IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁑ Peer-reviewed & Refereed journal ⁑ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.15183

Lung Vision:Early Detection and Classification of Lung Cancer

Shreedhar S Hirekurabar[1], Prof. Suma N R[2]

PG Student, Dept. of MCA, Bangalore Institute of Technology, Bengaluru-560 004, Karnataka, India[1] Assistant Professor, Dept. of MCA, Bangalore Institute of Technology, Bengaluru-560 004, Karnataka, India[2]

Abstract: Lung cancer remains a leading cause of cancer-related mortality, where early diagnosis is critical for improving survival outcomes. Computed Tomography (CT) imaging is commonly used for lung cancer screening; however, manual interpretation of CT scans is time-consuming and susceptible to diagnostic variability. This paper presents LUNG VISION, an automated lung cancer detection and classification system based on machine learning and deep learning techniques. The proposed framework includes image preprocessing, lung region segmentation, feature extraction, and classification. Preprocessing techniques such as resizing, normalization, and noise reduction are applied to enhance CT image quality. Machine learning classifiers including Decision Tree, Random Forest, and Gaussian Naive Bayes are implemented using Histogram of Oriented Gradients features. In parallel, deep learning models such as Convolutional Neural Networks, DenseNet, and ResNet are employed through transfer learning to automatically learn discriminative features from CT images. The system classifies CT scans into normal, benign, and malignant categories and provides severity-related insights to support clinical decision-making. Experimental results indicate that deep learning models achieve superior diagnostic accuracy and robustness compared to traditional machine learning methods. The system is deployed via a web-based interface to assist radiologists in early and reliable lung cancer diagnosis.

Keywords: Lung Cancer Detection, Computed Tomography (CT), Deep Learning, Convolutional Neural Network (CNN), Machine Learning.

I. INTRODUCTION

Lung cancer is one of the leading causes of cancer-related mortality worldwide, largely due to late-stage diagnosis and limited early symptoms. Early detection and accurate classification are essential for improving patient survival rates and enabling effective treatment planning.

Computed Tomography (CT) imaging is widely used for lung cancer screening because it provides high-resolution cross-sectional images of lung tissues. CT scans allow early identification of lung nodules and abnormal growth patterns. However, manual interpretation of CT images is time-consuming, subjective, and highly dependent on radiologist expertise. The increasing volume of medical imaging data further increases diagnostic workload and the likelihood of delayed or inaccurate decisions.

Recent advances in artificial intelligence, particularly Machine learning and Deep learning, have shown significant potential in medical image analysis. Traditional machine learning methods rely on handcrafted feature extraction and classifiers such as Decision Tree and Random Forest, offering reasonable performance but limited capability in capturing complex image patterns. Deep learning models, especially Convolutional Neural Networks (CNNs), automatically learn hierarchical features directly from image data, resulting in improved accuracy and robustness for image classification tasks.

This paper presents LUNG VISION, an automated lung cancer detection and classification system that integrates machine learning and deep learning techniques for CT image analysis. The proposed system preprocesses CT scans, segments lung regions, and classifies images into normal, benign, and malignant categories. By automating the diagnostic process, LUNG VISION aims to reduce diagnostic variability, support radiologists in early detection, and improve the reliability and efficiency of lung cancer diagnosis.

1.1 Problem Statement

Early diagnosis of lung cancer remains a major challenge in modern healthcare due to the subtle nature of early-stage symptoms and the complexity of medical image interpretation. Computed Tomography (CT) imaging is commonly used for lung cancer screening, but the manual analysis of CT scan images requires significant time and expert knowledge. As the number of medical imaging studies continues to grow, radiologists face increased workload, leading to potential diagnostic delays and variability in interpretation.

DOI: 10.17148/IJARCCE.2026.15183

Most existing computer-aided diagnostic approaches depend on handcrafted feature extraction and conventional machine learning models, which are often insufficient for capturing intricate spatial and textural patterns in lung CT images. These methods typically demonstrate limited accuracy and reduced adaptability across diverse datasets. Consequently, there is a strong need for an automated lung cancer detection and classification system that leverages deep learning techniques to accurately differentiate between normal, benign, and malignant lung conditions, while improving diagnostic efficiency and supporting early clinical decision-making.

1.2 Proposed Methodology

The proposed LUNG VISION system follows a structured and automated pipeline for lung cancer detection and classification from CT scan images. The overall methodology consists of image acquisition, preprocessing, lung segmentation, feature extraction, classification, and result visualization.

