Deep Learning Based Lung Cancer Prediction Using Convolutional Neural Network

Mahum Fareed Khan
Electronics and Telecommunication
Symbiosis Institute of technology
Pune Maharashtra, India

Percy Dutta
Electronics and Telecommunication
Symbiosis Institute of Technology
Pune Maharashtra, India

Professor Amit Kukker
Electronics and Telecommunication
Symbiosis Institute of Technology
Pune Maharashtra, India

***Abstract*—Lung Cancer is a type of Cancer which is the most common and deadly. In recent years a significant increase in lung cancer cases has led to the need for early prediction which can not only improve the survival rate of the patient but also allow for a wide range of treatment options. Early-stage lung cancer prediction and detection involve various methods like Image testing in which CT scans, Chest X-rays, and PET scans are used to detect suspicious lesions in the lungs or Biomarker tests or Using machine learning and Deep learning models for prediction and detection. This paper provides a framework in which a Convolution neural network model is applied to the dataset consisting of CT scan images of lungs to predict if the provided lung image is showing any abnormality or not.**

***Index Terms*—Cancer, lung cancer, AI and ML models, CNN, image processing, Deep learning.**

I. Introduction

A category of illnesses known as cancer is characterized by the body's abnormal cells growing and spreading out of control. Through the circulation or lymphatic system, these abnormal cells can infect neighboring tissues, develop a mass or lump recognized as a tumor, and possibly spread to different parts of the body. There are over 100 various kinds of cancer, each with its own unique characteristics and variety of treatment choices, and they can appear in almost every organ or tissue in the body.

One form of cancer that begins in lung cells is lung cancer. It is one of the most prevalent and lethal types of cancer in the world. Non-small cell lung cancer (NSCLC) and small cell lung cancer (SCC) are the two main kinds of lung cancer. About 85% of cases of lung cancer are NSCLC, the most prevalent type; SCLC, although more uncommon, is a more aggressive variety. Since the incidence and prevalence other lung cancer may differ by geographical locations and population, a number of factors have been linked to the recent rise in lung cancer cases:

Use of tobacco: The primary risk factor for lung cancer is smoking tobacco. Lung cancer cases have surged as a result of the worldwide tobacco epidemic, particularly in nations where smoking rates have increased or stayed high. In certain areas, efforts to limit tobacco usage have been successful, but less so in others.

Environmental Factors: Lung cancer can be brought on by exposure to environmental carcinogens such secondhand smoke, radon gas, asbestos, and different occupational dangers. The danger has not totally been reduced despite continued efforts to decrease these exposures.

Improvements in Lifestyle: Alterations in lifestyle, such as food and exercise habits, can affect the chance of developing lung cancer. An higher likelihood of developing cancer of the lung is related to poor eating practices, inactivity, and obesity.

Specific genetic or molecular variables have been shown in studies that increase the risk of developing lung cancer. These variables may differ from person to person and may make a person more susceptible to the illness.

Due to the above Reasons, there is an urgent need for early detection of cancer in both lung and other types of cancer for a number of reasons.

1. Improved Survival Rates: Early cancer discovery frequently results in more effective treatment options and improved survival rates. When lung cancer is found in its early stages, it may be possible to remove the tumor surgically or get alternative curative therapies.

2. Treatment Options: A larger variety of treatments are available with early diagnosis. Palliative care and symptom management may be the extent of treatment for cancer in its final stages.

3. Quality of Life: By restricting the quantity of surgery or other therapies necessary and lowering the negative effects of harsh treatments, early identification can help preserve a patient's quality of life.

4. Decreased Healthcare expenditures, which include Cancer diagnosis at an advanced stage are often followed by additional and expensive treatments, which increases the cost of healthcare. Potentially reducing the financial burden on patients and healthcare systems is early detection.

5. Screening and Prevention: Screening programs can identify those who are more inclined to pick up lung cancer, such as smokers, and those with a family history of the disease, through early prediction and diagnosis. These individuals can then be closely watched and encouraged to take preventive measures such as quitting smoking.

Various Methods that are needed for early cancer detection and prediction include identifying presumed lung abnormalities, and imaging techniques like CT scans, PET scans, and chest X-rays are used. Using biomarker testing, it is possible to find particular molecules or compounds in the blood, sputum, or lung tissue that could be signs of cancer. Risk assessment and screening programs for people at higher risk, like smokers who are currently smoking or have recently quit. the use of machine learning and artificial intelligence to search through medical data for early indications of cancer.

