IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# Detecting stage of lung cancer based on tumor size - By using segmentation and feature extraction in medical image processing

**S.Divya[*1], S.Mani Swaroop[*3],Ruqaiyya [*4], G.Sravya[*5]**

**Mrs. P. S. Mayura Veena (Ph.D.) *[2],**

[*1,3,4,5] Students, Department of Electronics and Communication Engineering,

[*2]Assistant Professor, Department of Electronics and Communication Engineering,

Anil Neerukonda Institute of Technology & Sciences,

Sangivalasa, Bheemunipatnam mandal, Visakhapatnam district-531162, Andhra Pradesh, India

***Abstract***

Cancer is an extensive global and universal disease nowadays which pretends to be the utmost cause for a large impermanence rate among men and women every era. Approximately 80-85% of the people who get affected by cancer are being succumbed to death. Recognition of cancer at the first stage is the only aspect in front of us to give proper treatment. Among numerous types of cancers, lung cancer is a very fearful and complicated one. Lung cancer means the growth of tumor cells briskly and having chances of spreading those cancer cells to other organs which in turn damaging other normal tissue cells of the body. Noticing tumor prematurely can help to cure the disease completely and it becomes pivotal to find out whether the tumor has been changed to cancer or not, if the prognostication is made at an initial stage, then countless lives that are at risk could be rescued and accurate prediction can help the doctors to start their treatment at the earliest. In this paper, we have proposed a simple, easy, and precise method for accurate prediction of the stage of cancer using CT images of the lungs in Image processing. For this process, a CT image will be considered, and then the image will be pre-processed for noise removal. Further segmentation is done to identify and separate desired tumor nodule and extraction of morphological features such as area, perimeter, eccentricity, and diameter is carried out under feature extraction. Finally, the classification of lung cancer into different stages based on the size of tumor results has been proposed using MATLAB which is more accurate and less time-consuming when compared to other lung cancer prediction systems. The method proposed in this paper to detect a tumor in the lungs is simpler when compared to applying other difficult algorithms.

***Keywords*** **- Lung cancer, Impermanence, pivotal, briskly, prognostication.**

## I. INTRODUCTION

Cancer is defined as a group of cells in the human body growing uncontrollably in a massive number leading to the development of tumors. There is a myth that cigarette is the main cause to evolve lung cancer. But the truth is all smokers may not have lung cancer and not all people having lung cancer smoke. Lung cancer may cause due to air pollution, inheritance where family background may have lung cancer, exposure to any harmful radiation gases, carcinogens, and of course smoking also[8]. Cancer cells from the breast, kidney, or any other organs can be carried away in the blood or the lymph fluid to the lungs. Suppose if cancer cells from the breast spread to the lungs, it is metastatic breast cancer, not complete lung cancer, but the lungs also get affected here. In this way, lung cancer can be caused due to numerous reasons.

Lung cancer varies differently from person to person, depending on the size of the tumor and the stage it is in. Stage I is considered as when the cancer is restricted to the lung[1]. Stage II is when the cancer is limited to the chest[1]. Stage III is when the tumor grows larger and appears in the CT scan. Stage IV is confined to spreading cancer cells to other parts of the body and growth of tumors in other parts as well[1]. Analyzing CT scan images of lungs and predicting the stage of cancer based on tumor requires a high level of skill and concentration, and is possible only by expert doctors or radiologists. CT stands for computerized tomography where the passing of X-rays inside the human body takes place. There are many other image processing methods and techniques such as MRI, Ultrasound, DEXA, X-ray, and PET, but CT scan is best recommended because of three main reasons[9]. One is due to CT scans can completely examine not only bones as like X-ray does but also soft organs like lung tissues. The second one is CT scanning is painless, cost-effective, accurate, fast, simple, less sensitive to patient movement, and X-rays used in standard CT scans have no immediate side effects. The third reason is Computed tomography (CT) images have better clarity, low noise, and distortion[1]. Thus, CT scan images are preferred in this paper and taken as input. CT scan images are downloaded from the Internet from the Cancer Imaging Archive database[1]. Here one CT image having tumor regions is taken in this paper. In the same way, any CT image can be taken and can be tested. As medical images may have noise, the CT image is passed to the second step, i.e., pre-processing where the median filter is used in this paper for the elimination of noise [8]. Then conversion of an image into binary followed by segmentation is done where CT image is partitioned into some sets of pixels and obtaining of tumor (white) region pixels i.e., extraction of the required and interested region (which are the white regional pixels in different places present in the lungs, eliminating the other part of the lung which is not affected takes place). By the end of this step, just some white pixel areas/regions grouped will be separated which particularly means that final declaration of whether it is a tumor or still few cancer cells grouped in less number which not yet developed into the tumor ( that means the cancer cells size still less than tumor size) cannot be predicted. But, after segmentation, for analyzing CT scan images, it became simpler and complexity was reduced. Various simple segmentation techniques such as morphological operations like dilation, erosion, opening to apply big mask were used in this paper[2]. The last step is feature extraction to decide whether those white regional pixels were still initial cancer cells or tumor is decided here and area, perimeter, centroid, eccentricity, and diameter of those white regional pixels are extracted and based on the radius value of the white grouped pixels, classification of lung cancer into various stages is decided in this paper. Based on a fixed eccentric value as prescribed by the medical standards, only regions having a greater eccentric value than the standard value will be considered as exact tumors and based on radius value of tumor, stages of cancer are determined and it becomes easier for the radiologists/doctors to easily find the stage of cancer instantly. Proper medication like radiotherapy or surgery can be done if the tumor's size (radius value is much higher than the standard medical value) is large.

