Journal of Analysis and Computation (JAC)
(An International Peer Reviewed Journal), www.ijaconline.com, ISSN 0973-2861
Volume XVIII, Issue II, July-December 2024

STUDY OF AN ADVANCE MULTICLASS LUNGS DETECTION USING MACHINE LEARNING

Brijesh Kumar Singh[1], Aman Kabra[2], Akhilesh Yadav[3], Vinay khatri[4]

[1] Assistant Professor, Department of Information Technology, JECRC College, Jaipur
[2]B.Tech Student, Department of Information Technology, JECRC College, Jaipur
[3]B.Tech Student, Department of Information Technology, JECRC College, Jaipur
[4]B.Tech Student, Department of Information Technology, JECRC College, Jaipur

brijeshkumarsingh.it@jecrc.ac.in
amankabra.it24@jecrc.ac.in
akhileshyadav.it24@jecrc.ac.in
vinaykhatri.it24@jecrc.ac.in

ABSTRACT:

Lung cancer and its various subtypes complex genetic variations represent a major challenge precise diagnoses and selection of individual treatments. The Study examines the application of advanced machine learning Multiclass classification techniques for lung cancer subtypes base donhistological and genomic data. AND comprehensive datasets covering various sample types subtypes of lung cancer, including non-small cell lung cancer (NSCLC), such as adenocarcinoma and squamous cell carcinoma-cell lung cancer and small cell lung cancer (SCLC) and Rare subtypes were used. Study used advanced machine learning, which integrates deep learning algorithms and ensemble methods for distinguishing complex patterns within Data.[1] Convolutional neural networks (CNN) were used to extract complex features from histopathological images, captures detailed characteristics specific to each cancer Subtype. Genetic data were analyzed at the same time ensemble methods such as Random Forest and Gradient enhancer that allows integration of multiple genomes functions. The developed model showed exceptional accuracy, Sensitivity and specificity in the classification of various lung tumors subtypes insights genetic markers and main histological features of the subtype Rank. analysis with advanced machine learning techniques for classifying different lung cancer subtypes based on extensive data collection, including histopathological tests images and genome profiles[2]. convolution neuron networks (CNNs) were used to extract complex spatial data. features from histological images that capture subtle nuances specific to each subtype. At the same time team methods, including Random Forest and Gradient Boost were used to analyze multivariate genomic dataholistic

understanding of the genetic landscape. Our results showed exceptional performance Integrated model, characterized by high precision and sensitivity in the classification of non- small cell lung cancer (NSCLC) subtypes. such as adenocarcinoma and squamous cell carcinoma, small -cell lung cancer (SCLC) and rare subtypes. Models The prediction function has been further improved Validity analysis, explanation of the most important genetic markers and histological features essential for subtype differentiation.

Keywords- Machine Learning, CNNs, RF, CAD, NSCLC, SCLC, KNN, DL, PCA.

[1] INTRODUCTION

Lung cancer represents an enormous global health challenge, represents a significant percentage of cancer deaths. According to the World Health Organization, the number is estimated, there were 2.21 million new cases of lung cancer in 2020andisthereforethe most frequently diagnosed cancer in the world [1]. The complexity of lung cancer is not simple with high incidence but also in its various subtypes that require it Precise classification for effective treatment planning. Traditional methods of diagnosing lung cancer such as: biopsies and CT scans have limitations. Biopsies during final, can be invasive, expensive and sometimes impractical, especially for patients with health problems. Although CT scanning is widely used, there is a heavy reliance on it Radiologists' knowledge is not always subtle shades that can differentiate cancer subtypes Multiclass lung cancer refers to the classification system used to classify different types of lung cancer based on their characteristics histological features or genetic features. Lung cancer is here is a type of cancer that starts in the lungs and can spread to other people body parts. There are two main types of lung cancer: non-small cell lung cancer (NSCLC) and small cell lung cancer tumors(SCLC).These type sare then divided into subtypes, Introduction of a multi-class classification system.

1. **Non-small cell lung cancer (NSCLC):** NSCLC is the most common type of lung cancer in about 85% of cases. It contains several subtypes such as adenocarcinoma, squamous cell carcinoma and large cell carcinoma. Adenocarcinoma is the most common subtype and often occurs in the outer areas of the lungs. Squamous cell carcinoma usually develops in bronchi.
2. **Small cell lung cancer (SCLC):** SCLC is less common, a common but very aggressive type of lung cancer. Tends to spread quickly to other parts of the body. SCLC is and is often caused by smoking is characterized by small cells under the microscope.
3. **Other rare subtypes:** Except NSCLC e.g SCLC, there are rare subtypes of lung cancer, including carcinoid and sarcomatoid tumors tumors. Carcinoid tumors grow slowly and generally does not spread quickly. Sarcomatoid cancer is a rare and aggressive subtype contains both cancer and epithelial cells sarcoma-like cells.

