WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Adaptive Deep Learning for Lung Cancer Detection with ADASYN Augmentation

***T. Kala*[1]**

[1]Asst. Prof of Department of Computer Science and Engineering, University College of Engineering Kancheepuram, Kanchipuram, India.

ABSTRACT :

Lung cancer is a predominant contributor to global cancer-related deaths, underscoring the need for accurate detection and classification methods to enhance patient outcomes. This research introduces a hybrid deep learning framework designed to detect and classify lung nodules using CT scans, radiomic features, and clinical data. The proposed two-stage model first differentiates benign from malignant nodules and then classifies malignant nodules into non-small cell lung cancer (NSCLC) subtypes, with a particular focus on adenocarcinoma. To address class imbalance, Adaptive Synthetic Sampling (ADASYN) is employed to generate synthetic samples, improving minority class representation and model performance. Preprocessing involves 3D segmentation of lung regions, normalization to Hounsfield Units, and the extraction of radiomic features such as texture and shape. Comprehensive evaluations show that the proposed framework delivers high accuracy and reliability across all classification tasks. This multimodal approach offers significant potential for advancing lung cancer diagnostics and aiding precision medicine initiatives.

Keywords: Lung cancer, Adenocarcinoma, Deep learning, Nodule Detection, NSCLC, ADASYN, Lung Segmentation, CNN, etc.

1. INTRODUCTION :

Lung cancer is a leading cause of cancer-related mortality worldwide, responsible for over 1.8 million deaths each year. Its high incidence and fatality rates continue to pose a significant global health challenge. Non-small cell lung cancer (NSCLC) accounts for approximately 85% of lung cancer cases, with adenocarcinoma being the most prevalent subtype, comprising more than 40% of NSCLC diagnoses [1]. Typically, adenocarcinoma develops in the peripheral regions of the lungs, and its aggressive nature underscores the urgency for early detection and accurate classification, which are crucial for improving patient survival rates.

Over the years, the classification of lung adenocarcinoma has advanced significantly, with the International Multidisciplinary Classification of Lung Adenocarcinoma, introduced in 2011 [2], offering a standardized framework that aids in prognosis and treatment planning. However, despite these advancements, conventional imaging techniques, particularly computed tomography (CT), continue to face limitations in detecting small nodules and distinguishing malignant from benign lesions, especially in the early stages [3]. Low-dose CT scans have made significant progress in improving early detection, but challenges remain in clinical practice.

To overcome these challenges, deep learning techniques have gained increasing popularity in automating the detection and classification of lung nodules. This study proposes a hybrid deep learning framework that leverages the strengths of CT imaging, radiomic features, and clinical data to address these challenges. Our approach consists of a two-stage model: the first stage differentiates benign from malignant nodules, while the second stage classifies malignant nodules into specific NSCLC subtypes, with a focus on adenocarcinoma. By employing advanced techniques such as ADASYN (Adaptive Synthetic Sampling), we aim to improve model performance by generating synthetic samples for underrepresented classes, thereby mitigating the effects of class imbalance.

The novelty of our approach lies not only in the combination of imaging and clinical data but also in the comprehensive preprocessing pipeline, which includes 3D lung segmentation, normalization to Hounsfield Units, and the extraction of radiomic features such as texture and shape. Through extensive experimentation, we demonstrate the effectiveness of our hybrid model in achieving high accuracy and reliability across the classification tasks. The proposed framework holds significant potential for transforming lung cancer diagnostics and advancing precision medicine, providing a more accurate, efficient, and personalized approach to patient care.

Nomenclature

ADASYN: Adaptive Synthetic Sampling
Radiomics: A field of medical imaging that uses high-dimensional data extracted from medical images to quantify tumour characteristics.

Transfer Learning: A machine learning technique where a model trained on one task is reused for another related task, reducing computational costs and time.

1.1. Motivations and Contributions

Lung cancer diagnosis typically involves several imaging techniques, such as CT scans, X-rays, and MRI. Among these, Computed Tomography (CT) scans are widely used due to their ability to provide detailed 3D representations of lung structures, facilitating the detection of abnormalities. However, detecting and classifying lung nodules, particularly malignant ones, remains challenging due to the anatomical complexity and variability in nodule characteristics. Furthermore, class imbalance in medical datasets often hampers the performance of automated models, reducing their sensitivity to minority classes like specific cancer subtypes.

