International Journal of Scientific Research and Engineering Development-– Volume 5 Issue 3, May-June 2022

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

A Review of the most Recent Lung Cancer Detection Techniques using Artificial Intelligence and Machine Learning

Abhishek*, Charu Jain**, Ankit Garg***

*(Department of Computer Science and Technology, Amity University, Haryana
Email: abhitherage16@gmail.com)

**(Department of Computer Science and Technology, Amity University, Haryana
Email: cjain@ggn.amity.edu)

***(Department of Computer Science and Technology, Amity University, Haryana
Email: agarg1@ggn.amity.edu)

----------**********----------

Abstract:

Lung cancer is one of the dangerous and ardous to detect sort of cancer. Thousands of people across the world are high-flown by this every year, and if not detected within the early stages of sickness then the chances of survival of patients are fewer. It generally causes death for both of the genders therefore there is an urgent need to accurately examine the lung nodules. Altough numerous methods have been incorporated to detect the existence of cancer in lung nodules in the early stages. This paper aims to present a comparative and in-depth analysis of various techniques for the detection of lung cancer using artificial intelligence and machine learning. Many methods have been developed in the past years to detect lung cancer , out of which many utilizes CT scan images and a few have used X – ray images. After conducting the study it is found that CT scan images are much suited to get accurate results. The results collected using deep learning based algorithms showed higher than those that were implemented using classical machine learning algorithms.

Keywords — Lung Cancer Detection , Artificial Intelligence , Machine Learning , Deep Learning , SCLC and NSCLC.

----------**********----------

I. INTRODUCTION

Lung cancer can either spread through the windpipe or the main airway in the lungs.[1]It can be caused by spreading of specific cells in the lungs in an uncontrolled way. Generally, those who suffer from emphysema have more probability to get affected with lung cancer.[2]The two major forms of lung cancer identified are Small-cell lung carcinoma (SCLC) and non-small-cell lung carcinoma (NSCLC) .SCLC is a fast growing cancer type which is often caused due to smoking tobacco products like Cigarette. NSCLC is less commonly found and the rate with it spreads is slow. If a patient is found having the characterstics of both cancer types then it is termed as mixed small cell/large cell cancer.[3]The main factor that makes this disease most deadly is the spread of cancer cells without prior symptoms. Early detection of cancer is a critical task but made possible by imaging technologies like low-dose computed tomography.[4]

Lung cancer is bring about by a tumor called nodule found in the airways of the respiratory system. These lung nodules takes the form of a spherical object that are in direct contrast with X-rays. If

 ©IJSRED: All Rights are Reserved

these lung nodules detected in an early stages then the chances of survival drastically improves. But ,interpretation of these photographs is a cumbersome task and time consuming.[5] It is made easier by the used of Computer-aided diagnosis (CAD).[6]Analysis of CT scan images is done using artificial intelligence and machine learning techniques which make the early detection of lung nodules possible.[7] The system developed for this task are referred as decision support system which analyses the images using preprocessing, segmentation, feature extraction, classification represented in Figure 1.[8]

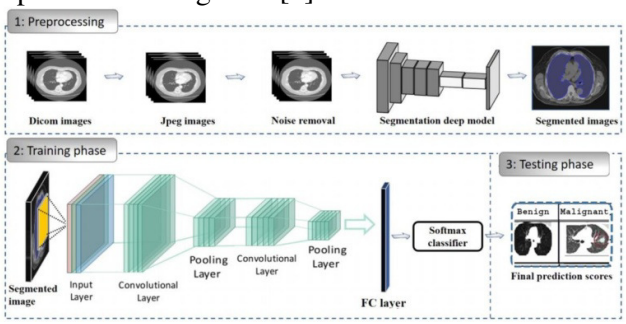

Figure 1. Framework for lung nodules detection

Numerous system have been developed to detect cancer but they lack in accuracy so there is an urgent need to develop such systems to attain 100% accuracy. It is found that the pulmonary identification and classification of cancer was done using machine learning techniques which consisted of image processing. We have studied the most recent systems developed for cancer detection and selected the best out of them and then analysed.

