Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776732

LUNG CANCER DETECTION USING SUPPORT VECTOR MACHINE (SVM) CLASSIFIER AND IMAGE PROCESSING

Ashwini Pawar[1], Sankirti Shiravale[2], Swati Shekapure[3], Geetha Chillarge[4]

[1]PG Student, Department of Computer Engineering, Marathwada Mitra Mandal College of Engineering Pune, India, [2,3,4]Assistant Professor, Department of Computer Engineering, Marathwada Mitra Mandal College of Engineering Pune, India.

Abstract. Lung cancer is quite possibly the most risky illness in the world. Various determinations and recognition of cancer in the lungs have been made utilizing different arrangement methods and information investigation. In this paper, a lung cancer detection framework utilizing a support vector machine and image processing are utilized to group the presence of lung cancer in CT scan images and blood tests. The work introduced in this paper is centered on the plan and improvement of a framework for primer conclusion and recognition of lung cancer from CT scan images. Arrangement of various tumor types is in this way imperative to guarantee higher endurance rates. The cycle of arrangement of lung cancer is continually testing. The framework recognizes the various phases of lung cancer which will help the specialists to identify lung cancer precisely and quickly from a lot of information.
Keywords: Lung Cancer Detection, Support Vector Machine (SVM), CT scan Image, Image Processing, Classification.

1 INTRODUCTION

Lung cancer is the chief reason for malignant cancer- related death. Lung cancer can start in the windpipe, primary aviation route, or lungs. It is brought about by unchecked development and the spread of certain cells from the lungs. Individuals with lung illnesses, for example, emphysema and past chest issues have more opportunity to be determined to have lung cancer. Overuse of tobacco, cigarettes and beedis, are the significant danger factor that prompts lung cancer in Indian men; notwithstanding, among Indian ladies, smoking isn't so normal, which shows that there are different variables which lead to lung cancer. Other danger factors incorporate openness to radon gas, air-contaminants, and synthetic compounds in the work environment. A malignancy that begins in the lung is essential lung cancer though those which start in the lung and spread to different pieces of the body are optional lung cancer. The size of the tumor and how far it has spread decides the phase of malignant growth. A superior comprehension of danger variables can assist with forestalling lung cancer infection.

Symptoms of lung cancer may not prompt critical objections until the sickness is very exceptional. The movement of lung cancer without symptoms is the main factor that makes this cancer growth so risky. Very nearly a fourth of patients have cancer with no side effects. A great many people discover that lung cancer is brought about by lung X- rays because of another illness. The early finding is vital in lung cancer. Since lung cancer frequently can quickly spread to the bones, liver, brain, and adrenal organs. Nonetheless, with the recently evolved lung cancer treatment strategies, the normal future and quality have expanded. Today, with progress

in imaging methods, for example, support vector machines, lung cancer can be recognized at a beginning phase. The proposed framework executing a lung cancer discovery from CT scan images input is prepared by the framework for finding lung cancer.

2 LITERATURE SURVEY

Mehdi Hassan Jony, et al. [1] describe Lung cancer is such an ailment that can be portrayed as one of the trading off ones and has become incredibly testing these days. All around, the cells that pass on threat cover with each other and it's hard to recognize in the hidden stage. However, the reality of the situation is, early areas bring an answer; it can reduce the destruction unquestionably. All through the assessment, we use a technique (GLCM); The Grey Level Co-occurrence to isolate the lung photos of the affected. The acknowledgment of the sporadic Lung image can be perceived by the Support Vector Machine (SVM).

Deep Prakash Kaucha, et al. [2] describe Image getting ready techniques are by and by regularly used in the clinical field in the early era of contamination. This assessment hopes to improve the exactness, affectability, and identity of the early area of lung cancer through a mix of image planning systems and data mining. The Computed Tomography (CT) checks if the image of the lungs is pre-readied and the Region of Interest (ROI) separated, held, and compacted using a DWT (Discrete Waveform Transform) strategy. The ensuing ROI image is rotated into four sub frequencies, bunches LL, HL, LH, and HH. Again, the LL sub repeat is rotated into four sub-gatherings, applying a 2- level DWT to the ROI based image. Further, features, for instance, entropy, co-association, energy, change, and homogeneity are isolated from the 2-level DWT images using a GLCM (Gray Level Co-occasion Matrix) with gathering influence through an SVM (Support Vector Machine). Request perceives whether the CT image is normal or dangerous. The Lung Image Database Consortium dataset (LIDC) has been used for planning and testing purposes for this examination. A Receiver Operating Characteristics (ROC) twist is used to separate the show the structure. For the most part, the system has a precision of 95.16%, affectability of 98.21%, and expresses 78.69%.