## II.PROPOSED METHODOLOGY

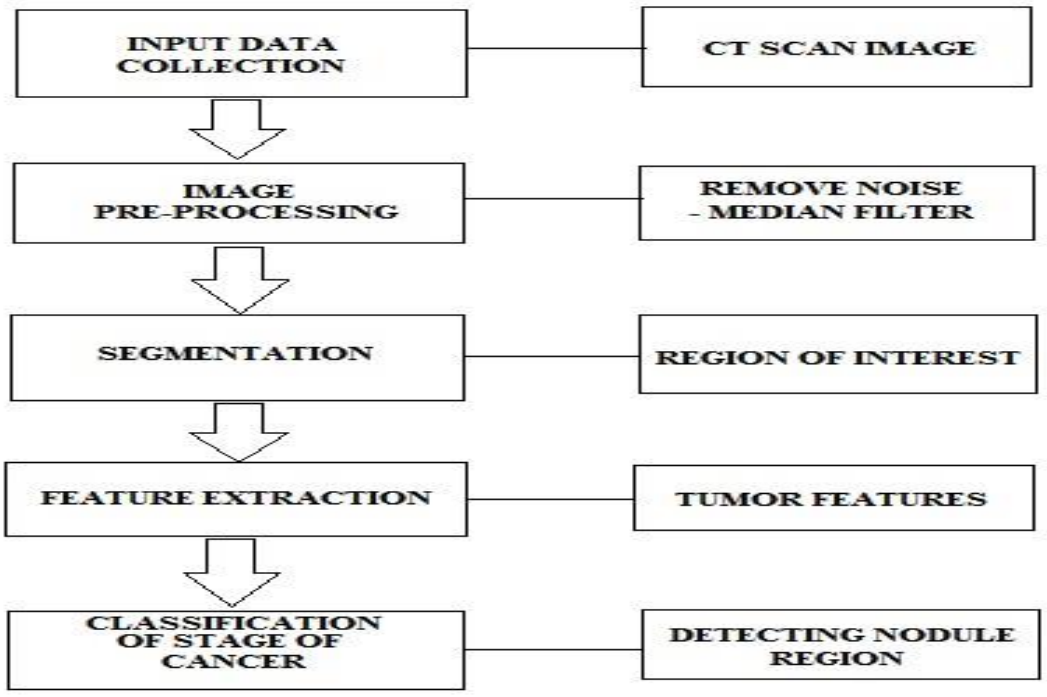


fig 1.) implemented methodology

## III. PRE-PROCESSING

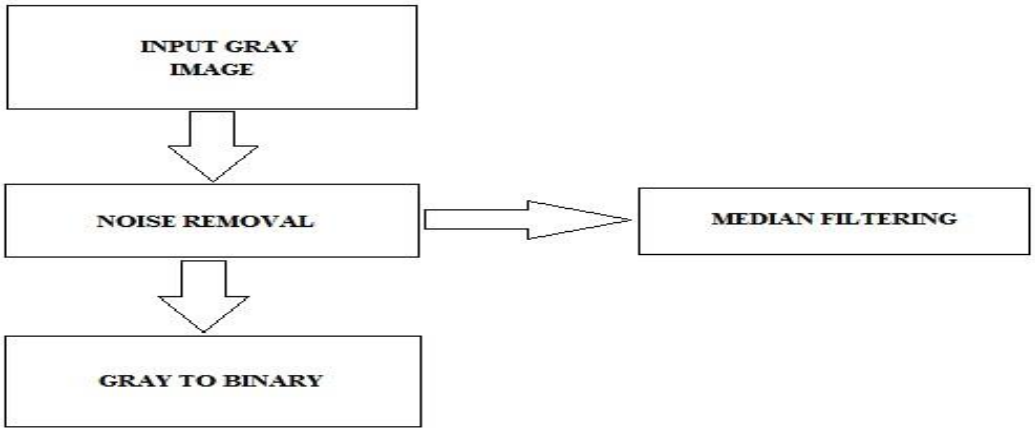


fig 2.) flow chart for pre-processing

Pre-processing is the second step to enhance image quality by suppressing the noise where the median filter is used in this paper. Generally, CT scan images have salt pepper noises(impulse noise) which can be effectively eliminated using the median filtering approach. As the median gives the mid-value of the pixels, no new extra pixels will be created and sharp edges of the CT scan image are preserved.[1] Then conversion of an image into binary using certain threshold took place because CT scan image contains different intensities of colors like gray, black, white and therefore to simplify the next steps, converting into two intensities i.e., into black and white took place here. Moreover, color, RGB, gray images may use more processing power, huge memory, and time to process. For this reason, images are transformed into binary for processing, which needs less computing power and takes less time for processing.

