Detection of lung cancer using image processing and machine learning

Venkatesh V [p1], Gayathri M J[s2], Jeyavarshaa M[s3] , Akila Umesh[s4].

Assistant Professor[p1], UG Student[s2,s3,s4]
Sri Krishna College of Technology Coimbatore, Tamil Nadu.
venkatesh.v@skct.edu.in[p1], 19tucs041@skct.edu.in [s1],19tucs058@skct.edu.in[s2] ,
19tucs007@skct.edu.in[s3].

ABSTRACT:

The Lungs are the part of respiratory system in human body. The function of the lungs is the process of extracting oxygen from air and passing it to the red blood cells. With the advent of time various types of cancers are found. Lung cancer is one of the serious type which has the highest mortality rate. It is treatable if it is identified in the early stages. The cancer nodules are identified using image processing techniques which uses various methods to extract the information from the X-Ray without any loss of information. To identify the cancer nodules, the system also use the information from the history of cancer. Nowadays the reason for the lung cancer are due to the increase in pollution in city and unhealthy habits. For this reason, the model uses machine language and image processing to process images of CT scan or X-rays to detect cancer nodule. The detection system helps to identify the early stage of the lung cancer and also identify in which stage the cancer nodules is there
by using the dataset that is taken from the Iraq Oncology teaching hospital/ National centre for cancer disease and also by using the image processing techniques.

Keywords: Image Processing Technique, X-Ray, Computer Tomography (CT), Smoothening techniques, Classification Techniques, Machine language.

I.INTRODUCTION:

Over the past few years lung cancer is considered as major health issues among the humans. Lung cancer is the deadlier disease and a primary concern of high mortality rate [1].

It is very rare to detect the lung cancer before the age of 45 years. It is generally possible for lung cancer to be detected in the age of 55 to 70. The history of research says that the survival rate of the lung cancer is less than 14%. If the lung cancer is detected at the early stage than the survival rate of the patients increases [2].

Screening methodology has helped to predict lung cancer, but early cancer detection and accuracy are difficult to maintain. Thus, a computer-aided automatic detection (CAD) process must be implemented in a clinical centre to develop an effective cancer prediction system that uses optimized and intelligent technology. It is quite unlikely that an X-Rays may accidentally detect the affected nodules in its early stages[3].

It is quite possible to provide treatment, if it is identified and analyzed in the early stages of the disease. Powerful technology are used to get information and that information are given to doctors which is used to assist them in developing the medicine for the lung cancer. To diagnose the cancer, there is a need for preprocess and trained data such as X-Rays or computed Tomography (CT) scan which is then processed through various neural network and machine learning techniques to give the accurate output [4].

In this research, the system can detect the early stage of cancer nodules and can overcome the disadvantages that are faced by the preceding system, which are used for detecting the nodules and exploit the robust noise by using filtering methods such as Bilateral filtering(extension of Gaussian filtering) and Classification techniques such Computational Neural Network(CNN) and Artificial Neural Network(ANN).This analysis shows that the method that are used in the system achieves the objectives of researchers which are up to the mark to detect lungs cancer nodule.

II.LITERATURE SURVEY:

Literature review is carried out throughout the whole project to gain skills and to improve the knowledge needed to complete this project. This section concentrates on the basic concepts and all fundamental theories which related to this project.

Patra et al. investigated that RBF classifier gives the accuracy of 81.25% on lung cancer data and also predicted that suitable feature selection method and integrated approach with learning process approach in RBF will be improved [1].

Witschi et al. has researched about the history of lung cancer in the human body. He also studied that how cancer nodes occur and replicate [2].

Chaudhury et al. proposed a system that uses the machine learning techniques for the detection of breast cancer in early stage3].

Furat et al. investigated that conventional neural network significantly enhances the CT images which is used to obtain the grain wise segmentation to identify the stage of the cancer nodules.[4]

Deppen et al. researched about Predicting lung cancer prior to surgical resection in patient with lung nodules[5].

Feigenbaum et al. investigated the Probabilistic of lung nodule classification with belief decision trees, which gives the appropriate method of representing the diagnoses of multiple CT scans [6].

Aggrawal et al. invented a system using threshold values and characteristics which are used in image processing techniques to extract the features which shows the difference between the normal cancer nodules and malignant cancer nodules [7] .

