International Journal of Robotics and Control Systems
Vol. 5, No. 3, 2025, pp. 1701-1718
ISSN 2775-2658
http://pubs2.ascee.org/index.php/ijrcs

# Deep Learning Approach to Lung Cancer Detection Using the Hybrid VGG-GAN Architecture

Yuri Pamungkas [a,1,*], Djoko Kuswanto [a,2], Achmad Syaifudin [a,3], Evi Triandini [b,4], Dian Puspita Hapsari [c,5], Kanittha Nakkliang [d,6], Muhammad Nur Afnan Uda [e,7], Uda Hashim [e,8]

[a] Department of Medical Technology, Institut Teknologi Sepuluh Nopember, Surabaya, 60111, Indonesia
[b] Department of Information System, Institut Teknologi dan Bisnis STIKOM Bali, Denpasar, 80234, Indonesia
[c] Department of Informatics Engineering, Institut Teknologi Adhi Tama Surabaya, Surabaya, 60117, Indonesia
[d] Department of Health System Management, Valaya Alongkorn Rajabhat University, Pathum Thani, 13180, Thailand
[e] Department of Electrical and Electronics Engineering, Universiti Malaysia Sabah, Kinabalu, 88400, Malaysia
[1] yuri@its.ac.id; [2] djoko.kuswanto@its.ac.id; [3] syaifudin@its.ac.id; [4] evi@stikom-bali.ac.id; [5] dian.puspita@itats.ac.id; [6] kanittha.nak@vru.ac.th; [7] nurafnan@ums.edu.my; [8] uda@ums.edu.my
* Corresponding Author

ARTICLE INFO

**Article history**
Received May 12, 2025
Revised June 18, 2025
Accepted July 07, 2025

**Keywords**
Lung Cancer Detection;
Generative Adversarial Networks;
VGG Convolutional Neural Network;
Data Augmentation;
Medical Image Classification

ABSTRACT

Lung cancer ranks among the primary contributors to cancer-related deaths globally, highlighting the need for accurate and efficient detection methods to enable early diagnosis. However, deep learning models such as VGG16 and VGG19, commonly used for CT scan image classification, often face challenges related to class imbalance, resulting in classification bias and reduced sensitivity to minority classes. This study contributes by proposing an integration of the VGG architecture and Generative Adversarial Networks (GANs) to improve lung cancer classification performance through balanced and realistic synthetic data augmentation. The proposed approach was evaluated using two datasets: the IQ-OTH/NCCD Dataset, which classifies patients into Benign, Malignant, and Normal categories based on clinical condition, and the Lung Cancer CT Scan Dataset, annotated with histopathological labels: Adenocarcinoma, Squamous Cell Carcinoma, Large Cell Carcinoma, and Normal. The method involves initial training of the VGG model without augmentation, followed by GAN-based data generation to balance class distribution. The experimental results show that, prior to augmentation, the models achieved relatively high overall accuracy, but with poor performance on minority classes (marked by low precision and F1-scores and FPR exceeding 8% in certain cases). After augmentation with GAN, all performance metrics improved dramatically and consistently across all classes, achieving near-perfect precision, TPR, F1-score, and overall accuracy of 99.99%, and FPR sharply reduced to around 0.001%. In conclusion, the integration of GAN and VGG proved effective in overcoming data imbalance and enhancing model generalization, making it a promising solution for AI-based lung cancer diagnostic systems.



## 1. Introduction

Lung cancer continues to be among the most common and fatal types of cancer globally. Based on the most recent estimates from the American Cancer Society, approximately 234,580 new lung

cancer cases and 125,070 related deaths are anticipated in the United States by 2024 [1]. Although there have been notable improvements in lung cancer therapies over the years, the overall outlook for patients is still grim, with a five-year survival rate hovering around 20% [2]. A major factor contributing to this low survival rate is the fact that many individuals are diagnosed when the disease has already progressed to a late stage, making treatment less effective. Detecting the disease early is vital to improving recovery prospects for lung cancer patients. If identified during a localized phase, the five-year survival rate can rise significantly to 63.7% [3]. Nevertheless, only about 27% of lung cancer cases are currently identified at this early stage, with the majority being diagnosed after the cancer has advanced [4]. Widely used screening techniques, such as low-dose CT, still face limitations in identifying small lung nodules, which are often indicative of early-stage lung cancer.

In recent years, the rapid development of AI, particularly in deep learning, has introduced promising possibilities for the early detection of lung cancer. Among these, CNNs have emerged as a powerful tool for automatically and accurately interpreting complex patterns in medical imaging. Specific CNN architectures, such as VGG-16, have been widely utilized for classifying medical images and have demonstrated strong capabilities in distinguishing between malignant and benign lung nodules. For instance, Pandian et al. (2022) implemented a method combining CNN and GoogleNet, with VGG-16 serving as the foundational layer, to detect lung cancer from CT scans [5]. Their model achieved an impressive classification accuracy of 98%, highlighting CNN's potential to enhance the speed and accuracy of early lung cancer diagnosis. Additionally, Sait (2023) introduced a deep learning-based model for detecting lung cancer using PET or CT scans, incorporating data augmentation and optimized training strategies like quantization-aware training [6]. This model reached 98.6% accuracy, demonstrating its viability for integration into clinical workflows to aid in early detection. Further, Klangbunrueang et al. (2025) assessed the performance of several CNN models, ResNet50, VGG16, and MobileNetV2, for lung cancer classification from CT images [7]. Their study revealed that VGG16 achieved the highest accuracy at 98.18%, reaffirming its suitability and effectiveness for clinical applications in lung cancer detection.

