International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 11 Issue 6, June-2022, Impact Factor: 7.957

Lung Cancer Detection Using Medical Data Analytics

Mrs. J. Sathiya Jothi[1], Dr. K. Mohan[2], Mr. R. Chandru[3], Mr. S. Hariharan[4], Mr. V. Kaviyarasan[5]

[1]Assistant Professor, IT, AAMEC, Kovilvenni, Thiruvarur Dist., TN, India
[2]Associate Professor, IT, AAMEC, Kovilvenni, Thiruvarur Dist., TN, India
[3,4,5] B.Tech, IT, AAMEC, Kovilvenni, Thiruvarur Dist., TN, India

ABSTRACT

Lung cancer is among the foremost fatal sickness in developed countries, and early identification of the sickness is tough carcinoma. identification and treatment has been one in all the foremost intimidating challenges humans have encountered in recent decades. Early neoplasm identification can still save a huge quantity of lives round the world on a daily. It describes a technique for classifying respiratory organ tumors as malignant or benign that mixes a CNN withthe AlexNet Network Model. AlexNet CNN is one in all the transfer learning models. As compared to accuracy achieved by standard neural network systems, the planned CNN achieves a high degree of accuracy, that is more practical.

Keywords—Lung cancer, Machine Learning (ML), DeepLearning (DL), Computed Tomography (CT) scan images.

INTRODUCTION

Cancer is one of the riskiest illnesses a human can ever had. Lung cancer is certainly one of them. Lung cancer is a sickness that occurs due to the uncontrolled cell appear mosteffectively in the superior stages. In recent times cancer is the most serious health problem for any human being in globe and detecting lung cancer in its early tiers may be verytough. Detection of Lung nodule in early level plays powerful role in early analysis of Lung cancer. However, authenticity determination is a hard key factor as the number of candidate nodules recognized based on primary screening is high in number. In 2019, 13% of newly identified cancer is contributed with the aid of Lung Cancer alone which stands to be the second, first is prostate cancer with 20%. Five-year survival charge of affected people is 19% [1]. When the Lung Nodule is detected early and treated in right time the five-year survival rate can be increased to 55%. Normally there are two kinds of imaging used for detection of lung Nodules which incorporates simple X-Ray and computed tomography (CT).Cigarette smoking is an extremely major source of lung carcinoma. Because of smoking 80% of lung carcinoma deathsstill happen in the world. The threat of lung carcinoma increases with both the quantity of smoking and the time span of smoking. Cigar and pipe inhaling, radon gases are the next supreme source of lung carcinoma. Further threats are vocational or environmental disclosure to pre-owned smoke, chromium, cadmium, radiation, air pollution, asbestos, diesel exhaust, arsenic, and some organic chemicals. Few occupational disclosures raise threats such as chimney sweeping, roofing, rubber manufacturing, painting, etc. Genetic history also plays an important part to form or initiate lung cancer, mostly in those who suffer from lung cancer at an early age.

Lung Carcinoma has two important types of carcinomas first became Non-Small Cell Lung Cancer (NSCLC) which in additioncategories as Adenocarcinomas, Squamous, and Large cell lung cancer. It was more frequent lung cancer, out of 10 cases8 are NSCLC sort of lung carcinoma. And next, changed into Small Cell Lung Cancer (SCLC). SCLC's are offensive to returns after the preliminary round of chemotherapy [2]. Lung carcinoma has alternate tiers: Stage 1: Cancerous cells within the lung, still no longer not growing on the outer side of the lung. Stage 2: Cancer developed in the lung and close by lymph nodes and size may increase. Stage 3: Cancerous cells elevate within the lung and lymph nodes in the median of the chest. Stage 3A: Cancer found in lymph nodes, even so only on the samepart of the chest from where cancer initially developed. Stage 3B: Cancer has spread to lymph nodes on the other side of the chest or to lymph nodes above the collarbone. Stage 4: Cancer has escalatedto everywhere in the lungs, or to other organs of the body. So this machine will take CT images as enter and technique on them using numerous method like photos enhancement, segmentation, feature extraction and hit upon degrees of most cancers.