Early detection, accurate diagnosis and personalization treatment plans are crucial in the management of multiclass lungtumors. Treatment options often include surgery, chemotherapy, radiation therapy, targeted therapy, Immunotherapy or a combination thereof, as needed Type and stage of cancer as well as general health of the patient. Research and advances in genomics have led to the definition of targeted targets therapies that specifically target cancer cells genetic mutations that improve the effectiveness of treatment and quality of life of patients diagnosed with multiclass disease lung cancer cases[2].The emergence of advanced machine learning techniques The offers a promising path to solving this challenge. Out of particularly uses the power of artificial intelligence Deep learning algorithms and ensemble methods, researchers can deal with the complex patterns contained within histological images and genomic data. These techniques enable the extraction of various features and sophisticated classification models.

[2] RELATED WORK

LUNG CANCER USING CAD APPROACH

Silva proposed a CAD approach as described in their, based on U-Net and ResNet-34 architectures. The method was applied to four different crosses cohort records. To evaluate their performance, the authors used the bone similarity coefficient (DSC). CBF is a measurement commonly used in medicine image segmentation tasks and measure the overlap between them predicted and informed truth segmentation with value between 0 and 1.[3] A DSC value of 1 indicates a perfect result Overlaps between predicted truth and ground truth Segmentation. Based on the results of Silva, although the CAD approach achieved an average DSC value of more than 0.93infourdifferentdatasetscoveringdifferentcohorts. The suggests that this method worked well and accurately Segmentation and delineation of the structures that interest us medical images. The method was verified by two radiations experts who identified some limitations, particularly in Consolidation file. However, a system has been developed The achieved over 99.3% accuracy and an F-score of 99.2%. The study employed four performance metrics to evaluate the disease. Additionally, the system was evaluate during two DL tools on three datasets.

In the work by Alsheikhy et ala manual process machine was developed for the detection of lung cancer. The process involved utilizing numerous CT images and a Gabor filter. Thedatasetusedconsistedof1,800images,900ofwhichare photos of children diagnosed with lung cancer. All had a size of 200 200 pixels and the data set was collected from IMBA's personal database. However The study also does not name any specific performance indicators gives results. This CAD system is average Accuracy 99.42%, with a maximum accuracy of 99.61%. Furthermore, the system achieves impressive results for others included performance measurements with recall, precision, and F values reach 99.76%, 99.88% and 99.82% respective.

LUNG CANCER USING ACNN APPROACH

Li and Zhao introduced an update of the CNN model to estimate the volume of the left ventricle from MRI images and achieve multi-view fusion. Benjamin and Jayasree proposed an image fusion method using translation-invariant wavelets and cascaded principal component averaging (PCA).[4] According to experiments, the fusion method outperforms the visual and quantitative evaluation framework. Aishwarya and Thangammal proposed an adaptive dictionary learning method for multi-view medical image fusion. For dictionary learning, blocks of useful information were separated by removing blocks of zero data and using a multiscale function to approximate the remaining parts of the image. This reduced the amount of calculation required while still producing a high quality image.

Likewise, Faruqui et al. used a CNN-based deep model intended to improve the accuracy of lung cancer CAD. Combine CT images with wearable medical IoT (MIoT) data to improve diagnostic capabilities. LungNet's unique 22-layer CNN architecture extracts features from both data sources, achieving high accuracy of 96.81% and low false positive (FP) rate of 3.35% in lung cancer classification into five classes. Outperforms similar CNN-based classifiers. LungNet also subclasses stage 1 and stage 2 lung cancer with an accuracy of 91.6% and a FP rate of 7.25%. LungNet ran on a central server and was trained on a balanced dataset of 525,000 images. The high accuracy, low FP rate and substage classification make it a promising solution in automatic lung cancer diagnosis systems. Nasser and Abu-Naser introduced an artificial neural network (ANN) model designed to detect lung cancer in the human body.[5] The model uses a set of symptoms as data to diagnose the disease. A lung cancer survey dataset was used to train and validate the model, achieving an accuracy of approximately 96.67% after more than 1,418,000 training cycles. However, the approach proposed in this paper surpasses the previously mentioned accuracy and achieves more than 99% accuracy with a small number of learning cycles, proving its superiority. In addition, the implementation time of the approach described in this article is shorter compared to the model presented in this article. The CAD system developed in this study showed high performance: accuracy, recall, precision and F-score reached 99.42%, 99.76%, 99.88% and 99, respectively.82% respectively. To achieve these results, the researchers adapted and implemented two DL tools, namely VGG-19 and LSTM.

