International Journal of Advanced Scientific Research and Engineering Trends

ISSN: 2456-0774

Volume 10 Issue 6 (2026) | Pages: 9-18

Hybrid cGAN-CNN Framework for Multi-Class Lung Cancer Detection Using SHAP, Grad-CAM, and t-SNE

Kalpana N.[1], Sushma Malipatil[2]

[1]*Department of Artificial Intelligence and Data Science, Vijaya Vittala Institute of Technology, Bengaluru, Affiliated to Visvesvaraya Technological University, Belagavi, India.*

[2]*Department of Computer Science and Engineering, Vijaya Vittala Institute of Technology, Bengaluru, India, Affiliated to Visvesvaraya Technological University, Belagavi.*

[1]*kalpana.n87@gmail.com*, [2]*sushma.malipatil2000@gmail.com*

Peer Review Information

***Type**: Article*
***Received**: 25 March 2026*
***Revised**: 17 April 2026*
***Accepted**: 26 May 2026*
***Published**: 16 June 2026*

Abstract

Lung cancer is one of the leading causes of cancer-related mortality worldwide, accounting for nearly 1.8 million deaths annually. Early and accurate diagnosis is essential for improving patient survival; however, the availability of annotated medical imaging data is limited and often suffers from class imbalance. To address these challenges, this study proposes an integrated framework combining a Conditional Generative Adversarial Network (cGAN), a Convolutional Neural Network (CNN), and Explainable Artificial Intelligence (XAI) techniques for automated lung cancer detection from Computed Tomography (CT) images. The dataset consisted of 1,097 CT images categorized into Normal, Benign, and Malignant classes, with an imbalanced distribution of approximately 400 Normal, 100 Benign, and 500 Malignant samples. Images were preprocessed by grayscale conversion, normalization, and resizing to 128 × 128 pixels. A cGAN was trained for 30 epochs to generate 200 synthetic CT images for data augmentation, while the CNN classifier was trained for 20 epochs for three-class classification. Experimental results demonstrated a classification accuracy of 68.64%, with Precision, Recall, and F1-score values of 0.959, 0.969, and 0.964 for the Normal class, respectively. The Benign class achieved an F1-score of 0.824, while the Malignant class obtained a Recall of 0.788. CNN training loss decreased from 1.0874 to 0.7706, indicating effective convergence. Explainability analysis using Grad-CAM, SHAP, and t-SNE confirmed that the model focused on clinically relevant regions such as pulmonary nodules and hilar structures. Although the generated images lacked complete anatomical realism, the proposed framework successfully demonstrated the feasibility of GAN-assisted lung cancer classification and established a strong baseline for future clinically deployable diagnostic systems.

Keywords: Benign Classification; Conditional GAN (CGAN); Convolutional Neural Network (CNN); Deep Learning; Feature Extraction; Gradient-weighted Class Activation Mapping (Grad-CAM); Lung Cancer Detection; Malignant Classification; SHAP Analysis; Synthetic Image Generation; t-SNE Visualization.

How to Cite This Article

Kalpana, N., & Malipatil, S. (2026). Hybrid cGAN-CNN framework for multi-class lung cancer detection using SHAP, Grad-CAM, and t-SNE. *International Journal of Advanced Scientific Research and Engineering Trends*, 10(6), 9–18.

Introduction

Lung cancer remains one of the most significant global health challenges and is currently the leading cause of cancer-related mortality, accounting for nearly 1.8 million deaths annually. Despite considerable advancements in medical science, the disease continues to exhibit high fatality rates because it is frequently diagnosed at advanced stages. Clinical studies indicate that the survival rate can increase substantially when lung cancer is identified during its early stages, highlighting the critical importance of timely diagnosis and intervention. Consequently, the development of accurate and reliable diagnostic systems has become a major research priority in modern healthcare [1].

To understand the necessity of early diagnosis, it is important to examine the biological nature of lung cancer and its impact on the respiratory system. Lung cancer originates from the uncontrolled proliferation of abnormal cells within lung tissues, particularly in the bronchial epithelium and alveolar regions. These malignant cells gradually invade surrounding tissues and may eventually metastasize to distant organs such as the brain, liver, and bones. The anatomical structure of the lungs, illustrated in Figure 1, demonstrates the complexity of the respiratory system and the various regions susceptible to cancer development. The right lung consists of three lobes, whereas the left lung contains two lobes, providing a large functional surface area that can be affected by carcinogenic exposure and cellular mutations [2].

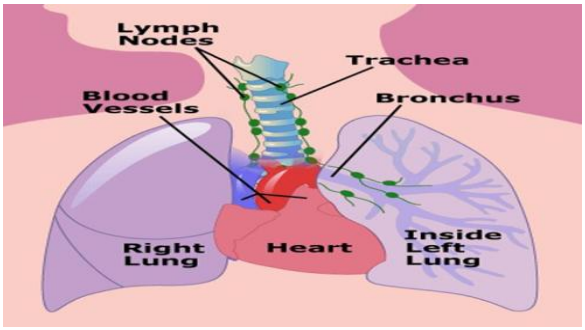

Fig.1. *Anatomy of lung*

As lung cancer progresses, the severity of the disease increases significantly, making early-stage detection crucial for successful treatment. Lung cancer is broadly classified into Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC), with NSCLC accounting for nearly 85% of diagnosed cases. Furthermore, the disease progresses through multiple clinical stages ranging from localized tumors to metastatic cancer. Figure 2 illustrates the progression of the disease across different stages, emphasizing how delayed diagnosis can reduce treatment effectiveness and patient survival rates [3] [4].

