applications Kongu Engineering
College,Perundurai-
638060,Erode,Tamil Nadu, India

Mr.Vasanth (PG Scholar)Department
of computer applications Kongu
Engineering College,Perundurai-
638060,Erode,Tamil Nadu, India

Abstract-Today's image manipulation is a highly challenging field in detecting the lung disease from lung disease Scan has become popular in our research area. In this paper, we'll look at, lung disease image is used to find out if you have a brain tumour. This system includes conversion to gray, test the brain, whether it is benign or malignant, Threshold Segmentation is the simplest segmentation method, canny operator for edge detection and morphological Based denoising used for removing the noise, clustering method combination of MRF and CRF, Watershed segmentation is the best a method for dividing pixels in a picture into groups based on intensity, by combining k-means with watershed Segmentation overcome the Over Segmentation and use texture Segmentation.

Keywords-K-means, watershed and Texture Segmentation, CannyOperator, MorphologicalBased denoising, MRF AND CRF.

INTRODUCTION

A lung disease is an abnormal growth of cells within the lung This disrupts lung processes like muscle control, sensation, memory, and other regular body functioning. There are two categories of people. of lung disease, primary and metastatic. The primary Tumors start in the brain and tend to stay there, whereas metastatic lung disease starts elsewhere in the body and travels to the brain.. There is two kinds of primary lung disease Benign and Malignant (mal="bad", Ignis="fire"). Lung disease does not contain cancer cells, whereas malignant disease contains cancer cells. A Lung Disease Scan is a radiological procedure that produces images using magnetism, radio waves, and a computer. the internal the framework of the lung disease.

PREVIOUS WORK

Liu et al^[1] proposed a method of aThe opening-closing operation and the distance transform are used to alter the watershed. It's a form of repeated Calculate erosion and dilation. It reflects the pixel's position in the

illustration and overcomes the over-segmentation and detects the brain tumor.

The unique picture segmentation method was described by Bala et al [2]. In this case, rather than using the Sobel operator as in previous marker controlled watersheds, the prewitt's operator is used to detect the edge. transform to control over-segmentation and detect the lung disease.

Rahman et al^[3] has discussed about the novel method of Watershed algorithm segmentation based on adaptive threshold and masking operation. Here the main aim is to overcome over segmentation and detect the lung disease.

Shan et al^[13] produce The mechanism for segmenting watersheds has been enhanced. The reconstruction filter is used in the morphological operators to reduce noise and provide complete information about the object contour. The morphological gradient is determined. to compensate for deficiencies traditional gradient operators as well as to produce a For watershed segmentation, a gradient map is appropriate.

Acharjya et al.[14] have developed a novel image segmentation method based on watershed algorithm, which is employed for distance transform. To identify edges and control over segmentation, the (LOG) Laplacian of Gaussian edge detector is utilised.

Fernando C. Monteiro [15] proposes a novel method for image segmentation. comprising using a spectral approach and a morphological algorithm to extract information from edges and regions the watershed. To begin, they use a bilateral filter to minimise the image's noise. Second, preliminary Region merging is used for segmentation followed by similarity between regions generation and The Multi-class Normalized Cut technique is used for graph-based region grouping [16]. As a dataset, the Berkley segmentation dataset is used. Mean shift, multi-scale graph-Some of the strategies employed are based

segmentation and JSEG. are all used in the technique comparison. The proposed strategy outperforms previous methods and produces superior results, according to the findings.

PROPOSED METHODOLOGY

The majority of wealthy countries' research has led to this conclusion. that 73% of brain tumor deaths occur, when compare to other cancers ie.47%. It is In the United States, a projected There are There are 23,380 adults (12,820 men and 10,560 women) with cancer. primary malignant tumours of the brain. In the year 2014, it is expected that 14,320 persons (8,090 males and 6,230 women) will die from this lung disease[5]. A total of 22,850 malignant lung tumours will be diagnosed (12,900 males and 9,950 females). Brain and spinal cord cancers will kill 15,320 persons (8,940 men and 6,380 women). In 2015, the likelihood of developing a malignant lung tumour in one's lifetime was less than 1% (about 1 in 140 for a male and 1 in 180 for a woman)[4]. .This lung disease is a notoriously difficult disease to diagnose. In the present work the implementation tool is MATLAB. It is used to improve the detection of lung diseasevery accurately. Image segmentation can be accomplished using a variety of methods..First Lung Disease is taken and the process can be done by the following steps.First step is the image acquisition. Here the Lung Disease is converted to the grayscale imagesThe input is pre-processed in this stage.LungDiseaseare tested to identify whether the Lung Disease is normal(Benign) or abnormal(Malignent).Then the Threshold Method uses to separate the background and ROI (region of Interest) i.e.Tumor portions.The tumor obtain doesn't have exact edges, so the canny edge detection method is used. Then a noise is removed by morphological Based de-noising. In clustering Method combination of theMRF and CRF were used. Finally, theThe major purpose of this project is to combination of K-Means and Watershed Segmentation is used to overcome Over Segmentation and texture Segmentation to detect the tumor portion of the input brainLung Disease.