Mohamed et al. (2023) integrated a CNN with the EOSA, a metaheuristic method, to enhance the performance of lung cancer detection [8]. Their method achieved an accuracy of 93.21%, demonstrating that incorporating optimization techniques into deep learning models can yield better results, particularly in complex lung cancer cases. In another study, Shah et al. (2023) introduced an ensemble strategy that combined three distinct CNN models to detect lung nodules, whether "malignant" or "benign" [9]. This ensemble approach reached a 95% accuracy rate, surpassing the baseline model and indicating that merging multiple CNN architectures can improve the detection of challenging lung nodules. Furthermore, Thangamani et al. (2024) proposed a weighted CNN approach to diagnose lung cancer based on gene expression data [10]. They employed Z-score normalization and selected genes using the LFCS optimization algorithm, which led to enhanced classification accuracy. This method highlights the promise of integrating genetic information with CNNs to increase the precision of lung cancer predictions. Despite these advancements, a notable limitation in the application of CNNs is their reliance on large volumes of high-quality medical imaging data for training, resources that are not readily accessible in many healthcare settings.

To address these limitations, recent studies have started integrating CNNs with GANs, which can produce realistic synthetic images that closely mimic the original data. This combination, often referred to as VGG-GAN, holds considerable promise in enhancing the precision of lung cancer diagnosis. GAN-generated synthetic data can increase the robustness of CNN models to variations in lung cancer image patterns, while overcoming the limitations of training data. This approach can theoretically increase the sensitivity and specificity in the early detection of lung cancer, thereby assisting doctors in making a more precise and rapid diagnosis. Thus, research into CNN deep learning approaches using the VGG-GAN architecture has high relevance in the current medical context. The successful development of this technology can not only increase the early detection rate of lung cancer, but also directly impact the prognosis and survival rate of patients. Therefore, further exploration of this method is essential to support public health strategies to meet the challenge of lung cancer.

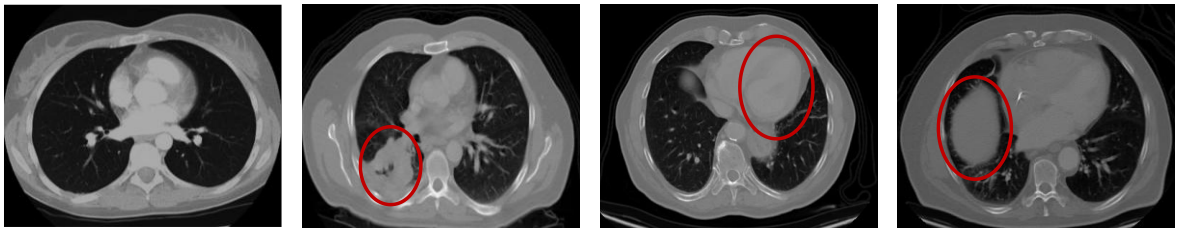
## 2. Dataset and Method

### 2.1. Dataset

This research employed two datasets: the IQ-OTH/NCCD lung cancer dataset [11] and the Lung Cancer CT Scan Dataset [12]. The IQ-OTH/NCCD dataset comprises medical CT scan images of the lungs and is commonly utilized in studies focusing on automated lung cancer classification through machine learning, particularly deep learning techniques. It contains a total of 1,097 images divided into three primary categories. Among these, 120 images depict benign tumors, which are not life-threatening but must be accurately identified to prevent incorrect diagnoses. The largest portion of the dataset, with 561 images, represents malignant tumors that pose serious health risks if not diagnosed and treated promptly. The remaining 416 images show normal lung tissue and serve as control data for model development. This dataset was gathered from two prominent healthcare institutions in Iraq: the Iraq-Oncology Teaching Hospital and the National Centre for Cancer Diseases. All images underwent thorough review by specialists in oncology and radiology to maintain high-quality standards. The CT scans offer high resolution, with slice thicknesses of around 1 mm, enabling clear visualization of lung anatomy. This level of detail greatly facilitates automated feature extraction by deep learning models. In parallel, the Lung Cancer CT Scan dataset offers a more detailed classification of lung cancer types based on histopathological characteristics. It includes 315 lung CT images sorted into four distinct categories: Adenocarcinoma (120 images), Large Cell Carcinoma (51 images), Squamous Cell Carcinoma (90 images), and 54 images of normal lung tissue. The inclusion of histopathological labels enhances the model's ability to learn and recognize specific visual traits associated with each cancer type, thereby improving the precision of automated diagnostic systems. IQ-OTH/NCCD lung cancer dataset shown in Fig. 1. Lung cancer CT scan dataset based on histopathological type in Fig. 2.