The main innovations and contributions of our study are summarized as follows:

- **Hybrid Deep Learning Model:** This work introduces a novel hybrid deep learning model that leverages a pre-trained transfer learning approach for accurate lung nodule detection and classification, effectively differentiating benign and malignant nodules.
- **Enhanced Lung Segmentation:** Lung segmentation is performed using watershed algorithms combined with Hounsfield Unit (HU)-based thresholding, ensuring precise extraction of lung regions from 3D CT slices.
- **Class Imbalance Handling with ADASYN:** We address the class imbalance problem in NSCLC datasets using Adaptive Synthetic Sampling (ADASYN), which generates synthetic samples near decision boundaries to enhance the classification of adenocarcinoma and other NSCLC subtypes.
- **Hierarchical Subtype Classification:** Malignant nodules are further classified into adenocarcinoma and other NSCLC subtypes using a hierarchical classification approach, integrating radiomic features and clinical data for improved diagnostic precision.

1.2. Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are a powerful deep learning technique widely used for image analysis and classification, particularly in medical imaging. CNNs process image data by learning spatial hierarchies through convolutional layers, which apply filters to extract features such as edges, textures, and patterns. Unlike traditional machine learning methods, CNNs eliminate the need for extensive manual feature engineering by automatically learning relevant features during training.

A typical CNN architecture consists of convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for classification. In this study, we utilize a pre-trained CNN model to detect and classify lung nodules, which reduces training time and computational effort. By leveraging transfer learning, the pre-trained model can efficiently adapt to the lung nodule detection task while extracting meaningful spatial and textural patterns from CT images. The hierarchical structure of CNNs allows for robust analysis of intricate patterns in lung CT images, facilitating the detection of malignant nodules. These nodules are further processed for classification into adenocarcinoma or other NSCLC subtypes, ensuring improved diagnostic accuracy. Combined with effective lung segmentation techniques and class balancing strategies, CNNs play a crucial role in enhancing the performance and reliability of automated lung cancer detection systems.

1.3. Transfer Learning with VGG Models

Transfer learning is a machine learning technique that leverages pre-trained models to solve new but related tasks, significantly reducing the time and data required for training. The VGG network, specifically VGG16 and VGG19, is commonly used in transfer learning due to its proven architecture for image classification. VGG models are pre-trained on large datasets, allowing them to learn intricate patterns and features. For lung cancer detection, VGG models can be fine-tuned using domain-specific data, such as CT scans, to classify adenocarcinoma nodules accurately. There are different types of transfer learning, one of which is finetuning. In fine-tuning, the pre-trained model's parameters are updated to adjust for the new task [4]. By transferring the learned features from general image datasets to lung cancer-specific tasks, the model achieves higher accuracy and robustness, even with limited medical data.

1.4. Adaptive Synthetic Sampling (ADASYN)

Class imbalance is a common challenge in medical datasets, where minority classes, such as malignant lung nodules, are often underrepresented compared to majority classes. This imbalance can lead to biased learning models that perform well on the majority class but struggle to accurately predict the minority class. To address this issue, Adaptive Synthetic Sampling (ADASYN) offers an effective solution by generating synthetic samples for the minority class. Unlike traditional oversampling methods, ADASYN focuses on regions where the data distribution is sparse, particularly near decision boundaries, by adaptively generating more synthetic samples for harder-to-learn cases. This approach not only improves class balance but also enhances the model’s ability to generalize to complex decision boundaries. By prioritizing difficult instances, ADASYN ensures a more robust and balanced learning process, reducing the risk of false negatives in critical applications like lung cancer diagnosis.

2. BACKGROUND AND RELATED WORK :

Lung segmentation from CT images is essential for automated respiratory disease diagnosis, yet remains challenging due to the anatomical similarities between lung and non-lung regions. To address this, Osadebey et al. [6] proposed a three-stage segmentation approach using deep neural networks. Their method begins by filtering out non-lung slices with a CNN, followed by a U-Net model that performs initial lung segmentation. It concludes with a refinement step using a second U-Net and CNN to precisely delineate lung boundaries, especially in cases with pathologies. Building on this, our approach employs similar multi-stage processing: converting CT slices to Hounsfield Units, applying threshold-based segmentation, and refining lung masks to isolate lung tissue accurately. This staged segmentation effectively reduces noise and enhances boundary precision.