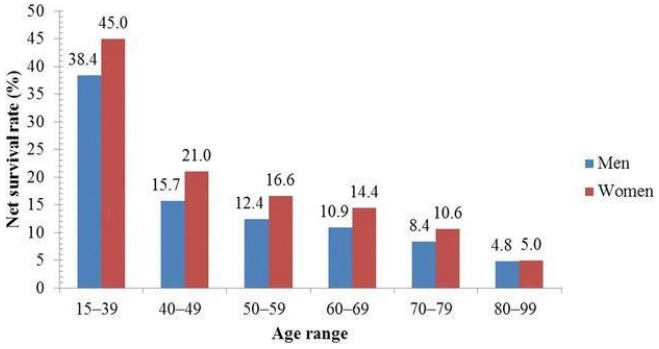

Fig 1. Prediction of Lung cancer

Several techniques are used to study lung cancer in patients. One of the simple method is to use a chest X-radiation (X-ray). Nowadays the Computed Tomography scan (CT scan)photographs is used for examination due to the CT scan images are too powerful and advanced versions of X-ray. Inthe CT scan, it takes images of internal organs, ertebrae, and the spine in 360 degrees which gives a clearer view. Lung cancer can cure, only if it's identified in the early stage.

So, Machine Learning (ML) or Deep Learning (DL)techniques are applied to detect or predict lung carcinoma like Support Vector Machine (SVM), Artificial Neural Networks (ANN), Convolutional Neural Network (CNN), and so on. This can be made a faster and more error-free diagnosis. Data quality is one of the issues that include the presence of noise, missing, or redundant data. After improving the data, the quality of the result will also improve. Many different techniques are available for data preprocessing that focuses on improving the data which is better suitable for ML or DL techniques.

RELATED WORKS

Chip M. Lyncha (1) In this paper, The abnormal image is detected by fastening on the tumour portion. The dataset in jpg format, composed of reckoned Tomography(CT)images. This is detected using the reckoned Tomography(CT) scanning and blood test reports. By blood test, the tumouris detected after the humans affected with a minimal span of4 times. So, to know the early stage of cancer, CT scanning isused. Nguyen Quoc Khanh Le(2) The end of the medical image analysis systems grounded on machine literacy styles is to ameliorate the delicacy and minimize the discovery time. EFFI- CNN uses lung CT checkup images from LIDC- IDRI and Mendeley data sets. EFFI-CNN has a unique combination of CNN layers with parameters. The EFFI- CNN achieved stylish results comparatively with ICDSSPLDCNN and EASPLD- CNN. RainerJ. Klement, PhD(3) In this paper, Three infrastructures were delved ADNN- DVH considered dosimetric information only; ADNN- com integrated multiomics information; and ADNN- com- joint combined RP2(RP grade- 2) and LC vaticination. Amulti-institutional RTOG0617 data set of 327 cases was used for external confirmation. ADNN- DVH yielded cross-validated cindicators (95 confidence intervals) of0.660(0.630-0.690) for RP2 vaticination and0.727(0.700-0.753) for LC vaticination, outperforming a generalized Lyman model for RP2(0.613(0.583-0.643)) and a generalized log-logistic model for LC(0.569(0.545-0.594)). The independent internal test and external confirmation yielded analogous results. José Raniery Ferreira Junior(4) DCNN was trained and estimated using stoked data and original data.298 images were aimlesslydivided into three groups. still, images taken from the same instance belonged to the same group. Using the DCNN, 70 of lung cancer cells were classified correctly. classification delicacy was loftiest for adenocarcinoma and smallest for scaled cell melanoma. This result may be related to the variation of images used for training. Syed saba raoof, M(5) The data was attained from the lung cancer section of Surveillance, Epidemiology, and End Results(foreseer)cancer registry.