## IV. SEGMENTATION

The primary purpose of segmentation is to remove or eliminate the irrelevant parts of the images and to extract required white regional grouped pixel parts ( which may be a tumor or less than tumor size i.e., initial small cancer cells lump formation). The first dilation is performed here to add pixels of sufficiently small size to bridge gaps in the binary image, because due to noise if present, some white regional large size tumors may appear as individual different small size cancer cells which may be mistreated as not a tumor. Hence by dilation, the pixels in the object are thickened and if there are any separated pixels of the tumor region due to noise can be combined and will be appeared as a tumor. In the same way, dilation is also done to remove small pixels present if any, so at end of this erosion step, the noise regions will not appear anymore and only tumor region pixels or group of few cancer cells pixels are visible.

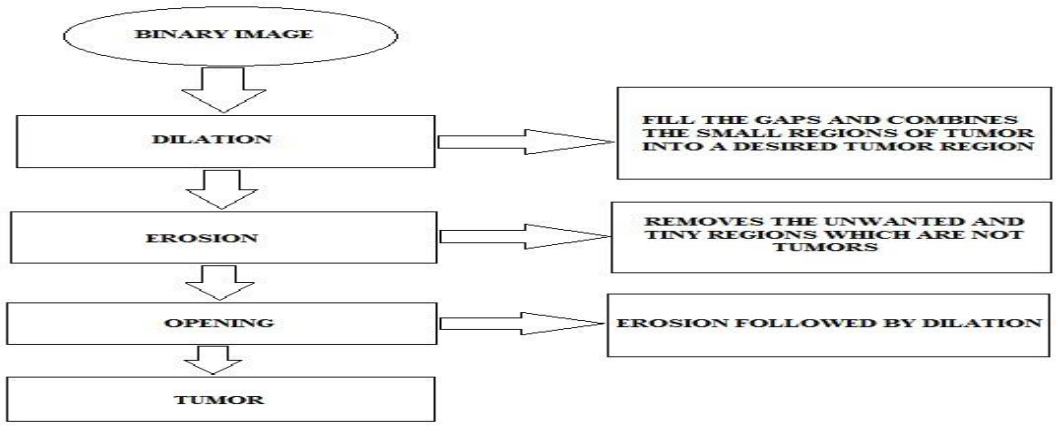


fig 3.) steps implemented in segmentation process

The next big mask is applied using the opening operation to fill the complete image and now subtracting or removing the eroded image from the big mask image, we got the exact tumors separated. But by the end of segmentation, we still cannot exactly tell whether those tumors are small(i.e., less gathering of cancer cells) or big (huge tumor). Hence feature extraction stage is necessary to classify cancer.