DETECTION OF LUNG CANCER USING AN ALGORITHM

Shimazaki developed and validated a DL model for lung cancer detection on chest X-rays using an algorithmic method. The training dataset included 629X-ray images with 652 nodules/mass, while the test dataset included 151 X-ray images with 159 nodules/mass. On an independent testing dataset, the model achieved a sensitivity of 0.73and an average false positive rate per image (mFPI) of 0.13.[6] However, the sensitivity of the model was lower for lung tumors with blind spots than for non-overlapping tumors. The Dice coefficient, which measures the similarity between predicted and truth masks, had an average value of 0.52form alignant lesions. This indicates that the performance of the model inaccurately identifying malignant lesions was moderate. Despite these limitations, the DL-based model showed promise in detecting lung cancer on chest radiographs at low FP rates, as indicated by low MFPI. Further research and improvements may be needed to increase the sensitivity of the model, particularly for lung tumors that overlap blind spots, and to improve the dice ratio for malignant lesions.

HasanandAlKabir[7] developed algorithms to determine the spread of cancer in a patient's lungs. These algorithms used image processing techniques and statistical learning methods. The evaluation was carried out using a dataset of 198 images from the Kaggle platform and achieved an accuracy of approximately 72.2%. In comparison, the approach analyzed in this article achieves a significantly higher accuracy of 99.42%. Furthermore, the algorithm proposed in this study achieves impressive results in providing recall, precision and F-score, reaching99.76%,99.88%and 99.82%, respectively.

These results demonstrate the superiority of the method presented in this study.

Furthermore, Bhatia et al. implemented a lung cancer detection algorithm using deep residual learning on CT images. The authors used U-Net and ResNet models to extract features and identify potentially cancer-proneregions. Several classifiers have been used to predict cancer, including XGboosting, randomforest(RF), and individual predictions. The algorithm achieved 84% accuracy on the LIDC-IDRI dataset.

DETECTION OF LUNG CANCER USING THE DL AND ML METHOD

To improve the efficiency of image classification, Bansal proposed an approach that integrates deep features generated with VGG19, a DL model, with other suitable feature extraction techniques such as SIFT, SURF, ORB and Shi- Tomasi corner detectionalgorithm.ThesefeatureswerethenusedalongwithvariousMLalgorithmsforclassification.Theempiricalresultsofthe study showed that the RF classifier combined with the cooperative features extracted using the integrated approach performed better than other classifiers with an accuracy of 93.73%. This suggests that using a combination of DL and classical features produces more consistent and accurate results than relying solely on a single feature extractor.

In, Toğaçar[8] introduced a DL model in which image classes were constructed based on the DarkNet-19 model. To create image classes, weak features were selected from the feature set obtained from the DarkNet-19model using an equilibrium foraging and manta optimization approach. These weak features were then separated from there stof the features etto create the optimal feature set. A support vector machine (SVM) approach was used to classify the relevant features generated by the two optimization algorithms used. The overall performance of the classifier reached an impressive 99.69%. The evaluation results showed a remarkable area under the curve(AUC)of99.3%.Furthermore,thistechnique has demonstrated high F-rate, precision, recall and accuracy with values up to 97.1%. This DL model developed by Tougaccar showcases the effectiveness of combining different optimization algorithms and strategies to enhance the categorization ability of the dataset. The impressive performance metrics, including AUC, *F*-measure, precision, recall, and accuracy, highlight the model’s capabilities in accurately classifying images. Second Dritsas and Mesutpropose the use ML methods, specifically the rotation forest model, to efficiently identify developing lung cancer.

[3] METHODOLOGY

This section presents the classification and accurate prediction of lung cancer using technology enabled by machine learning and image processing. First you need to collect photos. A geometric mean filter is then used to preprocess the images. This ultimately leads to better image quality. The K-Means method is then used to segment the images. This segmentation makes it easier to identify the area of interest. Categorization strategies based on machine learning are then used. illustrates the classification and prediction of lung cancer using technology enabled by machine learning and image processing.

