iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** IV **Month of publication:** April 2023

DOI: https://doi.org/10.22214/ijraset.2023.50960

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: 7.538
Volume 11 Issue IV Apr 2023- Available at www.ijraset.com

Lung Cancer Diagnosis Using Deep Learning: VGG-19

Ms. P. Nava Bhanu[1], Puram Vijaya Lakshmi[2], Konka Sushma[3], Maddirala Nandu Priya[4], Mudigonda Pavani[5], Morla Anush[6]
[1]*M.Tech, (Ph.D)*
[2, 3, 4, 5, 6]*Dept of Computer Science & Engineering, BWEC, Bapatla, Andhra Pradesh, India*

Abstract: ***Lung cancer is classified histologically into small cell and non–small cell lung cancers. The most common symptoms of lung cancer are cough, dyspnoea, haemoptysis, and systemic symptoms such as weight loss and anorexia. High-risk patients who present with symptoms should undergo chest radiography. If a likely alternative diagnosis is not identified, computed tomography and possibly positron emission tomography should be performed. If suspicion for lung cancer is high, a diagnostic evaluation is warranted. The diagnostic evaluation has three simultaneous steps (tissue diagnosis, staging, and functional evaluation), all of which affect treatment planning and determination of prognosis. The main aim of this project is to find out the whether the person having lung cancer or not and having a chance of occurring Cancer in the future. We can done this process by using the CT(Computed Tomography) Scan Images of the lungs of affected Person. And it also identifies the given input CT Scan Image belongs to which person, which means that the image belongs to Normal person or it belongs to Cancer person or it belongs to the person who is having the chance of occurring cancer in the future, Which is also known as the Pre-Malignant Classification of images is done by using various CNN Algorithm.***
Keywords: Computed Tomography (CT), Malignant Nodule, Classifier, Data Sets.

I. INTRODUCTION

Lung Cancer is of two types. Lung cancer occurs when a malignant (cancerous) tumour grows inside the lungs, in structures such as the bronchi (small tubes that connect the wind pipe to the inner surfaces of the lungs where gas transfer takes place). Like many other types of cancer, lung cancer is capable of spreading (metastasizing) to other parts of the body. In this case, cancer beginning in the lungs most commonly spreads to the brain, bones, adrenal glands and liver, via any of three mechanisms: direct extension, via the blood vessels, or via the lymph system. Direct extension occurs when a tumour grows rapidly in size such that it begins to touch an adjacent organ or structure, and then begins to penetrate itself into that adjacent organ or structure. Tumour cells are also able to get into the blood and lymph circulatory systems and travel, one by one, to distant structures. Lung cancer is now the most prevalent form of cancer affecting Americans with an estimated 222,500 new cases every year, according to the American Cancer Society (ACS, 2010). Beyond being the most common form of cancer, lung cancer is also often difficult to treat. As a result, lung cancer is the most deadly cancer with roughly 160,000 Americans dying from it every year. This is about 30% of all cancer deaths! (ACS, 2010). Although lung cancer is difficult to treat and cure, it is for the most part preventable. Lifestyle choices can be made which can almost eliminate your risk for getting the disease. Your decision to stop smoking and to eat a healthy diet featuring plenty of fresh fruits and vegetables can greatly decrease your risk.

A. Motivation

The motivation of this lung cancer diagnosis is to lower a chance of dying from lung cancer, which accounts for many deaths in people who currently smoke or formerly smoked. Lung cancer affects both men and women with the average age being 70 years. Younger adults can still develop the disease with approximately 7% of lung cancer cases occurring in people under the age of 55. To reduce this Lung Cancer Diagnosis is introduced to know the people there is a chance of occurring lung cancer in the future. If they know there is a chance of occurring cancer they should started the treatment to cure lung cancer.

B. Objective

The main objective of this project is to improve and save lives. And to achieve a cure for your cancer, allowing you to live a normal life span. Now -a-Days many people are not allowed them to live a normal life span due to lung cancer. Mainly people , who are having the habit of smoking, facing the lung diseases. So their families also facing the difficulties in day-to-day life through financially and in all aspects[Health]. To help those families we take it (lung cancer diagnosis) as an objective and Motivation.