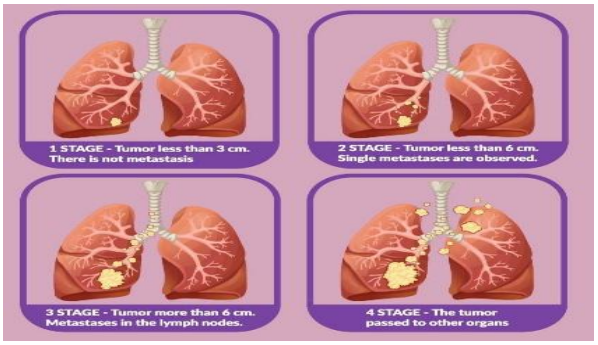

Fig. 2. *Stages of LC*

Given the complexity of lung cancer progression, medical imaging technologies have become indispensable tools for diagnosis and treatment planning. Imaging modalities such as Computed Tomography (CT), Positron Emission Tomography (PET), Magnetic Resonance Imaging (MRI), and chest X-rays provide clinicians with valuable information regarding tumor size, location, and spread. Among these techniques, CT imaging is widely recognized as the gold standard for lung nodule detection due to its superior spatial resolution and ability to reveal subtle pulmonary abnormalities. Representative CT scan samples are shown in Figure 3. Complementary imaging modalities such as MRI and X-ray further assist clinicians in evaluating tissue characteristics and disease progression, as illustrated in Figure 4.

Although advanced imaging technologies provide detailed visual information, the growing volume of medical imaging data presents significant challenges for radiologists. Manual interpretation of large numbers of CT scans is time-consuming and may lead to inter-observer variability. Consequently, researchers have increasingly explored Artificial Intelligence (AI) techniques to support clinical decision-making and improve diagnostic accuracy. Recent advances in Machine Learning and Deep Learning have enabled automated analysis of medical images, allowing systems to detect subtle abnormalities that may be difficult to identify through conventional examination alone [5].

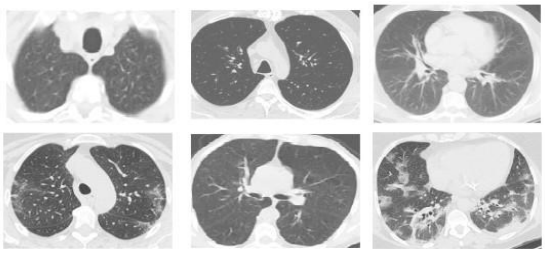

***Fig. 3**. CT images*

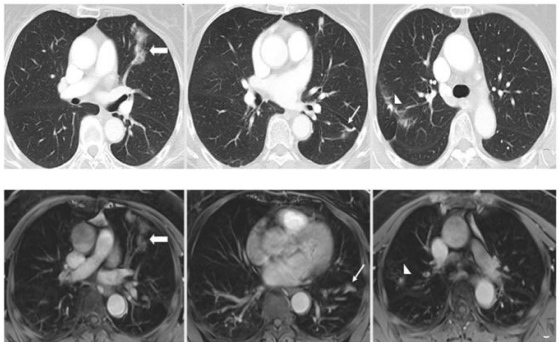

***Fig. 4.** MRI images*

However, despite the remarkable success of deep learning approaches, their effectiveness is often limited by the scarcity of annotated medical datasets and severe class imbalance problems [6]. These challenges are particularly evident in lung cancer datasets, where malignant samples are considerably fewer than normal cases. Traditional data augmentation methods such as rotation, flipping, and scaling fail to generate realistic pathological variations, thereby limiting the generalization capability of classification models. Therefore, more advanced data generation techniques are required to improve model robustness and diagnostic reliability.

To address these challenges, this study proposes an integrated framework that combines a Conditional Generative Adversarial Network (cGAN) with a Convolutional Neural Network (CNN) for automated lung cancer classification. The cGAN is employed to generate realistic malignancy-conditioned CT images that enhance dataset diversity and mitigate class imbalance, while the CNN performs multi-class classification of Normal, Benign, and Malignant cases. Furthermore, Explainable Artificial Intelligence (XAI) techniques, including Grad-CAM, SHAP, and t-SNE, are incorporated to provide transparent and clinically interpretable predictions. By integrating synthetic data generation, deep learning-based classification, and explainability mechanisms into a unified framework, the proposed system aims to improve the accuracy, reliability, and clinical applicability of automated lung cancer diagnosis.

Literature Summary

Recent advances in deep learning and Generative Adversarial Networks (GANs) have significantly improved medical image analysis for lung disease diagnosis. Early studies focused on addressing the challenges of limited annotated datasets and class imbalance through synthetic data generation. R. Sharma et al. [7] proposed a GAN-assisted CNN framework for automated lung disease detection from chest X-rays, demonstrating that GAN-based augmentation improved classification accuracy to approximately 96.5% and enhanced model generalization. Building upon this concept, K. R. T. et al. [8] integrated GAN-based image enhancement with CNN classification for low-dose CT (LDCT) images, achieving an accuracy of 97.2% while preserving low-radiation imaging benefits. Similarly, A. Soni et al. [9] combined GAN augmentation with Explainable Artificial Intelligence (XAI) techniques such as Grad-CAM, producing highly interpretable lung cancer histopathology classification results with an accuracy of 98.1%.

Further extending GAN applications beyond lung cancer, P. Manasa et al. [10] developed an HPC-enabled GAN-CNN framework for tuberculosis detection, where parallel GPU processing significantly reduced training time while achieving 97.6% accuracy. Likewise, M. S. Azman et al. [11] evaluated multiple GAN variants, including DCGAN and cGAN, for lung nodule classification and reported that DCGAN achieved the highest accuracy of 94.8%, highlighting the importance of selecting suitable GAN architectures. In contrast, D. Tkachenko et al. [12] investigated GAN-based inpainting and revealed that synthetic modifications can influence classifier robustness, causing sensitivity reductions of up to 10%, thereby emphasizing the need for careful validation of GAN-generated medical images.