Özge Günaydin, et al. [3] describe Lung cancer as such a difficult to examine and unsafe illness. It ordinarily causes passing in the two individuals so brisk precise examination of handles is more critical for treatment. Various procedures have been used for perceiving a harmful development to start with stages. In this paper, AI procedures are considered while recognizing the lung cancer in the lung handles. We applied Principal Component Analysis, K-Nearest Neighbors, Support Vector Machines, Naïve Bayes, Decision Trees, and Artificial Neural Networks AI procedures to perceive anomalies. We took a gander at all procedures, both ensuing to pre-handling and without pre-preparing. The test outcomes show that Artificial Neural Networks give the best result with 82,43% precision after image planning and Decision Tree gives the best result with 93,24% accuracy without image taking care of.

Ning Guo, et al. [4] in this undertaking, we propose a novel assessment of intra-tumor heterogeneity which has promising execution for finding and putting together lung cancer. We indicated the feasibility of performing SVM based harm putting together using distinctive image features in PET/CT. SVM examination and request with a blend of tumor heterogeneity and other effective features can amplify insightful exactness and improve tumor putting together in oncological practice.

Nidhi S., et al. [5] describe Dangerous development is conceivably the most real and wide contaminations that are responsible for a gigantic number of passing's reliably. Among all different kinds of threatening developments, lung cancer is the most inescapable sickness having the most raised demise rate. Enrolled tomography channels are used for the ID of lung cancer as it gives a quick and dirty image of the tumor in the body and tracks its turn of events. In spite of the way that CT is preferred over other imaging modalities, visual comprehension of these CT scan images may be a misstep slanted task and can cause delay in lung cancer distinguishing proof. Along these lines, image planning techniques are used extensively in clinical fields for starting stage ID of lung tumors. This paper presents a robotized approach for the disclosure of lung cancer in CT scan images. The figuring for lung cancer area is proposed using procedures, for instance, center filtering for image pre- preparing followed by division of lung district of premium using mathematical morphological exercises. Numerical features are enlisted from the isolated district of premium and used to portray CT look at images into normal and peculiar by using an assistance vector machine.

Moffy Vas, et al. [6] describe Malignant growth as the main driver for an enormous number of passing's around the world, out of which lung cancer is the reason for the most elevated death rates. PC tomography examination is utilized by radiologists to recognize disease in the body and track its development. Visual understanding of the data set can prompt malignant growth discovery at later stages, accordingly prompting late therapy of disease which just lifts up the malignancy passing rates. In this way, image preparing devices can be utilized for the early discovery of the disease. In this paper, a lung cancer recognition technique has been applied to arrange the unmistakable highlights. Dormancy Enhancement has been done and utilizing SURF (Speeded up Robust Features) calculation highlights, for example, entropy, co-connection, energy, and difference have been separated from the Saliency Enhanced calculation is propose utilizing numerical morphological tasks for division of the lung district of interest, from which Haralick highlights are separated and utilized for the order of disease by counterfeit neural organizations.

Snehal Dabade, et al. [7] describe Lung cancer is a harmful lung tumor described by controlled cell development in tissues of the lung. Lung cancer is the most widely recognized disease analyzed around the world. A greater number of passes occur because of lung cancer than some other sort of disease. For the endurance of persistent malignancy, early discovery and treatment are exceptionally useful and viable. For deciding the malignant growth cells from clinical images, different image preparing and delicate figuring methods can be utilized. CT-image has properties like high goal, better lucidity, low commotion, and low twisting. Due to this CT-images are most regularly utilized for image handling. It is the best strategy for the recognition of little knobs. In view of the early recognition of lung cancer, the odds of a patient's endurance are more. Thus, CAD frameworks for lung cancer have been proposed. The CAD framework includes three stages. Those mean areas: Pre-handling, division of the lung, and order of the knob applicants. In this paper, we propose a strategy for the division of extricated lung areas from human chest CT. That technique is the Artificial Neural Network classifier model. For upgrading the edge identification of lung area flaps, a mix of touch planes of every pixel is utilized.

