iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** IX **Month of publication:** September 2021

DOI: https://doi.org/10.22214/ijraset.2021.38258

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

Lung Nodule Detection using Segmentation Approach for Computed Tomography Scan Images

Henil Satra[1], Aryan Gupta[2]
[1, 2]*B.Tech Student, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, TN, India*

Abstract: ***Lung disorders have become really common in today's world due to growing amount of air pollution, our increased exposure to harmful radiations and our unhealthy lifestyles. Hence, the diagnosis of lung disorders has become of paramount importance. The commonly used Thresholding approaches and morphological operations often fail to detect the peripheral pathology bearing areas. Hence, we present the segmentation approach of the lung tissue for computer aided diagnosis system. We use a novel technique for segmentation of lungs from CT scan (Computed Tomography) of the chest or upper torso. The accuracy of analysis and its implication majorly depends on the kind of segmentation technique used. Hence, it is important that the method used is highly reliable and is successful in nodule detection and classification. We use MATLAB and OpenCV libraries to apply segmentation on CT scan images to get the desired output. We have also created a working proprietary user interface called "PULMONIS" for the ease of doctors and patients to upload the CT scan images and get the output after the image processing is done in the backend.***

Keywords: Lung nodule detection, Image Processing, Computed Tomography, Image Segmentation, Lung Cancer, Contour Segmentation, MATLAB, OpenCV, Computer Vision.

I. INTRODUCTION

Cancer is a deadly disease in which the body cells mutate and grow out of control. Cells are the building blocks of our body. Normal cells grow when our body needs them and die when our body does not need them. Sometimes, abnormal cells grow even when the body doesn't need them. This growth is cancerous and these abnormal cells form a lump or mass called a tumor. If these cancer cells are big enough, they have the capacity to spread to other regions of the body. Lung Cancer is one of the most harmful type of cancer owing to its smallest survival rate post diagnosis. Chances of survival from lung cancer are directly associated with its growth at its detection time. The patients' chances of survival are higher if the detection happens in the initial stages.

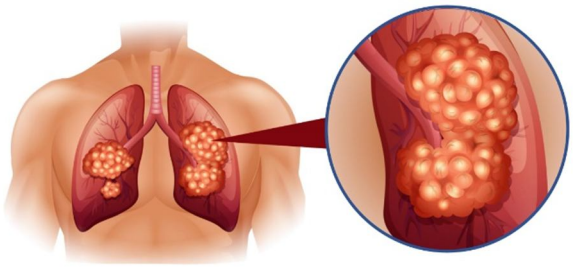


Fig-1: A typical lung cancer tumor

There are four stages of lung cancer and staging is completely based on the size of the tumor and the location of the lymph node. CT scan images are more efficacious than normal chest x-rays in detecting and diagnosing lung cancer. The region of interest is separated using image processing techniques like pre-processing and feature extraction. Using the algorithm, features like size, boundary and eccentricity are extracted from the images. These values are then compared with the normal values suggested by the healthcare professional.

In computer vision, image segmentation is a technique of breaking a digital image into numerous parts or segments. We use segmentation to make the analysis easier by converting the image into something more understandable and meaningful. The accuracy of the segmentation subsystem determines the success of the diagnosis. In the periphery of lungs, the gray scale values of the PBRs (Pathology Bearing Region) and chest region are nearly the same, hence it becomes difficult to detect periphery nodules or lesions. To overcome this conundrum, the segmentation approach we use also takes into account the convex edge of lungs in addition to gray scale values.

©IJRASET: All Rights are Reserved

The technique proposed by us can be used to preprocess the images of the lungs to improve the diagnostic performance of the computer system. The results of our technique show that lungs can be accurately segmented even in the presence of PBRs especially in the case of diseases like Lung Parenchyma. For implementation, we use MATLAB and OpenCV libraries. We also provide a UI system called "PULMONIS" for the easy implementation of our technique. Our custom made system is built on HTML, CSS, JavaScript, PHP, SQL and jQuery.



Fig-2: Home page of the PULMONIS system

II. LITERATURE REVIEW

1) The research paper "Image processing based detection of lung cancer on CT scan images" authored by Bariqi Abdillah, et al. lays the foundation of image processing methods for detection of lung cancer. Region growing approach, Marker Control Watershed (MCW) and MCW with Masking are the image segmentation techniques proposed by the research paper and the effectiveness of these methods are analyzed. First, the image of the lung is detected, then it is enhanced using a Gabor filter (a linear filter) and then it is segmented using the three techniques proposed. Furthermore, colour attribute for analysis of lung cancer using binarization was used which successfully determines whether a lung is normal or whether it has cancer. Drawback of this approach is the lack of pre-processing techniques like noise removal, image soothing which we implemented and it does not classify the extracted cancer as benign or malignant.

2) The paper "Lung cancer detection using CT scan images" by Suren Makaju, et al. proposes computer aided diagnosis which can be helpful for the doctors to identify cancerous cells in the lungs accurately. The paper evaluates many image processing techniques, analyses the current best model, finds limitations in it and finally proposes a new model of its own with a much higher accuracy. The procedure proposed by the paper is Median and Gaussian filter for image enhancement along with Watershed Segmentation, followed by Advanced Feature Extraction. Finally, it also classifies the cancer as benign or malignant using Support Vector Machine which is a supervised machine learning method. A drawback of the proposed approach is that it does not classify the cancer into different stages as stage I, II, III, IV of cancer.

