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.2021.38245

Lung Cancer Detection using Image Processing Techniques and its Classification

Aishwarya. R[1], Kushala. H[2], Shreesha Hathwar. GN[3], Flovia Mary Xavier[4], Prof. Sirasappa Y Pattar[5]
[1, 2, 3, 4]Department of Medical Electronics, BMS College of Engineering, Bangalore, India
[5]Professor, Department of Medical Electronics, B M S College of Engineering, Bangalore, India

Abstract: Lung cancer has been a major contribution to mortality rates world-wide for many years now. There is a need for early diagnosis of lung cancer which if implemented, will help in reducing mortality rates. Recently, image processing techniques have been widely applied in various medical facilities for accurate detection and diagnosis of abnormality in the body images like in various cancers such as brain tumour, breast tumour and lung tumour. This paper is a development of an algorithm based on medical image processing to segment the lung tumour in CT images due to the lack of such algorithms and approaches used to detect tumours. The work involves the application of different image processing tools in order to arrive at the desired result when combined and successively applied. The segmentation system comprises different steps along the process. First, Image pre-processing is done where some enhancement is done to enhance and reduce noise in images. In the next step, the different parts in the images are separated to be able to segment the tumour. In this phase threshold value was selected automatically. Then morphological operation (Area opening) is implemented on the thresholded image. Finally, the lung tumour is accurately segmented by subtracting the opened image from the thresholded image. Support Vector Machine (SVM) classifier is used to classify the lung tumour into 4 different types: Adenocarcinoma(AC), Large Cell Carcinoma(LCC) Squamous Cell Carcinoma(SCC), and No tumour (NT).

Keywords: Lung tumour; image processing techniques; segmentation; thresholding; image enhancement; Support Vector Machine; Machine learning;

I. INTRODUCTION

Cancer can be understood as uncontrolled cell growth having the ability to spread all over the body. Our body contains red blood cells (RBC's). Its main function is to supply fresh oxygen (O2) to all parts of the body with the help of blood flow. It is for this very reason why blood appears to be red in colour. In the lungs, tissue receives oxygen (O2) because of RBCs only. Lung cancer, also known as carcinoma is a malignant lung tumour which is an uncontrolled cell growth in tissues of the lung. If this cancer is not diagnosed, this growth penetrates beyond the lung by the process of 'metastasis' and is spread from one part of the body to another. The mortality rate of lung cancer is the highest among all other types of cancer. Lung cancer is one of the most widespread cancers in the world, with the least survival rate, with an increase in the number of deaths every year. Survival from lung cancer is directly related to its growth at its detection time. People do have a chance of survival if the cancer is detected in the early stage itself. Lung cancer is a disease where abnormal cells multiply and grow into a tumour. Cancer cells can be moved from the lungs in blood, or lymph fluid that surrounds lung tissue. Lymph flows through lymphatic vessels, which drain into lymph nodes located in the lungs and in the centre of the chest. Lung cancer usually gets concentrated toward the centre of the chest because the flow of lymph out of the lungs is toward the centre of the chest. Metastasis occurs when a cancer cell leaves the site where it began and moves into a lymph node or to another part of the body through the bloodstream. Cancer that starts in the lung is called primary lung cancer. There are several different types of lung cancer, and these are divided into two main groups: small cell lung cancer and non-small cell lung cancer which has three subtypes: Carcinoma, Adenocarcinoma and Squamous cell carcinomas. The most common symptoms of lung cancer include coughing (including blood coughing), weight loss, shortness of breath, and chest pains ranging from minute to severe. These cancer cells disturb the production and degradation of RBCs. Internally they change the structure and composition of plasma membrane i.e. the outer structure of the cell in such a way that these RBC's do not die as per the corresponding lifetime. As a result, the RBC count goes on increasing and a greater number of cells gets accumulated resulting in shortening of veins and arteries and ultimate bursting. This results in blood through cough etc. There are generally four stages of lung cancer; I to IV. These stages are based on tumour size and tumour and lymph node location. Presently, CTs are said to be more effective than plain chest x-ray in detecting and diagnosing lung cancer. The earlier the detection is, the higher the chances of successful treatment. An estimated 85% of lung cancer cases in males and 75% in females are caused by cigarette smoking.