II. MACHINE LEARNING

Machine learning is an art of learning computers created using algorithms which enables them to take decisions and provide the result to the user.[9] It is generally considered as a sub-field of artificial intelligence.[10]Machine learning finds its usefulness in classification of complicated data and decision making. It makes the use of algorithms to learn and produce the required outputs. It derives concepts from mathematical optimization that provides the tools, theory, and implementation, and is used in several computational activities where explicit algorithms cannot be programmed. Machine learning techniques are classified mainly into three types :

Supervised learning : This type of machine learning technique finds its usefulness in solving problems related to regression such as weather forecasting and, predicting life experiences based on algorithms like Linear Regression and Random Forest.[11]Supervised machine learning techniques can be used in voice recognition, digit recognition, fraud detection and diagnostics by making the use of algorithms like Support Vector Machines, Random Forest, Nearest Neighbour. It consists of two levels : the training phase and the testing phase. The dataset consisting of known labels is used for training purpose. The algorithm analyses the relation between input values and labels to predict the test values.[12]

Unsupervised learning : This machine learning method deals with the problems related to dimensionality reduction for visualizing large amount of data.It also serves the purpose of customer segmentation and targeted marketing.[13] This method does not use any labels rather the algorithms focuses on pattern recognition.

Reinforcement learning : In this type, the algorithm attempts to get the output of a problem based on the collection of tuning parameters till the optimum output is achieved. Deep Learning and Artificial Neural networks makes use of this machine learning technique. It finds applications in areas like robot navigation, AI gaming, real-time decisions, skill acquisition and more.

III. RELATED WORKS

Researchers have successfully deployed machine learning techniques in other fields using statistical methods to build prediction models. Lung cancer detection has earlier been carried out by making use of image processing along with techniques like deep learning and neural networks.

Carrillo et al.[14] proposed a system in which a study was carried out using multi-scale and multi-omic cancer data by fusing five multi-scale and one multi-omic modalities namely RNA-Seq, miRNA-

 ©IJSRED: All Rights are Reserved

Seq, whole-slide imaging, copy number variation, and DNA methylation. The technologies used were late fusion strategy along with machine learning techniques. The model was trained independently for each modality and output was obtained by fusion of gains in an increasing manner. The final model after making use of all modalities gave an accuracy of 96.82%.The results obtained clearly shows that the performance of single-modality decision support systems can be enhanced by implementing multi-scale and multi-omic nature of cancer data which ultimately helps to improve the process of diagnosis.

Md. Alamin et al.[15] stated that Cancer being a lethal disease is caused through combinations of genetically present disease and abnormalities present in the human body. Histopathological detection is the most crucial step to examine the best treatment to be given to the patient. If detected early can greatly reduce the mortality rate. In the proposed system a hybrid ensemble feature extraction model is being used to detect the presence lung and colon cancer. This is made possible by deep feature extraction and ensemble learning techniques performed on histopathological (LC2500) lung datasets. The study showed that the model can detect lung cancer with an accuracy of 99% making it applicable for clinical trials to enhance the diagnosis of this fatal disease.

Castillo et al.[16] suggested that KnowSeq is designed to form a powerful and scalable modular software that focus on the automation and assembly of bioinformic tools that are equipped with modern functionalities. It consisted of unified environment designed to perform analysis of complex genes to identify a specific disease .The process can be carried out using raw files available at reknowned platforms or can be passed by the users as well. A set of advance algorithms is used to select the most representative genes in the listed problem. KnowSeq is designed in a way to automatically generate a detailed report of the entire task. Study of biclass breast cancer and multiclass lung cancer were carried out to understand the efficiency of this method. The method showed an accuracy of 95% for lung cancer.

A. Rehman et al.[17] stated that major causes of death round the globe is due to lung cancer and approximately five million cases are reported annually, but early detection if made possible can enhance the diagnosis process. Computed Tomography (CT) scan images proved to deliver most appropriate information for lung infections. In the system proposed a cancer detection technique was formulated by use of machine learning techniques comprising feature extraction, fusion using LBP (Local Binary Pattern), and DCT (Discrete Cosine Transform), SVM (Support Vector Machine), and K-nearest neighbor. The model successfully obtained an accuracy of 93% and 91% for SVM and K-nearest neighbours as compared to the other state-of-the-art techniques.