3) The article "Lung cancer prediction using machine learning and advanced imaging techniques" authored by Timor Kadir and Fergus Gleeson, analyzes machine learning based lung cancer prediction models and highlights their strengths and weaknesses. The paper indicates that computer aided diagnosis techniques aim to provide a quantitative output related to the risk of lung cancer. The work also suggests that the most effective nodule classification approach is achieved using Convolutional Neural Networks (CNN) with Deep Learning. It also talks about the limitations of the training and data sets used. The paper however has some unanswered questions such as the optimal segmentation techniques which should be used and there is little emphasis on pre-processing techniques like noise removal.

4) The paper "Lung Cancer Detection Using Image Processing Techniques" authored by Mokhled Altarawneh focuses on image quality and accuracy as the core factors of their research achieved by applying pre-processing techniques based on Gabor and Gaussian filters. Further, two image segmentation techniques are implemented and their effectiveness is measured based on the time factor which is an important aspect of this research. Following the image segmentation techniques, feature extraction is done using binarization and masking to check whether the concerned lung is abnormal or not. This research however does not classify the found cancer as benign or malignant and does not give the size of the cancerous tumor.

5) The work "Automatic Lung Cancer Detection using Sobel & Morphological Operations" authored by Akanksha Soni, et al. proposes an automatic detection method with less processing time and good accuracy. This research primarily focuses on Sobel filtering used for edge detection followed by dilation which is a morphological operation to fill the focused region of interest and holes so as to get an enhanced and smoothened image of the cancerous area of the lung. This makes it easier for doctors to detect and extract the cancerous region of the lung. The method used in the paper is fairly efficient and still quite advanced in comparison to other techniques being used. This research however does focus on the optimal segmentation techniques that can be deployed.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

III. METHODOLOGY

Our algorithm for the proposed system has 3 main stages:

1) Pre-processing of the given CT Scan image
2) Performing morphological operations and segmentation
3) Finding the tumor type, size and distance between tumors.

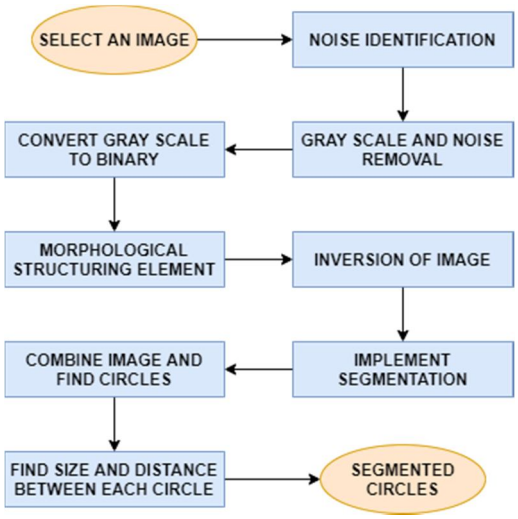


Fig-3: State diagram of our proposed technique

A. Identification of Type of Noise

The most common type of noise found in Lung CT scan images is salt and pepper noise. In this type of noise, certain pixels of the image are either black or white. We remove this type of noise my applying a median filter. This filter removes the salt and pepper noise without significantly affecting the sharpness of the image.

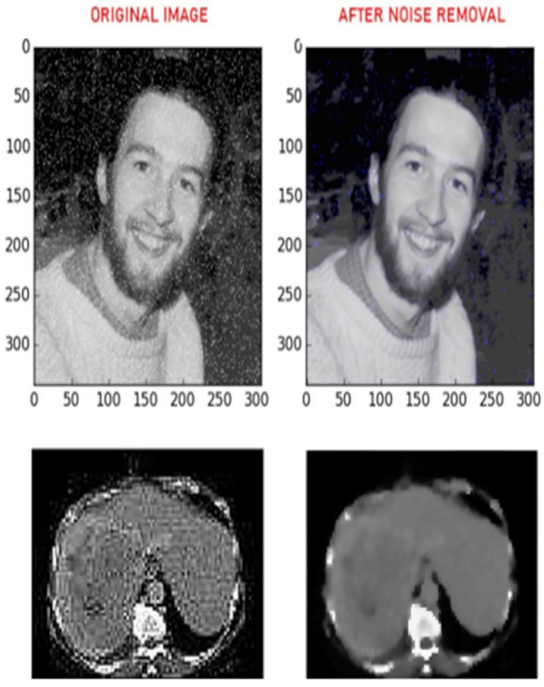


Fig-4: Noise removal after applying median filter

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

B. RGB to Grayscale Conversion and Noise Removal

The presence of noise in the image causes distortion within the image in terms of brightness and darkness. To overcome this, we first remove all the noise. We take the lung CT scan image as our input and after the image is read, we convert it from RGB (Red-Green-Blue form) to grayscale form. The gray scale image is one in which there are no colours and each pixel is a shade of grey depicting the intensity of light. The value of each pixel ranges from 0 to 255. This helps in removing the hue and saturated information while retaining the desired luminance.

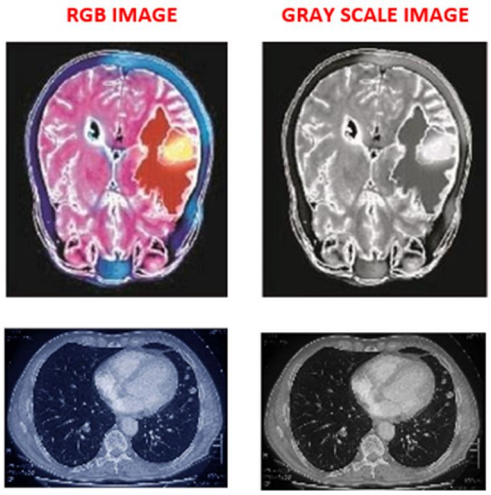


Fig-5: RGB to Gray Scale conversion

Once the image has been read and converted, we remove the noise using a custom median filter. The median filter gives the output pixel as the median value in the 3-by-3 region around the corresponding pixels within the image. This approach is similar to soothing techniques and by using corresponding pixels, we keep the image sharpness intact.