a. Normal b. Benign c. Malignant

**Fig. 1.** IQ-OTH/NCCD lung cancer dataset [11]



a. Normal b. Adenocarcinoma c. Squamous-cell Carcinoma d. Large-cell Carcinoma

**Fig. 2.** Lung cancer CT scan dataset based on histopathological type [12]

These two datasets have important value in the development of modern health technologies, especially in utilising the potential of AI for early and accurate detection of lung cancer. However, common challenges such as imbalance between classes remains a significant problem. This imbalance can cause bias in the classification algorithm, such that the category with the highest number of images may dominate the model's decision. In general, the open availability of these two datasets allows researchers around the world to develop and evaluate various ML and DL approaches. Thus, the IQ-OTH/NCCD and Lung Cancer CT Scan Datasets not only support the advancement of scientific research in the field of lung cancer early detection, but also open up opportunities for the development of automated diagnosis systems that are increasingly accurate, efficient, and able to reduce the workload of radiology doctors in healthcare facilities [13].

## 2.2. Generative Adversarial Network (GAN)

GAN is an innovative form of deep learning specifically designed to generate synthetic data of highly realistic quality [14]. This approach is fundamentally different from other conventional techniques as it involves a competitive process between two opposing neural networks, the generator network and the discriminator network [15]. Instead of simply memorising patterns from the original data, GAN creates a new representation that has similar characteristics, but is not a direct duplication of the original data [16]. In the context of class imbalance, GAN appears as an effective solution due to its ability to generate new data for classes that have a limited amount of data [17]. Class imbalance is often an obstacle in medical applications, especially image-based diagnosis such as lung cancer detection, where one class (e.g. malignant tumour cases) has much more data than other classes (e.g. benign or normal cases). This situation causes the classification model to be biased towards the majority class, thus decreasing the sensitivity towards the minority class [18]. This is where GANs play a critical role, generating additional synthetic minority class data to improve the dataset distribution and reduce bias in the classification. The main concept of GANs shown in Fig. 3.

The GAN architecture generally consists of two neural networks with different roles and functions [20]. The first network, called a generator, is responsible for converting random input (noise) into synthetic data that resembles the characteristics of the original data [21]. Generators are generally built using a series of deconvolution layers or transposed convolutional layers to generate visual data, such as medical images, from latent vector inputs. These layers gradually increase the resolution of the input until it reaches the desired image shape [22]. The second network, called a discriminator, is designed as a classifier that evaluates the realistic level of the received data, and provides feedback to the generator regarding the authenticity of the generated data [23]. Discriminators generally use CNNs to detect certain visual patterns in both synthetic and original data [24]. Through this process, the discriminator gradually hones its ability to recognise fake data generated by the generator.

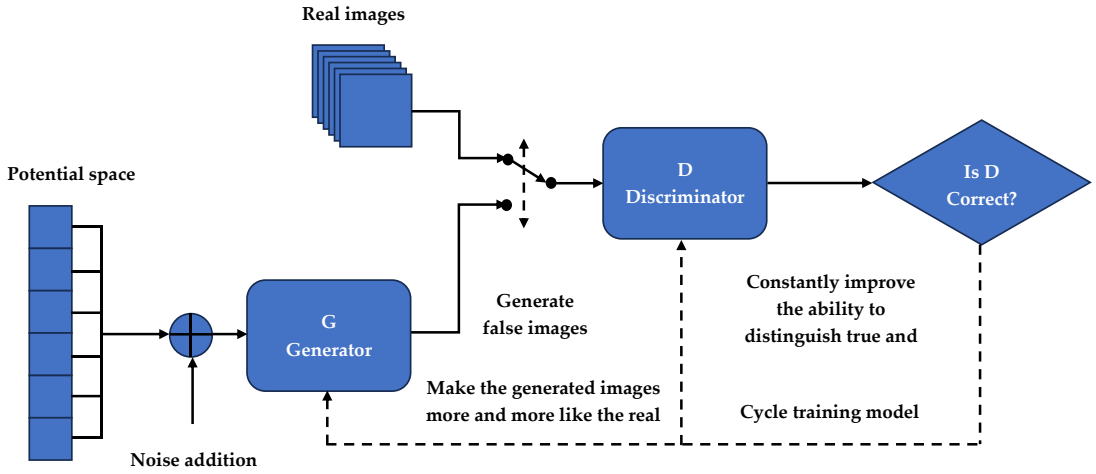



**Fig. 3.** The main concept of GANs [19]

The training of these two networks is done simultaneously through an adversarial or competitive mechanism [25]. The generator continues to improve its ability to create synthetic data that is increasingly difficult for the discriminator to distinguish, while the discriminator improves its ability to detect fake data [26]. This process proceeds iteratively until it reaches a Nash Equilibrium point, where the generated synthetic data is almost indistinguishable from the original data. Overall, the use of GAN to address data class inequality provides significant benefits [27]. In addition to generating additional data for realistic minority classes, GAN also helps enrich the variety of available data, which directly contributes to improving the generalisability and performance of classification algorithms in medical contexts and other deep learning applications [28].