Image processing is used to analyse images even if it is of the lowest or lowest quality. These operations do not increase probability of image information content, but they decrease it if entropy is an information measure. The main requirement of image processing is to improve pixel intensity by converting from discrete to digital image, segmenting to pixels, carrying out mathematical operations on pixels, and modifying images to a better quality. Pre- processing of CT images is the initial step in image processing followed by a segmentation process and ends with some morphological operations that are applied to detect the cancer spots/cells in the image. Also, it can be used to determine the amount of spreading of cancer i.e., what percentage of lung is affected with cancer. This paper has different sections with detailed explanation. Section II being the objective followed by section III which gives an overview of the literature survey done before implementation of the methods. Section IV is the methodology of our proposed system with all the steps explained in detail under it followed by section V, results and discussions and finally the conclusion under section VI.

II. OBJECTIVE

As we know the process of detection and classification of lung tumours can be automated. Our aim is to develop a method to detect the cancerous part and segment it out from the lung CT and classify the tumour into 4 different classes: Adenocarcinoma (AC), Large Cell Carcinoma (LCC) Squamous Cell Carcinoma (SCC), and No tumour (NT).

III. LITERATURE SURVEY

In [1], the proposed methodology has five phases. In phase one, lung cancer and non-lung cancer, images are collected from the lung cancer database. In phase two pre-processing is done by using the Median filter. Median filter is chosen as it preserves the edges. In Phase three, segmentation of the target image is done using Fuzzy C Means. Fuzzy C Means Clustering gave a better performance than K-means Clustering. In phase four, the features are extracted using GLCM (Gray Level Co-occurrence Matrix). In phase five, these extracted features are given to the SVM classifier for classification of lung cancer from normal lung. The SVM classifier achieved accuracy of 96.7% for detecting and classification of lung cancer.

In [2], an automated Computer Aided Diagnosing (CAD) system is proposed for detection of lung cancer from the analysis of computed tomography images. This system generally first segments the area of interest (lung) and then analyses the separately obtained area for nodule detection in order to diagnose the disease. Image processing techniques such as Erosion, Median Filter, Dilation, Outlining, and Lung Border Extraction are applied to the CT scan image in order to detect the lung region. Then the segmentation algorithm is applied in order to detect the cancer nodules from the extracted lung image. After segmentation, rule-based technique is applied to classify the cancer nodules. Finally, a set of diagnosis rules are generated from the extracted features.

In [3], a lung cancer detection system using image processing is used to classify the presence of lung cancer in CT- images. Here, MATLAB has been used in all the steps. In image processing procedures, processes such as image pre-processing, segmentation and feature extraction have been used. Various techniques are aimed to be used for segmentation.

In [4], An algorithm based on image processing is developed to segment the lung tumour in CT images due to the lack of such a method. This work involves the use of various image processing tools that give the desired results when all these tools are combined. Different stages are involved in order to reach the segmented result. First, Image pre-processing takes place where the image is enhanced and noise is reduced. In the next stage different parts in the images are separated to be able to segment the tumour later. In this stage, threshold was selected automatically. Finally, the lung tumour is segmented by subtracting the thresholded and the other image.

In [5], This paper proposes a lung cancer detection and prediction algorithm using multi-class SVM (Support Vector Machine) classifiers. Multi-stage classification was used for the detection of cancer. This system also predicts the probability of lung cancer. In every stage of classification, image enhancement and segmentation have been done separately. Image scaling, colour space transformation and contrast enhancement have been used for image enhancement. Threshold and marker-controlled watershed-based segmentation has been used for segmentation. For classification purposes, SVM binary classifiers were used. Our proposed technique shows higher degree of accuracy in lung cancer detection and prediction

In [6], This survey focuses on different techniques used to detect and classify the lung nodules which in turn will assist the domain experts for better diagnosis. Among many imaging modalities Computed Tomography (CT) being the most sought after because of its high resolution, isotropic acquisition which helps in locating the lung lesions. Since the volume of the CT scans are very large, Computer Aided Detection/Diagnosis (CAD/x) has more advantages in addition to manual interpretation with respect to speed and accuracy. This paper attempts to summarize various methods that have been proposed by several authors over the years of their research.