C. Gray Scale to Binary Conversion

We convert the Gray scale image into a binary image. A binary image has only two values in each pixel, i.e. 1 (white) and 0 (black). This helps in detecting objects within an image (circles). In our context, it will help us to detect tumors or nodules within the CT scan image of lungs. Binary conversion of a pixel depends upon its luminance level. Values above the set threshold become white (1) and values below the threshold become black (0). In case the value is too close to the threshold, the function checks the nearby cells/pixels in order to get the most accurate binary form of an image.

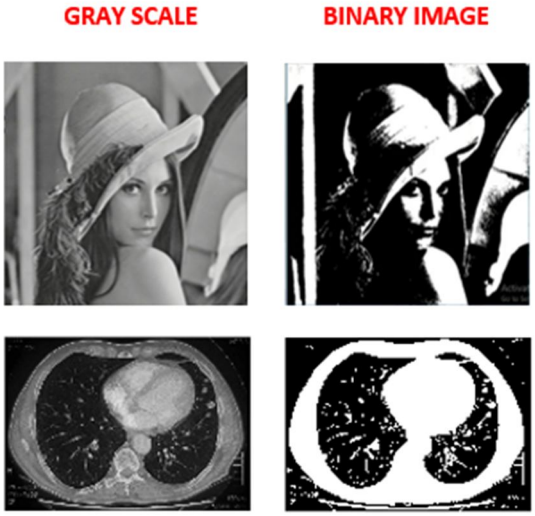


Fig-6: Gray Scale to Binary conversion

©IJRASET: All Rights are Reserved

D. Morphological Operations

In this step, we create a morphological structuring element in the form of a rectangle and this function creates a flat disk like shape with a specific radius. This is done twice to create two separate structures which are applied to the initial binary image. The process used here is called Opening of the image. Opening is achieved by erosion, followed by dilation and these operators are derived from Mathematical Morphology. The basic function of opening is exactly like erosion as it tends to remove some of the foreground (bright) pixels from the edges of regions. One image is applied to the subplot and given to the user interface while the other image is saved for further image processing.

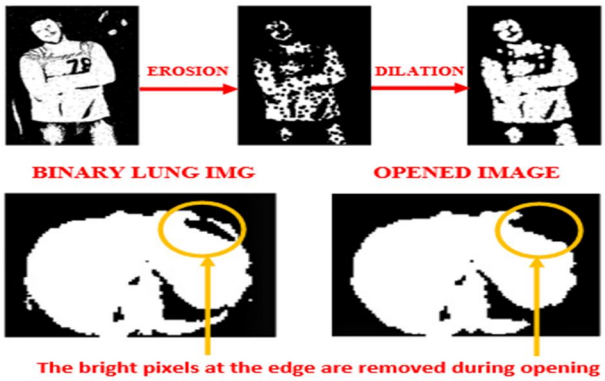


Fig-7: Morphological operation: Opening of an image

E. Inversion of the Opened Image

The next step is to invert the image. This function basically creates an array of the values in each pixel. The 1s become 0 and vice versa. This step initiates the segmentation process which occurs in the next step. We use inversion to overcome the lack of contrast in images and hence, we are able to focus better on the desired region of the image.

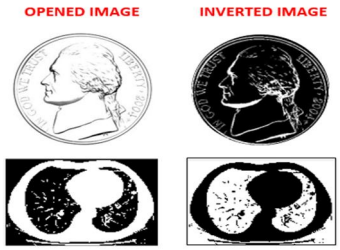


Fig-8: Inversion of an image

F. Initial Contouring and Segmentation

This step involves creating and specifying the initial contour using a mask. The mask is essentially a binary image which indicates the initial state of the active contour (boundary of the image regions). In our case, our case, the background is white and the black region's boundary becomes the contour. This allows for contour evolution which eventually paves way for segmentation to happen. Segmentation is the process of partitioning or breaking and image into number of segments or pieces based on some specific attribute, characteristic, property or constraint. We use Contour Based Segmentation in our model. Once the image is segmented, we have a clear way of finding the circles as now the foreground and background are clearly distinct and won't create any issue.

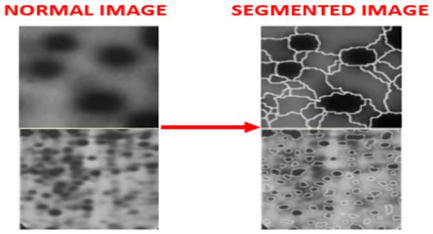


Fig-9: Contour based image segmentation

©IJRASET: All Rights are Reserved

G. Combination of Images and Finding Circles

This is one of the most important steps of the model. We attempt to accurately find tumors and circle them to point out their location. The inverted image and the contour segmented images are combined together to form a single image on which the circles are found. This function finds circles using properties like object polarity and sensitivity. The function looks for dark sensitivity within the image using the parameters of centres and radii with a sensitivity specification. We display the highlighted circles in green colour and also count the number of circles. These circles are actually tumors or cancer nodules. After counting and locating the circles, we segment the circles. This is done by taking a picture without circles to a picture with circles located, thus leading to an image with circles totally segmented away from the picture, which helps in finding the circle boundaries in a precise way. Now all the regions inside the circle are filled with green colour. We use MATLAB libraries for steps 3.1 to 3.7.