1.2.1 Image Acquisition

Lung CT scan images are collected from a publicly available dataset obtained from Kaggle. The dataset includes images categorized into normal, benign, and malignant classes to support multi-class classification.

1.2.2 Image Preprocessing

Preprocessing is performed to enhance image quality and ensure uniformity across the dataset. This includes resizing images to a fixed dimension, normalization to standardize intensity values, and noise reduction to remove unwanted artifacts. These steps improve model convergence and classification accuracy.

1.2.3 Lung Segmentation

Lung region segmentation is applied to isolate the region of interest from the background. This step reduces irrelevant information and ensures that the models focus only on lung tissues, improving diagnostic reliability.

1.2.4 Feature Extraction

For traditional Machine learning models, Histogram of Oriented Gradients (HOG) is used to extract discriminative texture and shape features from segmented lung images. These features are then used as input for classification algorithms.

1.2.5 Classification

Two parallel classification approaches are implemented:
Machine Learning Models: Decision Tree, Random Forest, and Gaussian Naive Bayes classifiers are trained using extracted HOG features.
Deep Learning Models: CNN, DenseNet, and ResNet architectures are implemented using transfer learning to automatically learn high-level spatial features directly from CT images.

1.2.6 Result Generation

The trained models classify CT images into normal, benign, or malignant categories. Performance is evaluated using accuracy, precision, recall, and F1-score. The system isdeployed through a web-based interface for real-time prediction and visualization.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁂ Peer-reviewed & Refereed journal ⁂ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.15183

II. SYSTEM DESIGN

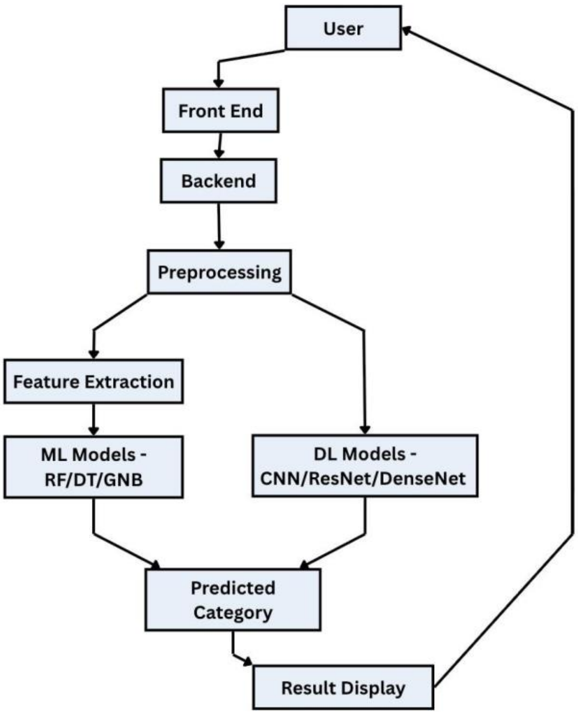

Fig -1: System Architecture

The system architecture illustrates the overall workflow of the proposed LUNG VISION lung cancer detection and classification framework. The process begins with the user, who interacts with the system through a web-based front-end interface. The front end allows the user to upload lung CT scan images and initiate the diagnostic process. These inputs are then forwarded to the backend, which acts as the core controller responsible for managing data flow and coordinating all processing modules.

Once the image is received by the backend, it is passed to the preprocessing module, where essential image enhancement operations such as resizing, normalization, and noise removal are performed. Preprocessing ensures that the input images are standardized and suitable for further analysis. After preprocessing, the workflow branches into two parallel paths to support both traditional machine learning and deep learning approaches.

In the first path, the preprocessed images undergo feature extraction, where discriminative features are extracted and supplied to machine learning classifiers such as Random Forest (RF), Decision Tree (DT), and Gaussian Naive Bayes (GNB). In the second path, the preprocessed images are directly fed into deep learning models, including Convolutional Neural Networks (CNN), ResNet, and DenseNet, which automatically learn high-level spatial features from the image data.