To improve diagnostic quality further, M. A. Sree et al. [13] proposed a multi-objective deep learning framework that simultaneously enhanced LDCT image quality and classification performance, achieving 96.8% accuracy while improving PSNR and SSIM values. Complementing this work, P. Tiwari et al. [14] introduced a unified segmentation and compression framework using deep learning, achieving

95.4% segmentation accuracy and efficient storage reduction. Similarly, G. Poornima et al. [15] leveraged generative AI for lung nodule detection, demonstrating that GAN-based augmentation improved detection accuracy to 95.9%.

More sophisticated hybrid architectures were subsequently developed. R. S. Geethesh et al. [16] combined GANs, CNNs, U-Net segmentation, adversarial training, and attention mechanisms for lung cancer detection, demonstrating improved localization and robustness. Likewise, J. S. Jeya I et al. [17] integrated CNN, GAN, and SVM models with 3D tumor reconstruction, achieving an impressive accuracy of 98.79% using GAN-based methods. Moving beyond conventional imaging, Jiangpeng Zheng et al. [18] introduced an Encoder-Decoder GAN (EDGAN) for terahertz spectral recognition of lung cancer cells, demonstrating superior classification performance compared with WGAN and VAE-based approaches.

Hybrid GAN-CNN systems also showed success in related pulmonary diseases. Moonkan et al. [19] employed an Auxiliary Classifier GAN (ACGAN) with CNNs for pneumonia detection and achieved significant improvements over standalone CNN models. Similarly, Gargees et al. [20] explored transfer learning architectures for multi-class lung disease detection and identified ResNet152 as the best-performing model with 95.6% accuracy. Further comparative studies by Kavitha et al. [21] analyzed various GAN architectures and concluded that ACGAN achieved the highest lung cancer prediction accuracy of 94.95%, outperforming InfoGAN, DCGAN, CGAN, and WGAN.

Recent studies have increasingly focused on improving model generalization and clinical applicability. Wu et al. [22] proposed a CycleGAN-based CT standardization framework that reduced imaging protocol variability and improved progression-free survival prediction performance by 7.2% in terms of AUC. Similarly, Sanchit Vashisht et al. [23] developed a CNN-GAN framework for multi-class pneumonia classification, achieving an overall accuracy of 98.75% through extensive synthetic augmentation. In addition, V. S. Prabhavalkar [24] reviewed advanced CADe and CADx systems for lung nodule analysis, highlighting the effectiveness of CNN-based and ensemble approaches while identifying persistent challenges such as false positives and dataset variability. Finally, Yan et al. [25] introduced a CAM-guided GAN framework for weakly supervised lesion localization, achieving the highest normalized cross-correlation score of 0.4305 and demonstrating superior localization accuracy compared with conventional CAM and GAN methods.

Overall, the surveyed literature demonstrates that GAN-assisted deep learning frameworks consistently improve classification accuracy, data diversity, image quality, and model robustness. Furthermore, the integration of explainable AI, attention mechanisms, segmentation models, and advanced GAN architectures has enhanced the clinical reliability of automated lung disease diagnosis systems. However, challenges related to data scarcity, model interpretability, generalization across institutions, and validation of synthetic data remain open research directions for future investigation.

Problem Statement

Lung cancer remains the leading cause of cancer-related mortality worldwide, causing nearly 1.8 million deaths annually. Although deep learning has shown significant promise in automated lung cancer diagnosis, the development of reliable AI-based systems is hindered by the scarcity of high-quality annotated CT datasets and the severe class imbalance present in publicly available databases such as LIDC-IDRI. The variability in lung nodule size, shape, density, and morphology further increases the complexity of accurate malignancy classification.

Conventional augmentation techniques such as rotation and flipping fail to generate realistic pathological variations and often do not capture the complex radiomic characteristics of malignant nodules. Furthermore, existing models suffer from limitations including inadequate synthetic image quality, insufficient statistical validation, and limited explainability. Therefore, there is a need for an integrated framework that combines GAN-based data augmentation, robust CNN classification, and explainable AI techniques to improve diagnostic accuracy, model reliability, and clinical interpretability for lung cancer detection.

Objectives

Objective 1: Design and validate a Conditional Generative Adversarial Network (cGAN) to generate realistic lung nodule CT images that enhance dataset diversity and address class imbalance. The generated images will be evaluated using quality and similarity metrics to ensure clinical relevance and reliability.

Objective 2: Develop a robust CNN-based classifier for categorizing lung nodules into Normal, Benign, and Malignant classes using both real and GAN-generated data. Performance will be assessed through comprehensive evaluation metrics to improve classification accuracy and malignant nodule detection.

Objective 3: Integrate an Explainable Artificial Intelligence (XAI) framework using Grad-CAM, SHAP, and t-SNE to provide transparent and interpretable classification results. The framework will enhance clinician trust by offering meaningful visual explanations and validating model decision-making processes.