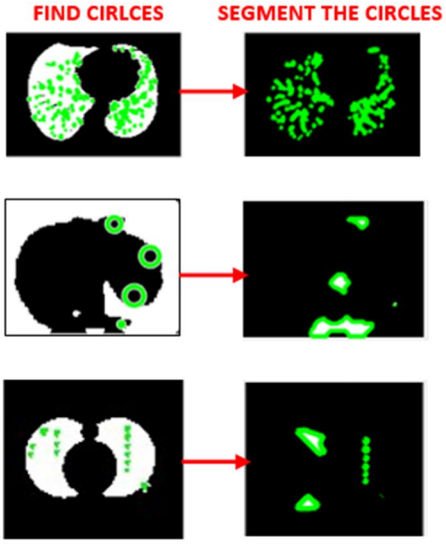


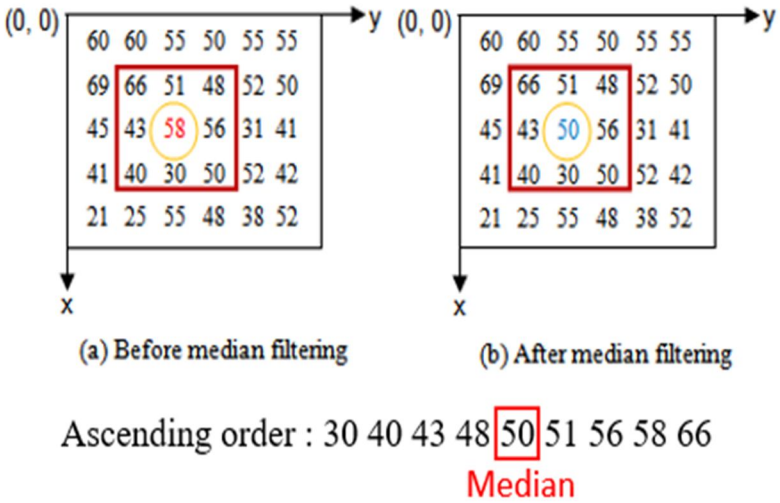
Fig-10: Locating the tumors in the image and circling them

H. Calculating type, size and Distance Between each Tumor

Based on the number of nodules or circles located, we classify the tumor as benign (harmless), pre-malignant and malignant. This is basically specifying the stage of the cancer. This is the final step of our model and we use OpenCV and NumPy libraries to implement it. We find the centres of each circle using contouring and then find the Euclidean distance between them. Since we make a boundary around the tumors, we also find the size by using pixel per metric ration and by applying the area formula.

IV. FORMAL CALCULATIONS

A. Median Filtering for Noise Removal



©IJRASET: All Rights are Reserved

B. RGB to Grayscale Conversion

$$\text{Grayscale Pixel (x,y)} = \text{red}*0.3 + \text{green}*0.59 + \text{blue}*0.11$$

C. Grayscale to Binary

If Pixel (x,y) >= 0.6 ← Threshold
Binary (x,y) = 1 ← White
Else Binary (x,y) = 0 ← Black

D. Distance Between Tumors

Euclidean Distance b/w (x_1,y_1) and (x_2,y_2) = $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$

V. OUTPUT AND RESULTS

We take the lung CT scan image as the input on our proprietary PULMONIS user-interface, run the backend on MATLAB and OpenCV and again redirect the output on our system. We have shown the outputs for 4 sample lung CT scan image inputs.

A. Sample Input : Lung image 1

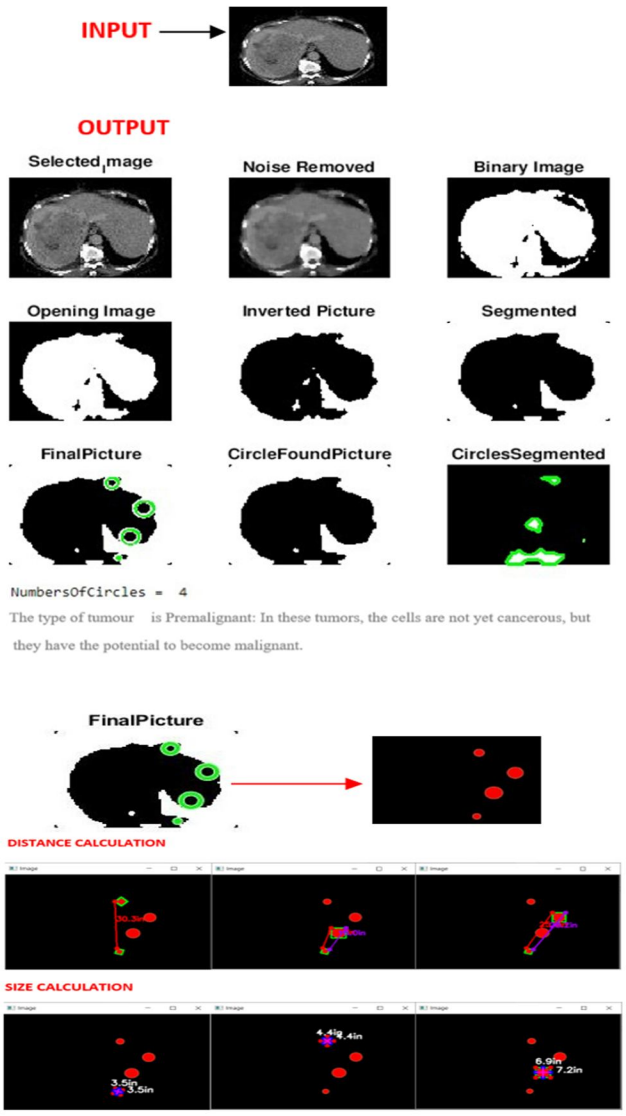


Fig-11: Output for Lung Image 1

©IJRASET: All Rights are Reserved

1) Output Results

a) We accurately find the location of tumors
b) Number of circles or nodules: 4
c) Type of tumor: Pre-malignant
d) Distance and size of tumors