Suren Makaju et al researched a model where new technique were applied for diagnosing the lung cancer in early stage and used the Median filter and Gaussian filter for pre-processing of X-Ray images rather than using other filter [8].

Berry et al developed the Prediction of large cell lung cancer prognosis by fully automated microscopic medical image features system, which has successfully distinguished long term survivors cells from small term survivors cells[9].

Nasser et al. developed an Artificial Neural Network (ANN) to detect the presence of lung cancer in human body, where symptoms were used as a data to diagnose the lung cancer in humans. This model was able to detect the cancer node with 96.675 accuracy [10].

Shimazaki et al. developed a system to detect the lung cancer via chest radiographs using deep learning-based model using segmentation method and five cross- validation approach [11].

Shakeel et al. proposed image processing and machine learning techniques to predict the stage of lung cancer. CT scans are used as a dataset to recognise the large cell lung cancer which is effectively applied to eliminate the noise and increase the quality of the lung image by multilevel brightness-preserving approach [12].

Boddu et al. researched that AI model may be effective as professional doctor. Deep and machine learning techniques shows excellent precision for non-cancerous cell from cancerous cell [13].

III. SOFTWARE USED IN THIS RESEARCH:

ANACONDA:

Anaconda is an open-source platform which allows the programmers to write and execute the code in programming languages like python and R. Anaconda software creates an environment for many different versions of python and package versions. It is used for machine learning, data science and deep learning. It provides advantages for simplification of package management and deployment.

JUPYTER NOTEBOOK:

Jupyter notebook is an open source web-based application for creating and sharing the computational documents. It enables the customers to convert the programming codes into other formats such as HTML and PDF. It is one of the applications of Anaconda software.

PYTHON:

Python is a high-level, general-purpose programming language which is used to build websites and software, automate tasks and conduct data analysis. Its supports multiple programming paradigms which includes structured and object-oriented, functional programming.

IV. IMAGEPROCESSING ARCHITECTURE:

The main objective of this annotations is to analyse and transform the input image into digital form and perform operation to get the information from it, here the proposed system uses the Computed Tomography (CT) or X-Ray images as input. Figure 1.0 depicts the stages performed for the system. The image processing techniques are performed in five steps, they are

i. Image Acquisitions
ii. Image Smoothening
iii. Image Enhancement
iv. Feature Extraction
v. Classification

IMAGE ACQUISITION:

Image acquisition is a process of extracting the images from external source (National centre for cancer Diseases/Iraq Oncology Teaching Hospital) and processing the extracts image into the enhancement phase.

IMAGE SMOOTHENING:

Image smoothening is an image processing technique that reduces and clears image noises. Averaging is often used to achieve the purpose of smoothing in special domain. Commonly used smoothing filters techniques include average smoothing, Gaussian smoothing, and adaptive smoothing.

IMAGE ENHANCEMENTS:

Image enhancement is a process of improving the quality and information of the original data. It makes the image more comprehensible and undoubtedly detect the data present in it. To get more accurate results in feature extraction and classification, the image should be enhanced.

FEATURE EXTRACTION:

Feature extraction is the process of transforming the raw pixel data values from an input image to a more meaningful and useful information i.e reducing the dimension of the larger number of pixel in an image.

CLASSIFICATION:

Classification is the process of identifying and analysing the entire image and extracting the information from image to get the final result on digital image.

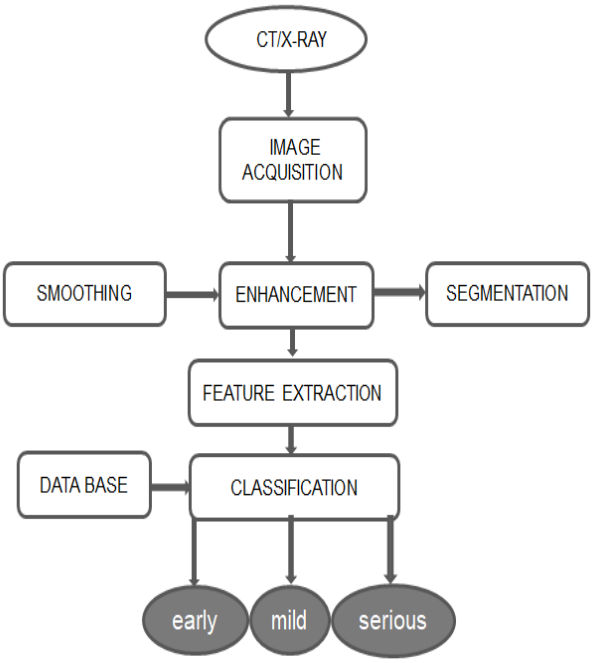

Fig 1.0 image processing architecture

IV.METHODS AND ALGORITHM:

The image is applied to various methods at the different stages of image processing to make the image unambiguous and significant. In image acquisition, the image is acquired from external source database (National centre for cancer Diseases / Iraq Oncology Teaching Hospital).

In image smoothing and enhancement Bilateral filtering which is an extension of Gaussian filter is used to improve contrast in image. It provides better quality of image without loss of information and reduces noise. It smoothens the edges and corners of image. Figure 2.0 depicts the image after enhancement.

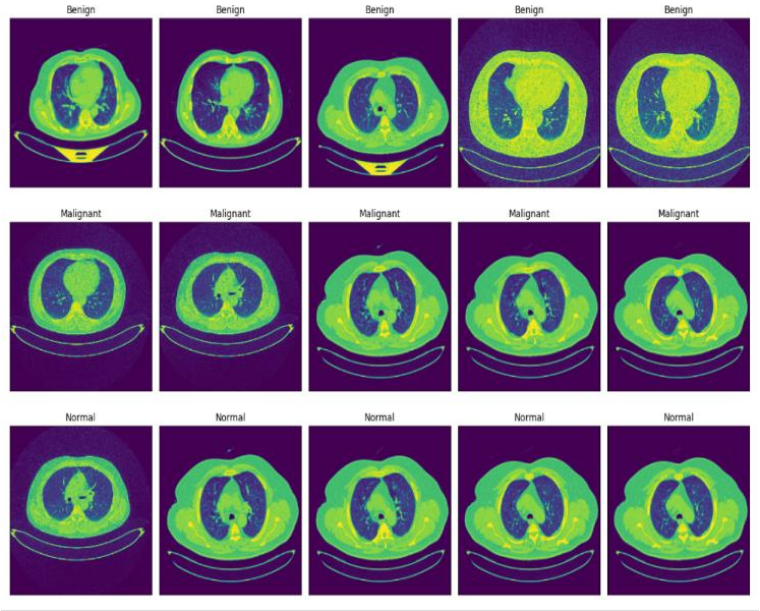

Fig 2.0 image after Bilateral Filtering

In feature extraction and classification method Artificial neural network (ANN) and Convolutional neural network(CNN) algorithm is used to identify and analyse the entire image and compare the accuracy of the image using these method. Figure 2.1 and 2.2 displays the model accuracy graph of CNN and ANN. It helps in comparison of training accuracy and model accuracy.