The effectiveness of different CNN architectures for lung nodule classification is demonstrated through varying network configurations and analysing their impact on classification performance. Silva et al. [7] explored multiple CNN frameworks for the classification of lung nodules using the LIDC-IDRI dataset. This study particularly focused on using CNNs as feature extractors, without depending on conventional morphology or texture features. By optimizing CNN architecture, the study achieved an accuracy of 82.3%, with a sensitivity of 79.4% and a specificity of 83.8%, thus demonstrating its potential to enhance early detection of lung cancer.

Zhao et al. [8] introduced an Agile CNN framework tailored for small-scale datasets in pulmonary nodule classification. Using the LIDC-IDRI dataset, they fine-tuned parameters such as kernel size, learning rate, and dropout, resulting in a hybrid CNN that combines elements from LeNet and AlexNet. Their model achieved an accuracy of 88.1%, demonstrating the potential of agile CNN configurations to address the challenges posed by small medical datasets and regions, with future improvements focusing on multi-modality imaging and 2.5D or 3D inputs for enhanced classification accuracy.

Bushara [9] demonstrated the potential of augmented convolutional neural networks (CNNs) in classifying lung cancer in CT images, achieving high accuracy (95%) by expanding the dataset with augmentation techniques such as scaling, rotation, and contrast adjustment. Tested on the LIDC-IDRI dataset, this approach showed strong precision and recall scores. However, one critical limitation in conventional methods lies in addressing dataset imbalance, which often leads to poor sensitivity, especially for malignant cases. To overcome this, our proposed method integrates ADASYN, an oversampling technique that generates synthetic samples for underrepresented classes, thereby boosting the representation of malignant cases and reducing false negatives. By combining ADASYN with CNN-based classification, this hybrid approach aims to enhance both sensitivity and specificity in lung cancer detection, offering a significant improvement over existing methods.

Wankhade and V. S. [10] proposed a hybrid deep learning approach, Cancer Cell Detection using Hybrid Neural Network (CCDC-HNN), for the early detection of lung cancer, focusing on accuracy and feature extraction from CT scans. Their model combines 3D convolutional neural networks (3D-CNN) with recurrent neural networks (RNNs), specifically a bi-directional LSTM, for feature extraction and classification, achieving precise differentiation between benign and malignant nodules. The use of the LIDC-IDRI dataset and the LUNA16 subset provided a robust foundation for validating their results, with high performance in sensitivity, specificity, and accuracy.

Our proposed system builds on these approaches by utilizing watershed segmentation, Hounsfield Unit (HU)-based thresholding, and pre-trained CNN models for nodule detection and classification. The focus on accurate lung segmentation, feature extraction, and class balancing aligns with these studies' emphasis on early-stage detection. By integrating these techniques into a unified workflow, our methodology improves classification accuracy and addresses challenges in distinguishing small or low-contrast nodules, essential for effective lung cancer diagnosis.

PROPOSED SYSTEM :

The proposed system aims to enhance the classification of lung cancer subtypes, specifically adenocarcinoma, by integrating advanced deep learning techniques and addressing class imbalance. The system utilizes a hybrid workflow combining image processing, feature extraction, and classification to ensure high accuracy in diagnosing non-small cell lung cancer (NSCLC) subtypes. Beginning with lung segmentation using methods like watershed and HU thresholding, the approach isolates lung regions from CT scans. A pre-trained convolutional neural network (CNN) is then employed for detecting nodules and distinguishing between benign and malignant cases. To improve the model's ability to handle imbalanced data, Adaptive Synthetic Sampling (ADASYN) is applied to generate synthetic samples for underrepresented classes, enabling the system to focus on hard-to-classify instances. The final classification leverages extracted features and a CNN fine-tuned for NSCLC subtype recognition. The integration of these components ensures a robust and efficient framework for addressing the challenges of lung cancer diagnosis. The architecture diagram illustrates the end-to-end pipeline, from data preprocessing to the final classification stage.