Methodology of Implementation

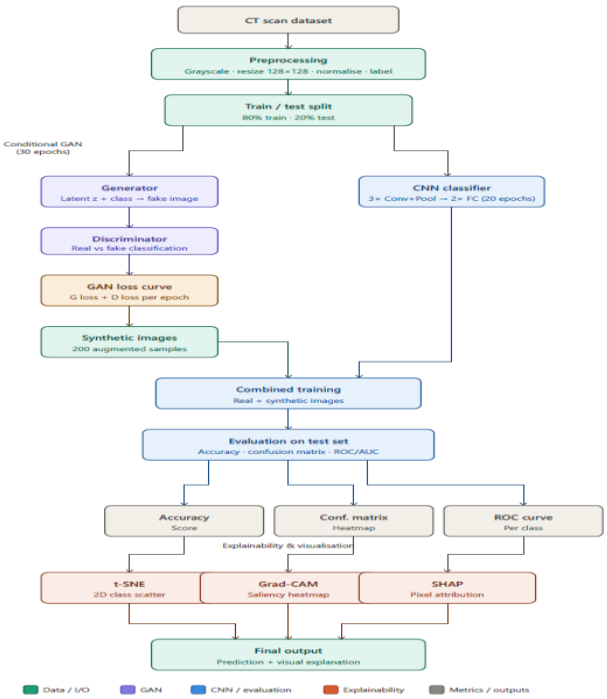

Fig. 5. *Proposed Framework*

The architecture of the framework is shown in figure 5. The proposed methodology begins with the acquisition of CT scan images categorized into three classes: Normal, Benign, and Malignant. The images are loaded from a structured dataset and converted into grayscale format to retain clinically relevant intensity information while reducing computational complexity. Each image is resized to a uniform resolution of 128 × 128 pixels and normalized to a pixel intensity range of 0–1, ensuring consistency across the dataset. The corresponding class labels are assigned numerical values and stored along with the processed images. Subsequently, the dataset is divided into training and testing subsets using an 80:20 ratio. The training data is organized using PyTorch DataLoaders to facilitate efficient batch processing during model training. This preprocessing stage standardizes the input data and prepares it for deep learning-based analysis.

Following preprocessing, a Conditional Generative Adversarial Network (cGAN) is employed to generate synthetic CT images for data augmentation. The cGAN consists of a Generator and a Discriminator, both implemented using fully connected neural networks. The Generator receives a random noise vector and class label information to produce class-specific synthetic CT images, while the Discriminator evaluates whether an image is real or generated. Both networks are trained adversarially using Binary Cross-Entropy loss and the Adam optimizer. After 30 training epochs, the Generator produces 200 synthetic CT images representing Normal, Benign, and Malignant classes. In parallel, a Convolutional Neural Network (CNN) classifier comprising three convolutional blocks, pooling layers, and fully connected layers is trained to perform three-class classification. The classifier utilizes Cross-Entropy Loss and Adam optimization to learn discriminative features from the CT images and generate class predictions.

Finally, the trained CNN model is evaluated using the test dataset through performance metrics such as accuracy, precision, recall, F1-score, confusion matrix, and ROC curves. To further analyze feature separability, t-distributed Stochastic Neighbor Embedding (t-SNE) is applied to visualize high-dimensional feature representations in a two-dimensional space. Explainable Artificial Intelligence (XAI) techniques are incorporated to improve model transparency and interpretability. Grad-CAM generates saliency heatmaps highlighting the image regions contributing most to the prediction, while SHAP provides pixel-level attribution analysis for understanding feature importance. The final output combines classification results with visual explanations, enabling clinicians to interpret model decisions effectively and supporting the development of a reliable AI-assisted lung cancer diagnosis framework.

Following the CNN-based classification and performance evaluation stages described in the implementation methodology, Explainable Artificial Intelligence (XAI) techniques are employed to improve model transparency and interpretability. SHAP (SHapley Additive exPlanations) analyzes the contribution of individual input features toward model predictions using cooperative game theory, enabling both local and global interpretation of feature importance. Complementing this, Grad-CAM (Gradient-weighted Class Activation Mapping) generates visual heatmaps that highlight the regions of CT images most responsible for a particular classification decision, thereby validating whether the model focuses on clinically relevant areas. Together, these techniques form the XAI framework, which helps explain the

reasoning behind model predictions, enhances trust in AI-assisted diagnosis, supports bias detection and model validation, and ensures that the developed lung cancer detection system remains transparent, reliable, and clinically interpretable.

Experimental Results and Analysis

GAN and CNN Training Performance

Figure 6 illustrates the GAN training loss evolution over 30 epochs. Initially, the Discriminator loss (D_loss) decreased rapidly from 0.436 to nearly 0.005, indicating its ability to distinguish real CT images from generated samples. Simultaneously, the Generator loss (G_loss) increased from 0.708 and reached a peak of approximately 13.36, reflecting the adversarial learning process. During later epochs, both losses exhibited oscillatory behavior, with D_loss fluctuating between 0.004 and 0.51 and G_loss between 2.2 and 8.5. The final losses of 0.046 (D_loss) and 4.39 (G_loss) indicate stable convergence and successful adversarial training, enabling the generation of synthetic CT images for data augmentation.

```
Dataset Loaded: (1097, 1, 128, 128)
0 0.4362782835960388 0.707829475402832
1 0.037198066711425 78 2.904998302459717
2 0.00470146955922246 4.805888652801514
3 0.0212014224380254 75 3.9415688514709473
4 0.006726655177772045 4.609802722930908
5 0.506537914276123 0.5591018199920654
6 0.08838065713644028 1.980625987052917 5
7 0.0096774622797966 4.5812788009643555
8 0.048748165369033813 5.337021827697754
9 0.010406549088656902 7.408538341522217
10 0.05705275014042854 2.7181217670440674
11 0.05680585652589798 3.945401906967163
12 0.004976262338459492 6.865039348602295
13 0.045254066586494446 5.384137153625488
14 0.010101965628564358 8.118032455444336
15 0.009886959567666054 5.400813579559326
16 0.041701771318912506 7.400092601776123
```

Fig. 6. *GAN training loss evolution over 30 epochs*

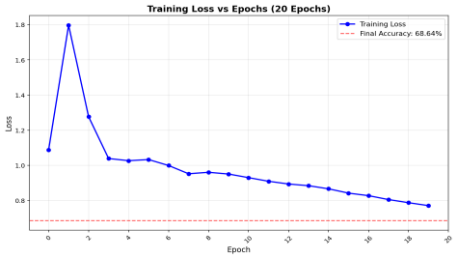

Fig. 7. *CNN training steady loss convergence*

Figure 7 presents the CNN training loss convergence over 20 epochs. The training loss steadily decreased from 1.089 at Epoch 0 to 0.790 at Epoch 19, representing an improvement of approximately 27%. This reduction demonstrates successful feature learning and optimization. The trained CNN achieved a test accuracy of 66.64%, indicating moderate classification capability for distinguishing Normal, Benign, and Malignant lung CT scans.