The outputs from both machine learning and deep learning models are combined to generate the predicted category, classifying the input CT scan as normal, benign, or malignant. Finally, the classification result is presented to the user through the result display module, completing the diagnostic cycle. This architecture ensures efficient processing, parallel model evaluation, and real-time result visualization, making the system suitable for practical clinical decision support.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.15183

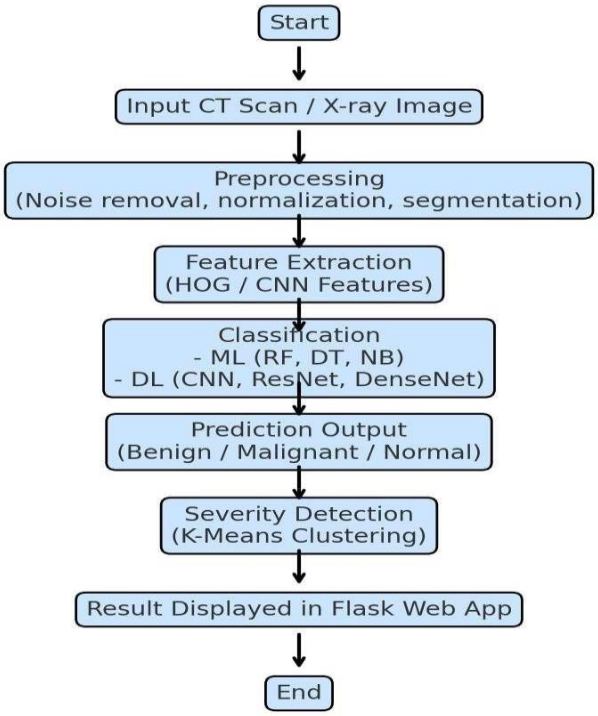

Fig -2: Flow Chart

The flowchart represents the complete operational workflow of the proposed LUNG VISION lung cancer detection and classification system. The process begins with the input of a lung CT scan or X-ray image, which is provided by the user through the system interface. Once the image is uploaded, it enters the preprocessing stage to prepare the data for accurate analysis.

During preprocessing, essential image enhancement operations such as noise removal, normalization, and lung region segmentation are performed. These steps improve image quality, reduce irrelevant background information, and ensure uniformity across the dataset. The preprocessed image is then forwarded to the feature extraction stage, where meaningful features are extracted. Histogram of Oriented Gradients (HOG) features are used for traditional machine learning models, while convolutional features are automatically learned when deep learning models are applied.

In the classification stage, both machine learning and deep learning approaches are employed. Machine learning classifiers such as Random Forest (RF), Decision Tree (DT), and Naive Bayes (NB) classify the image based on extracted features. In parallel, deep learning models including Convolutional Neural Networks (CNN), ResNet, and DenseNet perform classification directly on image data. The system predicts the category of the input image as normal, benign, or malignant.

Following classification, severity detection is carried out using K-Means clustering to assess the seriousness level of detected cancer cases. Finally, the diagnostic result and severity information are displayed to the user through a Flask-based web application, completing the workflow. This structured flow ensures accurate, efficient, and real-time lung cancer diagnosis.

III. SOFTWARE TESTING

3.1 Performance Evaluation of Machine Learning Algorithms

3.1.1 Random Forest Classifier

The Random Forest algorithm was trained using lung CT scan images after extracting relevant features through the Histogram of Oriented Gradients method. The classifier demonstrated strong predictive capability and maintained high accuracy on unseen test data, indicating good generalization.

Performance Metrics:

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.15183

Accuracy: 94%
Precision: 0.94 (benign), 0.95 (malignant), 0.94 (normal)
Recall: 0.93 (benign), 0.97 (malignant), 0.94 (normal)
F1 Score: 0.93 (benign), 0.96 (malignant), 0.94 (normal) Confusion Matrix:

	Predicted Benign	Predicted Malignant	Predicted Normal
Benign	382	12	21
Malignant	6	421	8
Normal	18	11	395

Table -1: RFC Confusion Matrix

3.1.2 Decision Tree Classifier

The Decision Tree model was also applied to the lung CT dataset to evaluate its classification capability. Compared to Random Forest, its performance was moderate.

Performance Metrics:

Accuracy: 70%
Precision: 0.67 (benign), 0.76 (malignant), 0.69 (normal)
Recall: 0.66 (benign), 0.77 (malignant), 0.68 (normal)
F1 Score: 0.66 (benign), 0.76 (malignant), 0.68 (normal)

Confusion Matrix:

	Predicted Benign	Predicted Malignant	Predicted Normal
Benign	284	59	78
Malignant	62	349	31
Normal	88	49	288

Table -2: DTC Comfusion Matrix

3.1.3 Gaussian Naive Bayes

The Gaussian Naive Bayes classifier recorded the lowest performance among the machine learning algorithms tested.