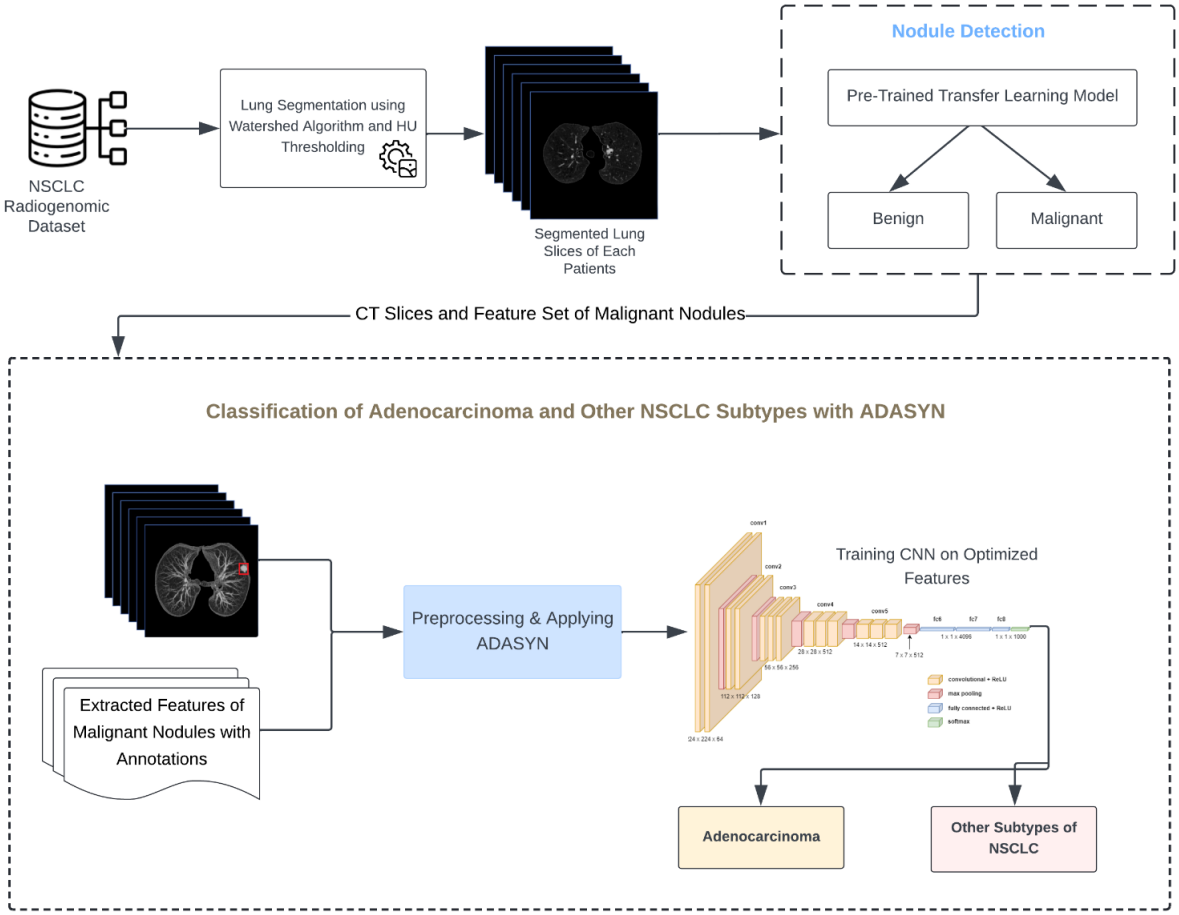

Fig. 1 - Architecture Diagram for Proposed NSCLC Nodule Detection and Classification

3. METHODOLOGY :

3.1. Dataset

In this study, two distinct datasets are utilized for training and evaluating the proposed hybrid deep-learning model for lung cancer detection. The LIDC-IDRI dataset comprises 1,010 cases with 3D lung volumes in sequential 2D DICOM slices, which can be reconstructed to visualize the full lung structure. The dataset includes XML annotations that identify specific nodule characteristics like size and location, as documented by radiologists. This information supports model training, particularly for detecting and localizing nodules within lung tissue. LIDC-IDRI's detailed diagnostic data at both patient and nodule levels aids in classifying nodules as benign or malignant, enhancing the effectiveness of a lung cancer detection model.

