**Lung Cancer Prediction using AI-ML**

*(13 size) A Project Based Learning Report Submitted in partial fulfilment of the requirements for the award of the degree*

*of*

**Bachelor of Technology**

**in The Department of Computer Science and Engineering**

**22AIP3305A DEEP LEARNING**

Submitted by

**2210030358: N. AKSHITA REDDY**

**2210030029: G. SUMA**

**2210030240: P. RAJA KEERTHANA**

Under the guidance of

DR.SAI SUDHA GADDE



Department of Computer Science and Engineering

Koneru Lakshmaiah Education Foundation, Aziz Nagar

Aziz Nagar – 500075

FEB - 2025.

**Introduction**

Lung cancer remains one of the most prevalent and deadly forms of cancer worldwide. Early detection plays a crucial role in improving survival rates, but traditional diagnostic methods often face challenges such as human error, limited specificity, and time-consuming procedures. This project explores the integration of machine learning techniques to enhance the accuracy and efficiency of lung cancer detection, utilizing advanced models to analyze medical imaging data effectively.

This study evaluates multiple Machine learning models, including logistic regression, decision trees, neural networks, and deep learning architectures like CNNs, ResNet, and AlexNet. While existing deep learning models such as ResNet and AlexNet have shown promise, achieving accuracies of 85.2% and 86.5% respectively, they still face challenges related to computational demands, interpretability, and overfitting. The proposed system aims to overcome these limitations by incorporating feature extraction, preprocessing techniques, and classification models tailored for early diagnosis.

The research findings indicate that logistic regression achieves the highest accuracy of 90.29%, making it a viable solution for lung cancer detection. By leveraging scalable real-time data processing and automation, this project not only enhances diagnostic precision but also reduces reliance on manual evaluation. The integration of machine learning and automation paves the way for more reliable, faster, and cost-effective lung cancer diagnosis, ultimately improving patient outcomes.

In conventional lung cancer detection, medical professionals rely on CT scans, biopsies, and radiographic analysis to identify abnormalities in lung tissues. However, these methods often require expert interpretation and are prone to variability in diagnosis. Machine learning models address these challenges by analyzing vast amounts of imaging data to detect patterns that may be imperceptible to the human eye. By automating feature extraction and classification, these models enhance the precision of lung cancer detection, enabling early intervention and improving the chances of successful treatment. Additionally, techniques such as data augmentation and transfer learning help overcome issues like limited dataset availability, ensuring robust and scalable diagnostic systems.

This project not only focuses on improving detection accuracy but also emphasizes the need for interpretability in AI-driven healthcare applications. Deep learning models, particularly CNNs, often function as "black boxes," making it difficult for clinicians to trust their decision-making processes. By incorporating explainable AI (XAI) techniques, this research aims to provide transparency in model predictions, allowing medical practitioners to understand the reasoning behind the diagnosis. Furthermore, addressing common challenges like noise reduction in medical images and model generalization ensures that the proposed solution remains effective across diverse datasets. Ultimately, this study contributes to the ongoing efforts to integrate AI in healthcare, making lung cancer detection more accessible, efficient, and reliable.

**Literature Review/** **Application Survey**

Lung cancer remains one of the most significant causes of mortality worldwide, necessitating the development of advanced diagnostic techniques for early detection and treatment. Traditional diagnostic approaches, including biopsy, radiographic imaging, and expert-based evaluation, often suffer from subjectivity, time consumption, and the risk of human error. As a result, the application of machine learning (ML) and deep learning (DL) in lung cancer detection has gained substantial traction in recent years. Various studies have demonstrated the effectiveness of ML and DL models in improving diagnostic accuracy, reducing false positives, and enhancing interpretability in medical imaging.

1. **Traditional Approaches and Their Limitations**

Conventional methods for lung cancer diagnosis involve the use of X-rays, CT scans, PET scans, and tissue biopsies. Radiologists and oncologists analyze these images to identify nodules and assess malignancy. However, these techniques pose several challenges, including:

1. **Subjectivity in Interpretation** – Variability among radiologists can lead to discrepancies in diagnosis.
2. **Time-Consuming Process** – Manually analyzing CT scan images is labour-intensive.
3. **High False Positive/Negative Rates** – Nodules may be misclassified, delaying diagnosis and treatment.
4. **Data Limitations** – Small sample sizes and data inconsistencies affect model reliability.

To address these limitations, automated techniques leveraging ML and DL have been proposed to enhance accuracy, speed, and consistency in lung cancer detection.