Performance Metrics:

Accuracy: 50%
Precision: 0.60 (benign), 0.55 (malignant), 0.45 (normal)
Recall: 0.18 (benign), 0.58 (malignant), 0.70 (normal)
F1 Score: 0.28 (benign), 0.56 (malignant), 0.55 (normal)

3.2 Performance Evaluation Of Deep Learning Models

3.2.1 Convolutional Neural Network (CNN)

The CNN model delivered outstanding performance on lung CT scan images by LEARNING complex spatial and hierarchical features automatically.

Performance Metrics:

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 1, January 2026
DOI: 10.17148/IJARCCE.2026.15183

Accuracy: 98%
Precision: 0.98 (all classes) Recall: 0.97 (all classes) F1 Score: 0.97 (all classes)

Confusion Matrix:

	Predicted Benign	Predicted Malignant	Predicted Normal
Benign	2589	21	26
Malignant	18	2664	20
Normal	23	17	2676

Table -1: CNN Confusion Matrirx

3.2.2 DENSENET
DenseNet also showed reliable performance, though slightly lower than CNN.

Performance Metrics:
- Accuracy: 89%
- Precision: 0.87 (benign), 0.90 (malignant), 0.91 (normal)
- Recall: 0.86 (benign), 0.89 (malignant), 0.90 (normal)
- F1 Score: 0.86 (benign), 0.89 (malignant), 0.90 (normal)

3.2.3 Residual Network (RESNET)

The ResNet model was evaluated to analyze its effectiveness in classifying lung CT scan images. By incorporating residual connections, ResNet enables deeper network training while addressing issues such as vanishing gradients. This makes it suitable for learning complex representations from medical imaging data.

Performance Metrics:
Accuracy: 84%
Precision: 0.83 (benign), 0.85 (malignant), 0.84 (normal)
Recall: 0.82 (benign), 0.84 (malignant), 0.85 (normal)
F1 Score: 0.82 (benign), 0.84 (malignant), 0.84 (normal)

IV. RESULTS AND DISCUSSION

The performance of the proposed LUNG VISION system is evaluated using both machine learning and deep learning models on lung CT scan images. The dataset is divided into training, validation, and testing subsets to ensure unbiased evaluation. Model performance is assessed using standard metrics including accuracy, precision, recall, and F1-score.

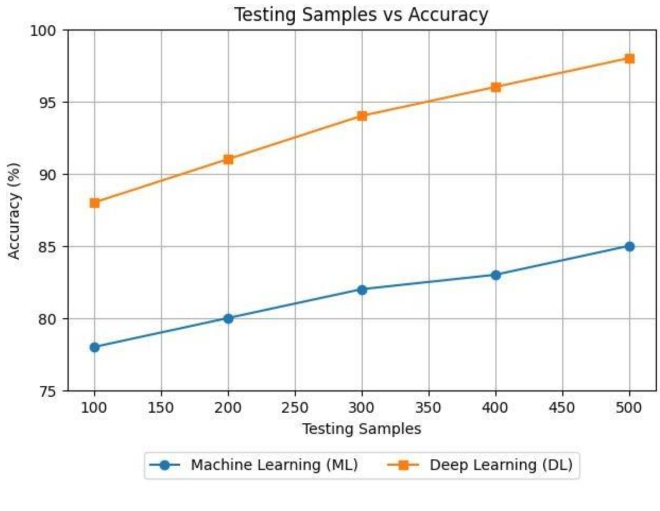

Chart -1: Accuracy Analysis for ML and DL Models

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.15183

4.1 Performance of Machine Learning Models

Traditional machine learning classifiers such as Decision Tree (DT), Random Forest (RF), and Gaussian Naive Bayes (GNB) are trained using Histogram of Oriented Gradients (HOG) features extracted from segmented lung images. Among these models, Random Forest demonstrates superior performance due to its ensemble nature, which reduces overfitting and improves generalization. Decision Tree provides interpretable results but shows comparatively lower accuracy, while Gaussian Naive Bayes exhibits limited performance due to its assumption of feature independence.

4.2 Performance of Deep Learning Models

Deep learning models including Convolutional Neural Network (CNN), DenseNet, and ResNet are trained using transfer learning. These models automatically learn hierarchical spatial features directly from CT images, eliminating the need for handcrafted feature extraction. Experimental results show that CNN-based models achieve significantly higher accuracy and robustness compared to traditional machine learning approaches. DenseNet and ResNet further enhance performance by improving feature reuse and gradient flow, leading to better classification of complex lung patterns.