R. Sathishkumar, et al. [8] In this PC time, we are thoroughly going with the computerization of everything, similarly, the clinical business is additionally mechanized with the assistance of

image handling and information investigation. The most ideal approach to control the passing brought about by malignancy is early location. The clinical image or a CT filter image is pre-handled. The difference of the image is expanded with the CLAHE Equalization strategy. At that point, it is divided with the assistance of the arbitrary walk division technique. In division, the three cycles will happen the ROI of the image is portioned, and afterward then the line remedy is finished. As the third part, the consistent pixel change is portioned. The arrangement is the significant part where the malignant and non- carcinogenic is related to the pre-prepared model. All the techniques utilized above arrangements with the conventional method of image preparing and information investigation. In the Future this exactness will be supported with the cutting edge XGboost calculation where less information is utilized to get high precision.

Kyamelia Roy, et al. [9] describe Distinctive information investigation and arrangement techniques have been utilized for the conclusion and discovery of lung carcinoma. Since the reason for lung carcinoma stay unlimited, circumvention gets unreachable, consequently early uncovering of tumor in lungs is the principal route for treatment of lung carcinoma. Thus, biomedical image preparing and AI is the interaction used to group the presence of lung carcinoma. The rationale of this investigation is to create exactness and decide explicit incentive for identification of lung carcinoma at a previous stage utilizing a combination of biomedical image handling strategy and Knowledge Discovery in Data. The image of the lungs that is acquired from the CT (Computed Tomography) filter images are being pre-handled and the division is being done in the Region of Interest (ROI), Random Forest images with the assistance of an SVM (Support Vector Machine) Classifier. Characterization recognizes whether the image given is sound or undesirable (carcinomic). The least target versus the Number of capacity assessment plot is utilized to consider the exhibition of the technique. This whole interaction has been finished utilizing two calculations– arbitrary woods calculation and SVM order. The best outcome is obtaining SVM characterization. This technique is all out 94.5% effective, 74.2% delicate, 66.3% review and 77.6% explicit.

Barath Narayanan, et al. [10] describe Lung cancer regularly displays its essence with the arrangement of respiratory knobs. PC Aided Detection (CAD) of such knobs in CT sweeps would be of significant assistance in lung cancer screening. The common CAD framework includes a competitor finder and a component based classifier. In this exploration, we examine and investigate the exhibition of Support Vector Machine (SVM) in light of a huge arrangement of highlights. We study the presentation of SVM as a component of the quantity of highlights. Our outcomes show that SVM is heartier and computationally quicker with an enormous arrangement of highlights and less inclined to over-preparing when contrasted with conventional classifiers. Moreover, we additionally present a computationally effective methodology for choosing highlights for SVM. Results are introduced for a freely accessible Lung Nodule Analysis 2016 dataset. Our outcomes dependent on 10-crease approval show that SVM based arrangement technique beats the fisher straight discriminant classifier by 14.8%.

Table 1: Literature Survey comparison table

Paper Title	**Author**	**Algorithm/ Techniques**	**Advantages**	**Disadvantages**