II. Literature Review

[1]This study offers an approach for identifying aberrant lung tissue growth. A tool with a greater chance of detection is considered in order to attain high accuracy.[2]This study shows that the DL algorithms have the potential to improve the accuracy and efficiency of lung cancer detection. However, further research is needed to address the challenges of training and deploying DL algorithms in clinical practice.[4]In this paper Lung cancer cell classification using artificial neural network (ANN) is a promising new method for early detection and diagnosis of lung cancer. ANNs are a type of machine learning algorithm that can be trained to recognize patterns in data.[5]The paper "Intelligent Breast Cancer Diagnosis Using Hybrid GA-ANN" by Fadzil Ahmad, published in the Proceedings of the Fifth International Conference on Computational Intelligence, Communication Systems and Networks in 2013, proposes a hybrid genetic algorithm (GA) and artificial neural network (ANN) approach for breast cancer diagnosis.[6] This study gives an overview of Lung cancer screening with CAD is an effective way to detect lung cancer early, when it is most treatable. However, it is important to weigh the risks and benefits of screening before making a decision.[7]Characterization of CT cancer lung image using image compression algorithms and feature extraction is a promising new method for early detection and diagnosis of lung cancer.[10]Deep learning is a type of machine learning that uses artificial neural networks to learn from data. Neural networks are inspired by the structure and function of the human brain, and they are able to learn complex patterns from data.Deep learning has been shown to be very effective for a variety of tasks, including image classification. This has led to a growing interest in using deep learning for lung cancer diagnosis.[11]The paper "Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks" by A. R. Mohamed, M. S. Kamel, and M. E. Abdel-Nasser, published in Measurement in 2019, proposes a new method for lung cancer detection from CT images using improved profuse clustering and deep learning instantaneously trained neural networks (DITNNs). The proposed method has the potential to be a valuable tool for lung cancer detection. The method is very accurate and can be used to detect lung cancer nodules in real-time. This could help radiologists to diagnose lung cancer earlier and more accurately.[12] This study gives an overview of Deep learning is still a relatively new technology, but it has the potential to revolutionize the way that cancer is diagnosed and classified. Deep learning-based CAD systems and other deep learning-based methods could help radiologists diagnose cancer earlier and more accurately, and they could also help improve the treatment of cancer patients

III. Methodology

Lung Cancer is one of the most common types of cancer and a dangerous one so early detection of cancer is necessary to reduce the impact of lung cancer as it provides the best opportunity for a successful treatment and improved outcomes. The below flowchart describes the framework for detection. The above flow diagram explains how the framework works. The first stage in lung cancer detection is data collection and preprocessing because getting accurate data for analysis is very important to obtain satisfying results. For this paper Kaggle dataset of chest CT scan is used. The dataset is divided into 3 subcategories:-

1. train dataset which is used for training.

2. test dataset used for testing the model.

3. Valid dataset used to check if our model is predicting accurately after training our model on the training dataset. each folder contains 3 folders of different chest cancer types(adenocarcinoma, large cell carcinoma,squamous cell carcinoma) and 1 folder of normal CT-Scan images (normal).

After acquiring the dataset, is then split into the train, test, and valid dataset and then we build the CNN model. numerous convolutional blocks, with batch normalization and max-pooling layers in between each pair of convolutional layers that defines a block. The parameters block1, block2, block3, and block4—which each have a boolean value indicating whether it should be included in the model—determine these blocks. Two convolutional layers with 64, 128, 256, and 512 filters each make up each block. These layers have ReLU activation,'same' padding, and a 3x3 kernel size. The specified boolean values are used to set the trainable parameter, which gives you control over whether a block should be trainable or frozen during training. A max-pooling layer and a batch normalization layer are added to downscale and normalize the features after each pair of convolutional layers. After Building the model the train data is fed as input and the hyperparameter considered is:- epochs ie taken as 50 and Batch size as 64 with image size as 224 x 224.

IV. Result

The above algorithm was then evaluated using various parameter such as Accuracy, Precision, F1 score, recall, confusion matrix.

Report	Accuracy	Precision Score	Recall	F1 Score
CNN Model	0.91	1.00	0.91	0.95

The above table represents the classification of how the CNN model has performed on the provided dataset.A graphical representation of the prediction is shown below.

From the above figure, we can conclude that with an increase in the epochs, loss decreases.