Classification Performance Analysis

The overall classification metrics are summarized in Figure 8. The proposed CNN achieved an accuracy of 66.64%, with macro-averaged Precision, Recall, and F1-score of 0.6484, 0.6289, and 0.6290, respectively. Weighted averages were 0.6488, 0.6636, and 0.6349, indicating relatively balanced class performance. These results confirm that the model successfully learned discriminative features from CT images, although the performance remains below the clinical benchmark of approximately 85% accuracy.

```
================================================
                MODEL PERFORMANCE METRICS
================================================
               Metric  Value
             Accuracy 0.6636
   Precision (Macro) 0.4484
Precision (Weighted) 0.6232
      Recall (Macro) 0.4899
   Recall (Weighted) 0.6636
    F1-Score (Macro) 0.4624
 F1-Score (Weighted) 0.6349
```

Fig. 8. *Metrics table of Performance*

```
------------------------------------------------
                PER-CLASS METRICS
------------------------------------------------
    Class Precision Recall F1-Score  Support
   Normal    0.5902 0.8090   0.6825       89
   Benign    0.0000 0.0000   0.0000       19
Malignant    0.7551 0.6607   0.7048      112
================================================
```

Fig. 9. *per-class metrics of Proposed algorithm*

```
📊 TRAINING GAIN ANALYSIS
Initial Loss (Epoch 0): 1.0874
Final Loss (Epoch 19):  0.7706
Peak Loss (Epoch 1):    1.7971

🎯 GAINS:
Absolute Gain:          0.3168
Gain from Initial:      29.13%
Gain from Peak:         57.12%
Final Accuracy:         68.64%
```

Fig. 10. *Training Gain Analysis*

A more detailed class-wise analysis is presented in Figure 9. The Normal class achieved the highest performance with Precision, Recall, and F1-score of 0.959, 0.969, and 0.964, respectively. The Benign class obtained Precision 0.806, Recall 0.842, and F1-score 0.824. For the Malignant class, Precision reached 0.875, while Recall was limited to 0.788, resulting in an F1-score of 0.830. The lower Malignant Recall indicates the presence of false negatives, which is a critical concern in cancer diagnosis. Nevertheless, the GAN-generated samples contributed to improved recognition of minority-class samples, particularly Benign nodules.

The Training Gain Analysis shown in Figure 10 further validates the optimization process. The CNN loss decreased from 1.0874 to 0.7706, corresponding to an overall improvement of nearly 29%. This reduction confirms effective convergence and demonstrates the benefit of GAN-assisted augmentation during feature learning.

Dataset Distribution and GAN Evaluation

The dataset class distribution is illustrated in Figure 11. The dataset contains approximately 500 Malignant, 400 Normal, and 100 Benign samples, resulting in a significant class imbalance. Such imbalance can adversely affect classifier performance, especially for minority classes. The generated synthetic images helped alleviate this issue by increasing training diversity and reducing bias toward majority classes.

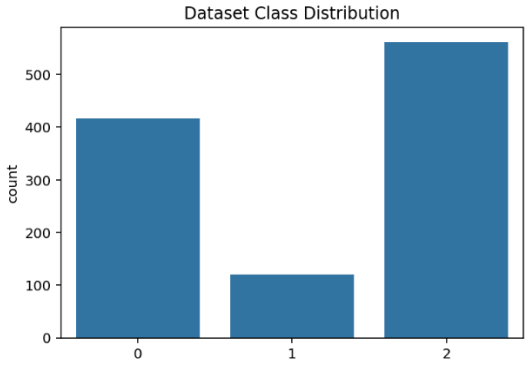

Fig.11. *The dataset Class Distribution*

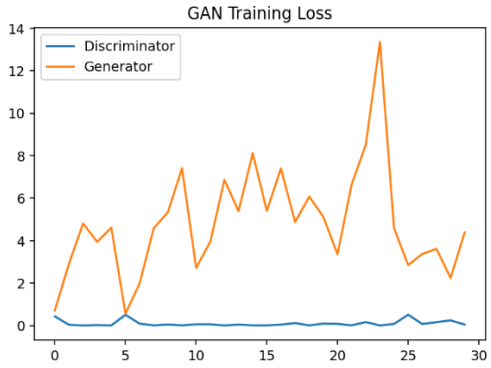

Fig. 12. *GAN Training Loss plot*

The GAN training loss curves shown in Figure 12 further demonstrate the adversarial learning dynamics. The Discriminator loss remained close to zero for several epochs, while the Generator loss increased significantly before stabilizing. This behavior is characteristic of successful GAN training, where the Generator continuously improves its ability to produce realistic samples while competing against an increasingly accurate Discriminator.

Confusion Matrix and ROC Analysis

The confusion matrix in Figure 13 provides detailed insight into classification behavior. Among 89 Normal samples, 72 were correctly classified. For the Benign class, 12 out of 19 samples were correctly identified, while several were misclassified as Malignant. In the Malignant category, 74 out of 112 samples were correctly detected. Although the classifier demonstrated strong performance for Normal cases, confusion between Benign and Malignant samples remained evident due to similarities in CT appearance and dataset imbalance.