Detection of Lung Cancer from CT Scan Images using GLCM and SVM (2018)	Mehdi Hassan Jony, et al. [1]	SVM, GLCM	Gives more straightforwardness and less mutilation.	Not suitable for a large data set.
Early Detection of Lung Cancer using SVM Classifier in Biomedical Image Processing (2017)	Deep Prakash Kaucha, et al. [2]	SVM	Exactness, affectability, and explicitness of early discovery of lung cancer detection	Disadvantages as it utilizes estimation just, barely ideal when high exactness results are required.
Comparison of Lung Cancer Detection Algorithms (2019)	Özge Günaydin, et al. [3]	SVM, ANN, Decision Tree	Accuracy results influenced modest quantities that can be disregarded to lessen time intricacy and save stockpiling.	Only clarify the examination of the algorithm.
SVM based Lung Cancer Diagnosis using Multiple Image Features in PET/CT (2015)	Ning Guo, et al. [4]	SVM	Great potential to expand indicative exactness and improve tumor arranging in oncological practice.	Malignant tumors are organically intricate and display significant spatial variety.
Detection of Lung Cancer in CT Images using Image Processing (2019)	Nidhi S., et al. [5]	Image Processing	More powerful in high dimensional space.	Not appropriate for the enormous informational collection. Less compelling and Expensive.
Lung cancer detection system using lung CT image processing (2017)	Moffy Vas, et al. [6]	ANN	Increasing the precision and speed of the lung cancer identification framework	Lungs differ in size and a fixed size organizing component can't be utilized to portion every one of the images.
A Review Paper on Computer Aided System for Lung Cancer Detection (2017)	Snehal Dabade, et al. [7]	CAD, ANN	Provide the state of the patient in the prior stage.	This paper presents the surveys on CAD. The high pace of bogus positive imprints.

Detection of Lung Cancer using SVM Classifier and KNN Algorithm (2019)	R. Sathishkumar, et al. [8]	SVM, KNN	Boosting estimation is used to extend the precision of the instrument.	Global least mistake is available.
A Comparative Study of Lung Cancer Detection using Supervised Neural Network (2019)	Kyamelia Roy, et al. [9]	Supervised Neural Network	Images are less misshaped, which is favorable.	Only clarify the examination of the algorithm.
Performance Analysis of Feature Selection Techniques for Support Vector Machine and its Application for Lung Nodule Detection (2018)	Barath Narayanan, et al. [10]	SVM	Well-characterized limit utilizing a huge arrangement of highlights at a lot quicker rate.	Less compelling and Expensive.
A Comparative Study of Lung Cancer Detection using Machine Learning Algorithms (2018)	Radhika P. R. et al. [11]	Machine Learning	There are less slip-ups. This cycle is time.	This measure is time taking. Accuracy is Low.
Predicting Lung Cancer Using Datamining Techniques With The Aid Of SVM Classifier (2018)	Dr. S.Senthil, et al. [12]	SVM	Framework will be available on the web, Patients from distant spots can benefit.	Taking much time
Lung Cancer Detection and Classification using Deep Learning (2018)	Ruchita Tekade, et al. [13]	Deep Learning	Malignancy level prediction	Less efficiency
A Study on Lung Cancer Detection by Image Processing (2016)	Weixing Wang, et al. [14]	Image Processing	Auto detecting nodule	Low accuracy
Multi-Stage Lung Cancer Detection and Prediction	Janee Alam, et al. [15]	SVM	High accuracy	Time consuming

Using Multi-class SVM Classifier (2018)				

3 Propose System

3.1 ARCHITECTURE

The work introduced in this paper is centered around the plan and advancement of a framework for starter finding and discovery of lung cancer from CT scan images. Order of various tumor types is accordingly imperative to guarantee higher endurance rates. The cycle of order of lung cancer is continually testing.

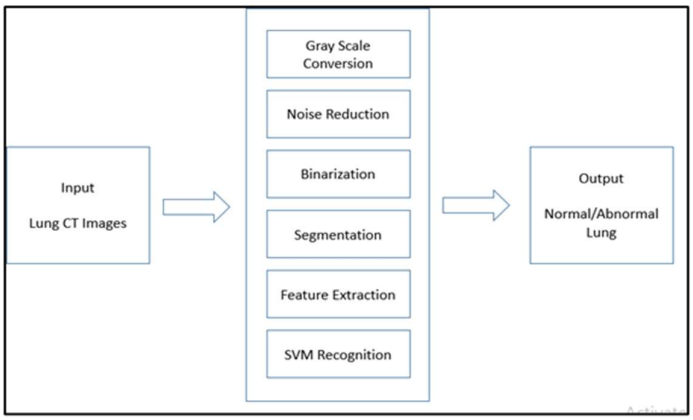

Fig 1: System Architecture

The framework identifies the various phases of lung cancer which will help the specialists to identify lung cancer precisely and quickly from a lot of information.[11].