©IJRASET: All Rights are Reserved

C. Existing System

Some of the existing methods for lung cancer diagnosis are Imaging tests where a CT Scan on your lungs reveal abnormal mass or nodules. Sputum Cytology which analyses Phlegm (Mucus Cells). Biopsy which is the most invasive method where abnormal cells are removed from your body to be analysed. This method is done in a number of ways such as bronchoscopy in which an incision is made at your neck and surgical tools are inserted behind your breastbone to take tissue samples. Computer aided detection systems.
The Main Objective of this Existing System is to find out the early stage of lung cancer and explore the accuracy levels of various machine learning algorithms. After a systematic literature study, we found out that some classifiers have low accuracy and some are higher accuracy but difficult to reached nearer of 100%. Low accuracy and high implementation cost due to improper dealing with DICOM images. For medical image processing many different types of images are used but Computer Tomography (CT) scans are generally preferred because of less noise. Deep learning is proven to be the best method for medical image processing, lung nodule detection and classification, feature extraction and lung cancer stage prediction. In the first stage of this system used image processing techniques to extract lung regions. The segmentation is done using K-Means. The features are extracted from the segmented images and the classification are done using various machine learning algorithm. The performances of the proposed approaches are evaluated based on their accuracy, sensitivity, specificity and classification time algorithms. After a systematic literature study, we found out that some classifiers have low accuracy and some are higher accuracy but difficult to reached nearer of 100%. Low accuracy and high implementation cost due to improper dealing with DICOM images. For medical image processing many different types of images are used but Computer Tomography (CT) scans are generally preferred because of less noise. Deep learning is proven to be the best method for medical image processing, lung nodule detection and classification, feature extraction and lung cancer stage prediction. In the first stage of this system used image processing techniques to extract lung regions. The segmentation is done using K Means. The features are extracted from the segmented images and the classification are done using various machine learning algorithm. The performances of the proposed approaches are evaluated based on their accuracy, sensitivity, specificity and classification time.

D. Proposed System

In the proposed system we are going to use the VGG-19 Architecture for Classification of Lung Cancer and CNN algorithm is used for image Classification.in this proposed system we are going to train the images from the KAGGLE Data Set. Train Data Sets and input image are the two inputs for Training , CNN. This gives the output as the given CT Scan lungs image is Normal or Abnormal.if it is Normal then we know that there is no Cancer and chance of Occuring Cancer or if it is Abnormal then there is a chance of Cancer and already having the cancer.either it is Normal or Abnormal it given as input to the VGG-19 Model Fitting. It gives the final output by giving the image is belongs to which person. When given the input as image it gives the output like the given image belongs to which person.(like normal, malignant, and Pre-Malignant (the person having a chance of ocurring cancer in the future.))

II. LITERATURE SURVEY

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy n company strength. Once these things are satisfied, then next steps is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system.

A. Lung Cancer Detection Using CT Scan Images [2018]

In this Paper[1] Lung Cancer Detection Using CT Scan Images ,the proposed system is used to detect the cancerous nodules from the lung CT Scan Image using watershed segmentation for detection and SVM for classification of nodule as malignant or benign.in this project detects the cancer with 92% accuracy which is higher than current model and classifier model . but in this project does not classifies the different stages of the cancer i.e, Stage-I , Stage-II.

B. Lung Cancer detection system using Image Processing and Machine Learning Techniques

In this paper[2] we have studied that the given input CT Scan Image is Segmented first and next classifies the image .for segmentation purpose it uses the OTSU Algorithm and for classification purpose it uses the Decision tree and CNN Algorithms. In this paper overcome the challenges of preceding methods which are used in detection system and exploit the robust noise filtering methods using autoencoder System.

On the basis of truthful study the comparative statement on classification techniques are expressed in various section . The project given brief discussion over the how Lung Disease Diagnosis research is progressing to reach the fulfilment of medical field using robust R-CAD System.