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

In [7], The main aim is to evaluate the various computer-aided techniques, analysing the current best technique and finding out their limitations and drawbacks and finally proposing the new model with improvements in the current best model. The method used was that lung cancer detection techniques were sorted and listed on the basis of their detection accuracy. The techniques were analysed on each step and overall limitation, drawbacks were pointed out. It is found that some have low accuracy and some have higher accuracy but not nearer to 100%. Therefore, our research targets to increase the accuracy towards 100%.

IV. METHODOLOGY

The block diagram of the proposed system is as follows:

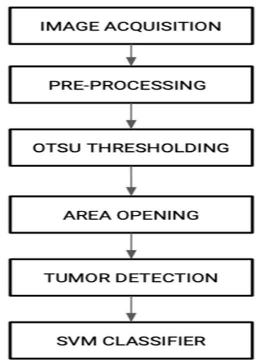


Fig 1. Methodology of the proposed system

A. Image Acquisition

Lung CT images from our dataset are given as input. Our dataset contained 613 Lung CT images which were classified into the four types. We used these two sample images to get the output and to check the working of the algorithm. As we can see the images have a lot of noise and can be a hindrance in getting the desired output. Thus, step by processes is done to get the desired output which is the detected tumour. These are the images which are fed into the classifier too and thus need to be processed in the best way to get accurate results/classification.

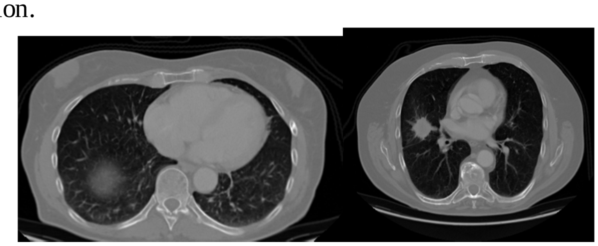


Fig 2. Images from the dataset

B. Pre-Processing

Pre-processing of the input CT images is an important process in the imaging processing algorithm. As we can see in the input CT images, there is a lot of salt and pepper noise which needs to be removed. Before doing that, we resize the image to a desirable value/ pixel and also convert the image into a grayscale image.

1) *Image Erosion:* Erosion is one of two fundamental operations in morphological image processing from which all other morphological operations are based. Erosion removes pixels on both operations. Morphological opening is useful for object boundaries.

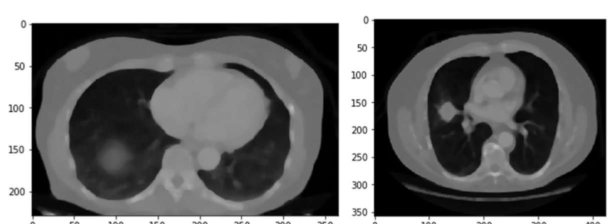


Fig 3. Images after Erosion

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

2) *Image Smoothening:* The median filter is a non-linear digital filtering technique, often used to remove noise from an image or signal. We used a Median filter(9x9) to smoothen the eroded image.

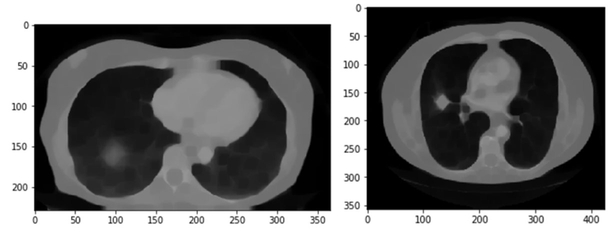


Fig 4. Images after applying Median filter

C. Otsu Thresholding

threshold value where the sum of foreground and background spreads is at its minimum. The images were then threshold by comparing Otsu's thresholding method that involves iterating through all the possible threshold values and calculating a measure of spread for the pixel levels on each side of the threshold, i.e., the pixels that either fall in foreground or background. The aim is to find the g pixel values to threshold values.

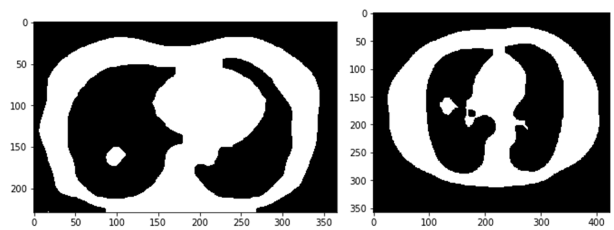


Fig 5. Thresholded Images

D. Area Opening

The opening operation erodes an image and then dilates the eroded image, using the same structuring element for both operations. Morphological opening is useful for removing small objects from an image while preserving the shape and size of larger objects in the image. We used a 3x3 kernel which we got by trial-and-error method.