1. **Machine Learning in Lung Cancer Detection**

Machine learning techniques, including supervised and unsupervised learning, have been widely used for lung cancer classification. Popular ML models include:

1. **Logistic Regression** – Used for binary classification tasks, logistic regression has been effective in distinguishing between cancerous and non-cancerous cases.
2. **Decision Trees and Random Forest** – These ensemble learning methods improve classification by reducing overfitting and increasing model generalization.
3. **Support Vector Machines (SVM)** – SVM has demonstrated success in medical image classification by mapping data points in high-dimensional spaces.
4. **K-Nearest Neighbours (KNN)** – Used for simple classification tasks, although its efficiency decreases with high-dimensional data.
5. **Gradient Boosting Algorithms (XGBoost, LightGBM, AdaBoost)** – These models enhance predictive performance by combining multiple weak learners.
6. **Deep Learning Models for Lung Cancer Detection**

Deep learning has revolutionized medical imaging analysis, enabling more accurate and scalable solutions. Some widely adopted DL techniques include:

1. **Convolutional Neural Networks (CNNs)** – CNNs excel in medical image analysis by extracting hierarchical features. Studies have shown that CNNs achieve higher accuracy than traditional ML models in CT scan classification.
2. **Residual Networks (ResNet)** – ResNet overcomes the vanishing gradient problem, allowing deeper architectures to be trained efficiently.
3. **AlexNet and VGGNet** – These architectures have been used for image recognition and segmentation in medical imaging.
4. **3D CNNs** – Unlike 2D CNNs, 3D CNNs analyze volumetric data, making them suitable for processing CT scan slices in three dimensions.
5. **Long Short-Term Memory Networks (LSTMs)** – Used in sequential data analysis, LSTMs improve classification when CT scan slices need temporal correlation.
6. **Comparison of Existing Studies**

Several studies have explored ML and DL models for lung cancer prediction.

1. **Jeon et al.** developed a comparative model for forecasting lung cancer mortality trends, emphasizing the role of smoking history in lung cancer risk prediction.
2. **El-Baz et al.** proposed a CAD system for early lung nodule detection, improving early-stage diagnosis.
3. **Silvana et al.** utilized YOLOv8 and TNM staging techniques for lung cancer subtype classification.
4. **Tao et al.** developed a risk factor-based model to predict lung cancer occurrence based on epidemiological data.
5. **LeCun et al.** pioneered CNNs for pattern recognition, laying the groundwork for modern AI-driven medical diagnostics.

Despite these advancements, challenges persist in model generalization, interpretability, and computational efficiency. Overfitting, data scarcity, and noise in medical images remain major hurdles.

1. **Application of Machine Learning in Clinical Practice**

ML applications in lung cancer detection extend beyond image classification:

1. **Computer-Aided Diagnosis (CAD) Systems** – These systems assist radiologists in detecting lung nodules and analyzing malignancy with greater precision.
2. **Predictive Analytics** – Machine learning models help assess risk factors such as smoking, genetic predisposition, and environmental influences.
3. **Survival Rate Prediction** – AI models predict patient prognosis based on historical data, aiding personalized treatment plans.
4. **Real-Time Monitoring** – ML-powered wearable devices track patient symptoms, enabling early intervention.
5. **Automated Image Segmentation** – Deep learning models segment lung nodules, improving the detection of malignant growths.
6. **Challenges and Future Directions**
7. **Data Availability and Quality** – The performance of ML models heavily depends on high-quality, labeled datasets. Data augmentation and transfer learning can mitigate this issue.
8. **Model Interpretability** – Black-box nature of DL models limits clinical trust. Explainable AI (XAI) techniques are needed to enhance transparency.
9. **Computational Complexity** – High-resolution CT scans require substantial computing power, which may limit deployment in resource-constrained settings.
10. **Bias in Training Data** – Models trained on imbalanced datasets may fail to generalize to diverse populations, necessitating bias mitigation strategies.
11. **Regulatory Compliance** – AI-driven diagnostic tools must adhere to medical regulations and ethical standards to ensure patient safety.

Machine learning and deep learning have significantly improved lung cancer detection, addressing challenges faced by traditional diagnostic methods. While logistic regression, decision trees, and CNNs have demonstrated promising results, ongoing research is essential to optimize model accuracy, interpretability, and scalability. The integration of AI-driven CAD systems, real-time monitoring, and predictive analytics in clinical settings has the potential to revolutionize lung cancer diagnosis, ultimately leading to better patient outcomes. Future research should focus on enhancing model reliability, reducing bias, and ensuring ethical deployment in healthcare applications.

**References**

[1] R. L. Siegel, K. D. Miller, and A. Jemal, ‘‘Cancer statistics, 2017,’’ Cancer J. Clin., vol. 67, no. 1, pp. 7– 30, Jan. 2017.

1. R. L. Siegel, K. D. Miller, and S. A. Fedewa, ‘‘Colorectal cancer statistics,’’ Cancer J. Clin., vol. 68, no. 1, pp. 7–30, 2018.

[3] J. Lortet-Tieulent, I. Soerjomataram, J. Ferlay, M. Rutherford, E. Weiderpass, and F. Bray, ‘‘International trends in lung cancer incidence by histological subtype: Adenocarcinoma stabilizing in men but still increasing in women,’’ Lung Cancer, vol. 84, no. 1, pp. 13–22, Apr. 2014.