The deep literacy models outperformed traditional machine literacy models across both bracket and retrogression approaches. achieved71.18 delicacy for the bracket approach withthe best- performing ANN model and13.50 Root Mean Squared Error(RMSE) and50.66 R2 value for the retrogression approach with the best- performing CNN model. point significance to probe the model interpretability, gaining farther sapience into the survival analysis models and the factors that are important in cancer survival period vaticination. K. Karthick(6) In this exploration, They proposed two way of process for diagnosing the presence of cancer either benign or nasty. In the first step, featuresare uprooted by using GLCM. In the alternate step, the lung cancercells are classified either benign or nasty by using Nearest Neighbour classifier. Experimental results demonstrated that the proposed approach performance is98.76 bracket delicacy for diagnosing the lung cancer data. Dakhaz Mustafa Abdullah(7) In this paper, both the identification and vaticination of lung cancer is done in early. The armature used is VGG which correspond of 16 layers. The input is passed through all the 16 layers in order to find the stylish accurate results and out of that the

delicacy and probability rate of VGG was set up more whereas for unborn work, VGG can also be compared with other infrastructures or differently the convolutional layers can be increased for better results. Ibrahim M. Nasser(8)) In this exploration, An artificial Neural Network for diagnose the presence or absence of lung cancer in mortal body movie was developed. The model was validated; it was96.67 accurate. This study showed that the neural network is suitable to diagnose lung cancer, so it can used as a diagnose tool by croakers . Smita Raut1(9) This paper proposed a new effective CT bracket system based on VGG16-T. For the first stage, we introduced a classification framework grounded on bettered VGG16- T to accommodate lungcancer typing. For the alternate stage, a boosting grounded classifier is trained to reduce FPs produced by the first stage. Three models are trained succession ally, and the training data of each model includes the mis- classified excrescences by the former model. Finally, the weight of the soft max loss functionis set, and the loss of a many types of discriminant crimes is increased to palliate the order imbalance problem of the dataset. Rashmee jihad(10) This study proposes a lung cancer opinion system grounded on reckoned tomog- raphy (CT) checkup images for the discovery of the complaint. The proposed system uses a successional approach to achieve this goal This classifier, including Haralick and LBP features, is latterly applied to the entered dataset from the CNN classifier. Eventually, if the point based method also doesn't descry cancer, the case study is healthy; else, the case study is cancerous. Chandi S. Wanigasooriya (11) To descry cancer colorful image processing ways have been founded and applied like median-wiener sludge in the preprocessing stage. In the bracket Back Propagation model, SVM(Support Vector Machines), Forward Neural Networks, complication Neural Networks are used to descry whether the bump is cancerous or Although, there are numerous similar ways which are available these days but there's still need to further develop early discovery to ameliorate delicacyleading to better survival rate.

Table I. Comparison of Lung Carcinoma Detection Techniques

S.no	Author, Title, Year	Techniques	Data Set	pros	cons
1.	Automated Classification of Lung Cancer Types from Cytological Images Using DeepConvolutional Neural Networks Atsushi Teramoto,Tetsuya Tsukamoto, YukaKiriyama, and Hiroshi Fujita,2017	deep convolutional neural network (DCNN),	*Image Dataset.* Seventy-six (76) cases of cancer cells were collectedby exfoliative or interventional cytology under bronchoscopy	It will indicate that DCNN is useful for the classification ofthe lung cancer in cytodiagnosis.	The number of images of squamous cell carcinoma was less than that for adenocarcinom aand data augmentation improved classification performance by 15%
2.	Lung cancer survival period prediction andunderstanding: Deep learning approaches Shreyesh Doppalapudi, Robin G. Qiu, Youakim Badr,2020	Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Recurrent NeuralNetworks (RNN)Random Forest classifier, Naive Bayes, SVM Classifier	Surveillance, Epidemiology, and End Results (SEER)	Accuracy is more and theerror rate is less.	There is a low performance of the random forestclassifier on the middle class.
3.	Integrating Multiomics Information in Deep Learning Architectures for Joint Actuarial Outcome Prediction in Non-Small Cell LungCancer Patients After Radiation Therapy Sunan Cui, PhD, Randall K. Ten Haken, PhD,and Issam ElNaqa,,2021.	ADNN-DVH, ADNN-com, andADNN-com- joint	multi-institutiona l RTOG061 7 data set of 327patients	ADNN-com achieved an even better performanc ethan ADNN-DVH	The dataset issmall