**MAIN TASK OF CREATING MASK**: To separate the tumor regions or group of small cancer cells still not yet developed into the tumor from the remaining area/region of lungs.

**Binarisation**: It is the process of converting a pixelated image into binary and the foremost thing is to select the threshold for binarisation. Based on formulas, the threshold is determined in this paper.

**Filling holes**: Also called dilation is done to gradually enlarge the boundaries of regions of foreground pixels (i.e. white pixels, typically) in a binary image.[6]

**Erosion**: It is done to erode the boundaries of regions of foreground pixels (i.e. white pixels, typically). Thus areas of foreground pixels shrink in size, and holes within those areas become larger. This step is important because it separates the lung nodules from the blood vessels. [6]

**Opening**: The next operation is an opening operation which is an erosion followed by a dilation. This step is done to keep the blobs connected to the walls of the lungs.

**Big mask**: Subtracting the eroded image from the big mask image gives the tumor region areas. Again by applying another opening operation, if the tumors detected having a size less than 55, then only those tumors are considered as exact tumors.

## V. FEATURE EXTRACTION

The ultimate objective of feature extraction is to take out required features from the target image. In this boundary tracing followed by labeling the number of tumors and then using region props, extracting regional properties or parameters like area, perimeter, centroid,eccentricity and diameter results are obtained and based on a single parameter value i.e., the radius of the tumor, the stage of the cancer is decided. Table I shows the parameters that were deduced from the extraction step. Table II shows the principle which was determined by the medical field radiologists and doctors for differentiating cancer stages based on the radius value ( R ) of the tumor [7].

table 1. parameters obtained

| REGION | AREA | PERIMETER | CENTROID | DIAMETER | ECCENTRICITY |
|---|---|---|---|---|---|
| 1 | 144.0 | 42.9 | 137.6 143.5 | 13.5 | 0.8 |
| 2 | 521.0 | 81.8 | 249.2 76.6 | 25.8 | 0.6 |

table 2. cancer stages classification based on radius value ( R )[7].

| STAGE OF CANCER | REQUIREMENT |
|---|---|
| I | R < 3 cm |
| II | 3cm < R < 7cm |
| III | R > 7cm |
| IV | R > 10 and above |

## VI. RESULTS

As stated in the preceding section, based on the stages of cancer data, various stages are determined for the tumor according to the radius value of the tumor. “Fig. 4” shows cancerous lungs CT scan image which was collected from Cancer Imaging Archive (CIA) database. The various experiments/processes proposed in the above sections for the lung cancer detection were implemented using matlab, which is necessary and suitable for better classification of the stage of cancer and accuracy in the process of prediction using segmentation and feature extraction. “Fig. 5 and 6” illustrates the preprocessing output results and “Fig. 7,8,9,10” shows the segmentation results and extraction of desired regions. “Fig. 11” exhibits numbering for the tumors detected and “Fig. 12” reveals the classification for the exact tumors detected.

The accuracy of the proposed model to detect the lung cancer stage is highlighted based on these results shown under.