Shakeel et al.[18] introduced a machine learning methodology that predicts the presence of non-small cell lung cancer using CT scan images dataset by application of multilevel brightness preservation technique which effectively examines each pixel and clears away the noise present thus improving the quality of lung CT scan image. Further, segmentation of affected region from the processed image is carried using by the use of deep neural network layers. Then, using the intelligent-generalized rough set technique to hybrid spiral optimization, the effective qualities are chosen, and those features are categorised using a classifier that works as a group. The proposed strategy enhances lung cancer prediction rates, which are examined using MATLAB-based findings such recall logarithmic loss, accuracy, F-score, and mean absolute error.

A modified AlexNet (MAN) is proposed to evaluate lung anomalies in evaluated images by Bhandary et al.[19]. Chest X-Rays and lung CTs are the two types of pictures considered. The proposed MAN is tested individually on these two picture datasets. The X-Ray thoracic is assessed as normal during the initial diagnostic process, and the pneumonia class is determined. The proposed deep learning method has a 96 percent accuracy rate when

 ©IJSRED: All Rights are Reserved

compared to the other techniques discussed in this research.

Li et al.[21] used lung segmentation and rib elimination to preprocess the CXR images. Patches were extracted for each pixel in the lung field, and three CNNs were trained at varying picture resolutions. Finally, all of the data collected at various resolutions was combined using the fusion function approach. In the four fusing methods tested, the total fusion method performed the best. The proposed method can identify 99 percent of lung nodules in the JSRT database. The proposed method was precise, long-lasting, and applicable to clinical practise.

To detect lung cancer, Bhatia et al.[20] employed deep residual learning with CT scans. The scientists devised a pre-processing pathway for highlighting cancer-prone lung areas and extracting characteristics using UNet and ResNet models. To forecast how probable a CT scan is to be malignant, the feature set is given to multiple categories, including XGBoost and Random Forest. The accuracy of this study was 84 percent higher than LIDC-previous IRDI's attempts.

The scientific community has clearly given lung cancer a lot of attention, as evidenced by the state-of-the-art literature review. The majority of approaches have relied on traditional machine learning and neural network methodologies. Others applied deep learning techniques to both types of photos (CT and X-Ray). As a result, the focus of this research is on examining several methodologies in order to determine the optimal methodology for detecting lung cancer.

Table 1. Comparison among state-of-the-art Lung Cancer detection Methods

Ref.	Year	Methods	Datasets	Results
Carrilo et al.	2022	Multi-scale and multi-omic fusion with machine learning	Cancer Genome Atlas (TCGA) from GDC portal	Achieved accuracy of 96.82%
Md. Alamin et al.	2022	Feature extraction and ensemble learning	Histopathological lung and colon datasets	Accuracy of 99% for lung cancer
Castillo et al.	2021	DEG extraction, Feature selection,	Raw files available at well known platforms	Accuracy 95% for lung cancer considering 6 genes
A. Rehman et al.	2021	Feature extraction, Fusion using LBP, DCT, SVM, K-nearest neighbours	CT scan images datasets	91% for SVM and 93% for K-nearest neighbor
Li et al.	2020	multi-resolution patch-based CNNs were trained for lung nodule detection	Japanese Societyof Radiological Technology (JSRT) database	The method candetect 99% lung nodules on JSRT database
Bhandary et al.	2020	MAN is used to classify chest X-Rays images and EFT is used to classify thelung CT images.	Chest X-Ray And Lungcancer (LIDC- IDRI)	DL accuracy is96% for X-Ray images while the accuracy is 97.27% for CT images
Shakeel et al.	2020	improved deep neural network and ensemble classifier.	Database of cancer imaging archive (CIA) dataset	The proposed system detected cancer with maximum accuracy.
Bhatia etal.	2019	A number of classifiers like XGBoost and Random Forest are used.	Dataset of LungImage Database Consortium image collection (LIDC-IDRI)	accuracy 84%