4.3 Comparative Analysis

A comparative analysis between machine learning and deep learning models indicates that deep learning architectures consistently outperform traditional classifiers across all evaluation metrics. CNN-based models demonstrate improved sensitivity in detecting malignant cases, which is critical for early diagnosis. The results confirm that integrating deep learning techniques with effective preprocessing and segmentation significantly enhances lung cancer classification accuracy.
Overall, the experimental findings validate the effectiveness of the proposed LUNG VISION system as a reliable and efficient diagnostic support tool for early lung cancer detection.

V. CONCLUSION

This paper presented LUNG VISION, an automated lung cancer detection and classification system based on machine learning and deep learning techniques using CT scan images. The proposed framework integrates image preprocessing, feature extraction, and classification to accurately identify lung conditions and support early diagnosis. By combining traditional machine learning models with advanced deep learning approaches, the system achieves improved diagnostic performance and reliability.

Experimental results demonstrate that deep learning models significantly outperform conventional machine learning methods, achieving a maximum classification accuracy of 98%. The close alignment between training and validation accuracy indicates good generalization and minimal overfitting. The inclusion of severity detection further enhances clinical relevance by providing additional insights beyond basic classification.

The deployment of the system through a Flask-based web application enables real-time image analysis and result visualization, making it suitable for practical clinical support. Overall, the proposed system reduces diagnostic workload, minimizes subjectivity in image interpretation, and improves the efficiency and accuracy of lung cancer detection. The results highlight the potential of artificial intelligence–driven solutions in assisting healthcare professionals and improving early lung cancer diagnosis.

FUTURE ENHANCEMENT

The Lung Vision lung cancer detection and classification system has demonstrated reliable performance using CT scan images; however, further improvements can be made to enhance its accuracy, scalability, and clinical usefulness. The following enhancements outline potential directions for future development.

• Incorporation of larger and diverse datasets
Expanding the dataset with CT scan images collected from multiple hospitals and clinical environments can improve model generalization and reduce bias across different patient demographics and disease stages.

• Integration of 3D CT scan analysis
Future versions of the system can be extended to process full 3D CT scan volumes instead of individual slices, enabling richer spatial feature extraction and more accurate diagnosis.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁘ Peer-reviewed & Refereed journal ⁘ Vol. 15, Issue 1, January 2026
DOI: 10.17148/IJARCCE.2026.15183

• Severity and stage level prediction
The system can be enhanced to predict lung cancer stages and severity levels, which would support clinicians in treatment planning and prognosis assessment.

• Use of Generative Adversarial Networks (GANs)
GANs can be incorporated to generate high quality synthetic CT scan images, increasing dataset size and diversity. This data augmentation approach can help improve model robustness and classification accuracy.

• Hybrid ensemble models
Combining multiple deep learning architectures such as CNN, DenseNet, and ResNet into an ensemble framework can further enhance prediction accuracy and reduce individual model limitations.

• Explainable AI implementation
Integrating explainable AI techniques such as Grad CAM or heatmap visualization can improve transparency by highlighting important lung regions involved in the prediction process, increasing clinician trust.

• Real time clinical integration
Future work can focus on integrating the system with hospital information systems and Picture Archiving and Communication Systems to support real time diagnostic workflows.

• Mobile and cloud based deployment
Deploying the system on cloud platforms or mobile applications can improve accessibility, particularly in remote or resource constrained healthcare settings.

• Multi disease detection capability
The system can be extended to detect other pulmonary conditions such as pneumonia, tuberculosis, or COVID 19 using the same CT imaging framework.