B. Sample Input : Lung image 2

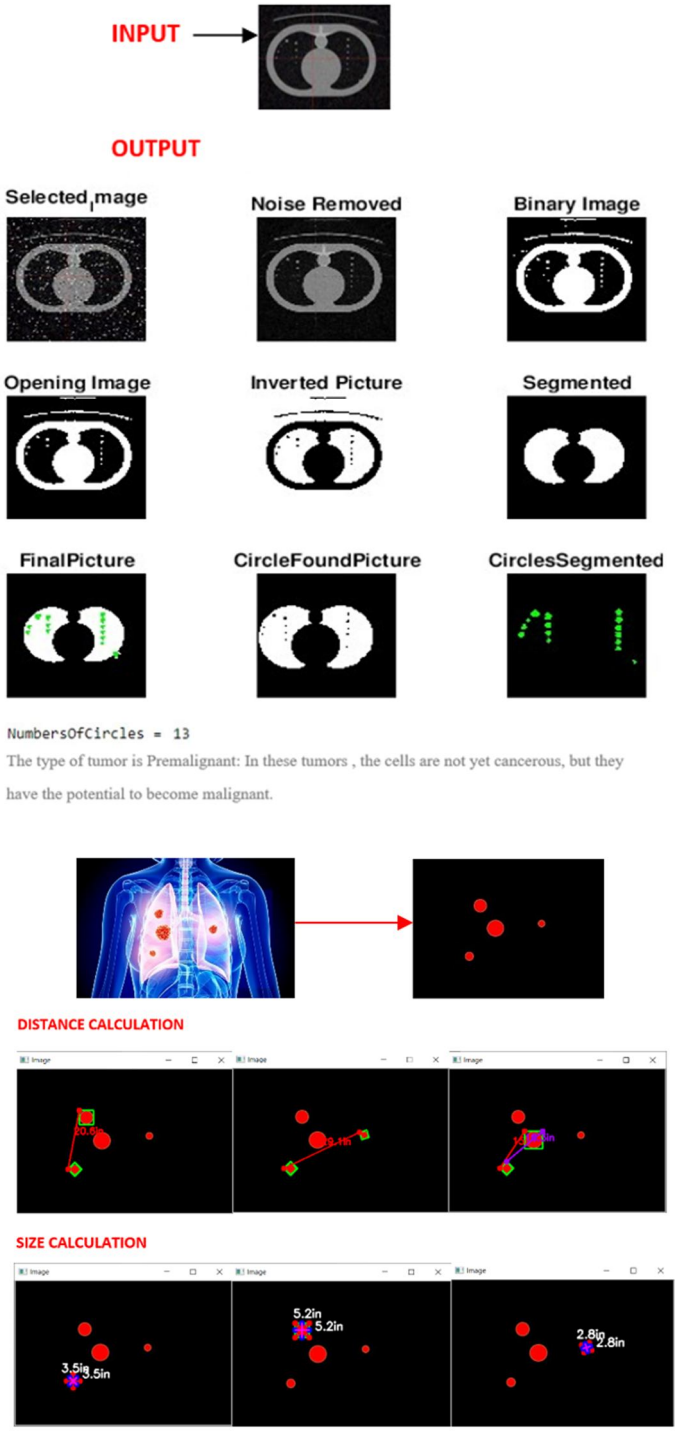


Fig-12: Output for Lung Image 2

©IJRASET: All Rights are Reserved

1) Output Results

a) We accurately find the location of tumors
b) Number of circles or nodules: 13
c) Type of tumor: Pre-malignant
d) Distance and size of tumors