[4] A. Wang, H. Wang, Y. Liu, M. Zhao, H. Zhang, Z. Lu, Y. Fang, X. Chen, and G. Liu, ‘‘The prognostic value of PD–L1 expression for non-small cell lung cancer patients: A meta-analysis,’’ Eur. J. Surgical Oncol., vol. 41, no. 4, pp. 450–456, Apr. 2015.

[5] R. L. Siegel, K. D. Miller, and A. Jemal, ‘‘Cancer statistics, 2015,’’ CA, Cancer J. Clin., vol. 60, no. 5, pp. 277–300, 2010.

[6] P. Naresh and R. Shettar, ‘‘Image processing and classification techniques for early detection of lung cancer for preventive healthcare: A survey,’’ Int. J. Recent Trends Eng. Technol., vol. 11, no. 1, p. 595, 2014.

[7] D. Sharma and G.Jindal, ‘‘Computer aided diagnosis system for detection of lung cancer in CT scan images,’’ Int. J. Comput.

Elect. Eng., vol. 3, no. 5, pp. 714–718, Sep. 2011.

[8]D. Bhatnagar, A. K. Tiwari, V. Vijayarajan, and A. Krishnamoorthy, ‘‘

[9] F. Passiglia, M. Calandri, and F. Guerrera, “Lung Cancer in Italy,” Journal of Thoracic Oncology, Volume 14, Issue 12, 2019, Pages 2046-2052, ISSN 1556-0864,

[10] R. L. Siegel, K. D. Miller and A. Jemal, "Cancer statistics 2017", Cancer J. Clin., vol. 67, pp. 7-30, Jan. 2017.

[11] Clark SB, Alsubait S. Non–Small Cell Lung Cancer. [Updated 2023 Sep 4]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: https[://www.](http://www.ncbi.nlm.nih.gov/books/NBK562307/)ncb[i.nlm.nih.gov/books/NBK562307/](http://www.ncbi.nlm.nih.gov/books/NBK562307/)

[12] I. Sluimer, A. Schilham, M. Prokop and B. van Ginneken, "Computer analysis of computed tomography scans of the lung: a survey," in IEEE Transactions on Medical Imaging, vol. 25, no. 4, pp. 385-405, April 2006, doi: 10.1109/TMI.2005.862753.

[13] Hochhegger B, Marchiori E, Sedlaczek O, Irion K, Heussel CP, Ley S, Ley-Zaporozhan J, Soares Souza A Jr, Kauczor HU. MRI in lung cancer: a pictorial essay. Br J Radiol. 2011 Jul;84(1003):661-8. doi: 10.1259/bjr/24661484. PMID: 21697415; PMCID: PMC3473490.

[14] A. El-Baz, G. Gimel’farb, R. Falk and M. A. El- Ghar, "A New CAD System for Early Diagnosis of Detected Lung Nodules," 2007 IEEE International Conference on Image Processing, San Antonio, TX, USA, 2007, pp. II - 461-II - 464, doi: 10.1109/ICIP.2007.4379192.

[15] K. Manikandan, "Blob based segmentation for lung CT image to im proving CAD performance," 2014 International Conference on Recent Trends in Information Technology, Chennai, India, 2014, pp. 1-6, doi: 10.1109/ICRTIT.2014.6996157.

[16] N. A. Pande and D. Bhoyar, "A comprehensive review of Lung nodule identification using an effective Computer-Aided Diagnosis (CAD) Sys tem," 2022 4th International Conference on Smart Systems and Inven tive Technology (ICSSIT), Tirunelveli, India, 2022, pp. 1254- 1257, doi: 10.1109/ICSSIT53264.2022.9716327.

[17] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition", Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, 1998.

[18] J. Jeon, T. R. Holford, D. T. Levy, E. J. Feuer, P. Cao, J. Tam, L. Clarke, J. Clarke, C. Y. Kong, and R. Meza, “Smoking and lung cancer mortality in the United States from 2015 to 2065: A comparative modeling approach,” Ann. Intern. Med., vol. 169, no. 10, pp. 684–693, Nov. 2018. doi: 10.7326/M18-1250

[19] Alaa Wehbe, Irene Minetti and Silvana Dellepiane , “Enhanced Lung Cancer Detection and TNMStaging Using YOLOv8 and TNMClassifier: An Integrated Deep Learning Approach for CT Imaging”, Diten, University of Genoa, Italy 2EBIT- ESAOTE Group, Genoa, Italy,doi: 10.1109/ACCESS.2024.DOI.

[20] Jinpeng Li,Yaling Tao and T. Cai, “Predicting Lung Cancers Using Epidemiological Data: A Generative-Discriminative Framework”, University of Chinese Academy of Sciences, Ningbo 315010, China,doi: 10.1109/JAS.2021.1003910.