From the above figure, we can conclude that with an increase in the epochs, Accuracy increases.

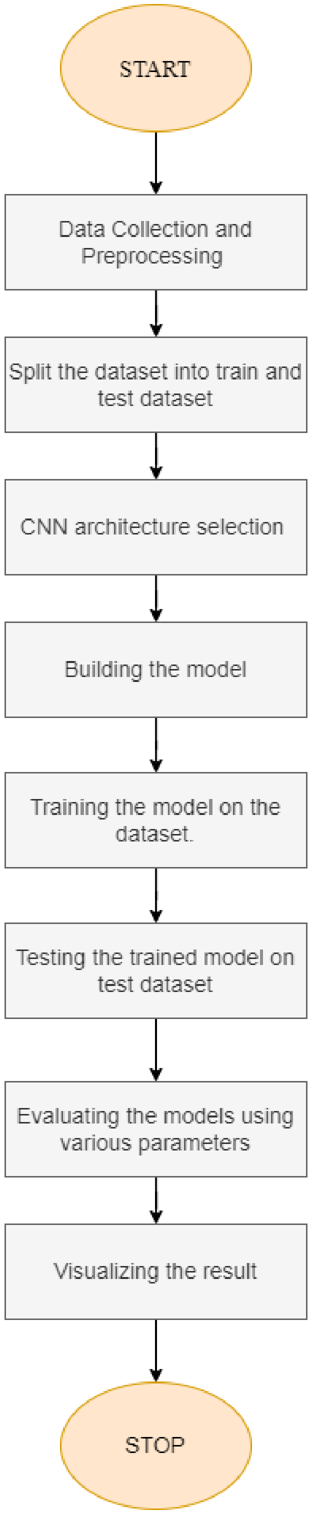

Fig. 1. Flow Diagram

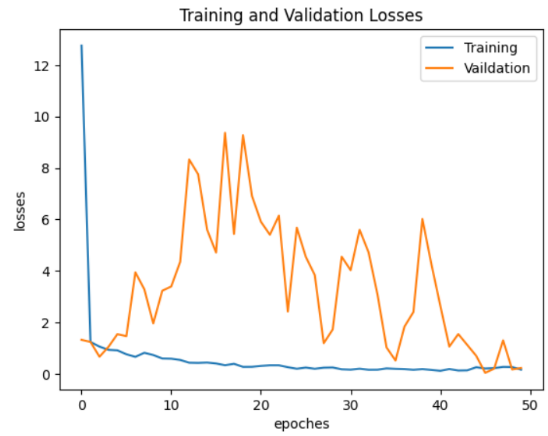

Fig. 2. Training and Validation losses

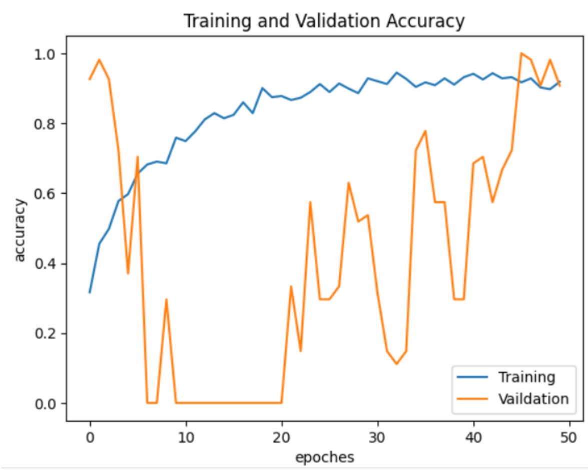

Fig. 3. Training and Validation Accuracy

V. CONCLUSION

A crucial field of medical study and application is lung cancer diagnosis, which has the potential to save lives and enhance patient outcomes. Early detection of lung cancer is crucial because it opens the door to more efficient treatment choices and raises survival rates. This Paper Provides a detailed explanation of how we can detect lung cancer using Deep learning CNN. As a custom-designed architecture was used for image classification Multiple convolutional blocks with batch normalization, max-pooling, and fully connected layers are part of this design.A max-pooling layer and a batch normalization layer are added to downscale and normalize the features after each pair of convolutional layers, and 91% accuracy was achieved.