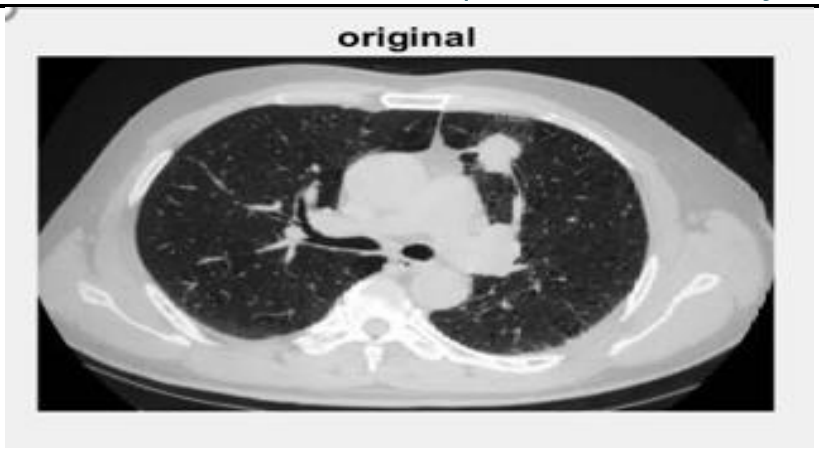


fig 4.) lung cancerous CT scan image

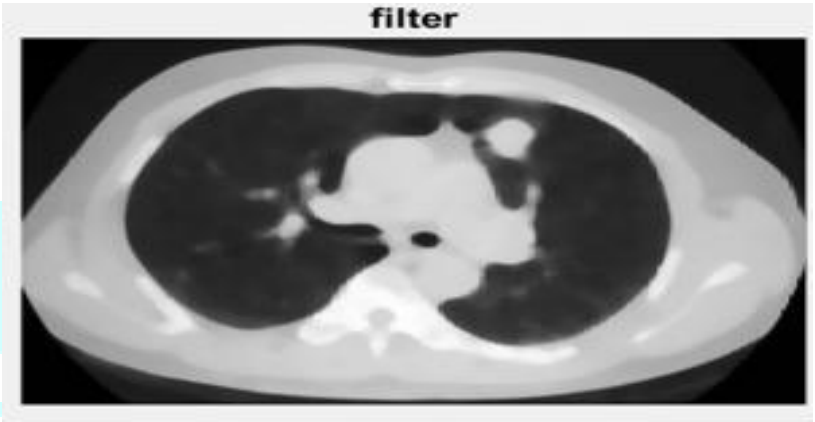


fig 5.) noise removal using the median filter

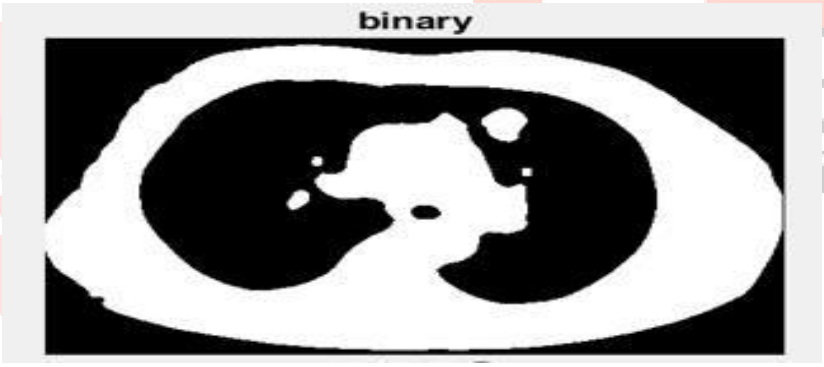


fig 6.) gray to binary converted image

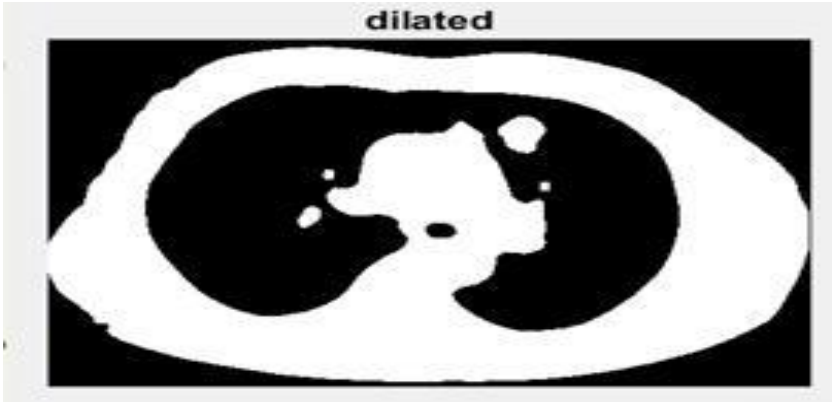


fig 7.) dilated image

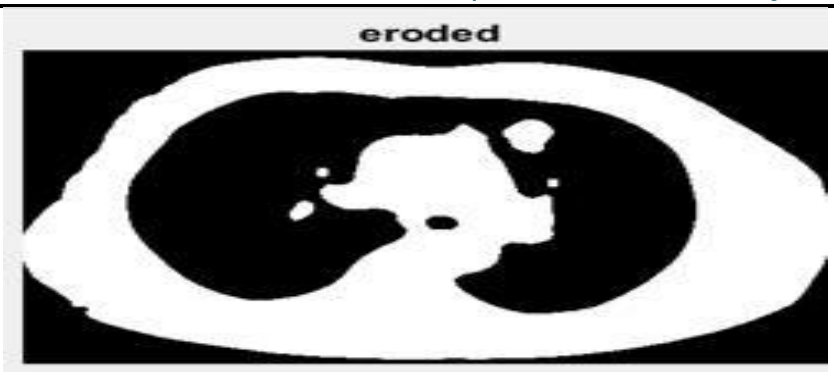


fig 8.) eroded image


Fig 9.) Big mask image

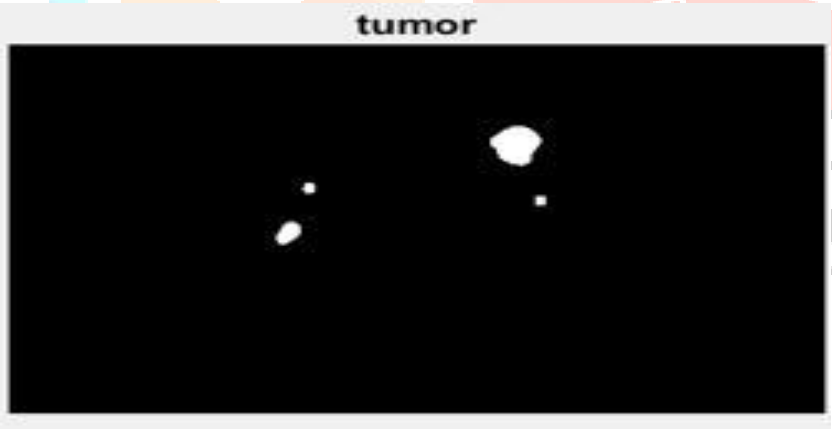


fig 10.) tumors or small lump of less sized cancer cells

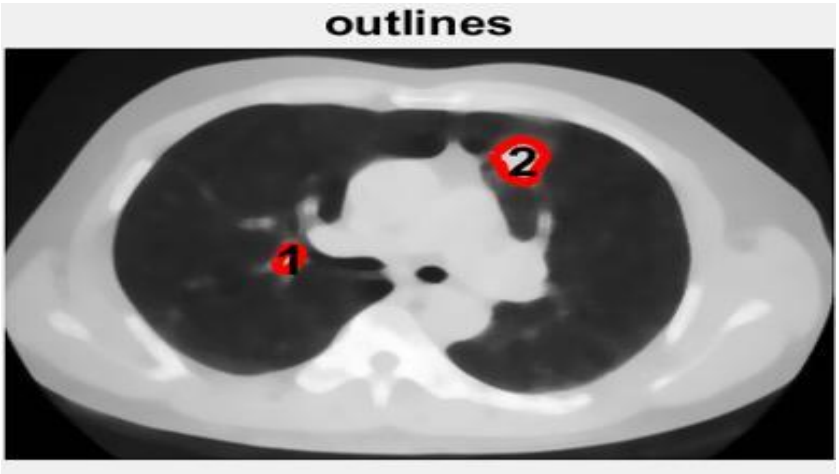


fig. 11.) outlines and numbering of tumors

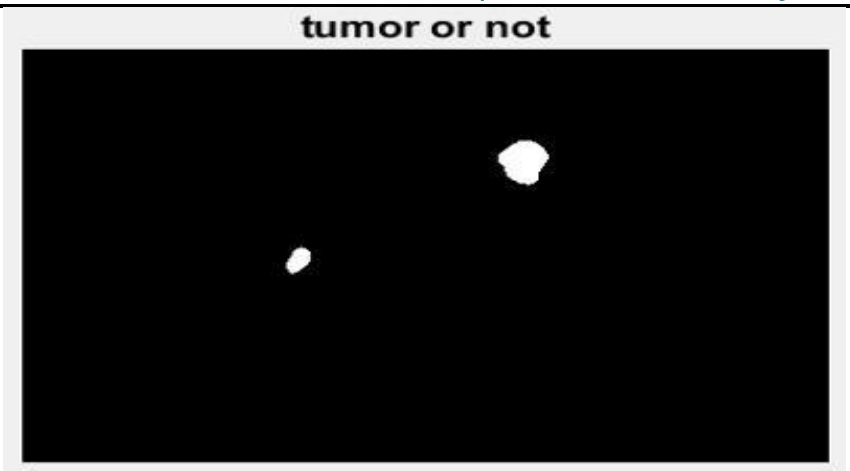


fig 12.) showing exact tumors

## VII. CONCLUSION

The study carried out in this paper entirely is about predicting the stage of cancer-based on any one of the four features extracted. From the results obtained it is observed that using the eccentricity value of the tumor part, the exact tumors are separated. The stages of cancer are classified based on the radius value of the tumor(radius value extracted from the diameter).According to the cancer stages classification table prescribed by the medical standards as mentioned above. A total of five features were extracted in this paper like area, perimeter, diameter, eccentricity, and centroid. Based on these outcomes, the result is an exact finding of cancer stage with accuracy. The motivation behind lung cancer identification is to help