3.2 PROPOSE SYSTEM WORKING STEPS

The lung nodules are expanded in limited scope in the lungs, which measure from 5 mm to 25 mm. The abnormal nodule size in images is more than 25mm. Surveying the size of cancer growth in organizing and entering the tissue encompassing it, just as the presence or absence of metastasis in lymph nodes or different organs. The stage from I to IV as per the intensity:

1) Stage I: Cancer is confined to the lung
2) Stage II and III: Cancer is confined to the chest
3) Stage IV: Cancer has spread from the chest to different parts of the body

Step i) Gray Scale Conversion- While changing over an RGB picture to grayscale, we need to take the RGB esteem for every pixel and make a binary worth mirroring the brilliance of that pixel.

FOR CONVERTING THE COLOR IMAGE INTO GRAYSCALE WE FIND THE AVERAGE OF IT, AND REPLACE THE R, G, B PIXEL VALUE WITH AVERAGE

AVERAGE = R + G + B / 3

Step ii) Noise Reduction- The noise removal technique is the way toward eliminating or decreasing the noise from the picture. The noise removal technique decreases or eliminates the visibility of noise by smoothing the whole picture leaving zones close to differentiate limits.
This filter is more effective while reducing the noise, maintaining edges and also removing 'salt and pepper' type noise.

Step iii) Binarization- Binarization is the way toward changing over a multi-tone picture into a bi-tional picture. On account of report pictures, it is regular to plan the forefront text pixels to dark and the remainder of the picture (foundation) to white.
Let input image is p(m,n), T can be defined as threshold value and the output image is q(m,n) of thresholding process therefore mathematically it can be expressed as q(m,n)=1, if p(m,n) greater than equal to T otherwise 0.[12]
The computerized image is partitioned into numerous sections. The fundamental target of division is to simplify and change the assignment of CT scan reproduction into more educational and look at it effectively in subtleties.

Step iv) Segmentation- Segmentation is the way toward dividing a digital picture into various portions. The objective of the segmentation is to streamline as well as change the portrayal of a picture into something that is more important and simpler to analyze.

Step v) Feature Extraction- The feature extraction procedure assumes a vital part in nearby image processing. Feature extraction methods are applied to get highlights that will be valuable in characterizing and acknowledging pictures. As features characterize the conduct of a picture, they show its place regarding capacity taken, productivity in order, and clearly in time utilization moreover.[13]
Feature Extraction is a fundamental advance/stage that utilizes calculations and methods to perceive the examples of an imitation. The Segmented yield is given as a contribution for the element extraction. The accompanying features are covered under include extraction, like Area, Perimeter, and Eccentricity, and these all are scalar quality.
Step vi) SVM Recognition- SVM is a supervised learning method that analyzes data which is used for classification analysis. For non-linearly separable datasets, SVM is more suitable since it reduces the misclassification rate.
SVM classifier used to characterize the straight and non- direct locales. A direct partition, classifier is utilized to isolate the influence and non-influenced locales inside the Image. It essentially utilizes delicate and hard edges. These direct conditions are homogeneous. In non-direct detachment, we will isolate the influence part or district by addressing a non-straight structure. These straight conditions are homogeneous.[14]

4 RESULTS

The expected result of lung cancer detection depends on various factors such as the stage of cancer, the patient's age and overall health, and the treatment options available.

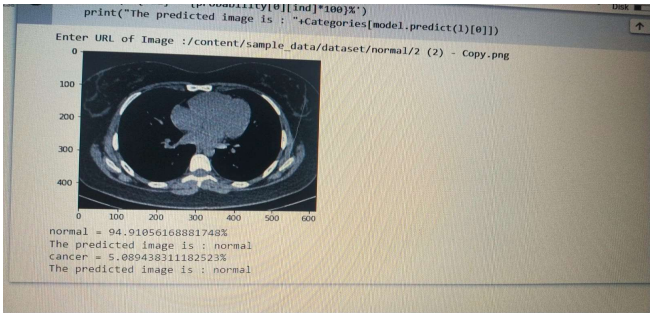

Fig 2: Expected Result

If lung cancer is detected at an early stage, the chances of successful treatment are higher. The treatment options for lung cancer may include surgery, radiation therapy, chemotherapy, targeted therapy, immunotherapy, or a combination of these treatments.