Image preprocessing plays an important role in the correct classification of clinical pictures. CT scans provide images with a variety of artifacts, including noise, which may be visible during these scans. These artifacts can be removed using image filtering methods. A geometric mean filter is applied to the input images to reduce noise. This is achieved using a method known as linear discriminant analysis (LDA), which reduces the space required for the initial data matrix. PCA and LDA are two examples of parallel transformation algorithms.[9] In contrast to supervised LDA, PCA is an unsupervised analysis method. In contrast to principal component analysis (PCA), latent dynamic analysis (LDA) aims to identify the feature subspace that maximizes class recovery. Overfitting can be avoided by placing more emphasison the repairability of data asses rather than the processing cost.

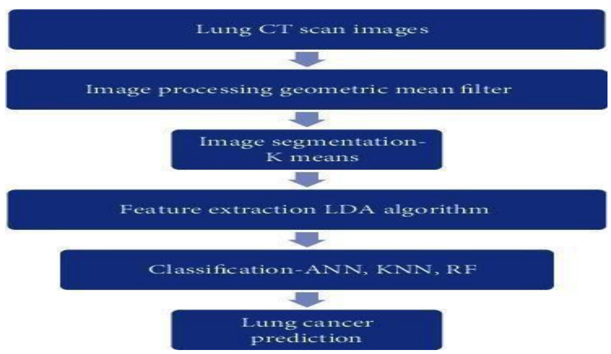

Fig. Classification and Perdition of ML Model Suggested method

The image fusion method proposed in this study is based on the MRR and DWT-PCAv methods. The MRR method is very accurate compared to the SRR method. The resulting an technique. In the present study, medical CT images are used as input.[10] The MRR technique is used to contrast images and compare different intensity levels in the initial stages of the registration process. It is observed that the proposed method provides better results than the SRR technique. The resulting image is perfectly aligned and provides more use full diagnostic information. The DWT-PCAv fusion mode is used to combine the two images after registration. Figure 1 shows the general structure of the proposed method. The resulting image is perfectly aligned and provides useful diagnostic data. DWT-PCAv fusion mode is used to combine multiple images after registration. Figure shows the source image is the improved image used in the next process of this study, and Fig. shows the structure of the proposed fusion process.

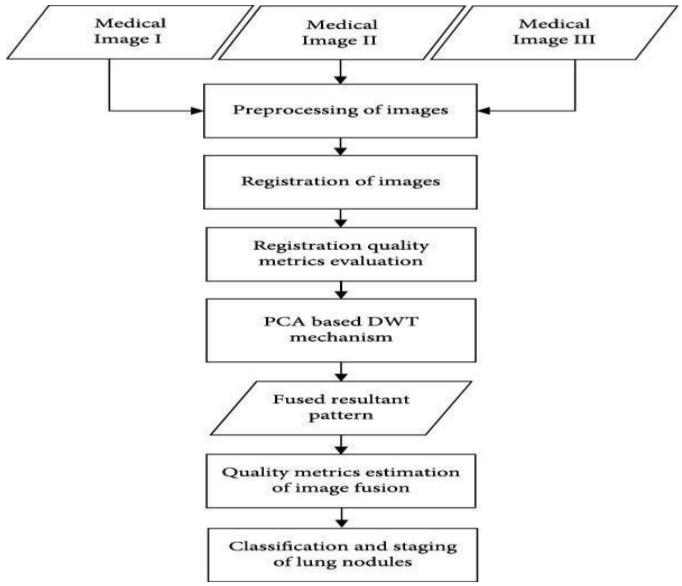

Fig. Recommended technique's block diagram

[4] RESULT

Statistical analysis of epidemiological characteristics and clinical symptoms from 842 cases in the first-layer subsystem: The comparisons of 14 features describing epidemiological features and clinical symptoms were detected (between 372 lung tumors and 470 benign lung diseases). Statistical analysis showed that there were significant differences between the two groups(P0.05) in chest tightness or chest pain, family history of cancer and lung cancer.[11] A dataset of 83 CT images from 70 different patients was used in the experimental study [*x*]. Images are preprocessed using the geometric mean filter. This results in improving image quality. Then, images are segmented using the *K*-means algorithm. This segmentation helps in the identification of the region of interest. Then, machine learning classification techniques are applied.