C. Predictive Modelling of Lung Cancer Illness by Continuous Monitoring.

In this paper[3] Palani and Venkatalakshmi have given predictive modelling of lung cancer illness by continuous monitoring. They did this by using fuzzy cluster-linked augmentation with a categorization. The fuzzy clustering approach is essential to the production of accurate picture segmentation. We instead utilized the fuzzy *C*-means clustering approach in order to accomplish our goal of further disentangling the characteristics of the transitional area from those of the lung cancer image. In this particular investigation, the Otsu thresholding method was applied in order to distinguish the transition area from the lung cancer representation. In addition to this, the right edge picture is utilized in conjunction with the morphological, thinning procedure in order to improve the presentation of the segmentation. The current Association Rule Mining (ARM), the conventional decision tree (DT), and the CNN are combined with a novel incremental classification technique in order to accomplish classification in an incremental fashion. In order to carry out the operations, standard images from the database were utilized, as well as the most recent data on the patient's health collected from IoT devices that were attached to the patient. The culmination of the research indicates that the predictive modelling system has become more accurate.

D. Determination of a CT Picture Contains Lung Cancer or Not

In this paper[4] Deep residual learning was utilized by Bhatia et al. in order to develop a method for determining whether or not a CT picture contains lung cancer. The researchers have devised a preprocessing pipeline by making use of the UNet and ResNet models. This pipeline is intended to highlight and extract features from sections of the lung that are cancerous. An ensemble of XG Boost and random forest classifiers is used to gather predictions about the likelihood that a CT scan is malignant. The results of each classifier's predictions are then pooled, and the final result is used to determine the likelihood that a CT scan is malignant. The LIDC-IRDI has an accuracy that is 84 percent higher than that of typical techniques.

E. An effective method for lung cancer diagnosis from CT Scan using deep learning-based Support Vector Machine

In this paper [5] The diagnosis of Early Stage Lung Cancer is Challenging due to it asymptotic nature. Lungs are air filled organs thoracic cavity [Chest] main part of the human respiratory system. CNN Automatically Detect the lung nodules. Here LUNA16 Data set is used,which contains CT scanned images in DICOM format. The data set contains a total of 888 samples**.** The proposed deep learning-assisted SVM-based model yields pulmonary nodule detection representing early-stage lung cancer. It is found superior to other existing methods including complex deep learning, simple machine learning, and the hybrid techniques used on lung CT images for nodule detection. Support Vector Machine algorithm has been used to classify or diagnosis multiple cancers based on a protein chip. The main purpose of Support Vector Machine is used to detect the presence of lung cancer in CT images.

III. SYSTEM MODEL

A. Data Exploration

Under the Data Exploration we are taking the KAGGLE Data Sets. Kaggle allows users to find and publish data sets, explore and build models in a web-based data-science environment, work with other data scientists and machine learning engineers, and enter competitions to solve data science challenges. From the KAGGLE Data Set we are taking the 1120 images. Belongs to the three different classes those are Normal, having a chance of occurring cancer in the future, Malignant.
https://www.kaggle.com/hamdallak/the-iqothnccd-lung-cancer-dataset

B. Modules

There are three modules in this project they are Training the data sets, Image classification using CNN algorithm and identifying the image belongs to which class.

1) *Training the Data Sets:* For training the Data Sets we are going to use the KAGGLE Data Sets. We are using it because Every day a new dataset is uploaded on Kaggle. Each dataset is a small community where one can discuss data, find relevant public code or create your projects in Kernels. Sometimes, you can also find notebooks with algorithms that solve the prediction problem in a specific dataset and the data science frameworks used for Kaggle competitions are surprisingly effective for similar real-life problems.

2) *Image Classification:* For image classification A Machine learning algorithm CNN [2] is used.it classifies the Images whether it is Normal or Abnormal.
3) *Identifying the Image Belongs to Which Class:* By fitting the model of VGG-19 Is used to identify the image belongs **to** which class (Norma, having a chance of occurring cancer in the future[Pre -Malignant], Malignant).

C. Algorithms and Techniques

For Image Classification Purpose we are using CNN. CNN is a type of deep learning model for processing data that has a grid pattern. CNN is a mathematical construct that is typically composed of three types of layers (or building blocks): convolution, pooling, and fully connected layers. The first two, convolution and pooling layers, perform feature extraction, whereas the third, a fully connected layer, maps the extracted features into final output, such as classification. As one layer feeds its output into the next layer. In this Project the CNN Algorithm is used to classify the image belongs to either Normal or Abnormal.