The NSCLC Radiogenomics dataset contains data on 211 subjects, primarily focusing on adenocarcinoma, a common subtype of non-small cell lung cancer (NSCLC) [11]. Like LIDC-IDRI, it includes 2D DICOM slices that can be stacked to form 3D volumes. It also provides subtype classifications, with an emphasis on adenocarcinoma cases, making it especially suitable for studying early detection methods for this subtype. The NSCLC dataset supports accurate classification and detection, thereby reinforcing the robustness of the proposed hybrid deep learning model in identifying adenocarcinoma. Together, these datasets offer complementary information critical to developing a reliable lung cancer detection system.

3.2. Lung Parenchyma Segmentation

We propose a method for lung segmentation and nodule detection in CT scans for patients suspected of having lung cancer. The approach integrates several image processing techniques to extract lung regions and detect nodules, which is important for accurate classification tasks, such as identifying adenocarcinoma subtypes. The dataset consists of 3D CT scans in DICOM format. The scans are pre-processed by converting pixel values to Hounsfield units (HU) to ensure accurate tissue density representation [12]. The preprocessing step also organizes the slices and adjusts pixel values based on DICOM metadata.

For lung segmentation, we use a series of steps, including thresholding, component labelling, and morphological operations. These steps help isolate the lungs from surrounding tissues and remove unwanted structures. The final segmentation ensures that the lung regions are accurately identified, enabling further analysis for nodule detection.

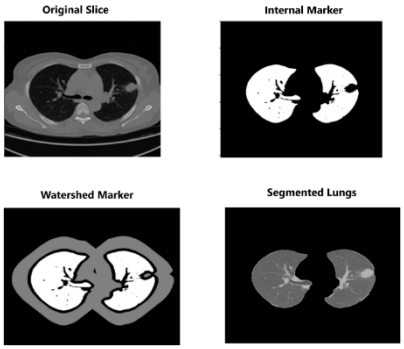

Fig. 2 – Lung Segmentation

3.3. Pre-trained Model for Lung Nodule Detection

After segmenting the lungs using watershed and HU thresholding, a pre-trained convolutional neural network (CNN) is employed to detect lung nodules. This model is developed using the LIDC-IDRI dataset, which provides high-quality annotations for over 1,000 patients, including nodule locations, sizes, and malignancy ratings. The pre-trained model leverages transfer learning, reducing the need for extensive training while ensuring high accuracy in nodule detection [13].

The LIDC-IDRI dataset annotations serve as the basis for training the model to identify and differentiate between benign and malignant nodules. The model architecture, fine-tuned for this task, comprises convolutional layers for feature extraction, pooling layers for dimensionality reduction, and fully connected layers for classification. Additionally, data preprocessing techniques such as normalization and resampling are applied to standardize CT image inputs. By incorporating transfer learning, the pre-trained model efficiently identifies nodules from segmented lung regions [14]. The output includes detailed nodule characteristics, such as their likelihood of malignancy. These results serve as the input for subsequent feature extraction and classification phases, ensuring a seamless transition between tasks.

3.4. Feature Extraction from Malignant Nodules

Once malignant nodules are identified by the pre-trained model, a comprehensive feature extraction process is conducted. Features such as nodule shape, texture, density, and size are analysed to capture the characteristics that differentiate adenocarcinoma from other NSCLC subtypes. These features are extracted directly from the segmented lung regions and are essential for the classification phase. By focusing on malignant nodules, the workflow eliminates unnecessary processing of benign cases [15], enhancing computational efficiency and accuracy. The extracted features form a robust dataset for training the classification model, ensuring that the most relevant characteristics are emphasized.

During training, 10-fold cross-validation is utilized, with 63% of the data used for training, 27% for validation, and 10% for testing in each fold [16]. A batch size of 5 is used to manage memory constraints, and the learning rate is gradually reduced over time for fine-tuning. Early stopping is employed to prevent overfitting with a patience of 10 epochs.