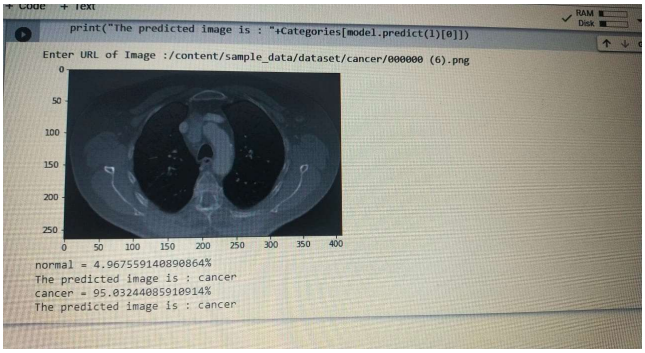

Fig 3: Lung Cancer detection

In some cases, lung cancer may have already spread to other parts of the body, which can make it more difficult to treat. The prognosis of lung cancer also depends on whether it is a small cell or non-small cell lung cancer.

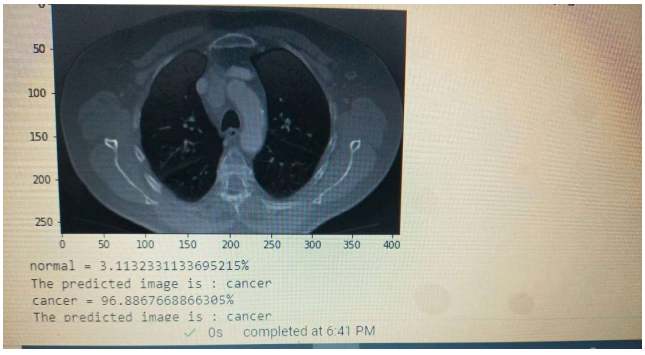

Fig 4: Expected result lung cancer detection

Once the model is selected, it needs to be trained on the data and validated [15]to ensure that it is accurate and reliable. This may involve using techniques such as cross-validation to evaluate the model's performance on different subsets of the data.

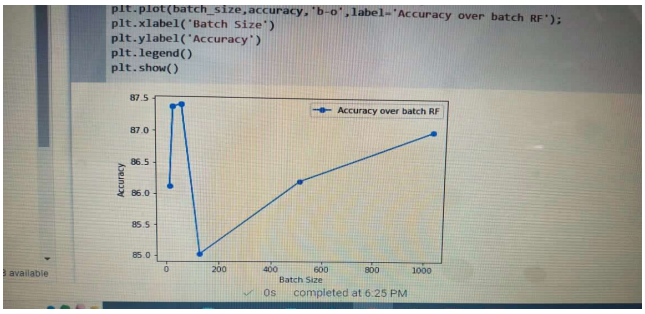

Fig 5: Accuracy of the result

Finally, the model can be deployed for use in clinical practice.This could involve integrating it with existing medical imaging software or developing a new software application specifically for lung cancer detection.

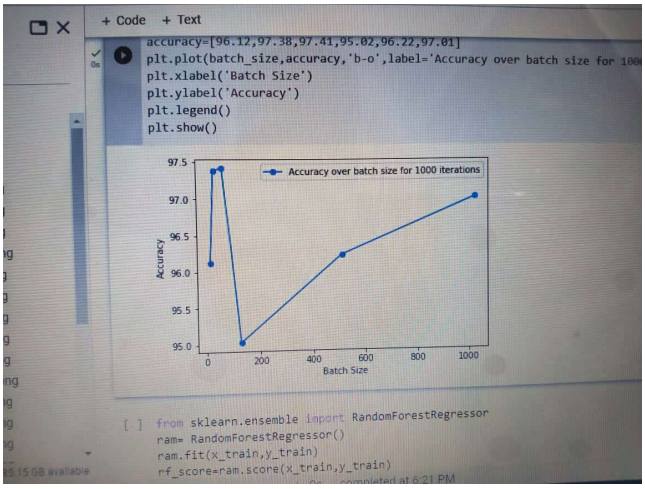

Fig 6: Accuracy of the result

5 CONCLUSION

This paper talks about the lung cancer detection framework by utilizing the methods of Image Processing. Here the proposed model is created utilizing Genetic Optimization, and SVM calculation utilized to include choice and arrangement. This paper is an augmentation of image preparation utilizing lung cancer discovery and produces the aftereffects of highlight extraction and highlight determination after division.