A. Image Acquisition:

Lung disease Images that are obtained by using lungdisease converting RGB to Gray level. In Weighted method give a gray color image by using the formula

$$(0.3*R)+(0.59*G)+(0.11*B)$$

I.e. based on the intensities 30% of red, 59% of green and 11% of blue will give the clear gray scale image

B.Pre-Processing:

On this Pre-Processing stage the input Lung Disease are tested to identify whether the Lung Diseaseis Benign or Malignant. The grayscale Two-dimensional graphics are displayed. matrix 0 to 255, with 0 indicating total black and 255 indicating total white thus we can separate the background and ROI and then the edges are detected using the Canny Operators. Noise is removed by Morphological based denoising-non-local means^{[9]. [12]} is When compared to other filters, it is more efficient and gives superior results.. Next in the clustering methods the combination of MRF and CRF reduces both the overlapping and noise issues with high accuracy^{[10],[11]}. A countable set of random variables is called a Markov Random Field. (RV) which is often used in the image access applications. It has the Markov property described as a unidirectional graph. It is similar to Bayesian networks. The differences between Markov networks are a type of undirected network that can be used to solve problems. acrylic, whereas Bayesian networks are directed and acrylic. As a result, a Markov network can express It can depict dependencies that a Bayesian network can't (such circular relationships), but not dependencies that a Bayesian network can (such as induced dependencies). A Markov chain random field's underlying graph might be finite or infinite.. CRF is a popular probabilistic method for the prediction of a a vast number of input photos combined with classification and graphical modelling.

C. Processing Stage:

Lung disease diagnosis entails two major steps:

- A. Image Classification and
- B. Image Segmentation.

On the subject of image classification, diagnosis of Lung Diseaseis based on Similarity measurement. In the The diagnosis of a brain tumour is dependent on image segmentation. size and shape.

D. Post-processing Stage:

The classification For processing, the methods listed below are used.

In computer vision, image segmentation is the process of splitting a digital image into numerous sections (sets of pixels, also known as super pixels). It's used to find things and define boundaries in the environment photographs. The majority of segmentation algorithms are based on one of two fundamental features: intensity levels.

Partitioning an image based on sudden shifts in intensity is known as discontinuity.

Partitioning an image into sections that are similar is known as similarity.



Identical based on a set of predetermined criteria According to [Gonzales], some of the methods for segmenting an image are as follows:

Methodology

A. Threshold Method: The Optimal Threshold is the To distinguish the ROI from the background, use this simple yet efficient method. ie After that, segmentation is accomplished by grouping All pixels with an intensity greater than the threshold are assigned to one class, while all other pixels are assigned to another. The threshold value is used to determine the threshold value at which the grayscale image will be converted to binary image format..^{[6],[7]}

B. K-means and watershed segmentation:One of the most effective approaches for grouping pixels in a picture based on their intensities is watershed segmentation. It's a morphologically mathematical operating tool for checking output segmentation. techniques rather than input segmentation techniques due to its drawbacks of over segmentation and under segmentation. The combination of K-Means and watershed segmentation is used to overcome over segmentation. [8]

The algorithm K-means clustering classifies objects into k groups based on attributes/features, where k is a positive integer. The data is grouped (clustered) by lowering the Euclidean distance between it and the cluster centroid. The purpose of k-means clustering is to cluster the data. Convert a two-dimensional report into three-dimensional images. The brain has three dimensions because it is a mass body for calculating centroid, and we need to calculate centroid in three dimensions to get more accurate results.