4.	VGG16-T: A Novel Deep ConvolutionalNeural Network with Boosting to Identify Pathological Type ofLung Cancer in Early Stage by CT Images Shanchen Pang, Fan Meng, Xun Wang, Jianmin Wang , Tao Song, XingguangWang,Xiaochun Cheng,2020	VGG16-T neuralnetwork	125 lung cancer patientsin early stage ranging in 10 year from Shandong Provincial Hospital.	The data set is enhanced by revolving, shifting and reproducing theoperations to avoid its inherent imbalance	A more accurate model is obtainedby repetitively training tumors which are more difficult to classify
5.	Novel computer-aided lung cancer detectionbased on convolutional neural network-based andfeature-based classifiers using metaheuristics Zhiqiang Guo, Lina Xu,Yujuan Si, NavidRazmjooy,2021	CNN	Lung CT-Diagnosis dataset	they do not use differentiation to solve problems in thealgorithm	Trials and error-based methodologies used for CNNs will have many drawbacks
6.	Morphological feature extraction and KNG-CNN classification of CT images for early lung cancer detection Sanjukta Rani Jena,Selvaraj Thomas George,2020	Kernel based Non-Gaussian Convolutional Neural Network (KNG-CNN). KNG-CNN	LIDC-IDRI	Superior resultsfor performing further process such as feature extraction, classification and so on.	The data in thenew variables would appear like a random numbers to human eyes
7.	Asbestosis diagnosis algorithm combining thelung segmentation method and deep learning model in computed tomography image Hyung Min Kim, Taehoon Ko, In Young Choi,Jun-Pyo Myong,2021.	long-term recurrent convolutional network deep learning model	CT images of447 patients who had been examined at Seoul St. Mary's Hospital	Advances image processing technology have enabled precise segmentation of certain areas of the image.	The differential diagnosis of theadvanced asbestosis with idiopathic pulmonary fibrosis is difficult.
8.	Semantic segmentation and detection ofmediastinal lymph nodes and anatomical structures in CT data for lung cancerstaging,David Bouget,ArveJørgensen,GabrielKiss,HaakonOlav Leira,· Thomas Lango,2019	U-Net and Mask R- CNN.	CT volume	Fusing 2D networks results in increases pixel-wise segmentation results while enabling good instance detection.	The most serious problem is the inability to explain the reasoning processand the basis of reasoning.
9.	Weakly Supervised Deep Learning for WholeSlide Lung Cancer Image Analysis Xi Wang , Hao Chen, Caixia Gan, HuangjingLin, Qi Dou,2019.	CNN	large-scale dataset in collaboration with Sun Yat-sen University Cancer Center (SUCC),	Histopathology image analysis serves as the gold standard for cancer diagnosis	it is hardly possible to tacklethe WSI classification taskby one step,

10.	Deep Learning Methods for Lung Cancer Segmentation in Whole-Slide HistopathologyImages—The ACDC@LungHP Challenge 2019 Zhang Li ,Jiehua Zhang , Tao Tan, XichaoTeng, Xiaoliang Sun , Hong Zhao, Lihong Liu,Yang Xiao,2021	CNN	dataset of 150training images and 50test images from 200 patients	results showed the potentialityof using deep learning for accurate lung cancer diagnosis on WSI.	pre-processing for the label noise during the training stage is crucial since training data wasnot accurately labeled for test set.

