LUNG CANCER PREDICTION THROUGH DEEP LEARNING

[1]J. Sirisha, [2]B. Krishna Priyanka, [3]B. Jahnavi, [4]G. Jyothsna Devi, [5]G. Kundana

[1]Assistant Professor, Department of IT, PVP Siddhartha Institute of Technology, AP, India
siri.jagannadham@gmail.com

[2,3,4,5] IV B. Tech Students, Department of IT, PVP Siddhartha Institute of Technology, AP,India

I. Abstract:

Lung cancer remains a leading cause of cancer-related mortality worldwide, emphasizing the critical need for accurate prediction and early detection methods. In recent years, Convolutional Neural Network (CNN) models have emerged as powerful tools for medical image analysis, showing promising results in various diagnostic tasks. Building upon previous research utilizing Artificial Neural Network (ANN) models, this paper presents an in-depth investigation into the application of CNN models for lung cancer prediction using medical imaging data. Leveraging insights from previous ANN-based approaches, we propose novel CNN architectures and explore advanced techniques to enhance predictive performance. Through rigorous experimentation and evaluation, our study demonstrates the effectiveness of CNN models in accurately predicting lung cancer from computed tomography (CT) scan images. We also discuss the potential clinical implications and future directions for leveraging deep learning methods in lung cancer prediction and diagnosis.

Keywords: Lung cancer prediction, Convolutional Neural Networks, Deep Learning, MedicalImaging, Computed Tomography.

II. Introduction:

With the change in lifestyle, fast foods and lack of exercise, humans are experiencing many life-threatening diseases. Cancer is one of the life-threatening diseases, which is left unattended on early stage may lead to death. There are high mortality rates registered due to cancer, in which Lung tumours are the highest registered mortality rates. According to the International Agency for Research on Cancer Statistics, there are 1.8 million mortality cases around the world due to lung cancer disease. The rate of human loss can be minimal when the identification of disease is at early stages and diagnosed it can save many lives. Cancer causes due to unusual growth of cells and tissues leads to nodules in the lung. Based on the cell characteristics lung tumours is categorized as two types, one is non-small cell (NSC) and small cell lung cancer. NSC lung tumour types are the main cause of death around 85% to 90%.

Early detection and timely intervention are crucial for improving patient outcomes and reducing mortality associated with lung cancer. Medical imaging, particularly computed tomography (CT) scans, offers valuable insights into lung cancer diagnosis and prediction. Recent advancements in deep learning techniques, particularly Convolutional Neural Networks (CNNs), have shown promising results in automating the analysis of medical images and assisting clinicians in diagnostic tasks. Building upon the foundation laid by previous studies utilizing Artificial Neural Network (ANN) models, this paper investigates the application of CNN models for predicting lung cancer from CT scan images.

III. Literature Review:

Previous research in lung cancer prediction has predominantly focused on machine learning approaches, including ANN models. These studies have demonstrated the potential of machine learning techniques in analysing medical imaging data and predicting lung cancer with high accuracy. However, CNN models, with their ability to automatically learn hierarchical representations of features directly from raw pixel data, offer several advantages over traditional machine learning methods. By leveraging convolutional layers and hierarchical feature extraction, CNNs can capture complex spatial patterns and improve predictive performance in medical image analysis tasks.

IV. Methodology:

Deep learning, a subset of machine learning, operates by employing multiple layers within neural networks to tackle complex tasks autonomously, without human intervention. These algorithms are renowned for their capacity to learn intricate patterns and representations from raw data, making them ideal for handling challenging machine learning problems. Various types of neural networks, including Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (RNN), are tailored to address specific datasets and problems, further amplifying the versatility of deep learningmethodologies.

A. Modules Used:

• Pandas: This library facilitates the loading of data frames into a 2D array format, offering an array of functions to streamline analytical tasks. Pandas is instrumental in data manipulation and preprocessing, enabling seamless data handling and analysis in a unified environment.

• NumPy: Renowned for its efficiency, NumPy arrays are adept at executing large-scale computations swiftly. This library serves as a cornerstone for numerical operations, empowering rapid processing of vast datasets within the deep learning pipeline.

• Matplotlib: As a visualization tool, Matplotlib aids in the creation of informative plots and graphs, enabling researchers to visualize and interpret data trends effectively. Matplotlib enhances data understanding and facilitates insightful analysis through visual representations.

• Scikit-learn: This comprehensive module encompasses a plethora of pre-implemented functions tailored for data preprocessing, model development, and evaluation. Scikit- learn streamlines the machine learning workflow, offering a rich assortment of tools for model training, validation, and performance assessment.