Tumor detectionIn this case, we are essentially using scanned images of lung disease. We calculate the height, width, and number of pixels of the scanned digital image of lung disease as we process it, which is then converted to grayscale.As a result, it is converted to 3D, i.e. the pixels of that image are converted to 3D, and the clustering concept is applied to that image.The number of clusters was difficult to determine during clustering because a predetermined number of clusters is required by the K-means algorithm. We also define the first mean value for clustering. After the clustering process, we finally get the output where we get the result whether that image contains tumour or not. The use of an automated system has increased the accuracy of detecting tumours.

Watershed segmentation is a technique that uses gradients to segment data. It treats the image's gradient map as a map of relief As if it were a dam, it divides the vision. Catchment basins are names given to the segmented regions. Watershed segmentation is useful for a wide variety of picture segmentation issues It works well with images that have a greater intensity value. Watershed segmentation is caused by over segmentation. To gain control over segmentation, marker controlled watershed segmentation is used. The operator Sobel is appropriate for edge detection.

The sobel masks in matrix form are as follow:

$$M_x = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}, M_y = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 0 \\ -1 & 0 & 2 \end{bmatrix}$$

The equation of gradient magnitude used in marker controlled watershed segmentation is

$$M = \sqrt{M_x^2 + M_y^2} \tag{2}$$

$$\text{Angle}, \theta = \tan^{-1} \frac{M_y}{M_x} \tag{3}$$

D. Texture Segmentation: Texture is the regular repetition of an element or pattern on a surface that has some amount of variability in element appearance and relative position.The proposed system extracted texture features using two methods. The first technique uses a first order histogram as a local feature, whereas the second method uses a co-occurrence matrix as a second order textural feature. The procedure for extracting features is depicted in detail in Fig.2. 1. Features based on first-order histograms 2.Features based on co-occurrence matrices: Local features are based on histograms. These attributes do not take into account spatial information. Second order histogram based features are gray-level spatial concurrence matrix hd (i,j) based features. defined for this purpose..

$$\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} [p(i,j)]^2 \tag{10}$$

$$\text{Correlation: } \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{11 p(i,j) \mu_x \mu_y}{\sigma_x \sigma_y} \tag{11}$$

$$\text{Inertia: } \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i-j)^2 p(i,j) \tag{12}$$

$$\text{Absolute Value: } \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} |i-j| p(i,j) \tag{13}$$

$$\text{Inverse Difference: } \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{p(i,j)}{1+(i-j)^2} \tag{14}$$

$$\text{Entropy: } H = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} p(i,j) \log_2 [p(i,j)] \tag{15}$$

$$\text{Maximum Probability: } \max_{i,j} p(i,j) \tag{16}$$



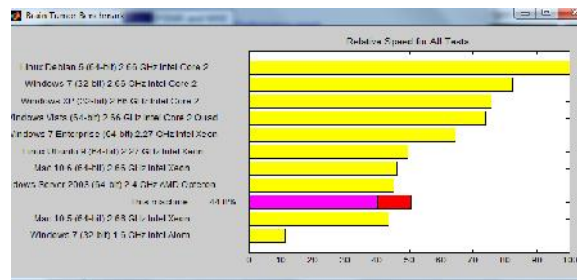
Determine the tumor's size. The disease region's area is the following equation was used to calculate: Tumor site equals the total the number of pixels in the tumour multiplied by A_x (5) $A = V \times H$ (6) Where A denotes the pixel's area. The horizontal dimension of the image is denoted by H, and the vertical dimension is denoted by V. $H = 1/\text{image resolution on the horizontal plane}$ $V = 1/\text{the image's vertical resolution}$

BENCHMARK FOR BRATS SET-UP

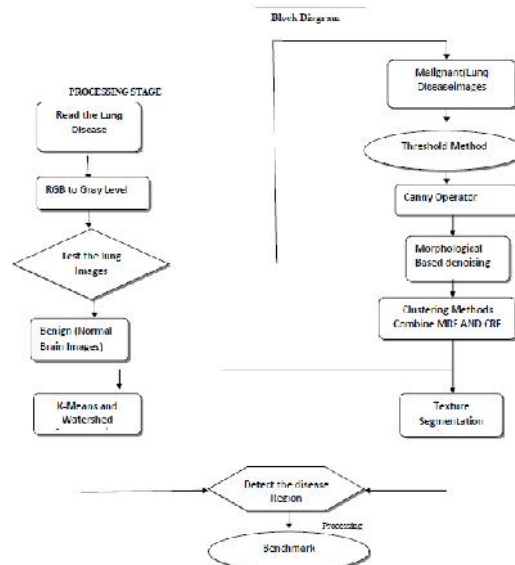
In 2012 and 2013, the BRATS standard was presented as two satellite challenge workshops in connection with the MICCAI conferences. Both tasks, as well as the imaging data and human annotation procedure, were set up with participating teams., validation processes, and online resources for comparing the numerous algorithms, are all described here. New submissions are still being accepted through the BRATS web tools, which new groups can collect training and test data, as well as submit segmentations for automatic classification rating against all previous submissions. A single point of contact for benchmarks, in addition to the most recent BRATS-related initiatives