EXSITING SYSTEM

Previously, various ML and DL algorithms were used for image classification or prediction like Random Forest Decission Tree, ANN, SVM, K-Mean, CNN, etc. Some of the algorithms used with feature selection or feature extraction methods for improving the result. For lung cancer detection some combination of algorithms was used such as Bayesian Network with extended Markov Blanket (eMB) method (feature selectionmethod) which gave 80% accuracy Similarly, SMO with Info gain ranking method selection method and the combination of both gave 91% accuracy. Online Streaming Feature Selection was used with different algorithms but the most and better result was given by SVM-SF i.e., 88 28%.These results also depend onthe datasets such as the type of dataset they used, size, quantity, and quality of datasets, etc. For image classification, the main issue with the DL is it holds more time to train the Artificial Neural Network. Howbeit, CNN exceptionally good at this typeof classification. The accuracy of CNN with x-ray image was 95%, 2D CNN with scan image has 97% and fusion of five 3D CNN with ensemble learning has 97.35% accuracy which was much better than the other algorithms. The existing machine provide the accuracy of 86.58% and they make use of the recordsset of CT photos from a hundred twenty-five sufferers of lung cancer in early degree. The information set is superior by using revolving, moving and reproducing operations to avoid its inherent imbalance

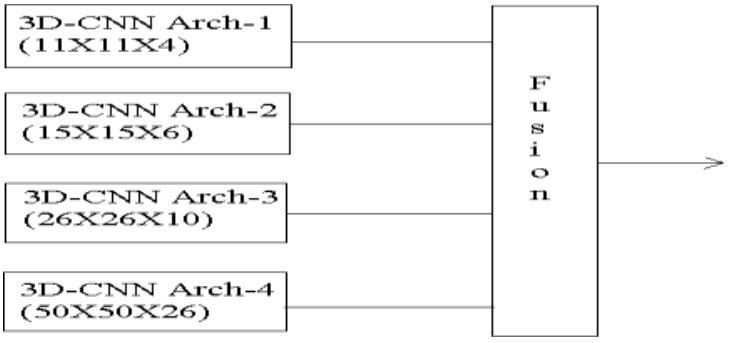

Fig. 2. 3D CNNs Fusion an existing system [14]

PROPOSED SYSTEM

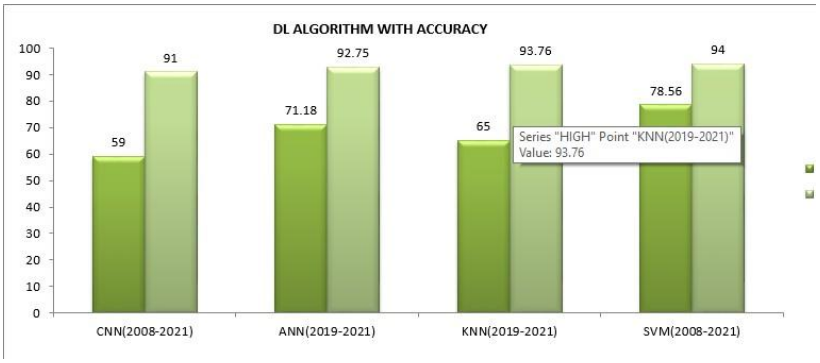

Fig. 3. Accuracy of Deep Learning Algorithm

In the proposed system the algorithm named as Alexnet CNN will be used to improve the accuracy stage of the lung cancer detection. The gadget has been skilled using a lung image of varying styles and sizes of cancerous tissues.