IV. DISCUSSION AND ANALYSIS

This study focuses on machine learning-based lung cancer detection strategies. Whereas the majority of the works examined in the literature were based on CT scan pictures, with some using X-ray images. In

 ©IJSRED: All Rights are Reserved

both circumstances, the technique for detecting lung cancer goes through the following stages: -

Pre-processing consists of the following steps: - The CT scan image or X-ray image is used as input in the first phase of pre-processing. Then image processing techniques such as de-noising, thresholding, binarization, normalisation, and zero centering will be used. Then there's segmentation, which divides the CT scan image into comparable and distinct parts. Many segmentation methods have been used in the literature, including Region Growing, Marker Controlled Watershed, and Marker Controlled Watershed with Masking. It has been found that using the masking method to watershed produced better results. Finally, the features can be extracted in order to prepare for the classification step that follows.

Classification: During this phase, the retrieved features are passed into a classifier to determine whether they are normal or cancerous. Multi-layer perceptron (MLP), SVM, Nave Bayes, Neural Network, Gradient Boosted Tree, Decision Tree, k-nearest neighbours, multinomial random forest classifier, nave Bayes, stochastic gradient descent, and ensemble classifier have all been utilised by researchers in the literature. Table 1 shows that achieved the greatest accuracy result of around 97 percent using a multi-class SVM classifier and marker-controlled watershed-based segmentation for image segmentation. On the other hand, all of the works that used Deep Learning methods yielded high accuracy results, with employing multi-resolution patch-based CNNs achieving the highest result of around 99 percent.

CONCLUSIONS

It is advantageous to be detected with lung cancer at an early stage since therapy can then be started to prevent the disease from becoming detrimental. As a result, this work offers a comprehensive review of various machine learning algorithms for classifying lung cancers using CT scan or X-ray pictures. Many classifiers, such as MLP, SVM, Nave Bayes, Neural Network, Gradient Boosted Tree, Decision Tree, k-nearest neighbours, multinomial random forest classifier nave Bayes, stochastic gradient descent, and ensemble classifier, have been utilised by researchers in the literature. As a result, and based on the broad survey conducted in this work, it can be determined that methods utilising deep learning techniques produced higher accuracy results than other traditional machine learning techniques. Using multi-resolution patch-based CNNs, the best performance was around 99 percent.

REFERENCES

[1] Yu, K. H., Lee, T. L. M., Yen, M. H., Kou, S. C., Rosen, B., Chiang, J. H., & Kohane, I. S. (2020). Reproducible Machine Learning Methods for Lung Cancer Detection Using Computed Tomography Images: Algorithm Development and Validation. Journal of medical Internet research, 22(8), e16709.

[2] Radhika, P. R., Nair, R. A., & Veena, G. (2019, February). A Comparative Study of Lung Cancer Detection using Machine Learning Algorithms. In *2019 IEEE International Conference on Electrical, Computer and Communication Technologies (ICECCT)* (pp. 1-4). IEEE.

[3] Hussain, L., Rathore, S., Abbasi, A. A., & Saeed, S. (2019, March). Automated lung cancer detection based on multimodal features extracting strategy using machine learning techniques. In Medical Imaging 2019: Physics of Medical Imaging (Vol. 10948, p. 109483Q). International Society for Optics and Photonics.

[4] Günaydin, Ö., Günay, M., & Şengel, Ö. (2019, April). Comparison of lung cancer detection algorithms.In 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT) (pp. 1-4). IEEE.

[5] Jothilakshmi, R., & SV, R. G. (2020) Early Lung Cancer Detection Using Machine Learning And Image Processing.

[6] Zebari, D. A., Zeebaree, D. Q., Abdulazeez, A. M., Haron, H., & Hamed, H. N. A. (2020). Improved Threshold Based and Trainable Fully Automated Segmentation for Breast Cancer Boundary and Pectoral Muscle in Mammogram Images. IEEE Access, 8, 203097-203116.

[7] Zeebaree, D. Q., Abdulazeez, A. M., Zebari, D. A., Haron, H., & Hamed, H. N. A. (2021) Multi-Level Fusion in Ultrasound for Cancer Detection Based on Uniform LBP Features.