Detecting such failures would necessitate a thorough examination of every component each pipeline for processing — for instance, in a learning-based approach to intensity normalisation, the feature Extraction, spatial regularisation, and classification algorithms. While this form of research is incredibly valuable, it necessitates a meticulous experimental design, which is difficult to achieve in a competitive benchmark like BRATS, post hoc on a broad range of algorithms offered by multiple parties.

Another issue in the current study, which is shared by other benchmarks, is the inability to choose adequate overall assessment criteria that can be utilised to directly compare the results of other studies. All competing algorithms are ranked. To obtain our overall final ranking in different tumour locations, we used average Dice scores., despite publishing Results for sensitivity, specificity, and Hausdorff distance are separated. However, as the Festa method's findings in "active tumour" segmentation illustrate, the correct choice of evaluation metric is critical. is critical.

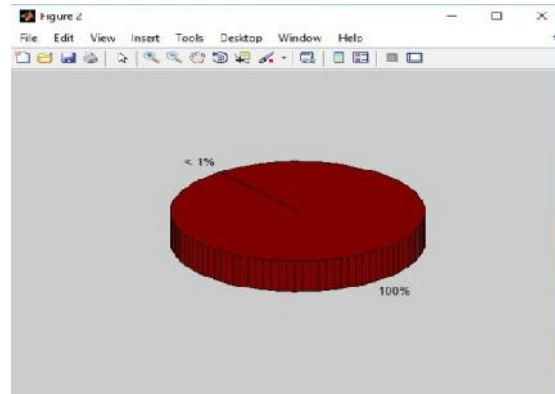


BenchMark



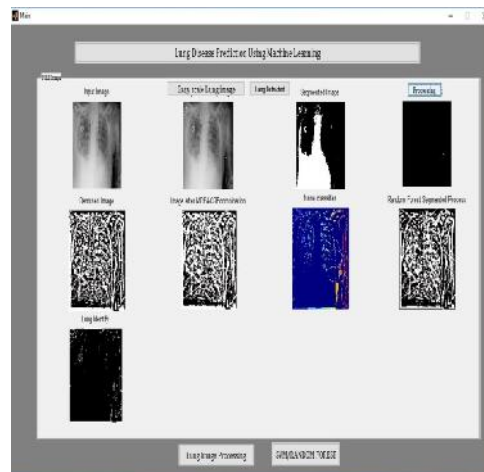


Lung Disease Analysis



Machine Type	1.0	1.1	1.2	1.3	1.4	1.5
Linux CentOS 5 (64-bit) 2.66 GHz Intel Core 2	0.2462	0.0922	0.1419	0.2600	0.2541	0.2199
Windows 7 (32-bit) 2.66 GHz Intel Core 2	0.2729	0.1185	0.136	0.2425	0.2923	0.2511
Windows XP (32-bit) 2.66 GHz Intel Core 2	0.2183	0.1127	0.1103	0.2202	0.2026	0.2453
Windows Vista (64-bit) 2.66 GHz Intel Core 2	0.2185	0.1025	0.1164	0.2725	0.2026	0.2453
Windows 7 Enterprise (64-bit) 2.77 GHz Intel Xeon	0.2695	0.0971	0.1471	0.2134	0.2177	0.2165
Linux Ubuntu 9 (64-bit) 2.77 GHz Intel Xeon	0.2673	0.0954	0.1478	0.2478	0.2177	0.2095
Mac 10.6 (64-bit) 2.66 GHz Intel Xeon	0.2735	0.1165	0.1199	0.2330	0.2179	0.2531
Windows Server 2003 (64-bit) 2.4 GHz AMD-Opteron	0.1838	0.1322	0.2701	0.2535	0.2508	0.2016
Mac OSX 10.5 (64-bit) 2.4 GHz AMD-Opteron	0.1454	0.1122	0.1473	0.2087	0.2195	0.2081
Mac 10.6 (64-bit) 2.4 GHz AMD-Opteron	0.2025	0.1425	0.1782	0.2137	0.2081	0.2087
Windows 7 (32-bit) 1.0 GHz Intel Atom	1.3047	0.7013	0.0214	1.0210	2.0777	1.8173