## 2.3. VGG Architectures

In the landscape of modern deep learning architectures, VGG occupies an important position as an example of systematising convolutional network design based on modular principles [29]. It is not

built with complex filter variations or residual connections as ResNet is, but instead with repetitions of simple 3×3 convolutional blocks [30]. This gives it the advantage of ease of reconstruction and reusability, especially in scientific experiments. The success of VGG lies not only in its performance in the ImageNet competition, but also in its role as a "backbone" in transfer learning for various other domains such as lesion detection in lung CT scans [31]. CNN architectures shown in Fig. 4.

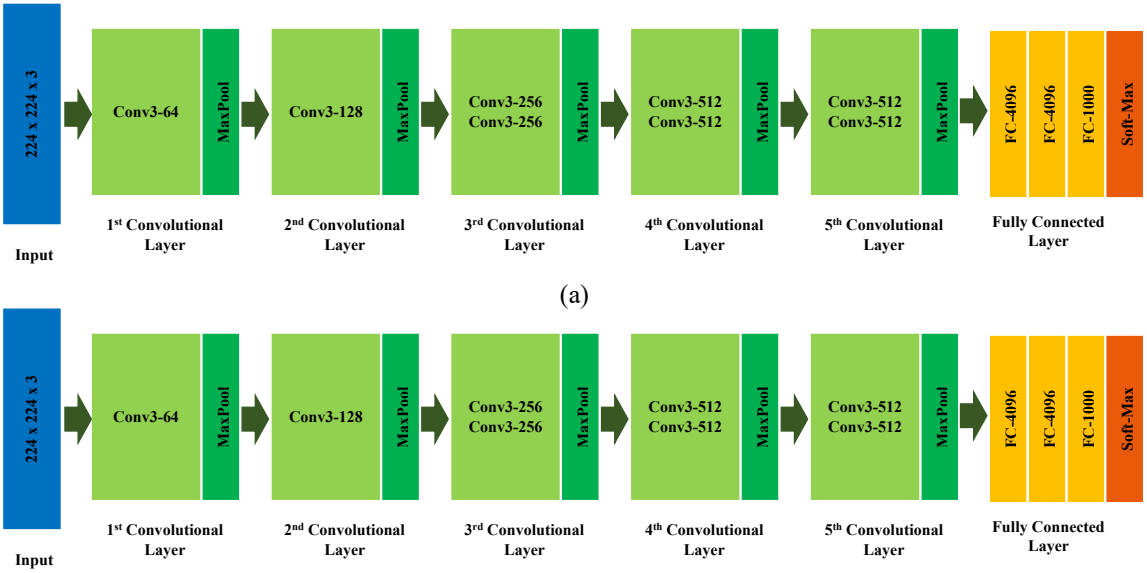



**Fig. 4.** CNN architectures: (a) VGG16 and (b) VGG19 [25]

VGG16 is the most widely used and studied structure, both as a base model and as a feature extractor in hybrid architectures [32]. Consisting of 13 convolutional layers, this architecture forms a complexity gradation of progressively deeper convolutional blocks, especially in blocks 3-5 which each consist of three consecutive layers [33]. VGG16 shows superior performance in distinguishing malignant from benign lesions due to its ability to recognise repetitive morphological patterns scattered at various levels of resolution. In the medical context, this model is even often used not only for classification, but also for segmentation when combined with a decoder such as the FCN or U-Net architecture [34]. As for VGG19, it takes the VGG architecture to the highest level of complexity with 16 more evenly distributed convolutional layers, specifically with four convolutions per block at the end. This addition aims to increase the depth of non-linearity, allowing the network to encode more complex visual structures [35].

In clinical or research applications with high-resolution images, such as lung CT with multilevel intensity mapping, VGG19 is able to store detailed information of abnormal tissue texture, including microscopic patterns or subtle transitions between tissue layers [36]. However, the consequence is a much larger computational burden and longer training requirements, so its use is often restricted to systems with high GPU support or as a pretrained, finetuned model. Conceptually, the strength of VGG lies not in form innovation, but rather in structural discipline [37]. Each model is built consistently, allowing direct comparison between variants for experimental purposes [38]. In the context of developing an intelligent system for lung cancer detection, the VGG architecture provides a stable basis for evaluating the extent to which the depth of the network is able to distinguish subtle visual anomalies that may be missed by shallower or overly complex networks [39]. Its modular approach also facilitates integration into medical pipelines, including GAN, segmentation, or end-to-end classification systems [40].