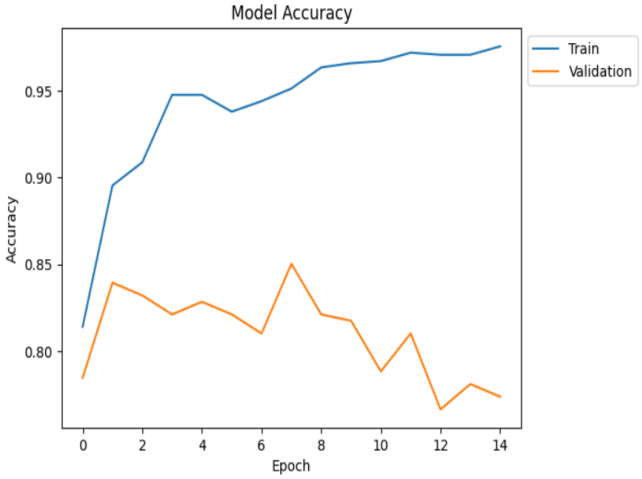

Fig 2.1 CNN model accuracy

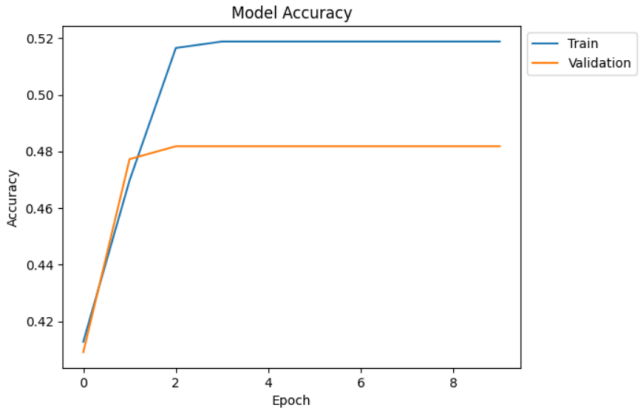

Fig 2.2 ANN model accuracy

The validation loss of both ANN and CNN model are 1.07 and 1.25 respectively.

VI.CONCLUSION:

Mostly the lung cancer are detected at the later stage, which is an ineffective stage for treatment. It leads to increase in the rate of death of the patient. As per the medical study early stage of cancer cell is treatable but detecting it in early stage is difficult. To overcome this, the proposed system is designed to overcome the challenges of preceding methods which are used in detection system and exploit the robust noise filtering methods using image acquisition, enhancement and Classification techniques such as Gaussian noise filtering algorithm for image acquisition and smoothing and Grayscale image histogram equalization algorithm for enhancement technique and classification techniques are ANN and CNN .The accuracy of the method that is used in classification techniques will be displayed in the graph . This analysis shows the effective use of different techniques such as filtering, ANN, CNN

and etc. so as to achieve the objectives that are up to the mark and to detect lungs cancer nodule stages such as early, mild and serious stage.

VII.REFERENCE:

1.Patra, R., 2020. Prediction of lung cancer using machine learning classifier. In Computing Science, Communication and Security: First International Conference, COMS2 2020, Gujarat, India, March 26–27, 2020, Revised Selected Papers 1 (pp. 132-142). Springer Singapore.

2.Witschi, H., 2001. A short history of lung cancer. Toxicological sciences, 64(1), pp.4-6.

3. Chaudhury, S., Krishna, A.N., Gupta, S., Sankaran, K.S., Khan, S., Sau, K., Raghuvanshi, A. and Sammy, F., 2022. Effective image processing and segmentation-based machine learning techniques for diagnosis of breast cancer. Computational and Mathematical Methods in Medicine, 2022.

4. Furat, O., Wang, M., Neumann, M., Petrich, L., Weber, M., Krill III, C.E. and Schmidt, V., 2019. Machine learning techniques for the segmentation of tomographic image data of functional materials. Frontiers in Materials, 6, p.145.

5. Deppen, S.A., Blume, J.D., Aldrich, M.C., Fletcher, S.A., Massion, P.P., Walker, R.C., Chen, H.C., Speroff, T., Degesys, C.A., Pinkerman, R. and Lambright, E.S., 2014. Predicting lung cancer prior to surgical resection in patients with lung nodules. Journal of Thoracic Oncology, 9(10), pp.1477-1484.

6. Zinovev, D., Feigenbaum, J., Furst, J. and Raicu, D., 2011, August. Probabilistic lung nodule classification with belief decision trees. In 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 4493-4498). IEEE.

7. Aggarwal, T., Furqan, A. and Kalra, K., 2015, August. Feature extraction and LDA based classification of lung nodules in chest CT scan images. In 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI) (pp. 1189-1193). IEEE.

8. Makaju, S., et al., Lungs Cancer Detections using CT Scan Image, Procedia Computer Science, 2018. 125: p. 107-114

9.Yu, K.H., Zhang, C., Berry, G.J., Altman, R.B., Ré, C., Rubin, D.L. and Snyder, M., 2016. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nature communications*, *7*(1), p.12474.

10.AURELIA, J.E. and RUSTAM, Z., 2020, November. Comparison of several Kernel Functions on Neural Network–Support Vector Machine as Classifier for Lung Cancer. In 3RD INTERNATIONAL CONFERENCE ON MATHEMATICAL AND RELATED SCIENCES: CURRENT TRENDS AND DEVELOPMENTS PROCEEDINGS BOOK (Vol. 20, p. 99).