C. Sample Input : Lung image 3

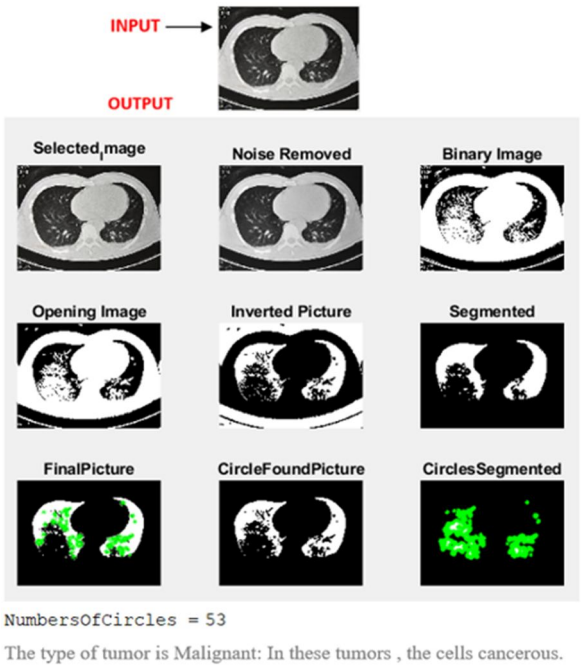


Fig-13: Output for Lung Image 3

D. Sample Input : Lung image 4

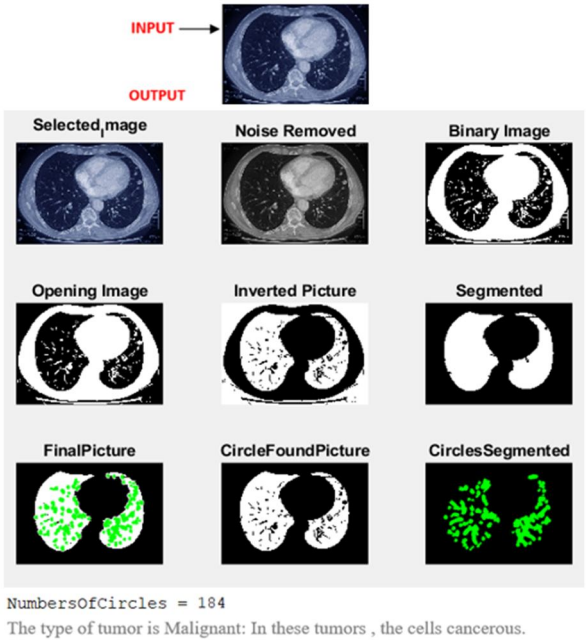


Fig-14 : Output for Lung Image 4

©IJRASET: All Rights are Reserved

VI. NOVEL FEATURES

We make use of pre-processing techniques to remove any kind of unnecessary and unrequired noise and contour based segmentation to accurately locate the tumors, circle them, find their size, type and distance between them. We accurately give the X and Y coordinates of the nodule locations which will help doctors immensely in surgical procedures and targeted chemotherapy. We provide a custom made UI system called PULMONIS to ease the image processing of lungs. This system has access to past records of patients which are stored and retrieved from SQL database. The main function of the system is to give the analysis output and display it juxtaposed with the input image for comparison.

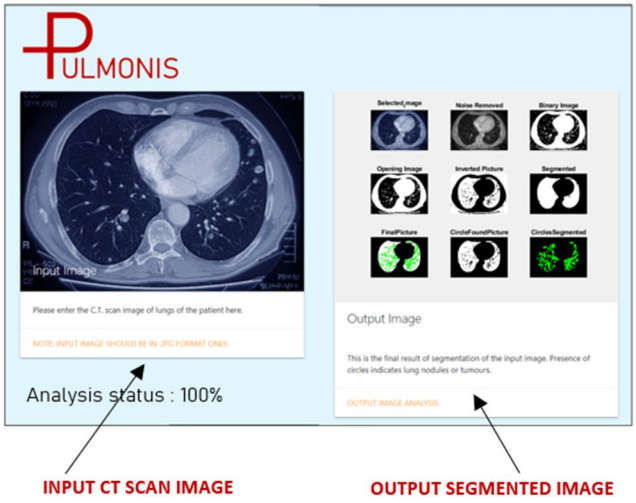


Fig-15: PULMONIS analysis page with input and output

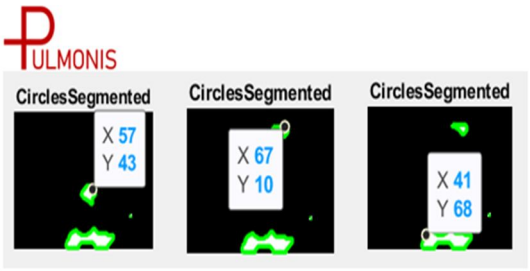


Fig-16 : Accurate (x,y) coordinates of the tumors

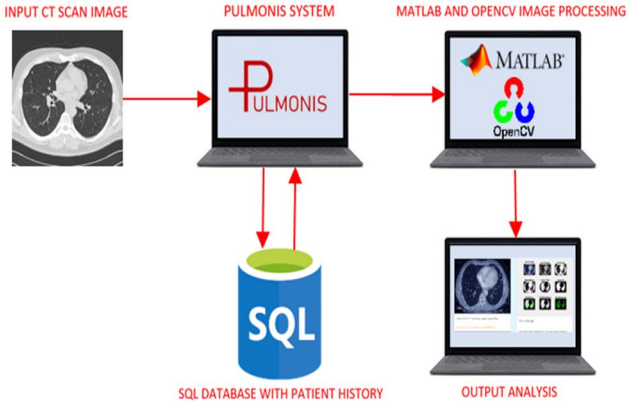


Fig-17: System architecture of PULMONIS

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue IX Sep 2021- Available at www.ijraset.com

VII. PERFOMANCE ANALYSIS

The segmentation methods are qualitatively and quantitatively assessed and compared based on three criteria: visual criteria, computational time and similarity index.

A. Visual Criteria

The segmentation done by active contour methods is compared with annotated images as reference by expert pulmonologists and radiologists. Contour based segmentation gives the best identification of lung nodules amongst other segmentation approaches. This is because we first make contours or boundaries to actively differentiate between regions before segmenting and finding nodules.

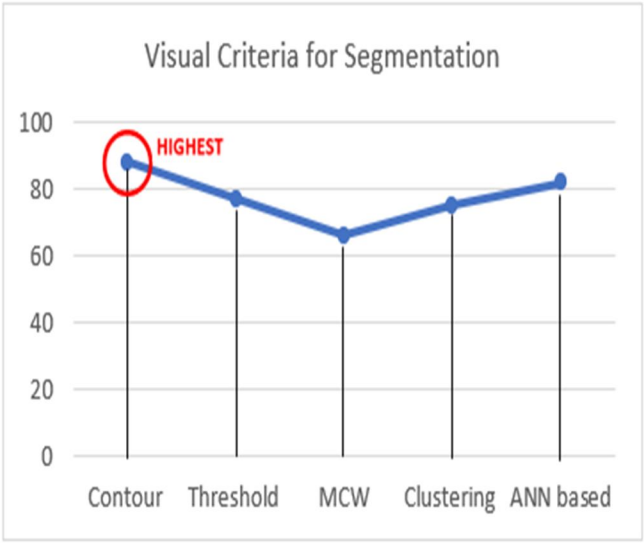


Chart-1: Visual criteria evaluation for different types of segmentation

B. Computational Time

This criteria represents the time taken for each algorithm to segment the region and essentially corelates to speed and efficiency of the model. We have achieved great time complexities in terms of our code and the worst time complexity of the model is of the order $O(n^2)$ due to the presence of nested for loops in the segmentation function. The average computational time in the backend for 4 lung images input was 4.24 seconds. The more the number of nodules, the more is the computational time to detect and circle them individually.

