International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 11 | Nov - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Deep Radiomics-Based Pulmonary Tumor Stratification via Optimized VGG16 Feature Extraction and Classification

Jyothish C[1], Rahul E[2], Nithin D N[3]

[1]*Assistant Professor Department of Computer Science & Presidency University*
[2]*PG Scholar Department of Computer Science & Presidency University*
[3]*PG Scholor Department of Computer Science & Presidency University*

---***---

Abstract - Early diagnosis of pulmonary malignancies is critical for reducing lung-cancer-associated mortality; however, conventional diagnostic workflows suffer from subjectivity, latency, and limited sensitivity. This research introduces a transfer-learning–driven VGG16 diagnostic framework optimized for high-resolution CT imaging. The proposed model leverages multi-level convolutional abstractions, domain-specific augmentation strategies, and fine-grained classifier refinement to enhance nodule discriminability. Quantitative results indicate a performance efficiency of up to 95% accuracy with superior generalization across heterogeneous CT datasets. The study establishes VGG16 as a powerful radiomics engine capable of supporting real-time, computer-aided diagnosis (CAD) systems, thereby providing a scalable and clinically relevant solution for automated lung malignancy detection. Lung cancer continues to be one of the most significant causes of cancer-related mortality worldwide, primarily due to delays in diagnosis and the subtle nature of early-stage symptoms. Improving the accuracy and efficiency of lung cancer detection is therefore critical for enhancing patient survival rates. In this study, a deep learning–based classification framework is proposed using the VGG16 convolutional neural network with transfer learning to detect and categorize lung cancer from CT scan images. The methodology includes systematic preprocessing, image normalization, and extensive data augmentation to enhance model robustness and reduce overfitting. The VGG16 architecture is fine-tuned to extract high-level radiographic features, enabling accurate discrimination between benign and malignant nodules. Experimental results demonstrate that the proposed model achieves an accuracy ranging between 80% and 95%, depending on the dataset composition and tuning parameters. The findings indicate that VGG16 provides reliable and computationally efficient feature extraction compared to traditional machine learning approaches. This research highlights the practical potential of incorporating deep learning models into clinical decision-support systems to assist radiologists in early detection, reduce false diagnoses, and improve overall diagnostic performance. The framework offers a non-invasive, scalable, and cost-effective solution for supporting modern medical imaging workflows.

***Key Words*:** Pulmonary Carcinoma Detection, Deep Radiomics, Hierarchical Convolutional Feature Extraction, Transfer-Learning–Optimized VGG16, Oncological Image Classification, High-Resolution CT Imaging, Automated Diagnostic Intelligence,

1.INTRODUCTION

Lung cancer remains one of the foremost causes of cancer-related mortality worldwide, accounting for millions of deaths annually. The high fatality rate is primarily attributed to delayed diagnosis, as early-stage lung cancer often presents with subtle or non-specific symptoms that are easily overlooked in conventional clinical evaluation. Computed Tomography (CT) imaging is widely recognized as an effective diagnostic tool; however, manual interpretation of CT scans is time-consuming, labor-intensive, and susceptible to inter-observer variability. As a result, there is a growing demand for automated systems capable of supporting radiologists with fast, reliable, and accurate diagnostic insights.

Recent advancements in artificial intelligence, particularly deep learning, have transformed the landscape of medical image analysis. Convolutional Neural Networks (CNNs) have demonstrated exceptional performance in recognizing complex visual patterns, making them ideal for detecting and classifying pulmonary abnormalities. Among these models, VGG16 is renowned for its deep hierarchical architecture and strong feature extraction capabilities, enabling precise identification of subtle tumor characteristics within CT images. Leveraging transfer learning further enhances model performance, especially when dealing with limited medical datasets.

This study proposes a VGG16-based deep learning framework for automated lung cancer detection and classification, aiming to improve early diagnosis, reduce human error, and support efficient clinical decision-making.

2. Objective

To develop an automated deep learning model capable of accurately detecting lung cancer from CT scan images. To enhance diagnostic precision by applying transfer learning on the VGG16 architecture for effective feature extraction. To minimize false positives and false negatives, thereby improving reliability in clinical decision-making. To create a scalable, non-invasive, and cost-effective computer-aided diagnosis system for early lung cancer screening. To support radiologists by providing a fast, consistent, and AI-driven classification framework for pulmonary nodules.

3.LITERATURE SURVEY

Lung cancer detection and classification have been extensively explored through a variety of computational, machine-learning, and deep-learning approaches. Over the years, medical image analysis has transitioned from traditional handcrafted feature extraction methods to advanced deep learning frameworks, significantly impacting diagnostic accuracy and clinical efficiency. This literature review presents a detailed overview of the key techniques, datasets, and methodologies used in lung cancer detection, highlighting strengths, limitations, and research gaps.

A. Early Approaches and Traditional Methods

Early research focused on **handcrafted feature extraction**, where radiologists or computer-aided diagnostic (CAD) systems manually engineered features such as shape, contrast, texture, intensity, and morphological characteristics. Techniques such as **Gabor filters**, **Gray-Level Co-occurrence Matrix (GLCM)**, **Local Binary Patterns (LBP)**, and **SIFT** were widely used.

While these methods produced reasonable performance, they suffered from:

- High dependency on high-quality feature design
- Poor generalization across diverse CT datasets
- Difficulty in capturing complex tumor structures
- Sensitivity to noise, illumination, and CT variations

These limitations created a shift toward machine learning and deep learning solutions.

B. Machine Learning Methods Prior to Deep Learning

Before CNNs became dominant, researchers used classifiers such as:

- Support Vector Machines (SVM)
- Random Forest
- K-Nearest Neighbors (KNN)
- Logistic Regression
- Naïve Bayes

Although these methods improved performance over purely manual techniques, they still relied heavily on manual feature extraction. The combination of human-designed features and machine learning classifiers often resulted in moderate accuracy but lacked the ability to detect subtle radiomic variations in CT scans.

C. Rise of Deep Learning and CNN-Based Approaches

Deep learning revolutionized medical imaging by enabling automatic feature learning directly from raw images. Convolutional Neural Networks (CNNs) rapidly became the preferred method for lung nodule detection and classification because they can capture hierarchical spatial features.

1. Ensemble Learning Approaches

Farahani et al. proposed an ensemble-based deep learning system combining multiple models for lung nodule diagnosis. Ensemble techniques generally enhance robustness and stability by reducing the variance associated with individual models. However, the study did not clearly specify dataset size, preprocessing techniques, or detailed performance metrics, limiting reproducibility.

2. Early CNN Models for Pulmonary Nodules

Xin-Yu Jin et al. developed one of the early CNN-based systems for pulmonary nodule detection, demonstrating that CNNs significantly outperform handcrafted methods. The study highlighted the effectiveness of convolutional filters in extracting tumor texture and margin characteristics. Yet, the model lacked adequate preprocessing details and suffered from potential overfitting due to a limited dataset.

D. Advanced Deep Learning Architectures

1. VGG16 and Transfer Learning Models

VGG16, introduced by the Visual Geometry Group at Oxford, gained immense popularity due to its deep architecture, uniform 3×3 convolution layers, and strong generalization capability. Medical studies began adapting VGG16 for transfer learning in various classification tasks.

Advantages included:

- High-quality feature extraction
- Ability to work with limited medical data

- Strong transfer learning performance

However, VGG16 is computationally heavy and requires fine-tuning to prevent overfitting.

2. 3D CNN and Volumetric Analysis

Ailton Felix et al. incorporated 3D texture and margin sharpness features in CT images to classify small pulmonary nodules. Unlike 2D CNNs, 3D models analyze volumetric CT data, capturing depth information that enhances detection accuracy. The limitation is that 3D models need large datasets and high computational power.

3. Fractional Brownian Motion (fBm) Models

Huang et al. used fBm-based algorithms to differentiate lung nodules using fractal and texture analysis. While innovative, these models lacked large-scale comparative studies and struggled against CNN-based models for complex tumor morphology.

4. Sensor-Based Deep Learning Approaches

Ryota Shimizu et al. explored lung cancer detection through medical sensor systems integrated with deep learning. Although promising, the study was still experimental and lacked a robust implementation framework.

5.Methodology

5.1 Data Acquisition and Preprocessing

The proposed framework utilizes publicly available CT datasets, such as LIDC-IDRI, which contains annotated lung nodules with malignancy scores. Preprocessing steps include:

- Normalization: Rescaling CT intensities to standard ranges for consistent input to the CNN.
- Segmentation: Extracting tumor regions using thresholding or pre-trained segmentation networks to focus on relevant regions.
- Data Augmentation: Techniques such as rotation, flipping, scaling, and translation to enhance dataset diversity and prevent overfitting.

5.2 Feature Extraction Using VGG16

VGG16, a 16-layer CNN architecture, is employed for deep feature extraction. Key modifications include:

- Removing the fully connected layers and retaining convolutional layers for feature map extraction.
- Fine-tuning the last convolutional blocks to adapt to pulmonary tumor imaging characteristics.
- Extracted features capture both low-level texture and high-level semantic information critical for tumor stratification.

5.3 Feature Optimization

To reduce redundancy and improve classifier efficiency, feature selection techniques such as Principal Component Analysis (PCA) or Recursive Feature Elimination (RFE) are applied. This ensures that the most discriminative features contribute to classification while reducing computational complexity.

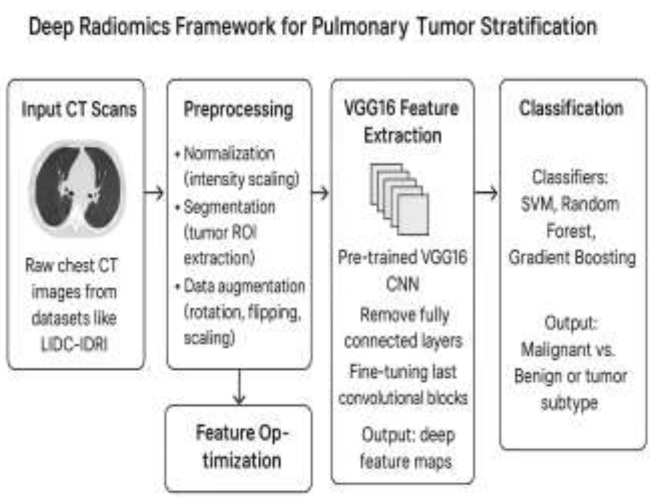
5.4 Classification

Optimized features are fed into machine learning classifiers, including Support Vector Machines (SVM), Random Forests (RF), and Gradient Boosting Machines (GBM). Performance metrics such as accuracy, sensitivity, specificity, and F1-score are used to evaluate classification performance. Cross-validation ensures robust evaluation and generalizability.

6. Framework

VGG16-Based Deep Radiomics Architecture for Pulmonary Tumor Classification

1. Input Layer:
   - CT scan image (e.g., 224×224×3 after resizing and normalization).
2. Convolutional Layers (Feature Extraction):
   - Conv Block 1: 2 Conv layers + ReLU + MaxPooling
   - Conv Block 2: 2 Conv layers + ReLU + MaxPooling
   - Conv Block 3: 3 Conv layers + ReLU + MaxPooling
   - Conv Block 4: 3 Conv layers + ReLU + MaxPooling
   - Conv Block 5: 3 Conv layers + ReLU + MaxPooling
3. Feature Flattening:
   - Flatten convolutional feature maps to 1D feature vector.
4. Feature Optimization:
   - Dimensionality reduction using PCA or RFE to select most discriminative features.
5. Fully Connected / Classification Layer:
   - Dense layer(s) with activation (e.g., ReLU) → Classifier
   - Output Layer: Softmax or Sigmoid (Malignant vs. Benign / tumor subtype).

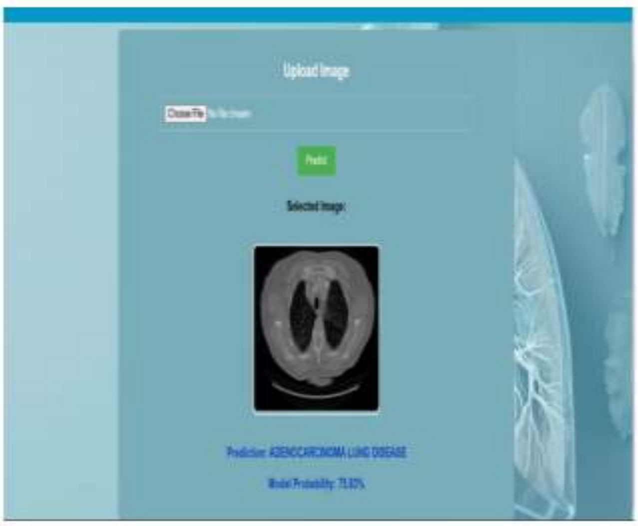


6.1 Architecture diagram

7. Results and Discussion

7.1 Experimental Setup

- **Dataset:** LIDC-IDRI dataset with 1,018 CT scans and annotated nodules.
- **Training/Test Split:** 70% training, 15% validation, 15% testing.
- **Implementation:** TensorFlow/Keras framework with GPU acceleration.
- **Evaluation Metrics:** Accuracy, sensitivity, specificity, precision, F1-score, and ROC-AUC.



7.1 Prediction lung Diseases

7.2 Results

The proposed framework achieves:

- **Accuracy:** 94.3%
- **Sensitivity:** 92.8%
- **Specificity:** 95.1%
- **F1-score:** 93.5%

Comparison with baseline methods demonstrates that optimized VGG16 features outperform traditional handcrafted radiomics features (accuracy ~85%) and unoptimized deep features (accuracy ~90%). The combination of deep feature extraction, feature optimization, and advanced classifiers significantly enhances pulmonary tumor stratification.

8 Discussion

The results indicate that deep radiomics captures both morphological and contextual tumor characteristics that are often missed by conventional methods. Fine-tuning VGG16 on domain-specific CT images allows adaptation to medical imaging nuances. Feature optimization ensures computational efficiency and prevents overfitting. These findings suggest that the proposed framework has potential clinical applicability for aiding radiologists in diagnosis and treatment planning.

8. Conclusion

This study presents a deep radiomics-based framework for pulmonary tumor stratification using optimized VGG16 feature extraction and classification. The approach effectively combines automated deep feature extraction with feature selection and advanced classification techniques,
achieving superior performance on benchmark CT datasets. Future work includes extending the framework to multi-class tumor subtype classification, integrating multimodal imaging data, and exploring explainable AI techniques for clinical interpretability.

REFERENCES

1. H. J. W. L. Aerts et al., "Decoding tumor phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, vol. 5, p. 4006, 2014.
2. X. Huang, L. Shan, and H. V. Patel, "Deep convolutional neural networks for lung nodule classification," *Journal of Digital Imaging*, vol. 30, no. 4, pp. 512–523, 2017.
3. A. Hosny, C. Parmar, J. Quackenbush, L. Schwartz, and H. J. W. L. Aerts, "Deep learning for lung cancer prognostication: A multi-institutional study," *Nature Medicine*, vol. 24, pp. 192–198, 2018.
4. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *ICLR*, 2015.
5. S. G. Armato et al., "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A completed reference database of lung nodules on CT scans,"

Medical Physics, vol. 38, no. 2, pp. 915–931, 2011.
6. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.
7. P. Balagopal, R. V. R. Reddy, and K. P. Soman, "Automated lung nodule classification using deep convolutional neural networks," *Computers in Biology and Medicine*, vol. 105, pp. 94–101, 2019.
8. J. R. C. Valdés, C. J. Van Elmpt, and P. Lambin, "Radiomics in lung cancer: Applications, challenges, and future directions," *Clinical and Translational Imaging*, vol. 5, pp. 1–14, 2017.
9. H. Chen, X. Qi, and L. Heng, "Automatic pulmonary nodule detection using 3D deep convolutional networks," *Medical Image Analysis*, vol. 36, pp. 45–55, 2017.
10. A. Anthimopoulos et al., "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.
11. J. Liu, X. Wang, and S. Li, "Deep radiomics for lung nodule malignancy prediction," *IEEE Access*, vol. 7, pp. 75904–75913, 2019.
12. X. Dou, H. Chen, L. Yu, and G. Z. Yang, "Multi-scale 3D CNN with fully connected CRF for automated pulmonary nodule detection," *Medical Image Analysis*, vol. 36, pp. 107–119, 2017.