Fig.2 Process of LungDisease detection



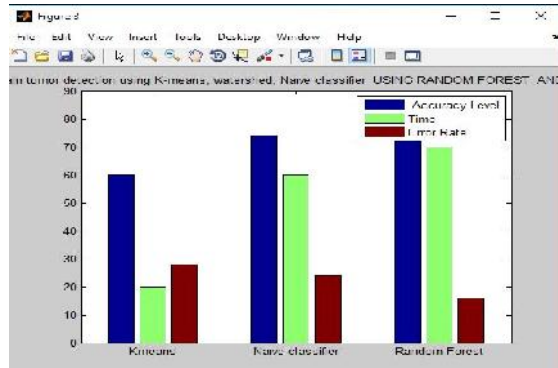


Fig: 3BenchMark (6 -Time Process)

EXPERIMENTAL RESULTS ANALYSIS AND COMPARISON

The division of an image into regions with comparable attributes such as grey level, colour, texture, and so on, brightness, and contrast is known as segmentation. The primary goal is to detect the tumour present in the patient brain, which is very useful for the diagnosis, which is very easy to take treatment for the survival of the patients. For the experiment, 20 Lung Disease images obtained from the clinical laboratory are included for to test the code. Fig2. Describe the whole process of the Lung Disease detection. There are two most common error options, namely the PSNR (peak signal noise ratio) and MSE (mean square error) are displayed in the graph. above fig.3. The result for the K-Means, Watershed and Texture segmentation are shown below and table.1 also clearly shows the implemented on different images taken from the laboratory for our research work. Calculation of MSR and PSNR are shown the accuracy of the proposed algorithm we report on the set-up and outcomes of the experiment Multimodal Lung Disease Segmentation Benchmark (BRATS) was held in conjunction with the MICCAI conferences in 2012 and 2013. Twenty cutting-edge tumour segmentation methods were used on a collection of 65 images. multi-contrast Lung Disease of Low- and high-grade glioma patients were manually annotated by up to four raters, and 65 other scans were created using tumour markers image simulation software..

Quantitative assessments revealed a significant difference between human raters in tumour sub-region segmentation (Dice scores) ranging from 74 to 85 percent), highlighting the challenging task of segmenting tumour sub-regions.

Different algorithms scored better in different sub-regions (with results comparable to human inter-rater variability), but no single method ranked first overall. first in all sub-regions. at the same time. combining a few good algorithms Segments consistently outperformed all individual methods when utilising a vote with a hierarchical majority, indicating that methodological improvements are still possible. The BRATS Data and manual annotations are still available to the public. as an ongoing benchmarking resource through an online evaluation system.

CONCLUSIONS AND FUTURES WORK

We presented the BRATS Lung Disease segmentation benchmark in this paper. We had the greatest outpouring of public support. dataset was made available for this project, and a large number of cutting-edge Lung Disease segmentation methods



Volume 5- Issue 2, Paper 14 August 2022

were evaluated. Our findings show that even segmenting a brain tumour is difficult. Existing algorithms can currently reach raters. Dicescores This figure exceeds 80% for the entire disease segmentation. Segmenting the disease The core region, particularly the active core region, proved more difficult to study in high-grade gliomas, with Dicescores reaching 70% and 60%, respectively. For all tumour regions, no single algorithm outperformed the others. taken into account In contrast, the flaws in the most advanced algorithms for each individual region were within the range of human inter-rater variability. An important finding in this study is that combining different segmenters improves performance significantly. Decisions were made

We were able to consistently outperform By giving specified groupings of algorithmic segmentations a hierarchical majority vote, the best individual segmentation algorithm for every single segmentation task This means that further advances (and, eventually, practical implementations) can be acquired by figuring out how to combine several separate algorithms, either through a majority vote or other ways fusion strategies, in addition to pushing the bounds of individual disease segmentation algorithms.