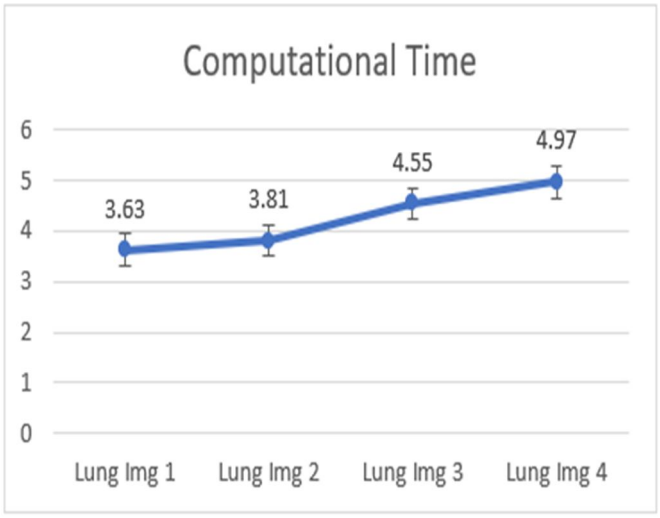


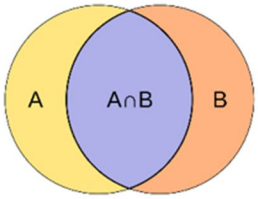
Chart-2: Computational Time for 4 sample inputs

©IJRASET: All Rights are Reserved

C. Similarity Criteria

This criterion measures the similarity between the reference and segmented image. The quality of output is measured by parameters like Dice coefficient, Hausdorff distance and Peak signal-to-noise ratio (PSNR).

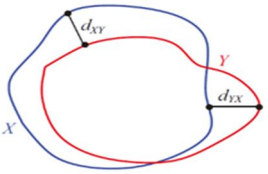
1) *Dice Coefficient:* Dice coefficient compared the segmented part with the reference region from the annotated image and gives a value from 0 to 1. If the value is 1, it means that the segmented region is more similar and if it is 0, it means they vary significantly.



$$\text{Dice Coefficient (A,B)} = \frac{2 \times |(A \cap B)|}{|A| + |B|}$$

Fig-18: Formula for Dice Coefficient

2) *Hausdroff Distance:* This is a metric to calculate the dissimilarity between two point sets. Distance transform is used to compute the Hausdroff distance of an image. It is used to control the progress of level set based algorithms and to find the quality of clusters.



$$D_1(A,B) = \max_{x \in A}\left(\min_{y \in B}(\|x-y\|)\right)$$

Fig-19: Formula for Hausdroff distance

3) *Peak signal-to-noise ratio (PSNR):* PSNR is the ratio between the maximum possible value of a signal and the power of distorting noise that affects the quality of its representation. Since many signals have a very wide dynamic range, PSNR values are calculated in terms of logarithmic decibel scale. It uses Mean Squared Error (MSE) equation.

$$PSNR = 10log_{10}\left(\frac{MAX_I^2}{MSE}\right)$$

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2$$

Fig-20: Formula for PSNR

VIII. STATISTICAL ANALYSIS

The features like mean, variance and standard errors are calculated for segmentation. We also use different filters for various operations like morphological enhancement, removing holes, etc.

A. Erosion and Dilation

The boundary pixel assume the value of their corresponding neighbours. If a pixel's value is 0 and the boundary neighbour pixels are all 1s, then the pixels value changes from 0 to 1. This helps in making a defined boundary or contour to differentiate between regions and also helps in filling holes and irregularities.

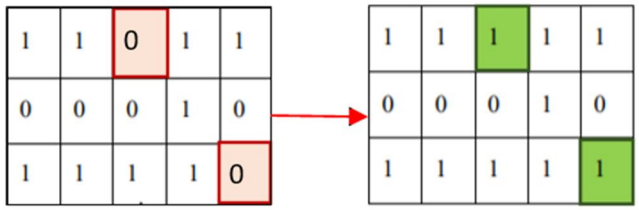


Fig-21: Changes in an image after erosion and dilation

©IJRASET: All Rights are Reserved

B. Mean, Variance and Standard Error

Mean is the average value which is computed by taking the sum of all outcomes divided by overall number of gray levels. Variance is the study of deviation of actual value against the predicted value. This deviation is an indication of the performance of method used. Standard error is defined as the measure of our prediction's accuracy. Estimated standard error is related to the sum of squared deviations of prediction (i.e. sum of squares of error). Y is the actual range and Y^1 is the predicted range and N is the total number of pairs of scores.

Mean Value $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x$

Variance Value $\sigma^2 = \frac{\sum(\chi - \mu)^2}{N}$

Estimated Error $\sigma_{est} = \sqrt{\frac{\sum (Y - Y')^2}{N}}$

Standard Error $SE = \frac{\sigma}{\sqrt{n}}$

Fig-22: Formulae for mean, variance and standard error

IX. CONCLUSION

Our model successfully accomplishes the task of finding lung nodules in CT scan images. The application of median filter to remove noise helps in the steps preceding actual segmentation. The opening of the image also contributes in the satisfactory results obtained as it helps to remove any kind of boundary based discontinuity. The contour based segmentation step is the most important step as it deals with finding the location of lung nodules. Our tests have proven that contour segmentation is fast efficient and accurate and it's results are at par with available advanced technologies.