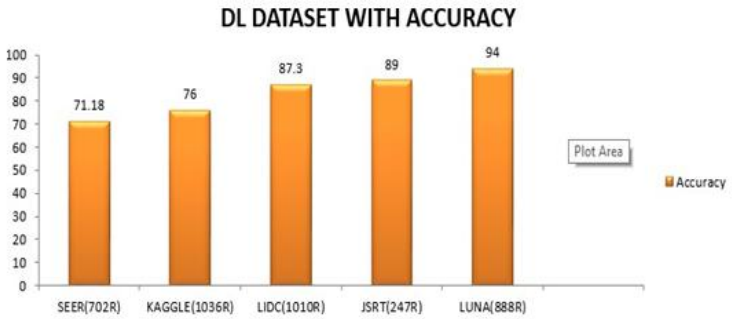

Fig. 4. Accuracy of Deep Learning Data set

IMPLEMENTATION

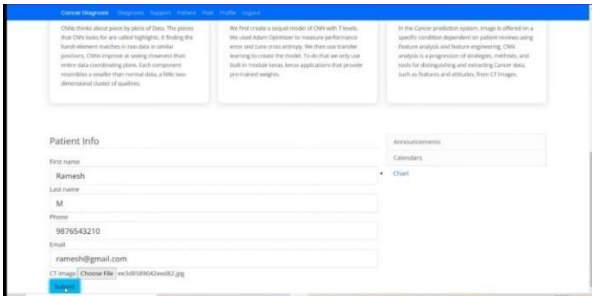

Fig. 5. Screenshot-1

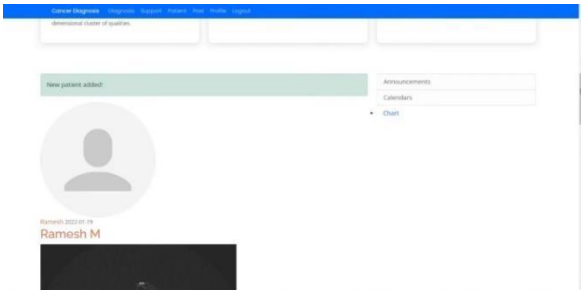

Fig. 6. Screenshot-2

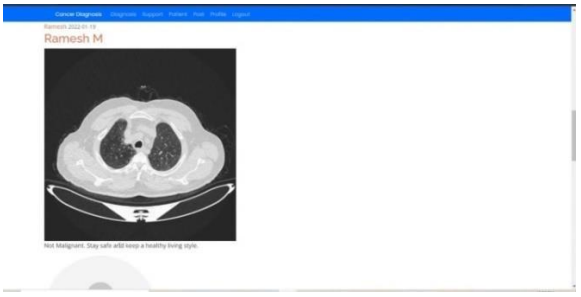

Fig. 7. Screenshot-3

Fig. 8. Screenshot-4

Fig. 9. Screenshot-5

REFERENCES

[1]. Chip M. Lyncha, Behnaz Abdollahib, Joshua D. Fuquac, Alexandra R. de Carloc," Prediction of lung cancer patient survival via supervised machine learning classification techniques"2017.
[2]. Nguyen Quoc Khanh Le 1,2,3,* , Quang Hien Kha 4, Van Hiep Nguyen 4,5 , Yung-Chieh Chen 6, Sho-Jen Cheng 6,*and Cheng-Yu Chen 1,2,6,7" Machine Learning-Based Radiomics Signatures for EGFR and KRAS Mutations Prediction in Non- Small-Cell Lung Cancer"2021.
[3]. Rainer J. Klement, PhD,*,y Michael Allga¨uer, MD,z Steffen Appold, MD,x Karin Dieckmann, MD,jj Iris Ernst, MD,{ Ute Ganswindt, MD,# Richard Holy, MD,**"Support Vector Machine-Based Prediction of Local Tumor Control After Stereotactic Body Radiation Therapy for Early-Stage Non-Small Cell Lung Cancer",2013
[4]. José Raniery Ferreira-Junior1,2•Marcel Koenigkam-Santos2 Ariane Priscilla Magalhães Tenório2,Matheus Calil Faleiros1•F ederico Enrique Garcia. Cipriano. Alexandre Todorovic Fabro2 Janne Näppi3"CT-based radiomics for prediction of histologic subtype and metastatic disease in primary malignant lung neoplasms"2019
[5]. Syed saba raoof, M A. Jabbar, Syed Aley Fathima" Lung Cancer Prediction using Machine Learning: A Comprehensive Approach"2020
[6]. Karthick, 2S.Rajkumar, 3N.Selvanathan, 4U.K.Balaji Saravanan, 5M.Murali, 6B.Dhiyanesh"Analysis of lung cancer detection based on the machine learning algorithm and IOT " 2021
[7]. Dakhaz Mustafa Abdullah&Nawzat Sdiq Ahamed "A Review of most Recent Lung cancer Detection Techniques using Machine Learning"2021
[8]. Ibrahim M. Nasser, Samy S. Abu-Naser"Lung Cancer Detection Using Artificial Neural Network "2019
[9]. Smita Raut1, Shraddha Patil2, Gopichand Shelke3"LUNG CANCER DETECTION USING MACHINE LEARNING APPROACH"2021
[10]. Rashmee jihad,Vijaya Ahire"Application of Machine Learning Techniques for the Diagnosis of Lung Cancer with ANT Colony Optimization"2015
[11]. Chandi S. Wanigasooriya, Malka N. Halgamuge, Azeem Mohammad "The Analysis of Anticancer Drug Sensitivity of Lung Cancer Cell Lines by using Machine Learning Clustering Techniques"2017