1) VGG-19 Architecture

VGG-19 is a convolutional neural network that is 19 layers deep. You can load a pre-trained version of the network trained on more than a million images from the ImageNet database. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. The network has an image input size of 224-by-224.

The network has 47 layers. There are 19 layers with learnable weights: 16 convolutional layers, and 3 fully connected layers. After every block, there is a maxpool layer that decreases the size of the input image by 2 and increases the number of filters of the convolution layer also by 2. The dimensions of the last three dense layers in block 6 are 4096, 4096, and 1000 respectively. Version of the network trained on more than a million images from the imagenet database.

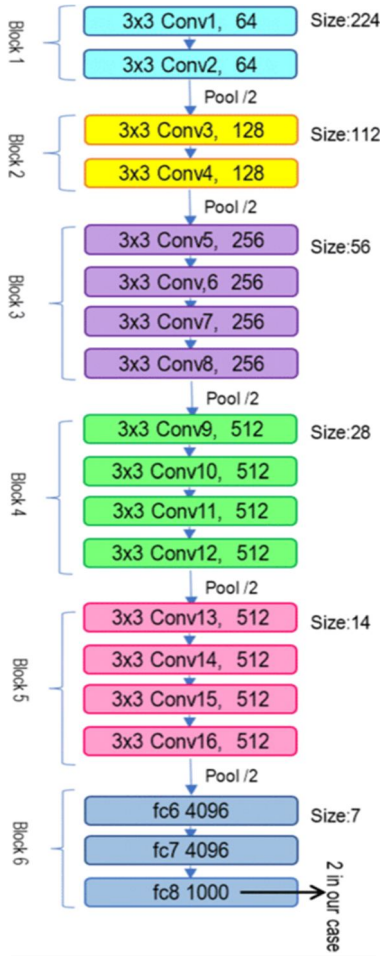


Fig 3.3.1 Architecture of VGG-19

IV. RESULTS AND ANALYSIS

Implementation of Training the Data Set. The training of normal case, Pre-Malignant Case and Malignant case is done in this section. For Training purpose we can take the Data Sets from the KAGGLE website. Because every day the data set is updated with the new data sets. The lungs which have the adenocarcinoma cells having the chance of occurring Cancer in the future. Large cell carcinoma and Squamous cell carcinoma are the cells which shows that the person has the Lung cancer. Normal Lungs Shows the Normal Case.

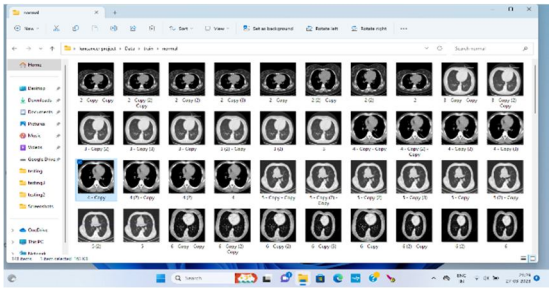


Fig 4.1: Training the Normal cases.

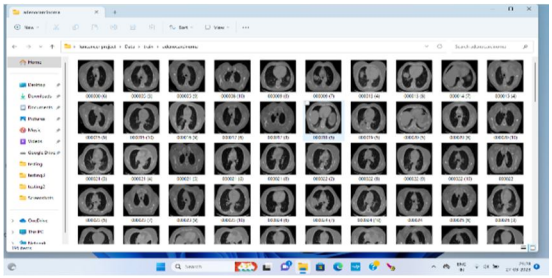


Fig4.2: Training the cancer ocurring in the future.[Pre-Malignant].

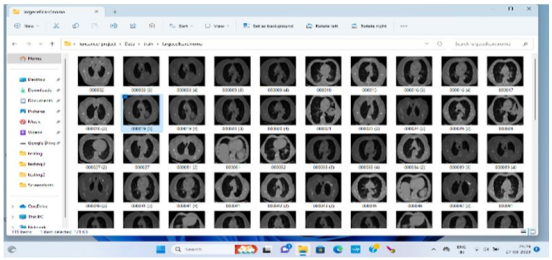


Fig4.3: Training the Malignant cases.