Finding the size, type and distance between the tumors is very essential as it helps the healthcare professionals during surgeries and operations for tumor removal. Stating the stage of cancer and the type of tumor gives added clarity to the output results. Finally, the user-interface PULMONIS built by us helps in easy processing of the image, concealing the backend, resolving patient history storage issues and commercializing of the model at a much larger scale. The computational time and space complexity achieved by us adheres to the international computer vision modelling standards. We can try using different types of segmentation approaches in the future like p-tile thresholding and watershed segmentation considering the recent advancement in these particular fields. We hope that our PULMONIS system and processing model helps the medical industry in some way.

REFERENCES

[1] Bariqi Abdillah, et al. "Image processing based detection of lung cancer on CT scan images." Journal of Physics: Conference Series **(2017)**
[2] Suren Makajua, et al. "Lung Cancer Detection using CT Scan Images." 6th International Conference on Smart Computing and Communications **(2017)**
[3] Kadir T. & Gleeson F. "Lung cancer prediction using machine learning and advanced imaging techniques." Translational Lung Cancer Research **(2018)**
[4] Altarawneh, Mokhled. "Lung Cancer Detection Using Image Processing Techniques." Leonardo Electronic Journal of Practices and Technologies **(2012)**
[5] Soni, Akanksha, Avinash Rai and Ekta Shivhare. "Automatic Lung Cancer Detection using Sobel & Morphological Operations." **(2019)**
[6] A. Chaudhary and S. S. Singh. "Lung Cancer Detection on CT Images by Using Image Processing." International Conference on Computing Sciences **(2012)**
[7] Shyamala S. and Dr. Pushparani M. "Pre-processing and segmentation techniques for lung cancer on CT images." International Journal of Current Research **(2016)**
[8] Badrul Alam Miah, et al. "Detection of lung cancer from CT image using image processing and neural network," International Conference on Electrical Engineering and Information Communication Technology **(2015)**

AUTHORS

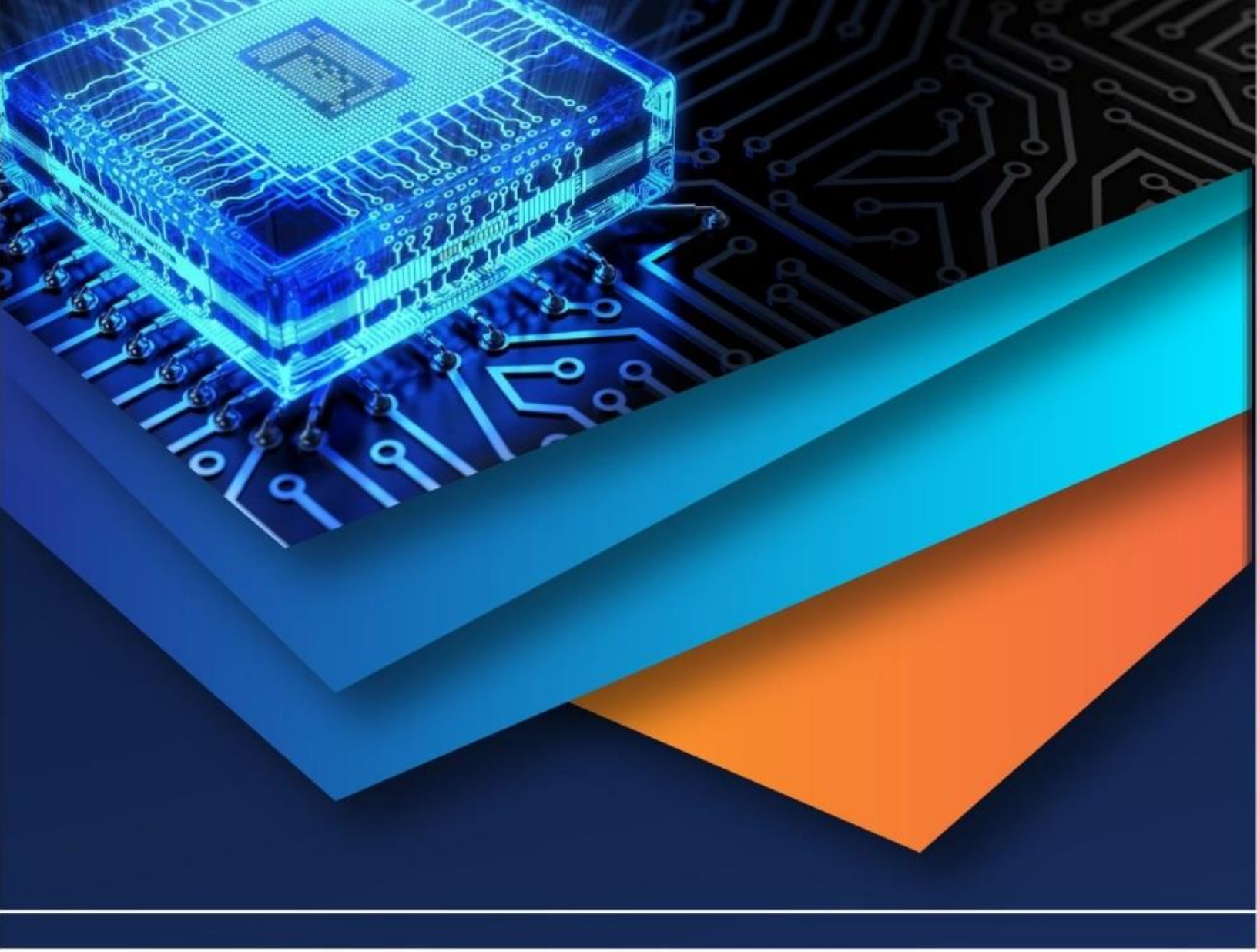


HENIL SATRA
B.TECH. CSE, VIT, VELLORE



ARYAN GUPTA
B.TECH. CSE, VIT, VELLORE

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)