[8] Saba, T. (2020). Recent advancement in cancer detection using machine learning: Systematic survey of decades, comparisons and challenges. Journal of Infection and Public Health, 13(9), 1274-1289.

[9] Khalaf, B. A., Mostafa, S. A., Mustapha, A., Mohammed, M. A., & Abduallah, W. M. (2019). Comprehensive review of artificial intelligence and statistical approaches in distributed denial of service attack and defense methods. IEEE Access, 7, 51691-51713.

[10] Ramos-Lima, L. F., Waikamp, V., Antonelli-Salgado, T., Passos, I. C., & Freitas, L. H. M. (2020). The use of machine learning techniques in trauma-related disorders: A systematic review. Journal of psychiatric research, 121, 159-172.

[11] Abdulqader, D. M., Abdulazeez, A. M., & Zeebaree, D. Q. (2020). Machine Learning Supervised Algorithms of Gene Selection: A Review. Machine Learning, 62(03).

[12] Zantalis, F., Koulouras, G., Karabetsos, S., & Kandris, D. (2019). A review of machine learning and IoT in smart transportation. Future Internet, 11(4), 94.

 ©IJSRED: All Rights are Reserved

[13] Sulaiman, D. M., Abdulazeez, A. M., Haron, H., & Sadiq, S. S. (2019, April). Unsupervised Learning Approach-Based New Optimization K-Means Clustering for Finger Vein Image Localization. In 2019 International Conference on Advanced Science and Engineering (ICOASE) (pp. 82-87). IEEE.

[14] Carrillo-Perez, Francisco, Juan C. Morales, Daniel Castillo-Secilla, Olivier Gevaert, Ignacio Rojas, and Luis J. Herrera. 2022. "Machine-Learning-Based Late Fusion on Multi-Omics and Multi-Scale Data for Non-Small-Cell Lung Cancer Diagnosis" *Journal of Personalized Medicine* 12, no. 4: 601.

[15] Md. Alamin Talukder, Md. Manowarul Islam, Md Ashraf Uddin, Arnisha Akhter, Khondokar Fida Hasan, Mohammad Ali Moni,Machine learning-based lung and colon cancer detection using deep feature extraction and ensemble learning, Expert Systems with Applications, Volume 205,2022,117695,ISSN 0957-4174.

[16] Castillo-Secilla, D.; Gálvez, J.M.; Carrillo-Perez, F.; Verona-Almeida, M.; Redondo-Sánchez, D.; Ortuno, F.M.; Herrera, L.J.; Rojas, I. KnowSeq R-Bioc package: The automatic smart gene expression tool for retrieving relevant biological knowledge. *Comput. Biol. Med.* 2021, *133*, 104387.

[17] Rehman, M. Kashif, I. Abunadi and N. Ayesha, "Lung Cancer Detection and Classification from Chest CT Scans Using Machine Learning Techniques," *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)*, 2021, pp. 101-104, doi: 10.1109/CAIDA51941.2021.9425269.

[18] Shakeel, P. M., Burhanuddin, M. A., & Desa, M. I. (2020). Automatic lung cancer detection from CT image using improved deep neural network and ensemble classifier. Neural Computing and Applications, 1-14.

[19] Bhandary, A., Prabhu, G. A., Rajinikanth, V., Thanaraj, K. P., Satapathy, S. C., Robbins, D. E., & Raja, N.S. M. (2020). Deep-learning framework to detect lung abnormality–A study with chest X-Ray and lung CT scan images. Pattern Recognition Letters, 129, 271-278.

[20] Bhatia, S., Sinha, Y., & Goel, L. (2019). Lung cancer detection: A deep learning approach. In Soft Computing for Problem Solving (pp. 699-705). Springer, Singapore.

[21] Li, X., Shen, L., Xie, X., Huang, S., Xie, Z., Hong, X., & Yu, J. (2020). Multi-resolution convolutional networks for chest X-ray radiograph-based lung nodule detection. Artificial intelligence in medicine, 103, 101744.

 ©IJSRED: All Rights are Reserved