### 2.4. K-Fold Cross-Validation

K-fold cross-validation is a model evaluation technique that randomly divides the dataset into K different groups (folds), each having approximately the same amount of data [41]. The purpose of this division is to ensure that all samples in the dataset can be used as both training and testing data in turn

[42]. This cross-validation process is performed in multiple iterations (K times). In each iteration, 1-fold will be used as test data (test set), while the other K-1 folds are used as training data (training set). After the iterations are completed, the average of the evaluation results from each iteration is calculated to provide an overview of the model performance [43]. Illustration of k-fold cross validation in splitting data shown in Fig. 5.

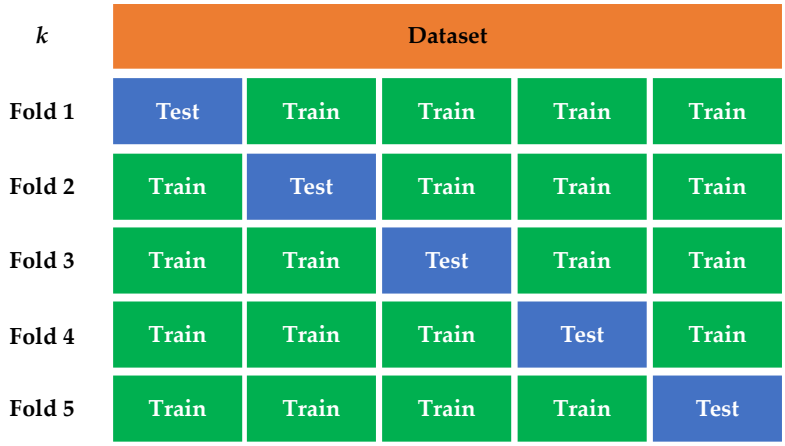



**Fig. 5.** Illustration of k-fold cross validation in splitting data [44]

This method provides great benefits as all the data in the dataset participate as test data in turn. Therefore, the evaluation results obtained from K-fold cross-validation tend to be more reliable and realistic than when relying on a single train-test split [45]. However, there are some aspects that need to be considered. Too small a value of K, such as K=2 or 3, may lead to less accurate evaluation due to limited training data. On the other hand, too large a value of K (e.g. K=n, i.e. leave-one-out cross-validation) will increase the computational burden significantly [46]. In this study, K=10 is chosen as the optimal value, providing a good balance between evaluation performance and computational requirements.

### 2.5. Performance Metrics

When assessing the effectiveness of lung cancer detection systems that rely on medical imaging, selecting appropriate evaluation metrics is crucial to accurately measure the reliability and precision of the model's output [47]. Various performance indicators, such as precision, TPR, FPR, F1-score, and overall accuracy, are frequently used to provide a comprehensive view of the model's performance, especially considering the common issue of class imbalance in medical datasets [48]. Among these metrics, precision evaluates how well the model correctly identifies actual cancer cases out of all the instances it predicted as positive. This metric is computed using the formula.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

True Positive Rate (TPR), also known as sensitivity or recall, measures the ability of the model to capture all cancer cases that are actually present in the data. The formula is:

$$TPR = \frac{TP}{TP + FN} \tag{2}$$

The False Positive Rate (FPR) represents the proportion of false positive predictions out of all negative cases. The formula:

$$FPR = \frac{FP}{FP + TN} \tag{3}$$

F1-score is the harmonic mean of precision and recall. F1-score provides a balance between precision and TPR, so it is very useful when we want to avoid too extreme a trade-off between the

two metrics, especially in situations of unbalanced data (e.g. cancer cases are much less than normal data). Formula:

$$F1 - score = 2 \times \frac{Precision \times TPR}{Precision + TPR} \quad (4)$$

Accuracy is the most common metric, which measures the proportion of correct predictions out of all predictions made.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Overall, in developing a lung cancer detection model, a combination of these metrics should be considered together. Precision and FPR assess the model's level of vigilance against false positives, while TPR and F1-score assess the effectiveness in detecting actual cancer cases. Accuracy, while still relevant, should be used with caution and not as the sole reference in clinical decision-making.

## 3. Results and Discussion

### 3.1. The Use of GAN on the Lung Cancer Datasets

This research is focused on developing a lung cancer detection method that integrates the GAN approach with the VGG convolutional network architecture. The use of GAN aims to increase the diversity and balance of training data distribution [49]. By synthesising medical images that resemble the original image, the GAN approach is expected to improve model training stability and lung cancer detection accuracy [50]. Meanwhile, the VGG architecture is used as the backbone in the deep visual feature extraction process, thanks to its convolutional hierarchical structure that has been proven effective in various medical image recognition tasks [51]. In this study, two independent datasets are used to test the generalisation and reliability of the model. The first dataset is the IQ-OTH/NCCD Lung Cancer Dataset, which consists of lung CT scan images and is categorised into three classes based on the patient's clinical condition, namely Benign, Malignant, and Normal [11]. This dataset is highly relevant for evaluating the model's performance in differentiating cancer conditions based on their progressivity. The second dataset used is the Lung Cancer CT Scan Dataset, which offers classification based on the histopathological type of lung cancer. This dataset includes four classes, namely Adenocarcinoma, Large Cell Carcinoma, Squamous Cell Carcinoma, and Normal, which provides more complex classification challenges due to the similarity of visual features between cancer subtypes [12].