For performance comparison, three parameters, accuracy, sensitivity, and specificity, areused:

Accuracy=TP+TNTP+TN+FP+FN,

Sensitivity=TPTP + FN,

Specificity=TNTN+FP

where TP is true positive, TN is true negative, FP is false positive, and FN is false negative. Results of different machine learning predictors are shown in Figures Figures333–5. The accuracy of ANN is better.

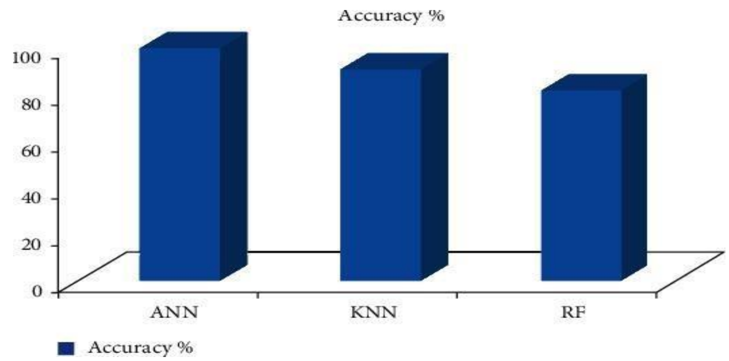

Fig. Accuracy of machine learning techniques for lung cancer detection.

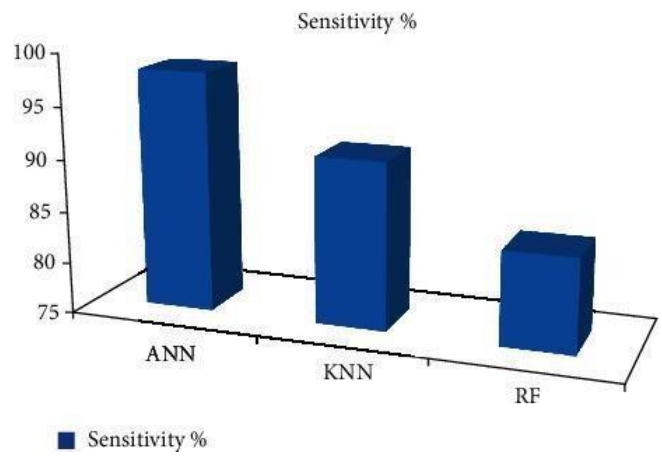

Fig. Sensitivity of machine learning techniques for lung cancer detection

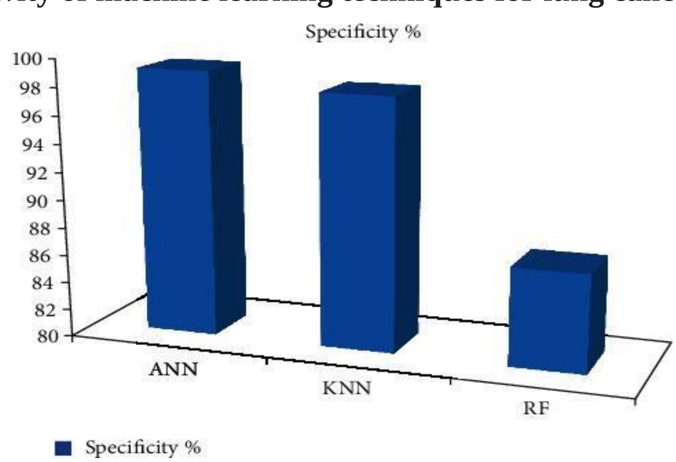

Fig. Specificity of machine learning techniques for lung cancer detection

[5] CONCLUSION

In summary, the development of an advanced multiclass lung cancer staging system using machine learning techniques represents an important step towards improving the accuracy, speed and accessibility of lung cancer diagnosis. The solution presented here leverages state-of-the-art deep learning architectures, extensive data preprocessing and interpretation methods and provides a robust framework for classifying different types and stages of lung cancer based on 3D CT images. The model's high accuracy, achieved through a careful class imbalance approach and advanced feature extraction, provides clinicians with a powerful tool for early diagnosis and accurate classification. Key achievements:

Accuracy and Reliability: The developed model shows a high level of accuracy in classifying different types and stages of lung cancer, providing reliable results that are crucial for rapid medical intervention.

Interpretability: Integrating interpretability techniques such as saliency maps and feature importance measures increases model transparency and provides clinicians with valuable insight into decision making.

Real-time Diagnosis: Implementing an API solution enables real-time diagnosis, facilitating instant medical decision- making and improving healthcare efficiency.