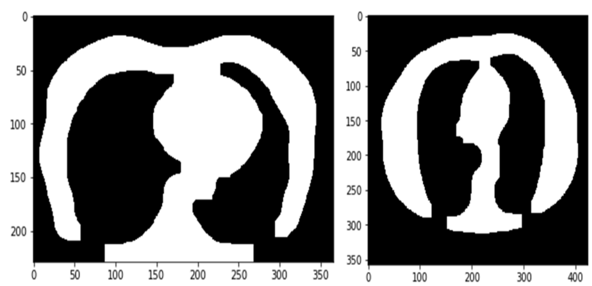


Fig 6. Opened Images

C. Tumour Detection

The opened image was subtracted from the thresholded image to obtain the cancerous part.

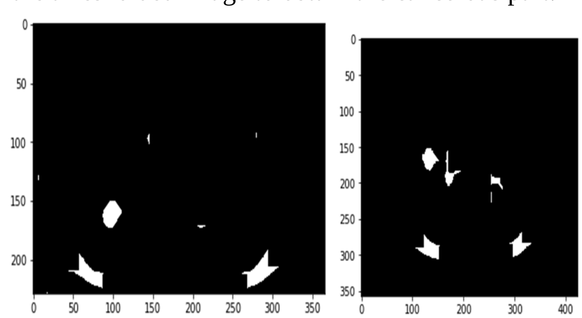


Fig 7. Images with detected cancer

D. Support Vector Machine

Support Vector Machine (SVM) is a simple Supervised Machine Learning Algorithm used for classification and/or regression. It is more preferred for classification but is sometimes very useful for regression as well. SVM basically finds a hyper-plane that creates a boundary between the types of data. In 2-dimensional space, this hyper-plane is nothing but a line. At first approximation SVMs find a separating line (or hyperplane) between data of two classes. SVM is an algorithm that takes the data as an input and outputs a line that separates those classes if possible. For multiclass classification, the same principle is utilized after breaking down the multi classification problem into multiple binary classification problems.

Our Dataset contains 613 images which were trained using this model, out of which 80% was train data and 20% test data. After the model/classifier gets trained, using logistic regression we predict the probability of an image to be part of a particular class by testing. After that, a single CT image is taken as the input and the class it belongs to is predicted by the model. There are four classes namely, Adenocarcinoma (AC), Large Cell Carcinoma (LCC) Squamous Cell Carcinoma (SCC), and No tumour (NT) and the class which the input image belongs will be assigned the highest percentage of probability and therefore it is concluded that the image is of that particular class. We got an accuracy of 86% on our classifier and this was measured using the confusion matrix. In Fig 8., the first image gave the highest percentage for LCC and the second gave the highest percentage for SCC therefore they are classified as Large Cell Carcinoma (LCC) and Squamous Cell Carcinoma (SCC), respectively.

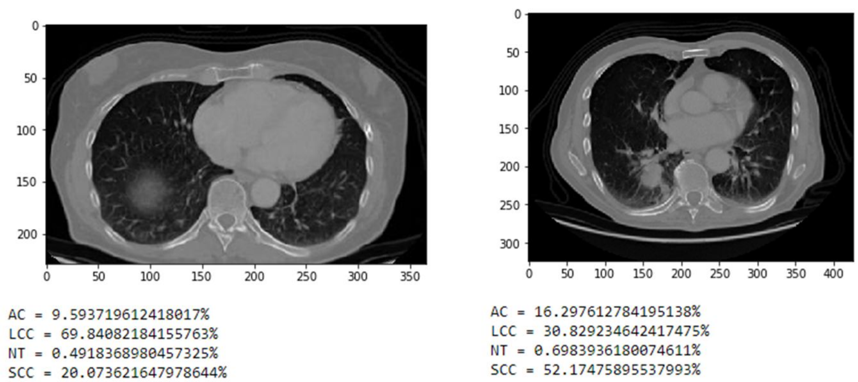


Fig 8. Results after classification using SVM.