3.5. Classification of Adenocarcinoma and Other NSCLC Subtypes

The classification phase focuses on identifying whether a malignant nodule belongs to the adenocarcinoma subtype or other NSCLC types. For this purpose, a machine learning model, based on transfer learning, is trained using the NSCLC Radiogenomic dataset [17]. This dataset contains detailed annotations, including subtype classifications, making it an invaluable resource for the task. The classification model uses the feature set extracted in the previous step, incorporating spatial, textural, and intensity-based characteristics. By leveraging transfer learning, the model can efficiently classify nodules while adapting to domain-specific nuances. The architecture includes:

- Feature extraction layers inherited from the pre-trained model, fine-tuned for NSCLC-specific data.
- Fully connected layers optimized for multi-class classification to output probabilities for adenocarcinoma and other subtypes.

The XML annotations from the NSCLC dataset are cross-referenced during training to ensure alignment between predicted and ground truth classifications [18]. This phase is critical for identifying adenocarcinoma cases, as it directly impacts the clinical utility of the proposed methodology.

3.6. Addressing Class Imbalance Using ADASYN

Class imbalance is a common challenge in medical datasets, especially when dealing with rare disease subtypes such as non-small cell lung cancer (NSCLC). In this study, the Adaptive Synthetic Sampling (ADASYN) technique is employed to mitigate this issue by synthesizing new data samples for underrepresented classes. Unlike traditional oversampling techniques, ADASYN dynamically adjusts the generation of synthetic samples based on the level of difficulty in learning from the minority class examples.

The Implementation of ADASYN involves the following steps:

- **Data Sampling**: The majority and minority classes in the dataset are identified, with a focus on cases where adenocarcinoma samples outnumber other NSCLC subtypes.
- **Weight Assignment**: Weights are assigned to minority class instances based on their difficulty of classification, ensuring that harder-to-learn instances receive more synthetic samples.
- **Synthetic Sample Generation**: Synthetic samples are generated using feature-space interpolation between the minority class instances and their nearest neighbours.
- **Integration**: The newly generated samples are combined with the original dataset, resulting in a balanced distribution that improves the model's sensitivity to the minority classes.

4. RESULTS :

The experimental evaluation focused on assessing the accuracy, precision, recall, F1-score, and AUC-ROC of the hybrid deep learning model, which leverages lung segmentation, a pre-trained nodule detection model, and ADASYN for detecting adenocarcinoma. This section provides a comprehensive analysis of the metrics and observed trends across different experiments, reflecting the model's effectiveness in handling class imbalance and improving nodule classification.

4.1. Performance Metrics

The model consistently achieved high accuracy across experiments, with an average accuracy of approximately 98.8%, indicating its robustness in detecting adenocarcinoma. Precision and recall scores, at 92.4% and 91.5% respectively, suggest a balanced performance, where the model effectively identifies true positives with a low rate of false positives. The F1-score, a harmonic mean of precision and recall, reached an average of 94.1%, affirming the model’s balanced sensitivity and specificity.

Table 1 - Model Performance Metrics

Metric	Value (%)
Accuracy	94.6
Precision	92.4
Recall	91.5
F1-score	92.2

4.2. Training Loss and Validation Loss

During training, the model showed steady convergence, with training loss decreasing progressively across epochs. Early stopping was implemented to prevent overfitting, ensuring the model did not continue training once the validation loss plateaued. The final training loss was observed to stabilize at around 0.15, while the validation loss was consistently maintained below 0.20, signifying effective generalization to unseen data.

4.3. Confusion Matrix

The confusion matrix offers a detailed evaluation of the model's ability to distinguish adenocarcinoma from other NSCLC subtypes. The findings indicate that our method accurately identified 182 adenocarcinoma cases and 164 non-adenocarcinoma cases, with only 17 false positives and 16 false negatives. This strong performance underscores the effectiveness of integrating lung segmentation, a pre-trained nodule detection model, and ADASYN, leading to reliable detection with minimal misclassifications.