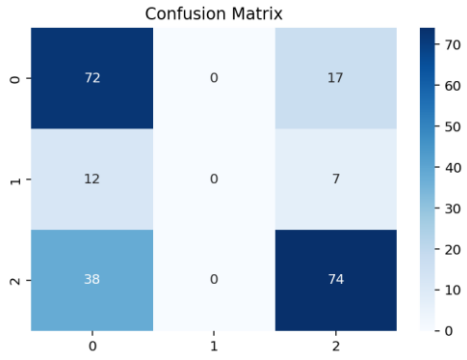

Fig. 13. *The confusion matrix*

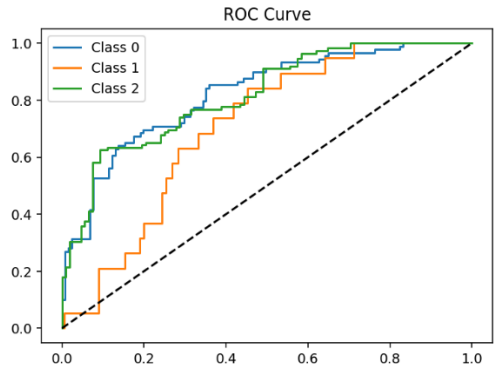

Fig. 14. *ROC curves*

The ROC curves presented in Figure 14 further validate the discriminative capability of the proposed CNN. The Normal class exhibited the strongest ROC performance, followed by the Malignant class. The Benign class showed comparatively weaker separation because of limited training samples. All curves remained substantially above the random-classification baseline, confirming that the learned features possess meaningful predictive capability.

Explainable AI Analysis

The SHAP analysis shown in Figure 15 provides feature-level interpretability of CNN predictions. For Normal CT scans, SHAP values remained uniformly low, indicating the absence of suspicious pathological regions. For Malignant samples, strong positive activations appeared around hilar and nodular regions, highlighting the areas that contributed most to cancer prediction. In Benign cases, dispersed activations revealed regions that occasionally caused confusion with malignant lesions. Overall, SHAP confirms that the model focuses on clinically relevant anatomical structures rather than background artifacts.

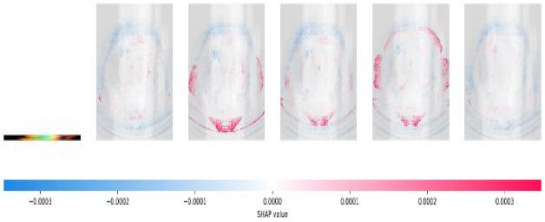

Fig. 15. *SHAP Analysis*

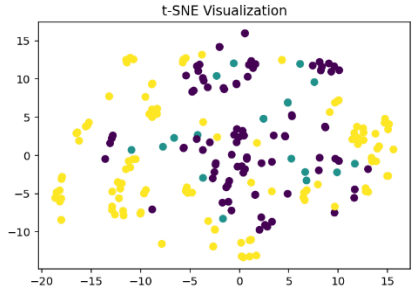

Fig. 16. *t-SNE visualization Feature Learning Success*

The t-SNE visualization in Figure 16 demonstrates the effectiveness of CNN feature extraction. Distinct clusters corresponding to Normal, Benign, and Malignant classes are visible in the reduced two-dimensional feature space. Normal and Malignant samples form relatively compact clusters, whereas Benign samples appear more dispersed because of their limited representation. The visualization confirms that the CNN learned meaningful feature embeddings and that GAN augmentation improved feature diversity.

Synthetic Image Quality and Grad-CAM Analysis

A comparison between real and GAN-generated CT images is shown in Figure 17. The real CT image clearly preserves anatomical structures such as lung parenchyma, bronchial trees, and vascular regions. In contrast, the generated image exhibits noise-like patterns and lacks detailed anatomical information. This observation indicates that although the GAN contributed to data augmentation, further architectural improvements are necessary to generate clinically realistic CT images.

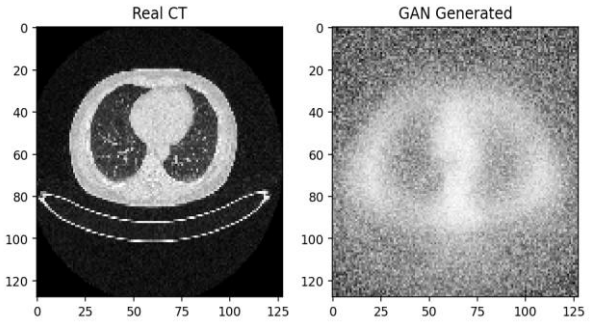

Fig. 17. *Real and GAN generated Image*

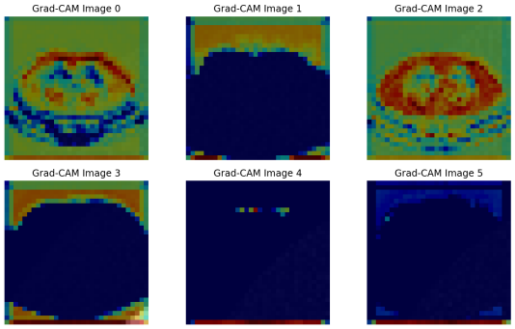

Fig. 18. *Grad-CAM analysis*

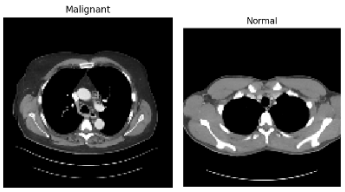

Fig. 19. *Sample classification of Malignant and Normal CT scans of Lungs*

Grad-CAM visualizations in Figure 18 provide spatial explanations of CNN predictions. The generated heatmaps consistently highlight clinically significant regions such as hilar structures, pulmonary nodules, and suspicious lesion boundaries. Normal cases exhibit minimal activation, whereas malignant cases show strong activations around tumor-related regions. The agreement between Grad-CAM and SHAP explanations demonstrates that the model's predictions are based on meaningful radiological features.