V. RESULTS AND DISCUSSIONS

This study is to develop a new image processing approach for the segmentation of the lung tumour compared to other studies. In this study different image processing techniques are used. Firstly, images were enhanced by using image smoothing techniques like erosion and median filtering. Then, the tumour is segmented by using Otsu thresholding. This technique was used because of the difference in Gray-levels intensities of images which require different thresholds for different images. The useless connected pixels were removed from the threshold image to make it ready for the last technique. The last step is the extraction of the tumour which is the main part of the system. The threshold images, which contain tumour, and other lung structures and the other images which only contain lung structures without tumour. This non tumour image which is clean was subtracted from the threshold images in order to extract the tumour. The tumour was successfully detected using this sequence of techniques.

Many challenges were encountered when developing this system. One of them is the selection of the right technique for the final tumour extraction. Thresholding was first selected to be used in which different thresholds were used to segment different areas of the image. This technique had some advantages; however, some tumours were not segmented using it. In addition, some images were detected to have tumours but actually they don't which decreases the accuracy and sensitivity of the system. The reason for this incorrect segmentation was the difference in intensities of tumour and other parts in each image. Thus, Otsu's thresholding was used where the threshold is set automatically depending on each image intensity. This technique was definitely good in segmenting an image, by keeping the tumour and clavicle in an image since they have almost the same intensities. Therefore, another technique was used to remove the tumour from the images keeping only the clavicle. After that the two images were subtracted and the tumour was segmented. Experimentally, the developed approach showed a great efficiency in segmenting tumours in a CT image using the above discussed technique. This developed lung tumour segmentation approach outperforms many of the researches listed in the state of art in terms of accuracy, sensitivity, and specificity.

SVM classifier was used to classify the tumour into four types i.e., Adenocarcinoma (AC), Large Cell Carcinoma (LCC) Squamous Cell Carcinoma (SCC), and No tumour (NT). We got an accuracy of 86% on our classifier and this was measured using the confusion matrix.

VI. CONCLUSION

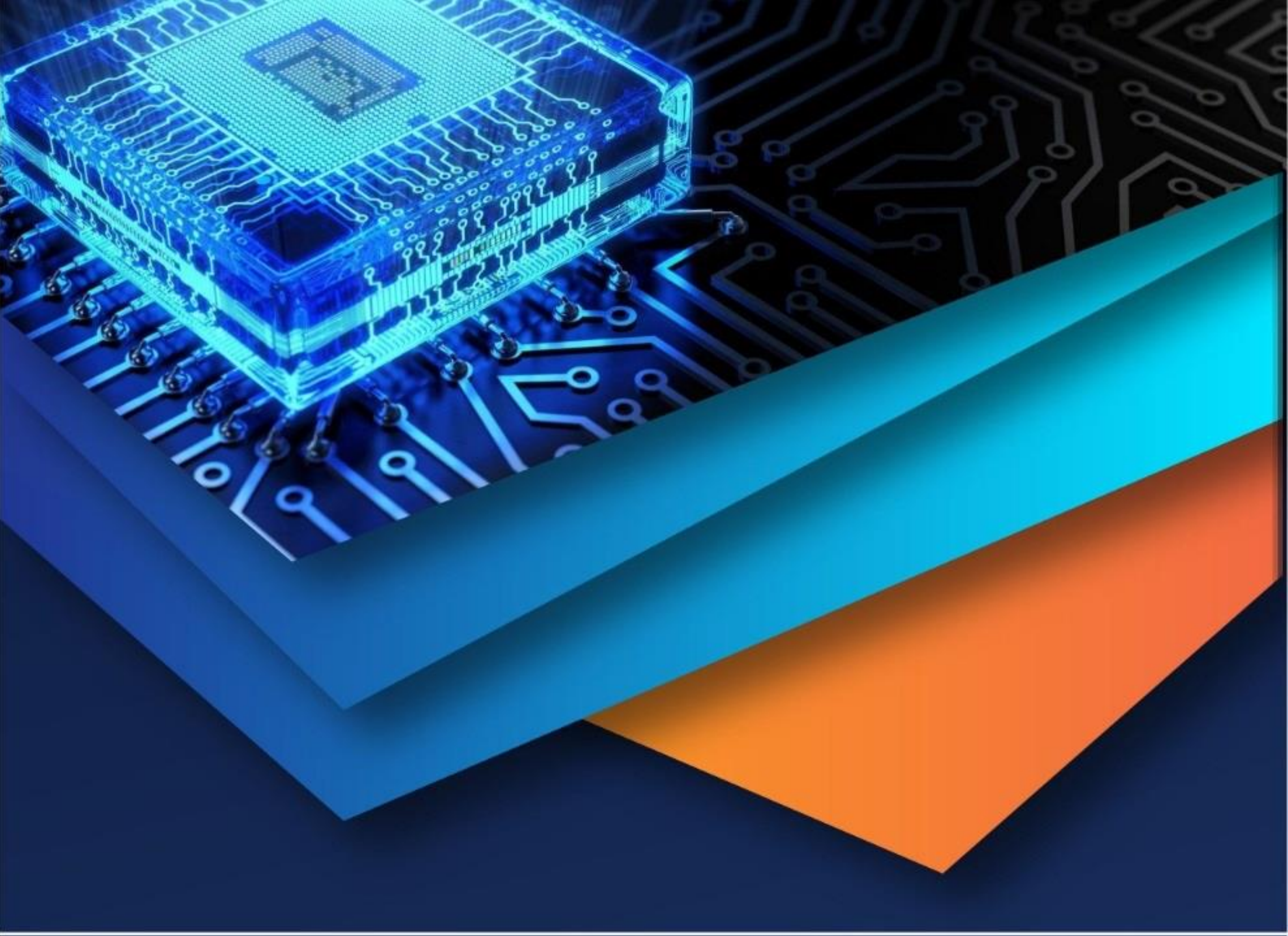
The lung tumour was fairly detected after using our proposed methodology. A total of 613 images were trained using the SVM classification and regression model. We were able to classify the tumour into 4 types: Adenocarcinoma (AC), Large Cell Carcinoma (LCC) Squamous Cell Carcinoma (SCC), and No tumour (NT). We designed an algorithm that gave 86% accuracy. This process makes the process of detection and classification easier using image processing techniques.