Implementation of identifying the image belongs to which class. Here in this Section we can identify the Cases of the Lung Cancer. There are totally three cases which are Normal Case ,Malignant Case and Pre-Malignant case. Where as Normal Case Means person does not have the Lung Cancer .Malignant Case Means Person have the Lung Cancer. And Pre-Malignant Means person having the chance of Occuring cancer in the future.

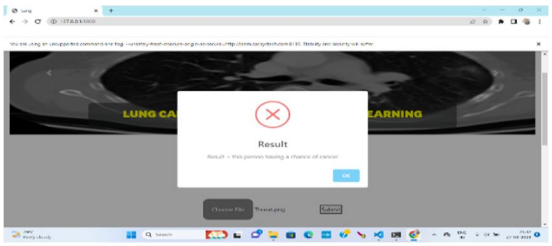


Fig 4.4: Identifying the image belongs to the Pre-Malignant Case (person having the Lung cancer in the future.)

Here in this picture the Pre-Malignant Case is Identified.

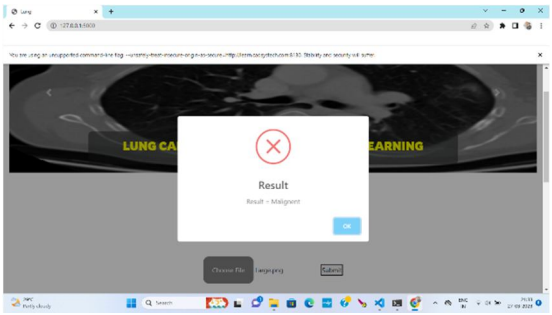


Fig 4.5 Identifying the image belongs to the malignant case.

Here in this Screen the malignant Case is identified. Malignant means person already having Lung cancer. For this purpose we train the Large cell carcinoma and Squamous cell carcinoma Data Sets.

V. CONCLUSION

The death rate can be decreased by the diagnosis of early lung cancer. But there is lot of difficulty in the discovery of early lung tumors classification. In this paper, a diagnosis of lung cancer using a hybrid deep neural network with adaptive optimization procedure has been proposed. The main goal of this is to distinguish the malignant lung nodules and to categorize the lung tumour whether it is malignant or benign. The different processes like preprocessing, segmentation, feature extraction, feature selection and classification are performed for the lung cancer detection.

VI. FUTURE WORK

In the future, the proposed approach will be extended by the diagnosis of lung cancer using advanced methods and algorithms in less time complexity.

REFERENCES

Here some of the references we take for this project reference.

[1] Smita Raut1, Shraddha Patil2, Gopichand Selke", Lung Cancer Detection using Machine Learning Approach", International Journal of Advance Scientific Research and Engineering Trends(IJASRET),2021.

[2] N.Camarlinghi, "Automatic detection of lung nodules in computed tomography images: Training and validation of algorithms using public research databases", Eur. Phys. J. Plus, vol. 128, no. 9, p. 110,Sep. 2013.

[3] R. L. Siegel, K. D. Miller, and A. Jemal, ``Cancer statistics, 2016", CA,Cancer J. Clin., vol. 66, no. 1, pp. 730, 2016. http://modelheelephant.blogspot.com/2017/11/detecting-and-classifying-nodulesin.html,2017