The IQ-OTH/NCCD Dataset and Lung Cancer CT Scan Dataset used in this study have a common problem in the medical domain, which is the imbalance in the amount of data between classes. In the IQ-OTH/NCCD dataset, the image distribution for each class shows a significant difference in number. In some cases, the number of samples for the “Malignant” category dominates, while the “Benign” and “Normal” classes are much less. This imbalance causes the model to learn better on the majority class and ignore or misclassify the minority class, which is often clinically important. Similarly, in the Lung Cancer CT Scan Dataset, categories based on histopathological types such as Adenocarcinoma (AC), Large Cell Carcinoma (LC), Squamous Cell Carcinoma (SC), and Normal (NL) are not evenly distributed. Cancer types such as Large Cell Carcinoma usually have fewer samples than Adenocarcinoma or Squamous Cell Carcinoma. This imbalance increases the risk of classification bias, where the model more often predicts the class with the most data, and reduces the model's sensitivity to less common cancer types. To overcome these issues, this study utilizes the GAN method as a strategy for synthetic data augmentation. GAN is applied to produce additional images for underrepresented classes, helping to create a more balanced dataset. The subsequent section presents the data distribution in the IQ-OTH/NCCD and Lung Cancer CT Scan datasets before and after implementing the GAN technique.

Fig. 6 shows the data distribution of the IQ-OTH/NCCD Lung Cancer Dataset before and after the augmentation process using GANs. The left pie chart shows the original distribution of the dataset,

which is clearly imbalanced between classes. The Malignant class dominates with 561 samples (51.14%), followed by Normal with 416 samples (37.92%), while the Benign class consists of only 120 samples (10.94%). This imbalance indicates that models trained without data balancing are likely to be biased towards the majority class and have low classification performance towards minority classes such as Benign. The right diagram shows the distribution results after applying data augmentation using GANs. Through this approach, the number of samples in each class (Benign, Malignant, and Normal) is balanced to 100,000 samples each, or about 33.33% of the total dataset. This process is done by generating new synthetic images in classes that have a limited number of samples, so that all classes now have equal representation in the model training process.

Fig. 7 shows the distribution of data in the Lung Cancer CT Scan Dataset, both before and after augmentation using GANs. The pie chart on the left shows the original distribution of data by histopathological type of lung cancer, which includes four classes: NL (Normal Lung), AC (Adenocarcinoma), SC (Squamous Cell Carcinoma), and LC (Large Cell Carcinoma). It can be seen that the number of samples is not evenly distributed, with the AC class dominating with 120 samples (38.10%), followed by SC with 90 samples (28.57%). Meanwhile, the LC and NL classes are much smaller, with only 51 (16.19%) and 54 (17.14%) samples respectively. This imbalance creates a potential bias in the classification model, where the model is more inclined to recognise the majority class and tends to ignore minority classes such as LC and NL. In response to this imbalance, a GANs-based augmentation process was performed to selectively generate additional synthetic data on minority classes. The right diagram shows the result of the process, where the amount of data in each class has been increased to 100,000 samples per class, resulting in a balanced distribution (25%) for each category. With this balance, the deep learning model used can learn equally from all classes without significant bias.