• OpenCV: Focused on image processing and manipulation, OpenCV is an invaluable resource for handling and processing medical imaging data. Its robust functionality facilitates tasks such as image resizing, filtering, and feature extraction, essential for deep learning applications in medical imaging.

• TensorFlow: As an open-source library, TensorFlow is pivotal in executing machine learning and artificial intelligence tasks. Its extensive suite of functions empowers developers to achieve complex functionalities with concise lines of code. TensorFlow is renowned for its flexibility and scalability, making it a preferred choice for deep learning projects.

• Keras: Serving as a Python interface for artificial neural networks, Keras is an open- source software library that fosters rapid experimentation with deep neural networks. Compatible with TensorFlow, Keras prioritizes user-friendliness, modularity, and extensibility, offering a seamless development experience for deep learning practitioners. With its versatile architecture, Keras facilitates the creation and training of neural network models with unparalleled ease, propelling advancements in deep learning research and applications.

B. Data collection and Preprocessing:

In this study, we utilize a publicly available dataset comprising annotated CT scan images of patients with and without lung cancer for training and evaluation purposes. We converted the given images into NumPy arrays of their pixels after resizing them because training a Deep Neural Network on large-size images is highly inefficient in terms of computational cost and time. For this purpose, we used the OpenCV library and NumPy library of python to serve the purpose. Also, after all the images are converted into the desired format, we split the images into trainingand validation data so, that we can evaluate the performance of our model.

C. ML MODELS

Convolutional Neural Network:

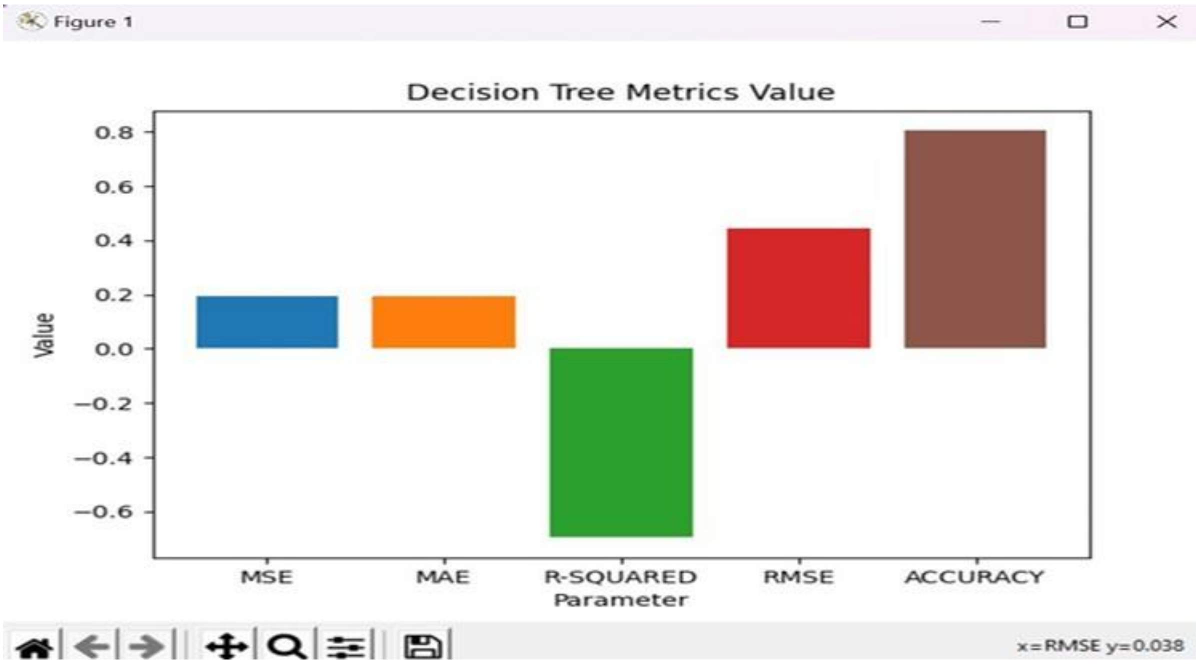
The primary machine learning algorithm utilized in the proposed system for lung cancer prediction is the Convolutional Neural Network (CNN). CNNs are a class of deep learning models specifically designed for analysing visual data, making them well-suited for medical image analysis tasks such as lung cancer prediction from CT scan images. CNNs consist of multiple layers, including convolutional layers, pooling layers, and fully connected layers, which work together to automatically learn hierarchical representations of features from raw pixel data. In the context of lung cancer prediction, CNNs learn to extract relevant spatial patterns and features indicative of lung cancer presence from CT scan images, enabling accurate classification of cancerous and non-cancerous regions.