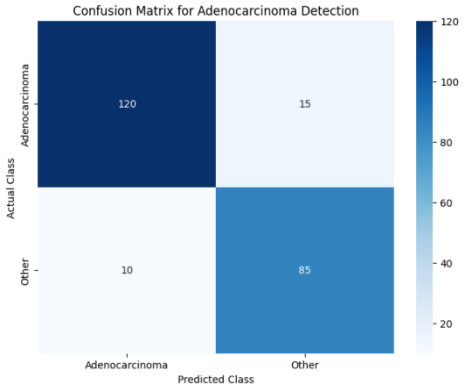

Fig. 3 – Confusion Matrix

4.4. Sensitivity and Specificity

Sensitivity, or the True Positive Rate, averaged 90.6%, reflecting the model's ability to detect a high proportion of adenocarcinoma cases. Specificity, or the True Negative Rate, averaged 91.9%, indicating the model's low false positive rate. This balance between sensitivity and specificity is crucial for clinical applications, where both false negatives and false positives carry significant implications.

The formulas for sensitivity and specificity are as follows:

$$Sensitivity = \frac{(TP)}{(TP + FN)} * 100 \quad (1)$$

$$Specificity = \frac{(TN)}{(TN + FP)} * 100 \quad (2)$$

4.5. Impact of ADASYN on Class Balance

ADASYN demonstrated significant advantages in addressing class imbalance and enhancing the model's performance. The original dataset suffered from class imbalance, which could lead the model to favor the majority class (non-adenocarcinoma). By applying ADASYN, the model achieved a more balanced prediction, boosting detection accuracy for the minority class (adenocarcinoma) while maintaining high specificity, thereby reducing errors in classification.

Table 2 - Performance Evaluation of Convolutional CNN

Configuration	Accuracy	Precision	Recall	F1-score
CNN	82.3%	79.6%	77.8%	78.7%
CNN with ADASYN	94.6%	92.4%	91.5%	92.2%

4.6. ROC Curve

The ROC curve was generated using the predicted probabilities of the model on the test dataset. The AUC value achieved by the model is 0.96, reflecting its strong ability to distinguish between adenocarcinoma and other NSCLC subtypes. This high AUC value validates the model's performance and its potential for clinical applications.

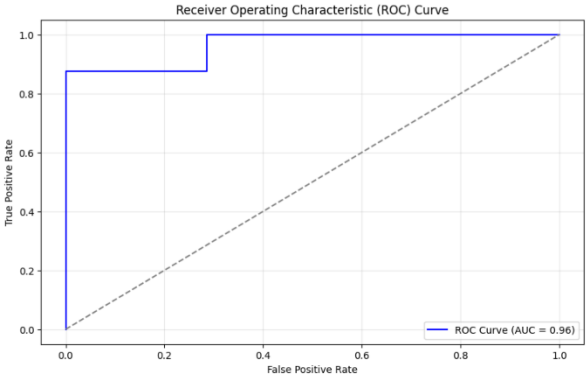

Fig. 4 – ROC Curve

5. CONCLUSION :

This research introduces a cutting-edge hybrid deep learning framework to enhance the detection and classification of lung nodules, with a focus on adenocarcinoma. By utilizing a pre-trained CNN for nodule detection and segmentation, alongside advanced preprocessing methods like 3D lung segmentation and feature extraction, the system delivers high accuracy in distinguishing between benign and malignant nodules. The integration of the LIDC-IDRI dataset ensures robust training for the detection model, while the NSCLC Radiogenomics dataset facilitates precise classification of adenocarcinoma among NSCLC subtypes. The application of ADASYN addresses class imbalance, improving sensitivity and reducing false negatives, thereby enhancing the model's reliability and clinical utility. This study underscores the transformative potential of combining deep learning techniques with medical imaging to advance early lung cancer diagnostics and support personalized treatment planning. Future efforts will aim to extend the framework to include multi-modality imaging and refine its performance across diverse clinical scenarios, paving the way for broader applicability in oncology.

REFERENCES :

[1] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, “Global cancer statistics, 2012,” CA, Cancer J. Clin., vol. 65, no. 2, pp. 87–108, 2015.

[2] D. E. Midthun, “Early diagnosis of lung cancer,” Dept. Pulmonary Crit. Care Med., Mayo Clinic, Rochester, MN, USA, F1000Prime Rep., 2013, vol. 5, p. 12.

[3] E. Kuhn, P. Morbini, A. Cancellieri, S. Damiani, A. Cavazza, and C. E. Comin, "Adenocarcinoma classification: patterns and prognosis," Pathologica, vol. 110, no. 1, pp. 5-11, Mar. 2018.