the radiologist and doctors to take an exact decision regarding the cancer stage instantly and this methodology can be used to prevent wrong predictions. The future scope could be to work on improving efficiency in diagnosis and prediction of the stage of cancer by any person simply by seeing CT scan report or without any assistance of doctors or radiologists.

## REFERENCES

[1] Nidhi S. Nadkarni and Prof. Sangam Borkar, “Detection of Lung Cancer in CT Images using Image Processing”, Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8.

[2] Diya Chudasama, Tanvi Patel, Shubham Joshi, and Ghanshyam I. Prajapati, " Image Segmentation using Morphological Images", International Journal of computer applications (0975-8887) Volume 117 – No.18, May-2015.

[3] Chiranji Lal Chowdhary, D.P. Acharjya, "Segmentation and Feature Extraction in Medical Imaging: A Systematic Review", International Conference on Computational Intelligence and Data Science (ICCIDS 2019).

[4] Gawade Prathamesh Pratap and R.P. Chauhan, “Detection of Lung Cancer Cells using Image Processing Techniques”, 1st IEEE International Conference on Power Electronics. Intelligent Control and Energy Systems (ICPEICES-2016).

[5] Parameshwar R. Hegde, Manjunath M. Shenoy, and B.H.Shekar, "Comparison of Machine Learning Algorithms for Skin Disease Classification Using Color and Texture Features", 978-1-5386-5314-2/18/$31.00 ©2018 IEEE.