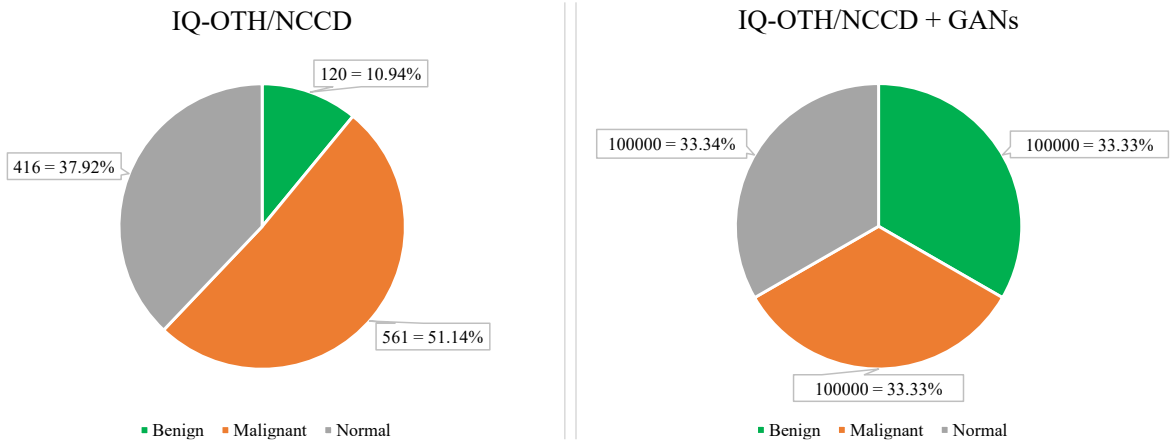



**Fig. 6.** Data distribution on the IQ-OTH/NCCD dataset before and after using GANs

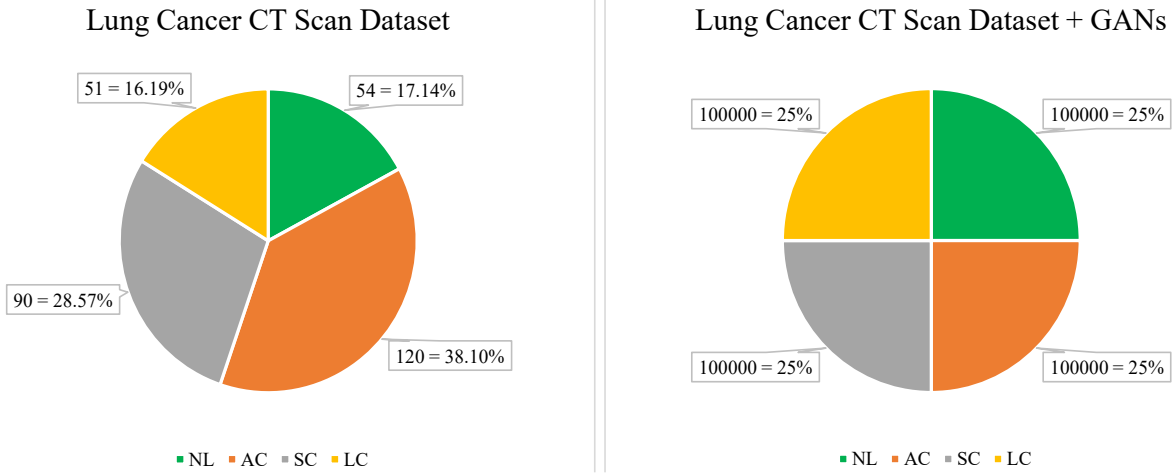



**Fig. 7.** Data distribution on the lung cancer CT scan Dataset before and after using GANs

### 3.2. Lung Cancer Detection on IQ-OTH/NCCD Dataset

This study employs the VGG model to detect lung cancer using the IQ-OTH/NCCD dataset, tested under two different scenarios: one without the use of Generative Adversarial Network (GAN) and the other incorporating GAN. The comparison aims to highlight the extent of performance enhancement achieved through the VGG model, particularly in addressing class imbalance issues within the dataset. The dataset is divided into separate subsets, allocating 80% for training, 10% for validation, and the final 10% for testing purposes. The IQ-OTH/NCCD dataset organizes its samples into three distinct categories: Benign, Malignant, and Normal. This structured division ensures that the model is trained effectively while maintaining an unbiased evaluation on unseen data. The clear categorization of classes also aids in assessing the model's ability to distinguish between different lung conditions. The following presents the VGG model's performance results in lung cancer detection using this dataset.

In Table 1 (without GAN), it can be seen that the model has inconsistent performance between classes. For example, the Benign class only has a precision of 30.65% and a TPR of 100%, but its F1-score is low at 46.91%, indicating an imbalance between precision and sensitivity. The Malignant class has a very low TPR (54.55%), indicating that the model struggles to recognise most malignant cancer cases despite its high precision (100%). Averaging the metrics shows that although the overall accuracy reached 82.58%, the imbalance between classes resulted in suboptimal performance, especially in the minority classes. In contrast, Table 2 (with GAN) shows a very significant performance improvement across metrics and classes. All classes show precision, TPR, F1-score and accuracy values close to 100%, with an overall average accuracy of 99.98%. This indicates that augmenting the synthetic data with GANs successfully balanced the data distribution and allowed the VGG16 model to learn more representatively for all classes. In addition, the FPR is very low (<0.01), indicating that the model has an excellent ability to minimise false negative prediction errors, which is crucial in the context of cancer detection.

In Table 3, which shows the performance of VGG19 without GAN, the evaluation results are relatively good. The average precision reached 91.21%, TPR was 91.94%, and F1-score was 91.54%. The Benign class, which generally has less data, showed slightly lower performance than the other two classes, with a precision of 80.00% and F1-score of 82.05%. Meanwhile, the Malignant and Normal classes show evaluation values that are close to optimal, especially Malignant with F1-score reaching 97.75%. However, the FPR value in the Malignant class (3.41%) is still relatively high, which indicates the possibility of false positives that are not ignored. In Table 4, after augmentation using GANs, there is a significant improvement in all evaluation metrics. All classes show almost perfect precision, TPR, and F1-score values (≥99.99%), and very low FPR (<0.005%), reflecting the very high classification accuracy of the model. Even the Benign class, which was previously most affected by data imbalance, now achieves 99.99% accuracy, indicating that the data augmentation process has significantly improved the model's ability to recognise patterns in the minority class.