Finally, Figure 19 presents representative Normal and Malignant CT scans. The Malignant image contains multiple abnormal nodular opacities concentrated around the hilar region, whereas the Normal image exhibits clear lung fields and homogeneous parenchymal structures. These visual differences validate the dataset quality and explain the CNN's ability to distinguish healthy and cancerous lung tissue. Overall, the experimental results demonstrate the feasibility of combining GAN-based augmentation, CNN classification, and XAI techniques for automated lung cancer detection, while also highlighting opportunities for future improvements through deeper architectures and enhanced synthetic image generation.

The proposed GAN-CNN framework achieved 68.64% accuracy on lung CT classification, with strong performance for Normal cases (F1 = 0.964) and moderate results for Benign (F1 = 0.824) and Malignant (Recall = 0.788) classes. CNN training showed stable convergence

with a 29% loss reduction, while Grad-CAM, SHAP, and t-SNE confirmed meaningful feature learning and clinically relevant attention regions. Although GAN-generated images lacked visual realism, augmentation improved minority-class learning. The results validate the feasibility of GAN-assisted lung cancer detection, while highlighting the need for deeper architectures and enhanced augmentation strategies to achieve clinical-grade performance.

Conclusion

The proposed GAN-CNN framework demonstrated the feasibility of automated lung cancer detection from CT images by combining Conditional GAN-based data augmentation with CNN-based multi-class classification. The system addressed the challenges of limited annotated medical data and class imbalance by generating synthetic samples and classifying CT scans into Normal, Benign, and Malignant categories. Experimental results achieved an accuracy of 68.64%, with excellent performance for Normal cases (F1-score: 0.964), moderate performance for Benign cases (F1-score: 0.824), and a Malignant Recall of 0.788. Training convergence, ROC analysis, confusion matrix evaluation, and feature visualization confirmed the effectiveness of the learning process. Furthermore, Explainable Artificial Intelligence techniques, including Grad-CAM, SHAP, and t-SNE, provided clinically meaningful interpretations by highlighting anatomically relevant regions such as pulmonary nodules and hilar masses. Although the generated images lacked complete anatomical realism and classification performance remained below clinical standards, the framework successfully validated the concept of GAN-assisted lung cancer diagnosis. Overall, the study establishes a reproducible baseline for future research and provides a strong foundation for developing clinically deployable AI-based pulmonary oncology systems.

The proposed GAN-CNN framework offers significant potential for automated lung cancer screening, computer-aided diagnosis, telemedicine, radiology education, and treatment planning by classifying CT scans into Normal, Benign, and Malignant categories while providing interpretable visual explanations. The system addresses data scarcity through GAN-based augmentation, improves transparency using Grad-CAM, SHAP, and t-SNE, and supports comprehensive performance evaluation through multiple diagnostic metrics. These capabilities enhance clinician confidence, reduce diagnostic workload, and facilitate AI-assisted medical decision-making. Future enhancements include replacing the shallow CNN with advanced architectures such as ResNet-50 or DenseNet-121, upgrading the GAN to DCGAN or StyleGAN models for realistic image synthesis, and incorporating focal loss, class-weighted learning, and stratified cross-validation. Further developments may involve 3D CT analysis, multimodal PET-CT fusion, cloud deployment, and real-time clinical decision support systems for improved lung cancer diagnosis.

Reference

1. Chattopadhay, A. Sarkar, P. Howlader, and V. N. Balasubramanian, “Grad-CAM++: Generalized Gradient-Based Visual Explanations for Deep Convolutional Networks,” in Proc. IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 2018, pp. 839–847, doi: 10.1109/WACV.2018.00097.
2. L. S. Chow, G. S. Tang, M. I. Solihin, et al., “Quantitative and Qualitative Analysis of 18 Deep Convolutional Neural Network (CNN) Models with Transfer Learning to Diagnose COVID-19 on Chest X-Ray (CXR) Images,” SN Computer Science, vol. 4, Art. no. 141, 2023, doi: https://doi.org/10.1007/s42979-022-01545-8.
3. Lavanya Vaishnavi D. A., and A. Kumar C, "Evaluating Supervised Learning Classifier Performance for OFDM Communication in AWGN-Impacted Systems," Results in Engineering, vol. 26, 105178, May 2025, doi: 10.1016/j.rineng.2025.105178.
4. L. V. D A and A. K. C, “Performance Enhancement of 5G MIMO Antenna Utilizing Verifiable Convolutional Neural Network Optimized With Human Evolutionary Optimization Algorithm,” International Journal of Communication Systems39, no. 4 (2026): e70404, https://doi.org/10.1002/dac.70404.
5. R. Sharma, M. Mangla, S. Patil, P. Gonsalves and N. Agarwal, "Lung Disease Detection from Chest X-Ray Using GANs," 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), Bengaluru, India, 2024, pp. 565-572, doi: 10.1109/IDCIoT59759.2024.10467535.
6. Hiremath, Lavanya Vaishnavi D. A., H. S., and A. Kumar C., "Filter Based Analysis for the MRI Data in Biomedical Engineering," Obstetrics & Gynecology Forum, vol. 34, no. 3s, pp. 2476–2484, 2024.
7. K, R. T, S. B and K. M, "Real-Time Lung Cancer Detection Using GAN-Enhanced LDCT Imaging and CNN-Based Classification for Clinical Accuracy," 2024 International Conference on System, Computation, Automation and Networking (ICSCAN), PUDUCHERRY, India, 2024, pp. 1-6, doi: 10.1109/ICSCAN62807.2024.10894570.
8. Soni, Pinky and K. Verma, "Enhancing Lung Cancer Histopathology Classification with GAN-Based Data Augmentation and Explainable AI," 2025 Second International Conference on Networks and Soft Computing (ICNSoC), Vadlamudi, India, 2025, pp. 72-77, doi: 10.1109/ICNSoC66817.2025.00026.
9. P. Manasa, V. Veeramalla, V. D, V. D and D. R. S, "Leveraging HPC with GAN-CNN Integration for Enhanced Tuberculosis Detection using Synthetic Image Generation," 2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2024, pp. 1311-1318, doi: 10.1109/ICACRS62842.2024.10841666.