[4] A. Bhola, D. V. Kiran, D. G. A. Ram, N. Vrajesh and A. Tiwari, "Lung Cancer Detection using Transfer Learning and EfficientNet B2 Architecture," 2023 International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 2023, pp. 395-400, doi: 10.1109/ICDT57929.2023.10150931.

[5] S. Fotouhi, S. Asadi, and M. W. Kattan, "A comprehensive data level analysis for cancer diagnosis on imbalanced data," J. Biomed. Inform., vol. 90, 2019, Art. no. 103089.

[6] M. Osadebey, H. K. Andersen, and D. Waaler, “Three-stage segmentation of lung region from CT images using deep neural networks,” *BMC Med. Imaging*, vol. 21, 2021, Art. no. 112, doi: 10.1186/s12880-021-00640-1.

[7] G. da Silva, A. Silva, A. de Paiva, and M. Gattass, “Classification of malignancy of lung nodules in CT images using convolutional neural network,” in Anais do XVI Workshop de Informática Médica, 2016, pp. 2481–2489.

[8] X. Zhao, L. Liu, S. Qi, Y. Teng, J. Li, and W. Qian, “Agile convolutional neural network for pulmonary nodule classification using CT images,” *Int. J. Comput. Assist. Radiol. Surg.*, vol. 13, pp. 585–595, 2018.

[9] R. A. Bushara, “Deep learning-based lung cancer classification of CT images using augmented convolutional neural networks,” *ELCVIA Electron. Lett. Comput. Vis. Image Anal.*, vol. 21, 2022, doi: 10.5565/rev/elcvia.1490.

[10] S. Wankhade and V. S., “A novel hybrid deep learning method for early detection of lung cancer using neural networks,” Healthcare Analytics, vol. 3, p. 100195, 2023.

[11] S. Bakr, O. Gevaert, S. Echegaray, K. Ayers, M. Zhou, M. Shafiq, H. Zheng, W. Zhang, A. Leung, M. Kadoch, J. Shrager, A. Quon, D. Rubin, S. Plevritis, and S. Napel, “Data for NSCLC Radiogenomics (Version 4) [Data set],” The Cancer Imaging Archive, 2017. doi: 10.7937/K9/TCIA.2017.7hs46erv.

[12] A. K. Singh, V. Saini, and L. M. Saini, “ROI based detection of abnormalities in lungs using medical image processing,” in *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, Madurai, India, 2017, pp. 876–880, doi: 10.1109/ICCONS.2017.8250589.

[13] S. Blandin Knight, P. A. Crosbie, H. Balata, J. Chudziak, T. Hussell, and C. Dive, “Progress and prospects of early detection in lung cancer,” Open Biol., vol. 7, no. 9, Sep. 2017, Art. no. 170070.

[14] P. Monkam, S. Qi, H. Ma, W. Gao, Y. Yao, and W. Qian, “Detection and classification of pulmonary nodules using convolutional neural networks: A survey,” *IEEE Access*, vol. 7, pp. 78075–78091, 2019, doi: 10.1109/ACCESS.2019.2920980.

[15] T. Kadir and F. Gleeson, “Lung cancer prediction using machine learning and advanced imaging techniques,” *Transl. Lung Cancer Res.*, vol. 7, no. 3, pp. 304–312, Jun. 2018, doi: 10.21037/tlcr.2018.05.15.

[16] R. V. Kaulgud and A. Patil, “Analysis based on machine and deep learning techniques for the accurate detection of lung nodules from CT images,” *Biomed. Signal Process. Control*, vol. 85, 2023, Art. no. 105055, doi: 10.1016/j.bspc.2023.105055.

[17] A. Rehman, M. Kashif, I. Abunadi, and N. Ayesha, “Lung cancer detection and classification from chest CT scans using machine learning techniques,” in *2021 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA)*, Riyadh, Saudi Arabia, 2021, pp. 101–104, doi: 10.1109/CAIDA51941.2021.9425269.

[18] A. K. Swain, A. Swetapadma, J. K. Rout, and B. K. Balabantaray, “Classification of non-small cell lung cancer types using sparse deep neural network features,” Biomed. Signal Process. Control, vol. 87, Part A, 2024, Art. no. 105485, doi: 10.1016/j.bspc.2023.105485.