REFERENCES

[1]. A. Gudur, H. Sivaraman, and V. Vimal, "Deep Learning-Based Detection of Lung Nodules in Ct Scans for Cancer Screening," International Journal of Intelligent Systems and Applications in Engineering, vol. 11, no. 7S, pp. 20–28, 2023. Available: https://ijisae.org/index.php/IJISAE/article/view/2925
[2]. R. Sharma and S. Verma, "Lung Cancer Detection Using Transfer Learning With Cnn Models," in Proc. IEEE India Conference, 2022.
[3]. A. Mehta and S. Sharma, "Lung Cancer Classification Through Image Preprocessing And Svm," International Research Journal of Engineering and Technology (IRJET).
[4]. P. Chaturvedi et al., "Prediction And Classification Of Lung Cancer Using Machine Learning Techniques," International Journal / Conference on Machine Learning Applications, 2021.
[5]. S. S. Raoof, M. A. Jabbar, and S. A. Fathima, "Lung Cancer Prediction Using Machine Learning: A Comprehensive Approach," in Proc. 2nd Int. Conf. Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 2020, pp. 108–115, doi: 10.1109/ICIMIA48430.2020.9074947.
[6]. Radhika Pra, "Comparative Study Of Lung Cancer Detection Using Machine Learning Algorithms," unpublished.
[7]. G. Paliwal and U. Kurmi, "A Comprehensive Analysis Of Identifying Lung Cancer Via Different Machine Learning Approach," in Proc. 10th Int. Conf. System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2021, pp. 691–696, doi: 10.1109/SMART52563.2021.9675304.
[8]. S. Z. Ahammed, R. Baskar, and G. N. Priya, "An Extensive Survey On Lung Cancer Detection Using Deep Learning Techniques," in Proc. IEEE North Karnataka Subsection Flagship Int. Conf. (NKCon), Vijaypur, India, 2022, pp. 1–6, doi: 10.1109/NKCon56289.2022.10126630.
[9]. S. Das and S. Majumder, "Lung Cancer Detection Using Deep Learning Network: A Comparative Analysis," in Proc. Fifth Int. Conf. Research in Computational Intelligence and Communication Networks (ICRCICN), Bangalore, India, 2020, pp. 30–35, doi: 10.1109/ICRCICN50933.2020.9296197.
[10]. L. Singh, H. K. Choudhary, S. Singh, A. K. Bisht, P. Jain, and G. Shukla, "Automated Detection Of Lung Cancer Using Transfer Learning Based Deep Learning," in Proc. Int. Conf. Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2022, pp. 500–504, doi: 10.1109/CISES54857.2022.9844372.
[11]. Y. Liu, H.-Y. Hsu, T. Lin, et al., "Lung Nodule Malignancy Classification With a 3D Attention-Gated

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⪫ Peer-reviewed & Refereed journal ⪫ Vol. 15, Issue 1, January 2026
DOI: 10.17148/IJARCCE.2026.15183

Convolutional Network," Journal of Translational Medicine.

[12]. A. A. Abd Al-Ameer, "Lung Cancer Detection Using Image Processing And Deep Learning," Indonesian Journal of Electrical Engineering and Computer Science, vol. 28, no. 2, pp. 987–993, doi: 10.11591/ijeecs.v28.i2.pp987-993.

[13]. N. Sudhir Reddy, "Intelligent Deep Learning Algorithm For Lung Cancer Detection And Classification," Engineering and Electrical Intelligence, vol. 12, no. 3, doi: 10.11591/eei.v12i3.4579.

[14]. S. Nageswaran et al., "Lung Cancer Classification And Prediction Using Machine Learning And Image Processing," Biomedical Research International, vol. 2022, Article ID 1755460, Aug. 2022, doi: 10.1155/2022/1755460.
(Retracted in Biomedical Research International, 2024, doi: 10.1155/2024/9851527)

[15]. D. Ardila et al., "End-To-End Lung Cancer Screening With Three-Dimensional Deep Learning On Low-Dose Chest Computed Tomography," Nature Medicine, vol. 25, no. 6, pp. 954–961, 2019, doi: 10.1038/s41591-019-0447-x.

[16]. D. Kumar, A. Wong, and D. A. Clausi, "Lung Nodule Classification Using Deep Features In Ct Images," in Proc. 12th Conf. Computer and Robot Vision (CRV), IEEE, 2015, doi: 10.1109/CRV.2015.25.

[17]. W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-Scale Convolutional Neural Networks For Lung Nodule Classification," in Information Processing in Medical Imaging (IPMI), LNCS 9123, Springer, 2015, pp. 588–599, doi: 10.1007/978-3-319-19992-4_46.

[18]. A. A. A. Setio et al., "Validation, Comparison, And Combination Of Algorithms For Automatic Detection Of Pulmonary Nodules In Computed Tomography Images: The Luna16 Challenge," Medical Image Analysis, vol. 42, pp. 1–13, 2017, doi: 10.1016/j.media.2017.06.015.

[19]. S. Hussein et al., "Tumornet: Lung Nodule Characterization Using Multi-View Convolutional Neural Network With Gaussian Process," in Proc. IEEE Int. Symp. Biomedical Imaging (ISBI), 2017, pp. 1007–1010, doi: 10.1109/ISBI.2017.7950669.

[20]. Y. Liu et al., "Radiomic Features From Lung CT Images Help Predict Tumor Stage And Patient Survival," Cancer Imaging, vol. 18, no. 1, p. 7, 2018, doi: 10.1186/s40644-018-0134-3.