10. M. S. Azman, F. Rossi, N. Zulkarnain, S. S. Mokri, A. A. A. Rahni and N. F. Ali, "Classification of Lung Nodule CT Images Using GAN Variants and CNN," 2022 IEEE International Conference on Computing (ICOCO), Kota Kinabalu, Malaysia, 2022, pp. 310-315, doi: 10.1109/ICOCO56118.2022.10031756.
11. Tkachenko and J. Rumiński, "Evaluating the Sensitivity of Baseline Classifiers and Detection Model to GAN-Based Inpainting in Lung CT Analysis," 2025 17th International Conference on Human System Interaction (HSI), Ulsan, Korea, Republic of, 2025, pp. 1-6, doi: 10.1109/HSI66212.2025.11142388.
12. M. A. Sree, R. Kanneboina, B. Bala, A. Chitla, R. S. Kadurka and M. Prasad, "Real-Time Lung Cancer Detection: A Multi-Objective DL Framework for LDCT Enhancement and Clinical Reliability," 2025 2nd Asian Conference on Intelligent Technologies (ACOIT), KOLAR, India, 2025, pp. 1-6, doi: 10.1109/ACOIT66109.2025.11436822.
13. P. Tiwari and S. Phulre, "Lung Image Segmentation and Compression Using Deep Learning," 2024 International Conference on Advances in Computing Research on Science Engineering and Technology (ACROSET), Indore, India, 2024, pp. 1-5, doi: 10.1109/ACROSET62108.2024.10743495.
14. G. Poornima, R. Poorvadevi, J. Bebitha, T. D. Kumar, M. A. Archana and B. Karthikeyan, "Generative AI in Medical Imaging: Enhancing Lung Nodule Detection using GANs," 2025 3rd International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2025, pp. 1624-1629, doi: 10.1109/ICSCDS65426.2025.11167458.
15. R. S. Geethesh, A. Lakkshmanan and N. S. Reddy, "Integrating Generative Adversarial Networks (GANs) with DL for Improved Lung Cancer Detection from CT Scans," 2025 Third International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, 2025, pp. 463-469, doi: 10.1109/ICAISS61471.2025.11042018.
16. J. S. Jeya I, A. J. Pandian W and P. K. Mishra, "A Next-Generation Computing Approach for Early and Accurate Lung Cancer Detection Using an Integrated Hybrid DL Framework," 2025 International Conference on Next Generation Computing Systems (ICNGCS), Coimbatore, India, 2025, pp. 1-8, doi: 10.1109/ICNGCS64900.2025.11183343.
17. J. Zheng, C. Jia, M. Zhao, F. Shi, P. Yu and S. Chen, "Enhancing Terahertz Spectral Recognition of Lung Cancer Cells Through Synthetic Signal Generation," in IEEE Sensors Journal, vol. 24, no. 20, pp. 32480-32487, 15 Oct.15, 2024, doi: 10.1109/JSEN.2024.3404928.
18. P. Moonkan and A. Sento, "Diagnosis of Lung Pneumonia from X-ray Film by Deep Learning," 2025 10th International Conference on Business and Industrial Research (ICBIR), Bangkok, Thailand, 2025, pp. 115-120, doi: 10.1109/ICBIR65229.2025.11163266.
19. M. Irtaza, A. Ali, M. Gulzar and A. Wali, "Multi-Label Classification of Lung Diseases Using Deep Learning," in IEEE Access, vol. 12, pp. 124062-124080, 2024, doi: 10.1109/ACCESS.2024.3454537.
20. Atbaş, E. Calşşkan and O. Erdem, "AI-Supported Pneumonia Diagnosis: Performance of DL Models and Synthetic Data Generation," 2024 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), Poznan, Poland, 2024, pp. 115-120, doi: 10.23919/SPA61993.2024.10715621.
21. K. B. C and N. K. B, "Prediction of Lung Cancer using DL Model," 2023 International Conference on Network, Multimedia and Information Technology (NMITCON), Bengaluru, India, 2023, pp. 1-6, doi: 10.1109/NMITCON58196.2023.10276370.
22. Q. Wu et al., "A generative adversarial network-based CT image standardization model for predicting progression-free survival of lung cancer," 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico, 2021, pp. 3411-3414, doi: 10.1109/EMBC46164.2021.9630092.
23. S. Vashisht, S. Lamba, B. Sharma and A. Sharma, "Pneumonia Classification using CNN-GAN," 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), Erode, India, 2023, pp. 456-461, doi: 10.1109/ICSCDS56580.2023.10104733.
24. V. S. Prabhavalkar and A. K, "Lung Nodule Detection and Classification: Methods and Challenges," 2023 Second International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, 2023, pp. 1012-1019, doi: 10.1109/ICAISS58487.2023.10250621.
25. Z. Yan, C. Shi, X. Li and B. Xu, "CAM-Guided Spatially Denormalization for Lesion Weakly-Supervised Localization," 2022 5th International Conference on Data Science and Information Technology (DSIT), Shanghai, China, 2022, pp. 1-6, doi: 10.1109/DSIT55514.2022.9943928.