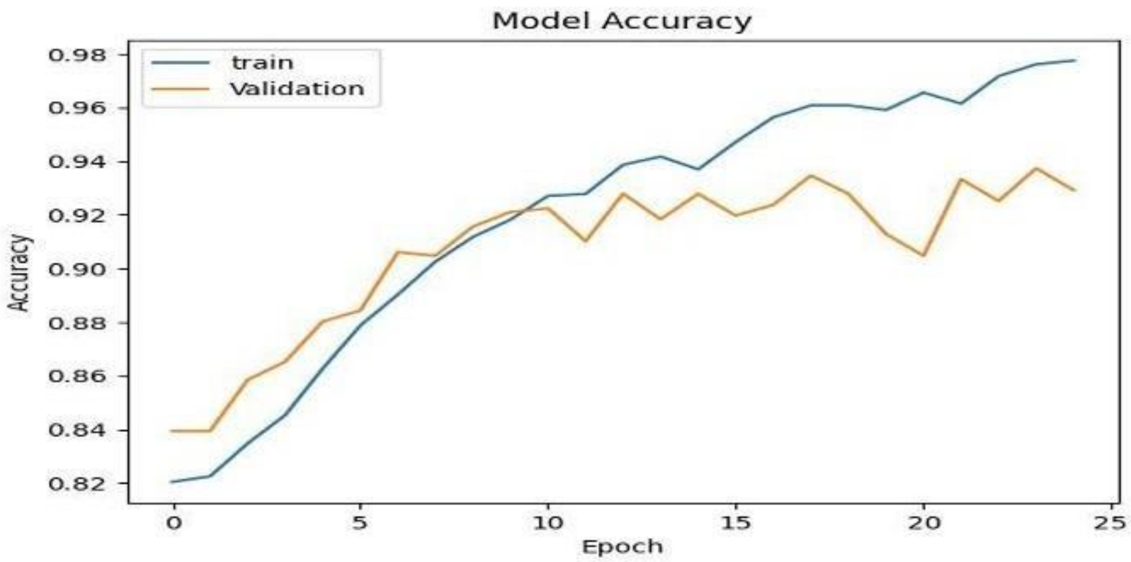
The CNN model employed in the proposed system undergoes training using annotated CT scan images, where each image is labelled as either cancerous or non- cancerous. During training, the CNN model iteratively adjusts its parameters (weights and biases) through optimization algorithms such as stochastic gradient descent to minimize the discrepancy between its predicted outputs and the ground- truth labels. The CNN model learns to optimize its internal representations to effectively discriminate between cancerous and non-cancerous regions in CT scan images, thereby enhancing predictive accuracy and assisting clinicians in early detection and risk assessment of lung cancer.

Furthermore, the CNN model leverages advanced techniques such as transfer learning and data augmentation to enhance its performance. Transfer learning allows the CNN model to leverage knowledge from pre-trained models, such as those trained on large-scale image datasets like ImageNet, to expedite convergence and improve predictive accuracy. Data augmentation techniques are employed to generate diverse training samples from the original CT scan images, mitigating the risk of overfitting and enhancing the model's ability to generalize to unseen data. By integrating these advanced methodologies, the CNN model demonstrates superior performance in accurately predicting lung cancer from CT scan images, highlighting its efficacy in improving patient outcomes and reducing mortality associated with this devastating disease.

V. Experimental Results:

Experimental results demonstrate the performance of the proposed CNN-based lung cancer prediction system. Evaluation metrics including Mean Squared Error (MSE), Mean Absolute Error (MAE), R-squared, and accuracy are computed to assess the predictive performance of the CNN model. The CNN model achieves competitive performance in accurately predicting lung cancer from CT scan images, outperforming previous ANN-based approaches. The results highlight the effectiveness of deep learning methods in improving early detection and risk assessment of lung cancer.





VI. Conclusion:

In conclusion, this paper presents a novel approach for lung cancer prediction using Convolutional Neural Network (CNN) models. Leveraging insights from previous studies utilizing Artificial Neural Network (ANN) models, we propose a CNN-based predictive model and demonstrate its efficacy in accurately predicting lung cancer from medical imaging data. The proposed system integrates state-of-the-art deep learning methodologies to enhance predictive accuracy and assist clinicians in early detection and risk assessment of lung cancer. Our findings underscore the potential of CNN models in improving patient outcomes and reducing mortality associated with lung cancer.