[12]. Dr. M. Kasthuri1, M. Riyana Jency2 "Lung Cancer Prediction Using Machine Learning Algorithms on Big Data: Survey " 2020
[13]. Arwinder Dhillon, Amrita Kaur, Ashima Singh "Application of Machine Learning for Prediction of Lung Cancer using Omics Data "2020
[14]. Sukhjinder Kaur "Comparative Study Review on Lung Cancer Detection Using Neural Network and Clustering Algorithm" 2015
[15]. Mr. R. Kishore "An effective and Efficient Feature Selection Method For Lung Cancer Detection " 2015
[16]. Atsushi Teramoto, Tetsuya Tsukamoto, Yuka Kiriyama, and Hiroshi Fujita, "Automated Classification of Lung Cancer Types from Cytological Images Using Deep Convolutional Neural Networks" 2017
[17]. Shreyesh Doppalapudi, Robin G. Qiu, Youakim Badr, " Lung cancer survival period prediction and understanding: Deep learning approaches" 2020
[18]. Sunan Cui, PhD, Randall K. Ten Haken, PhD,and Issam El Naqa,, "Integrating Multiomics Information in Deep
[19]. Learning Architectures for Joint Actuarial Outcome Prediction in Non-Small Cell Lung Cancer Patients After Radiation Therapy", 2021
[20]. Shanchen Pang, Fan Meng, Xun Wang, Jianmin Wang , Tao Song, XingguangWang,Xiaochun Cheng, "VGG16-T: A Novel Deep Convolutional Neural Network with Boosting to Identify Pathological Type of Lung Cancer in Early Stage by CT Images" 2020
[21]. Zhiqiang Guo, Lina Xu,Yujuan Si, Navid Razmjooy, "Novel computer-aided lung cancer detection based on convolutional neural network-based and feature-based classifiers using met heuristics", 2021

[22]. Detection Sanjukta Rani Jena, Selvaraj Thomas George, "Morphological feature extraction and KNG-CNN classification of CT images for early lung cancer detection", 2020
[23]. Hyung Min Kim, Taehoon Ko, In Young Choi, Jun-Pyo Myong, "Asbestosis diagnosis algorithm combining the lung segmentation method and deep learning model in computed tomography image" 2021
[24]. David Bouget, Arve Jørgensen, Gabriel Kiss, Haakon Olav Leira,• Thomas Lango, "Semantic segmentation and detection ofmediastinal lymph nodes and anatomical structures in CT data for lung cancer staging" 2019
[25]. Xi Wang , Hao Chen, Caixia Gan, Huangjing Lin, Qi Dou, "Weakly Supervised Deep Learning for Whole Slide Lung Cancer Image Analysis", 2019
[26]. Zhang Li ,Jiehua Zhang , Tao Tan, Xichao Teng, Xiaoliang Sun , Hong Zhao, Lihong Liu,Yang Xiao, "Deep Learning Methods for Lung Cancer Segmentation in Whole-Slide Histopathology Images—The ACDC @ Lung HP Challenge 2019", 2021