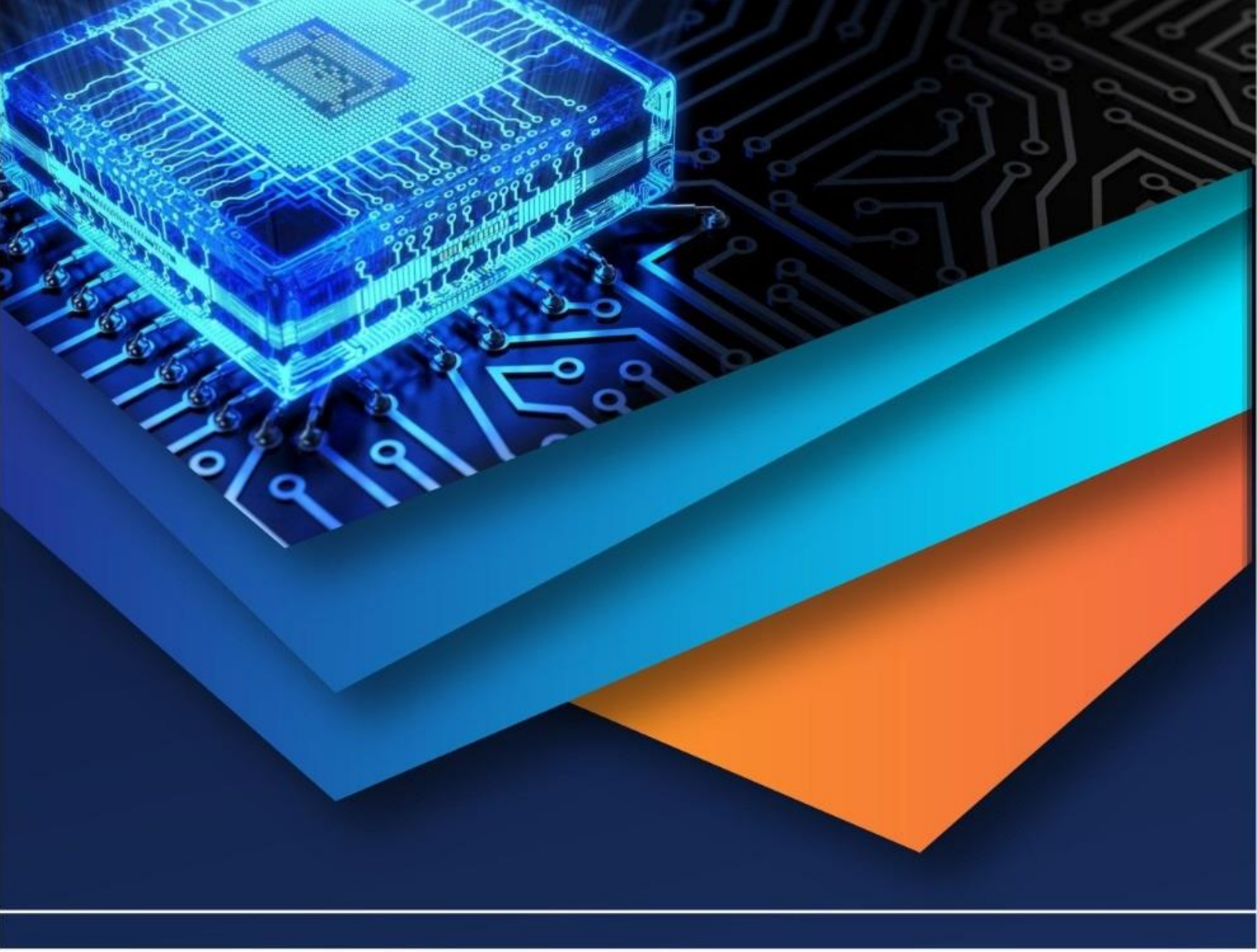
[4] Diego Riquelme and Moulay A. Akhloufi ,"Deep Learning for Lung Cancer Nodules Detection and Classification in CT Scans",www.mdpi.com,2020.

[5] Anita Chaudhary, Sonit Sukhraj Singh,"Lung Cancer Detection on CT Images by using Image Processing",IEEE,2012.

[6] Gawade Prathamesh Pratap, R.P. Chauhan, "Detection of Lung Cancer Cells using Image Processing Techniques", International Conference on Electronics, Intelligent Control and Energy Systems(ICPEICES),2016.

[7] Pooja R. Katre ,Dr. Anuradha Thakare,"Detection of Lung Cancer Stages using Image Processing and Data Classification Techniques",International Conference for Convergence in Technology,IEEE,2017

[8] Rituparna Sarma, Yogesh Kumar Gupta "A comparative study of new and existing segmentation techniques", ICCRDA, 2020.



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)