REFERENCES

[1].ImranNazir,1IhsanulHaq,1SalmanA.AlQahtani,2MuhammadMohsinJadoon,3andMostafaDahshan4Received6February2023;Revised6July2023;Accepted13July2023;Published16August2023AcademicEditor:EverardoVargas-Rodriguez View at:https://www.hindawi.com/journals/js/2023/6683438/

[2].“LungCancerClassificationandPredictionUsingMachineLearningandImageProcessing”Sharmila Nageswaran,1G.Arunkumar,2AnilKumarBisht,3ShivlalMewada,4J.N.V.R.SwarupKumar,5MalikJawarneh,6andEvans Asensocorrespondingauthor 7 Published online 2022 Aug 22 View at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9424001/

[3].ShuyinDuan1,HuiminCao1,HongLiu2,LijunMiao2 , Jing Wang2 , Xiaolei Zhou3 , Wei Wang1 , Pingzhao Hu4 , Lingbo Qu1,5 , Yongjun Wu1,6 Received:February10, 2020 Accepted: April 20, 2020Published:May 23, 2020 View at:https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7288961/

[4].G.Qu,D.Zhang,andP.Yan,“Medicalimagefusionbywavelettransformmodulusmaxima,”*OpticsExpress*,vol.9,no.4,pp.184–190, 2001.Viewat:PublisherSite|GoogleScholar

[5]. R. Tekade, “Lung nodule detection and classification using machine learning techniques,” *Asian Journal for Convergence in Technology*, vol. 4, 2018.View at: Google Scholar

[6].W.-J.ChoiandT.-S.Choi,“Automatedpulmonarynoduledetectionbasedonthree-dimensionalshape-basedfeature descriptor,” *Computer Methods and ProgramsinBiomedicine*,vol.113,no.1,pp.37–54,2014. Viewat:PublisherSite|GoogleScholar

[7].R.S.Valente,P.C.Cortez,E.C. Neto,J.M.Soares,V. H.C.deAlbuquerque,andJ.M.R.S.Tavares,“Automatic3DpulmonarynoduledetectioninCTimages:asurvey,”*ComputerMethodsandProgramsinBiomedicine*,vol.124,pp.91–107,2016. Viewat:PublisherSite|GoogleScholar

[8].L.Hussain,M.S.Almaraashi,W.Aziz,N.Habib,andS.U.R.SaifAbbasi,“Machinelearning-basedlungscancer detection using reconstruction independent component analysis and sparse filter features,” *Waves in Random and Complex Media*, pp. 1–26, 2021. Viewat:PublisherSite|GoogleScholar

[9]. T. V. Pyrkov, K. Slipensky, M. Barg et al., “Extracting biological age from biomedical data via deep learning: too muchof agood thing?” *Scientific Reports*, vol. 8, no. 1, pp. 1–11, 2018.Viewat:PublisherSite|GoogleScholar

[10].S.SandhiyaandY.Kalpana,“Anartificialneuralnetworks(ANN)basedlungnoduleidentificationandverification module,” Medico-Legal Update, vol. 19, no. 1, p. 193, 2019.

[11].D.PalaniandK.Venkatalakshmi,“AnIoTbasedpredictivemodellingforpredictinglungcancerusing fuzzycluster based segmentation and classification,” Journal of Medical Systems, vol. 43, no. 2, p. 21, 2019.

[12].S.Bhatia,Y.Sinha,andL.Goel,“Lungcancerdetection:a deeplearningapproach,”inSoft ComputingforProblem Solving, pp. 699–705, Springer, Singapore, 2019.

[13]. P. Joon, S. B. Bajaj, and A. Jatain, “Segmentation and detection of lung cancer using image processing and clustering techniques,” in Progress in Advanced Computing and Intelligent Engineering, pp. 13–23, Springer, Singapore, 2019.

[14].E.E.NithilaandS.S.Kumar,“SegmentationoflungfromCTusingvariousactivecontourmodels,”Biomedical Signal Processing and Control, vol. 47, pp. 57–62, 2019.

[15]. S.K.Lakshmanaprabu,S.N.Mohanty,K.Shankar, N. Arunkumar, and G. Ramirez, “Optimal deep learning model for classification of lung cancer on CT images,” Future Generation Computer Systems, vol. 92, pp. 374– 382, 2019.

[16].K. H. Yu, C. Zhang, G. J. Berry et al., “Predicting non-smallcelllung cancerprognosisby fully

automated microscopic pathology image features," Nature Communications, vol. 7, no. 1, p. 12474, 2016.