In future work, the proposed framework will assist with recommending diagnosing disease in various organs of individuals. The proposed procedure can be carried out for explicit tumors, i.e., a gathering of illnesses that help in lessening the development of unusual cells or spreading to different pieces of the body.

REFERENCES

1. Mehdi Hassan Jony, Fatema Tuj Johora, Parvin Khatun, Humayan Kabir Rana, “Detection of Lung Cancer from CT Scan Images using GLCM and SVM”, IEEE, 1st International Conference on Advances in Science, Engineering and Robotics Technology, 2019.

2. Deep Prakash Kaucha, P. W. C. Prasad, Abeer Alsadoon, A. Elchouemi, Sasikumaran Sreedharan, “Early Detection of Lung Cancer using SVM Classifier in Biomedical Image

Processing", IEEE International Conference on Power, Control, Signals and Instrumentation Engineering, 2017.

3. Özge Günaydin, Melike Günay, Öznur Şengel, "Comparison of Lung Cancer Detection Algorithms", IEEE, 2019.

4. Ning Guo, Ruoh-Fang Yen, Georges El Fakhri and Quanzheng Li, "SVM Based Lung Cancer Diagnosis Using Multiple Image Features in PET/CT", IEEE, 2015.

5. Nidhi S. Nadkarni, Prof. Sangam Borkar, "Detection of Lung Cancer in CT Images using Image Processing", IEEE, Proceedings of the Third International Conference on Trends in Electronics and Informatics, 2019.

6. Moffy Vas, Amita Dessai, "Lung cancer detection system using lung CT image processing", IEEE, 2017.

7. Snehal Dabade, Shubhangi Chaudhari, Sneha Jadhav, Arjun Nichal, "A Review Paper on Computer Aided System for Lung Cancer Detection", 2017 International Conference on Big Data, IoT and Data Science (BID) Vishwakarma Institute of Technology, Pune, Dec 20-22, 2017.

8. R.Sathishkumar, K.Kalaiarasan, A.Prabhakaran, M.Aravind, "Detection of Lung Cancer using SVM Classifier and KNN Algorithm", Proceeding of International Conference on Systems Computation Automation and Networking, 2019.

9. Kyamelia Roy, Sheli Sinha Chaudhury, Madhurima Burman, Ahana Ganguly, Chandrima Dutta, Sayani Banik, Rayna Banik, "A Comparative Study of Lung Cancer Detection using Supervised Neural Network", IEEE, 2019.

10. Barath Narayanan, Russell C. Hardie, Temesguen M. Kebede, "Performance Analysis of Feature Selection Techniques for Support Vector Machine and its Application for Lung Nodule Detection", IEEE, 2018.

11. Radhika P R, Rakhi.A.S.Nair, Veena G, "A Comparative Study of Lung Cancer Detection using Machine Learning Algorithms", IEEE, 2018.

12. Dr. S.Senthil, B.Ayshwarya, "Predicting Lung Cancer Using Data Mining Techniques With The Aid Of Svm Classifier", IEEE, 2018.

13. Ruchita Tekade, Prof. K. J. Rajeshwari, "Lung Cancer Detection and Classification using Deep Learning", IEEE, 2018.

14. Shekapure, Swati, Nikita Pagar, Bhagyashree Kulkarni, Dinesh Choudhary, and Priti Parkhad. "Predicting COVID-19 Pneumonia Severity based on Chest X-ray with Deep Learning." International Journal of Computer Applications 975: 8887.

15. Palwe, A., Shiravale, S. and Shekapure, S., 2022. Image Captioning using Efficient Net. Journal of Optoelectronics Laser, 41(7), pp.1259-1270