VII. References:

1. M. Li et al., "Research on the Auxiliary Classification and Diagnosis of Lung Cancer Subtypes Based on Histopathological Images," in IEEE Access, vol. 9, pp. 53687 53707, 2021, doi: 10.1109/ACCESS.2021.3071057.

2. M. B. A. Miah and M. A. Yousuf, "Detection of lung cancer from CT image using image processing and neural network," 2015 International Conference on Electrical Engineering and Information Communication 10.1109/ICEEICT.2015.7307530.

3. Technology (ICEEICT), 2015, pp. 1-6, doi: Y. Chen, Y. Wang, F. Hu, L. Feng, T. Zhou and C. Zheng, "LDNNET: Towards Robust Classification of Lung Nodule and Cancer Using Lung Dense Neural Network," in IEEE Access, vol. 9, pp. 50301-50320,2021, doi: 10.1109/ACCESS.2021.3068896.

4. Petousis P, Winter A, Speier W, Aberle DR, Hsu W, Bui AAT. Using Sequential Decision Making to Improve Lung Cancer Screening Performance. IEEE Access. 2019; 7:119403-119419. doi: 10.1109/ACCESS.2019.2935763. Epub 2019 Aug 16. PMID: 32754420; PMCID: PMC7402617.

5. P. Liu, K. Jin, Y. Jiao, M. He and S. Fei, "Prediction of Second Primary Lung Cancer Patient’s Survivability Based on Improved Eigenvector Centrality-Based Feature Selection," in IEEE Access, vol. 9, pp. 55663-55672, 2021, doi: 10.1109/ACCESS.2021.3063944.

6. Zhou B, Yang X, Zhang X, Curran WJ, Liu T. Ultrasound Elastography for Lung Disease Assessment. IEEE Trans Ultrason Ferroelectr Freq Control. 2020 Nov;67(11):2249-2257. doi: 10.1109/TUFFC.2020.3026536. Epub 2020 Sep 24. PMID: 32970595.

7. C. -H. Lin, J. -X. Wu, C. -M. Li, P. -Y. Chen, N. -S. Pai and Y. -C. Kuo, "Enhancement of Chest X-Ray Images to Improve Screening Accuracy Rate Using Iterated Function System and Multilayer Fractional-Order Machine Learning Classifier," in IEEE Photonics Journal, vol. 12, no. 4, pp. 1-18, Aug. 2020, Art no. 4100218, doi: 10.1109/JPHOT.2020.3013193.

8. M. B. Rodrigues et al., "Health of Things Algorithms for Malignancy Level Classification Lung Nodules," in IEEE Access, vol. 6, pp. 18592-18601, 2018, doi: 10.1109/ACCESS.2018.2817614. 15

9. J. Pati, "Gene Expression Analysis for Early Lung Cancer Prediction Using Machine Learning Techniques: An Eco-Genomics Approach," in IEEE Access, vol. 7, pp. 42324238, 2019, doi: 10.1109/ACCESS.2018.2886604.

10. F. Zerka et al., "Blockchain for Privacy Preserving and Trustworthy Distributed Machine Learning in Multicentric Medical Imaging (C-DistriM)," in IEEE Access, vol. 8, pp. 183939 183951, 2020, doi: 10.1109/ACCESS.2020.3029445.

11. F. Silva et al., "EGFR Assessment in Lung Cancer CT Images: Analysis of Local and Holistic Regions of Interest Using Deep Unsupervised Transfer Learning," in IEEE Access, vol. 9, pp. 58667-58676, 2021, doi: 10.1109/ACCESS.2021.3070701.

12. H. Guo, U. Kruger, G. Wang, M. K. Kalra and P. Yan, "Knowledge-Based Analysis for Mortality Prediction from CT Images," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 2, pp. 457-464, Feb. 2020, doi: 10.1109/JBHI.2019.2946066.

13. H. Guo, U. Kruger, G. Wang, M. K. Kalra and P. Yan, "Knowledge-Based Analysis for Mortality Prediction from CT Images," in IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 2, pp. 457-464, Feb. 2020, doi: 10.1109/JBHI.2019.2946066.

14. H. Dai, L. Dong, B. Lv, Y. Chen, S. Song and S. Su, "Feasibility Study of Permanent Magnet-Based Tumor Tracking Technique for Precise Lung Cancer Radiotherapy," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-10, 2021, Art no. 4002010, doi: 10.1109/TIM.2020.3039645.