A. Future Scope

SVM is not very ideal for multiclass classification. However, we got the desired results using SVM. Further to improve the accuracy, we can use Convoluted Neural Network (CNN) which is a more accurate method for multi class classification.

REFERENCES

[1] Ankita, R., et al. "Lung cancer image-feature extraction and classification using GLCM and SVM classifier." Int. J. Innov. Technol. Explor. Eng 8.11 (2019): 2211-2215.
[2] Sharma, Disha, and Gagandeep Jindal. "Identifying lung cancer using image processing techniques." International Conference on Computational Techniques and Artificial Intelligence (ICCTAI). Vol. 17. 2011.
[3] Chaudhary, Anita, and Sonit Sukhraj Singh. "Lung cancer detection on CT images by using image processing." 2012 International Conference on Computing Sciences. IEEE, 2012.
[4] Alam, Janee, Sabrina Alam, and Alamgir Hossan. "Multi-stage lung cancer detection and prediction using multi-class svm classifier." 2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2). IEEE, 2018
[5] Uzelaltinbulat, Selin, and Buse Ugur. "Lung tumor segmentation algorithm." Procedia computer science 120 (2017): 140-1
[6] K, Bhavanishankar; M.V, Sudhamani (2015). Techniques for Detection of Solitary Pulmonary Nodules in Human Lung and Their Classifications -A Survey. International Journal on Cybernetics & Informatics, 4(1), 27–40.doi:10.5121/ijci.2015.4103
[7] Makaju, Suren, et al. "Lung cancer detection using CT scan images." Procedia Computer Science 125 (2018):107-114
[8] Nagajyothi, D., et al. "Detection of lung cancer using SVM classifier." International Journal 8.5 (2020).
[9] Malik, Bhawana, et al. "Lung cancer detection at initial stage by using image processing and classification techniques." Lung Cancer 3.11 (2016).
[10] Gopi, K., and J. Selvakumar. "Lung tumor area recognition and classification using EK-mean clustering and SVM." 2017 International Conference on Nextgen Electronic Technologies: Silicon to Software (ICNETS2). IEEE, 2017.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY