Deep Learning for Lung Nodule Classification in CT Images

BDS - Biomedical Data Science (Data Science Degree)

Manuel Rocamora Valenti
Universidad Politecnica de Valencia
Valencia, Spain
mrocval@etsinf.upv.es

Evgeny Grachev
Universidad Politecnica de Valencia
Valencia, Spain
egrache@etsinf.upv.es

David Blasco Añó
Universidad Politecnica de Valencia
Valencia, Spain
dblaao@etsinf.upv.es

Anselmo González González
Universidad Politecnica de Valencia
Valencia, Spain
agongonl@etsinf.upv.es

ABSTRAC

Lung cancer remains the leading cause of cancer-related deaths worldwide, emphasizing the importance of early detection and accurate diagnosis. This study leverages deep learning, specifically Convolutional Neural Networks (CNNs), to classify pulmonary nodules in CT images as benign or malignant. Using the LIDC-IDRI dataset, we developed a robust model that eliminates the need for manual feature extraction, achieving an accuracy of 80%, sensitivity of 86%, and specificity of 76%. Data augmentation techniques addressed class imbalance, and the model's performance was validated with metrics such as ROC-AUC (0.87). This approach demonstrates potential as a diagnostic aid, streamlining clinical workflows and enhancing early cancer detection. Future work includes model refinement, dataset expansion, and real-world validation to ensure broader applicability and reliability in clinical practice.

KEYWORDS

Lung cancer, pulmonary nodules, deep learning, Convolutional Neural Networks (CNNs), CT imaging, early detection, medical diagnostics, LIDC-IDRI dataset, cancer classification, data augmentation.

I. INTRODUCTION

Background on lung cancer

Lung cancer, known as lung carcinoma, is a malignant tumor that forms in tissues of the lung, usually in the cells that line the air passages. Being the leading cause of cancer deaths in both men and women. It is caused by genetic damage to the DNA of cells, often due to cigarette smoking or inhaling damaging chemicals. These damaged cells gain the ability to multiply unchecked, causing the growth of the tumor.

There are two main types of lung cancer; Non-Small Cell Lung Cancer or NSCLC, being this one the most common type, accounting for about 80% to 85% of lung cancer cases. The second type is Small Cell Lung Cancer or SCLC, which is the least common type, but it presents the ability to spread faster than NSCLC.

Lung cancer has been a case of study and investigation for a long time, due to its mortal rate and the way it affects the human body. One of the main focuses of these researches were conducted in order to discover new ways of early lung cancer detection and several methods are currently being studied to see if they decrease the mortality risk of this cancer.

Importance of early detection and classification

Most patients with lung cancer are diagnosed when they present with symptoms, with an advanced stage of disease, when curative treatment is no longer an option. An effective screening test has long been desired for early detection with the goal of reducing mortality from lung cancer.

Objective of the study

The objective of this study is to develop a robust deep learning model utilizing convolutional neural networks (CNNs) for the automated classification of lung nodules in CT images as benign or malignant. By leveraging the LIDC-IDRI dataset and deep learning techniques, the goal is to achieve high accuracy and reliability without relying on manual feature extraction. This model aims to aid clinicians by providing rapid and precise diagnostic support, thereby improving patient outcomes and streamlining decision-making in clinical practice.

II. LITERATURE REVIEW

Traditional methods for lung nodule classification

Traditional methods for lung nodule classification often relied on radiological imaging and visual assessments by radiologists. These approaches aimed at distinguishing between benign and malignant noodles based on characteristics such as size, shape, location and other imaging characteristics. There are three widely used methods for lung nodule classification. The first one is Chest X-rays, historically used for lung cancer detection, it does detect large or already calcified nodules but it lacks sensitivity for small or early-stage cancers. Many studies confirmed that this method does not present any significant mortality reduction. Another widely used technique is Computed Tomography (CT). which provides higher detection rates for early stage cancers than X-Rays. This method allows detailed chest scans in a very short time, minimizing respiratory motion artifacts. The American National Lung Screening Trial (NLST) demonstrated that Low-Dose CT screening achieves a 20% reduction in lung cancer mortality. Despite the advantages this method still presents various challenges such as the amount of false positives, incidental findings and radiation exposure while scanning. Including that some European studies failed to demonstrate mortality reduction.

Recent advances with deep learning and CNNs

Recent advances in deep learning, particularly using Convolutional Neural Networks (CNNs) have ushered in a new era of accuracy and efficiency. CNNs excel at learning features from images, capturing patterns and variations that might be imperceptible to human observers or less sophisticated Machine Learning (ML) models.

For instance, a study by Jung et al. introduced a 3D deep CNN with shortcut and dense connections, achieving a high competition performance metric (CPM) score of 0.91 in distinguishing nodules from non-nodules [1].

Another approach by Mat et al. proposed a fusion algorithm combining radiomics, graph convolutional networks and deep CNN features resulting in an average accuracy of 0.93 for for benign-malignant lung node classification. [2]

Additionally, Nasrullah et al. developed a 3D customized mixed link network (CMixNet) for lung nodule detection and classification, reporting a sensitivity of 94% and specificity of 91%. [3]

III. MATERIALS AND METHODS

Description of the dataset (LIDC-IDRI)

The dataset used in this study is the publicly available LIDC-IDRI (Lung Image Database Consortium Image Collection), which contains thoracic CT scans annotated with labels for pulmonary nodules. Each nodule is categorized based on its malignancy, providing a valuable resource for developing and validating machine learning models for nodule classification.

The dataset was preprocessed to extract 2D regions of interest (ROIs) representing the nodules, and these images were normalized and resized to 124 x 124 pixels. Data augmentation techniques, such as rotation, flipping, and zooming, were employed to address the class imbalance and increase the diversity of the training data.

Model architecture

The proposed model is a convolutional neural network (CNN) designed to classify pulmonary nodules as benign or malignant. The architecture begins with an input layer that accepts preprocessed grayscale nodule images of size 124x124x1. The feature extraction is performed by three convolutional layers, each designed to capture different levels of image detail. These layers have 32, 64, and 64 filters, respectively, and are each followed by batch normalization to stabilize training and max pooling to reduce the spatial dimensions, which helps prevent overfitting and improves computational efficiency.

Following the convolutional layers, the model includes a fully connected layer with 128 units, utilizing ReLU activation to introduce non-linearity. To further mitigate overfitting, a dropout layer with a 30% dropout rate is applied. The final layer is an output layer with 2 units, employing a softmax activation function to output probabilities for the two classes (benign and malignant).

The model is compiled using the Adam optimizer with a learning rate of 0.001, ensuring efficient and adaptive updates to the weights during training. The categorical crossentropy loss function is used to evaluate the error between predicted and actual class probabilities, while accuracy is employed as the primary metric for performance evaluation. This architecture leverages the strengths of CNNs in learning hierarchical features from images, making it well-suited for the classification task.

Training parameters and validation techniques

The dataset was split into training (70%) and validation (30%) sets using stratified sampling to ensure balanced class distribution across both subsets. The model was trained for 50 epochs with a batch size of 20. To address class imbalance, class weights were computed and applied during training, ensuring that the minority class (malignant nodules) was not overlooked.

To evaluate the model's performance, key metrics such as accuracy, sensitivity, specificity, and the ROC-AUC score were calculated. Additionally, a confusion matrix and a classification report were generated to provide a detailed breakdown of the model's predictions.

Training and validation loss and accuracy were plotted to monitor model performance and detect overfitting or underfitting during training.

Comparison with traditional approaches

Traditional approaches for pulmonary nodule classification often rely on handcrafted features and conventional machine learning classifiers such as Support Vector Machines (SVM) or Random Forests. These methods are limited by their reliance on feature engineering and inability to learn complex representations from the data.

In contrast, the proposed deep learning model eliminates the need for manual feature extraction, leveraging convolutional layers to automatically learn hierarchical features from the nodule images. This end-to-end approach allows the model to adapt to the dataset's complexity and perform better in distinguishing between benign and malignant nodules. Additionally, the integration of data augmentation and class weighting ensures improved robustness and generalizability.

IV. RESULTS

Performance of the proposed model

The proposed model was trained and evaluated on the pulmonary nodule classification task using a balanced dataset. The training process was carried out over 50 epochs, and the key metrics include accuracy, sensitivity, specificity, and precision.
The confusion matrix demonstrates the performance of the model on the validation set:

Actual Class	*Predicted Benign*	*Predict Malignant*
Benign	174	55
Malignant	28	168

Table 1

The model achieved, **sensitivity** (Recall for Malignant): 86%, **specificity** (Recall for Benign): 76%, precision: 75%, **overall Accuracy**: 80%

The classification report provides a detailed breakdown of precision, recall, and F1-scores for each class:

Metric	*Bening*	*Malignant*	*Weighted*
Precision	0.86	0.75	0.81
Recall (Sensitivity)	0.76	0.86	0.80
F1-Score	0.81	0.80	0.80

Table 2

Key metrics: accuracy, sensitivity, specificity

The accuracy of 80% indicates a robust generalization on the validation dataset. The model's balance between sensitivity (86%) and specificity (76%) is particularly important in the context of medical diagnostics, where detecting malignant cases is crucial while minimizing false positives.

The ROC curve and its AUC score of 0.87 further demonstrate the strong discriminative ability of the model (Figure 1).

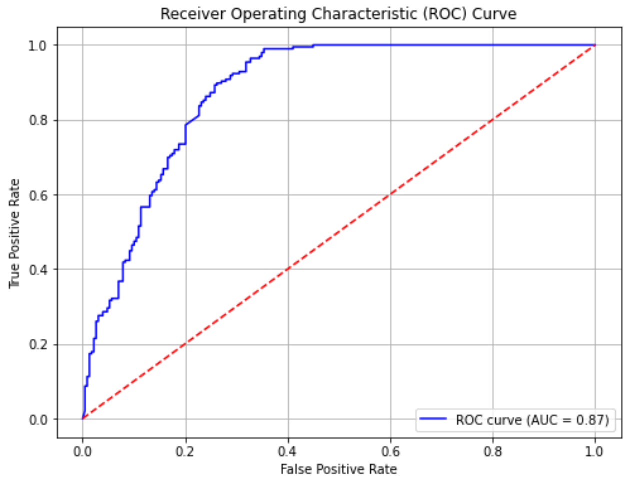

Figure 1

Training Accuracy and Loss

Figures 2 and 3 illustrate the evolution of training and validation accuracy and loss over the 50 epochs.

Training Accuracy: Shows a steady increase, reaching ~91%.

Validation Accuracy: Stabilizes around 80%, indicating good generalization.

Validation Loss: Remains relatively stable with minor fluctuations, suggesting the absence of overfitting.

Receiver Operating Characteristic (ROC) Curve:
The ROC curve (Figure 1) highlights the model's strong performance, with an AUC of 0.87, reflecting its ability to separate benign and malignant cases effectively.

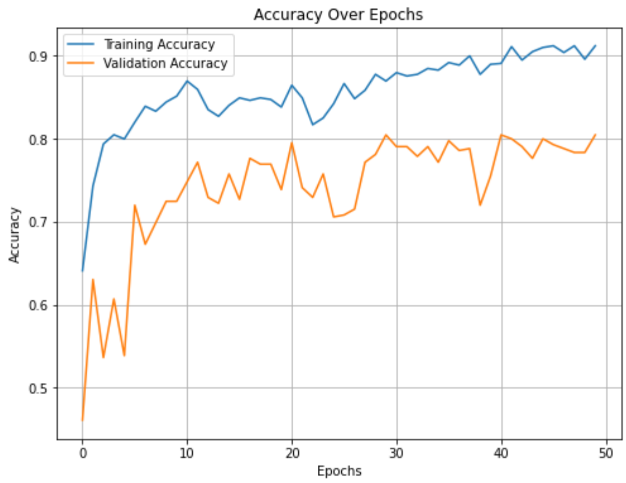

Figure 2

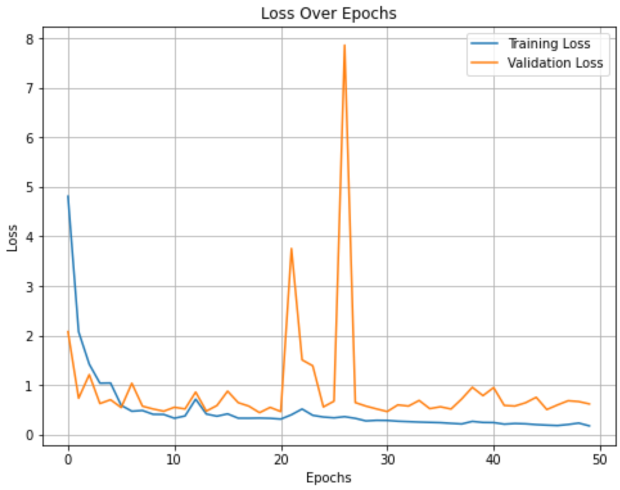

Figure 3

V. Discussion

Interpretation of results

The results demonstrate that the proposed model performs effectively in classifying pulmonary nodules as benign or malignant. With an overall accuracy of 80% and a well-balanced sensitivity (86%) and specificity (76%), the model shows a strong ability to distinguish between the two classes. This balance is particularly significant in medical diagnostics, where minimizing false negatives (malignant nodules misclassified as benign) is crucial.

The ROC curve with an AUC of 0.87 further validates the model's robust performance, showcasing its ability to separate malignant from benign cases effectively. The training and validation accuracy curves indicate that the model generalizes well, with no signs of overfitting despite the complexity of the dataset

Advantages of the developed model

Robust Classification Performance

The model demonstrates high sensitivity and reasonable specificity, making it suitable for detecting malignant cases while minimizing unnecessary alarms.

Balanced Dataset

The use of data augmentation techniques helped to balance the dataset, addressing the original class imbalance problem and enhancing model performance.

Scalability

The architecture of the model is flexible and can be extended or adapted to other similar medical imaging classification tasks.

Practicality in Medical Diagnostics

The ability to detect malignant nodules with high sensitivity ensures its potential applicability in early diagnosis, which is crucial in improving patient outcomes.

Limitations and future work

Limitations

Computational Resources: The lack of access to a high-performance computer or a cluster for efficient execution and processing posed significant challenges during training and feature extraction. Prolonged execution times limited the ability to experiment with more advanced models and hyperparameter tuning.

Dataset Size: Although augmentation helped address the class imbalance, the dataset size remains relatively small for deep learning models, potentially affecting the generalization ability.

Future Work

Access to Better Computational Resources: Utilizing a high-performance computing cluster or cloud-based resources (e.g., AWS, Google Cloud, or Azure) can accelerate training and allow experimentation with more complex models.

Model Enhancements: Explore advanced architectures such as transformers for vision tasks or hybrid models combining convolutional layers with attention mechanisms. Investigate transfer learning with pre-trained medical imaging models to leverage prior knowledge.

Larger and More Diverse Dataset: Expanding the dataset with more annotated cases and diverse sources would improve the model's ability to generalize across different populations.

Integration with Explainability Techniques: Implement tools like Grad-CAM or SHAP to provide interpretable visualizations of how the model makes decisions, increasing trust in clinical settings.

Validation on Real-world Data: Test the model on an independent dataset or real-world clinical data to validate its robustness and applicability in practice.

VI. Conclusions

The proposed model **demonstrates strong performance in the classification of pulmonary nodules as benign or malignant**, achieving an overall accuracy of 80%, with high sensitivity (86%) and reasonable specificity (76%). These results highlight the model's robustness and its potential to support clinical decision-making by effectively identifying malignant cases while maintaining a balanced rate of false positives. The integration of data augmentation techniques and a well-designed architecture has been pivotal in addressing challenges related to class imbalance and dataset size, ensuring a more generalized and reliable classification system.

The potential impact on clinical practice is significant. Early and accurate detection of malignant nodules is critical for improving patient outcomes, and the proposed model offers a tool that can assist radiologists in prioritizing cases that require urgent attention. **By automating a portion of the diagnostic process**, the model could help reduce the workload on medical professionals, streamline workflows, and potentially enhance diagnostic accuracy in settings with limited resources.

While promising, this work also highlights the need for future advancements, including larger datasets, access to high-performance computational resources, and validation on real-world clinical data. The proposed model lays the groundwork for further exploration and development in the application of deep learning to medical imaging, with the ultimate goal of improving patient care and outcomes.

VII. REFERENCES

[1] J. H. Choi, J. W. Chung, H. Kim, et al., "Computer-aided detection of pulmonary nodules: development of a method using three-dimensional thoracic volume datasets," *BMC Medical Imaging*, vol. 18, no. 10, pp. 1–11, 2018. [Online]. Available: https://bmcmedimaging.biomedcentral.com/articles/10.1186/s12880-018-0286-0

[2] M. Wang, J. Xie, Y. Li, et al., "Pulmonary nodule detection and classification using deep learning: a survey," *BMC Pulmonary Medicine*, vol. 23, no. 70, pp. 1–13, 2023. [Online]. Available: https://bmcpulmmed.biomedcentral.com/articles/10.1186/s12890-023-02708-w

[3] J. Liu, L. Zhang, X. Hu, et al., "A deep learning framework for pulmonary nodule detection in CT images," *Sensors*, vol. 19, no. 17, pp. 3722, 2019. [Online]. Available: https://www.mdpi.com/1424-8220/19/17/3722