[6] M. Bikromjit Khumancha, Aarti Barai, C.B.Rama Rao, “Lung Cancer Detection from Computed Tomography (CT) Scans using Convolutional Neural Network”, 10th ICCCNT 2019, July 6-8, 2019 – IIT Kanpur, India. ( IEEE – 45670).

[7] Mohd Firdaus Abdullah, Siti Noraini Sulaiman, Muhammad Khusairi Osman, Noor Khairiah A.Karim, Ibrahim Lutfi Shuaib, Muhamad Danial Irfan Alfamandu, "Classification of Lung Cancer Stages from CT Scan Images Using Image Processing and k-Nearest Neighbours", 2020 11th IEEE Control and System Graduate Research Colloquium ( ICSGRC 2020), 8 August 2020, Shah Alam, Malaysia.

[8] Nikita Benerjee, Subhalaxmi Das, " Prediction Lung cancer – in Machine Learning perspective", XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE Xplore.

[9] CMAK Zeelan Basha, B Lakshmi Pravallika, D Vineela, S Lakshmi Prathyusha, " An Effective and Robust Cancer Detection in the Lungs with BPNN and Watershed Segmentation", 2020 International Conference for Emerging Technology (INCET) Belgaum, India. Jun 5-7, 2020.

[10] Sanjukta Rani Jena, Dr. Thomas George, Dr. Narain Ponraj, "FEATURE EXTRACTION AND CLASSIFICATION TECHNIQUES FOR THE DETECTION OF LUNG CANCER: A DETAILED SURVEY", 2019 International Conference on Computer Communication and Informatics (ICCCI -2019), Jan. 23-25, 2019, Coimbatore, INDIA.

[11] Md. Badrul Alam Miah, Mohammad Abu Yousuf, " Detection of Lung Cancer from CT Image Using Image Processing and Neural Network", PID3702215.