11. Shimazaki, A., Ueda, D., Choppin, A., Yamamoto, A., Honjo, T., Shimahara, Y. and Miki, Y., 2022. Deep learning-based algorithm for lung cancer detection on chest radiographs using the segmentation method. Scientific Reports, 12(1), p.727.

12. Shakeel, P.M., Burhanuddin, M.A. and Desa, M.I., 2022. Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier. Neural Computing and Applications, pp.1-14.

13.Boddu, R.S.K., Karmakar, P., Bhaumik, A., Nassa, V.K. and Bhattacharya, S., 2022. Analyzing the impact of machine learning and artificial intelligence and its effect on

management of lung cancer detection in covid-19 pandemic. Materials Today: Proceedings, 56, pp.2213-2216
14. Zhao, L., Wang, H., Fu, J., Wu, X., Liang, X.Y., Liu, X.Y., Wu, X., Cao, L.L., Xu, Z.Y. and Dong, M., 2022. Microfluidic-based exosome isolation and highly sensitive aptamer exosome membrane protein detection for lung cancer diagnosis. Biosensors and Bioelectronics, 214, p.114487.
15.Li, X., Gong, P., Zhao, Q., Zhou, X., Zhang, Y. and Zhao, Y., 2022. Plug-in optical fiber SPR biosensor for lung cancer gene detection with temperature and pH compensation. Sensors.
16. He, L., Yu, X., Huang, R., Jin, L., Liu, Y., Deng, Y., Li, S., Chen, H., Chen, Z., Li, Z. and Xiao, P., 2022. A novel specific and ultrasensitive method detecting extracellular vesicles secreted from lung cancer by padlock probe-based exponential rolling circle amplification. Nano Today, 42, p.101334.
17. Hassanein, M., Callison, J.C., Callaway-Lane, C., Aldrich, M.C., Grogan, E.L. and Massion, P.P., 2012. The state of molecular biomarkers for the early detection of lung cancer. Cancer prevention research, 5(8), pp.992-1006.
18. Arya, S.K. and Bhansali, S., 2011. Lung cancer and its early detection using biomarker-based biosensors. Chemical reviews, 111(11), pp.6783-6809.
19. Li, X., Yang, T. and Lin, J., 2012. Spectral analysis of human saliva for detection of lung cancer using surface-enhanced Raman spectroscopy. Journal of biomedical optics, 17(3), pp.037003-037003.
20. Ehmann, R., Boedeker, E., Friedrich, U., Sagert, J., Dippon, J., Friedel, G. and Walles, T., 2012. Canine scent detection in the diagnosis of lung cancer: revisiting a puzzling phenomenon. European respiratory journal, 39(3), pp.669-676.
21. Ostroff, R.M., Bigbee, W.L., Franklin, W., Gold, L., Mehan, M., Miller, Y.E., Pass, H.I., Rom, W.N., Siegfried, J.M., Stewart, A. and Walker, J.J., 2010. Unlocking biomarker discovery: large scale application of aptamer proteomic technology for early detection of lung cancer. PloS one, 5(12), p.e15003.
22. Veronesi, G., Maisonneuve, P., Bellomi, M., Rampinelli, C., Durli, I., Bertolotti, R. and Spaggiari, L., 2012. Estimating overdiagnosis in low-dose computed tomography screening for lung cancer: a cohort study. Annals of internal medicine, 157(11), pp.776-784.
23. Ettinger, D.S., Akerley, W., Borghaei, H., Chang, A.C., Cheney, R.T., Chirieac, L.R., D'Amico, T.A., Demmy, T.L., Ganti, A.K.P., Govindan, R. and Grannis, F.W., 2012. Non–small cell lung cancer. Journal of the National Comprehensive Cancer Network, 10(10), pp.1236-1271.
24. Tanaka, F., Yoneda, K. and Hasegawa, S., 2010. Circulating tumor cells (CTCs) in lung cancer: current status and future perspectives. Lung Cancer: Targets and Therapy, pp.77-84.
25.Liu, Y.Z., Wang, Z., Fang, L.L., Li, L., Cao, J., Xu, X., Han, Y.L., Cai, Y., Wang, L.X. and Wang, M.R., 2012. A potentialprobe set of fluorescence in situhybridization for detection of lung cancer in bronchial brushing specimens. Journal of cancer research and clinical oncology, 138, pp.1541-1549.