REFERENCES

[1] Pandian, R., Vedanarayanan, V., Kumar, D. R., & RajaKuma, R. (2022). Detection and classification of lung cancer using CNN and Google net. Measurement: Sensors, 24, 100588. https://doi.org/10.1016/j.measen.2022.100588

[2]] Song, Zhao QingZeng, Luo Lei, XingKe Dou, XueChen, Using deep learning for classification of lung nodules on computed tomography images, J.Healthc. Eng. (2017) 1–7, 2017.

[3]] H. Abdel-Razeq, F. Attiga, A. Mansour, Cancer care inJordan, Hematology/Oncology and Stem Cell Therapy 8 (2) (2015) 64–70.

[4]] A. Abdulla, S. Shaharum, Lung Cancer Cell classificationmethod Using Artificial Neural Network, Information Eng Let, 2012, pp. 50–58, 2012

[5]] F. Ahmad, N. Isa, M. Noor, Z. Hussain, IntelligentBreastCancer diagnosis using hybrid GA-ANN, in Proceedings of the Fifth International Conference on ComputationalIntelligence, Communication Systems and Networks,2013, pp. 9–12.

[6] B. Al Mohammad, P. Brennan, C. Mello-Thoms, Areview of lung cancer screening and the role of computer-aided detection, Clin. Radiol. 18 (1) (2017) 46–51, 72(6): 433-442. A descriptive study. EuropJofOncology Nursing.

[7] R. Pandian, T. Vigneswaran, S. LalithaKumari, Characterization of CT cancer lung image using image Compression algorithms and Featu extraction,J. Sci. Ind. Res. 75 (December) (2016) 747–751

[8] Wang, L. (2022). Deep learning techniques to diagnose lung cancer. Cancers, 14(22), 5569. https://doi.org/10.3390/cancers14225569

[9] Eschmann, S.M.; Friedel, G.; Paulsen, F.; Reimold, M.; Hehr, T.; Budach, W.; Langen, H.J.; Bares, R. 18F-FDG PET for assessment of therapy response and preoperative re-evaluation after neoadjuvant radiochemotherapy in stage III non-small cell lung cancer. Eur. J. Nucl. Med. Mol. Imaging 2007, 34, 463–471.

[10] Hosseini, S. H., Monsefi, R., & Shadroo, S. (2023). Deep learning applications for lung cancer diagnosis: A systematic review. Multimedia Tools and Applications. https://doi.org/10.1007/s11042-023-16046-w

[11] Shakeel, P. M., Burhanuddin, M. A., & Desa, M. I. (2019). Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks. Measurement, 145, 702–712. https://doi.org/10.1016/j.measurement.2019.05.027

[12] R. Fakoor, F. Ladak, A. Nazi, and M. Huber, “Using deep learning to enhance cancer diagnosis and classification,” in Proceedings of the International Conference on Machine Learning, 2013

[13] H. Yang, H. Yu, and G. Wang, “Deep learning for the classification of lung nodules,” arXiv preprint arXiv:1611.06651, 2016.

[14] F. Ciompi, K. Chung, S. J. van Riel, A. A. A. Setio, P. K. Gerke, C.Jacobs, E. T. Scholten, C. Schaefer-Prokop, M. M. Wille, A. Marchiano et al., “Towards automatic pulmonary nodule management in lung cancer screening with deep learning,” arXiv preprint arXiv:1610.09157, 2016

[15] G. L. F. d. Silva, A. O. d. Carvalho Filho, A. C. Silva, A. C. d. Paiva, and M. Gattass, “Taxonomic indexes for differentiating malignancy of lung nodules on ct images,” Research on Biomedical Engineering, vol. 32, no. 3, pp. 263–272, 2016.

[16] “Kaggle data science bowl 2017.” [Online]. Available: https://www.kaggle.com/c/data-science-bowl-2017

[17] H. Demuth and M. Beale, “Matlab neural network toolbox user guide version 6. the mathworks inc,” 2009.

[18] D. Maji, A. Santara, P. Mitra, and D. Sheet, “Ensemble of Deep Convolutional Neural Networks for Learning to Detect Retinal Vessels in Fundus Images,” CoRR, vol. 1603.04833, 2016

[19] D. Sharma, T. G. Newman, and W. S. Aronow, “Lung cancer screening: history, current perspectives, and future directions,” Arch Med Sci, vol. 11, no. 5, pp. 1033–1043, 2015.

[20]] J.A.H. Bartholomai, et al., Lung cancer survival prediction via machine learning regression, classification, and statistical techniques, in: IEEE Access, 2018