**Table 1.** Performance of the VGG16 in lung cancer detection (IQ-OTH/NCCD dataset without GAN)

| Class | Performance Evaluation (%) | | | | |
|---|---|---|---|---|---|
| | Precision | TPR | FPR | F1-score | Accuracy |
| Benign | 30.65 | 100.00 | 27.39 | 46.91 | 75.57 |
| Malignant | 100.00 | 54.55 | 0.00 | 70.59 | 77.27 |
| Normal | 95.45 | 91.30 | 2.80 | 93.33 | 94.89 |
| **Average** | **75.37** | **81.95** | **10.06** | **70.28** | **82.58** |

**Table 2.** Performance of the VGG16 in lung cancer detection (IQ-OTH/NCCD dataset + GAN)

| Class | Performance Evaluation (%) | | | | |
|---|---|---|---|---|---|
| | Precision | TPR | FPR | F1-score | Accuracy |
| Benign | 99.93 | 100.00 | 0.030 | 99.97 | 99.98 |
| Malignant | 100.00 | 99.93 | 0.000 | 99.97 | 99.98 |
| Normal | 100.00 | 100.00 | 0.000 | 100.00 | 100.00 |
| **Average** | **99.98** | **99.98** | **0.010** | **99.98** | **99.98** |

**Table 3.** Performance of the VGG19 in Lung Cancer detection (IQ-OTH/NCCD dataset without GAN)

| Class | Performance Evaluation (%) | | | | |
|---|---|---|---|---|---|
| | Precision | TPR | FPR | F1-score | Accuracy |
| Benign | 80.00 | 84.21 | 2.55 | 82.05 | 96.02 |
| Malignant | 96.67 | 98.86 | 3.41 | 97.75 | 97.73 |
| Normal | 96.97 | 92.75 | 1.87 | 94.81 | 96.02 |
| **Average** | **91.21** | **91.94** | **2.61** | **91.54** | **96.59** |

**Table 4.** Performance of the VGG19 in Lung Cancer detection (IQ-OTH/NCCD dataset + GAN)

| Class | Performance Evaluation (%) | | | | |
|---|---|---|---|---|---|
| | Precision | TPR | FPR | F1-score | Accuracy |
| Benign | 100.00 | 99.99 | 0.000 | 99.99 | 99.99 |
| Malignant | 99.99 | 100.00 | 0.003 | 99.99 | 99.99 |
| Normal | 100.00 | 100.00 | 0.000 | 100.00 | 100.00 |
| **Average** | **99.99** | **99.99** | **0.001** | **99.99** | **99.99** |

From the comparison of these two tables, it can be seen that VGG19 has a better baseline performance than VGG16, even before augmentation. However, the application of GAN still provides a significant improvement, making the model almost perfect in detecting all three classes. This reinforces the important role of GAN-based data augmentation in correcting imbalanced class representations and improving the robustness and fairness of deep learning models in lung cancer diagnosis [52]. GANs not only increase the amount of data, but also broaden the spectrum of visual variation which is important for improving the generalisation of the model to new data [53].

### 3.3. Lung Cancer Detection on Lung Cancer CT Scan Dataset

As in the case of lung cancer detection using the IQ-OTH/NCCD dataset, this study also applies the VGG model to the Lung Cancer CT Scan dataset under two scenarios: one without GAN and another with GAN. The objective of these scenarios is to assess the impact of GAN in addressing class imbalance within the dataset and to compare the VGG model's performance before and after incorporating GAN. The training and evaluation process for the VGG model follows the same data partitioning strategy as used with the IQ-OTH/NCCD dataset, with 80% of the data allocated for training, 10% for validation, and 10% for testing. This approach ensures the model is trained on a substantial portion of data while maintaining reliability in testing with unseen samples. The Lung Cancer CT Scan dataset is categorized into four histopathological types: NL, AC, SC, and LC. The following section outlines the performance outcomes of the VGG model in detecting lung cancer using this dataset.

In Table 5, which represents the performance of the VGG16 model without GAN, it can be seen that although the Normal (NL) class gets very high evaluation results (Precision 100%, F1-score 98.31%, Accuracy 99.38%), the performance for cancer classes is still suboptimal. For example, in the Large Cell Carcinoma (LC) class, the precision value only reached 67.65%, and the FPR reached 8.09%, indicating a high misclassification to other classes. Even though the TPR was quite high (95.83%), the F1-score value was still below 80%. This reflects the imbalance in the amount of data between classes, which causes the model to be biased towards the more dominant class in the dataset. The overall average of the metrics also confirms this finding, with a precision of 89.93%, an F1-score of 90.30%, and an accuracy of 95.31%, which although high enough, does not yet reflect the fairness of classification across classes.

Meanwhile, Table 6 shows a very significant performance improvement after GAN was applied to the VGG16 model. All classes, including the previously low-performing ones such as LC, now show almost perfect precision, TPR, F1-score, and accuracy values (all ≥ 99.9%). The FPR value also decreased dramatically, averaging only 0.01%, indicating that the model very rarely made the mistake of classifying samples into the wrong class. The overall average of the metrics is also close to perfection, with an average precision and accuracy of 99.98%, proving that data augmentation with GANs not only augments the data, but also improves the representation of visual feature distribution

across classes [54]. This improvement highlights the effectiveness of GAN in addressing class imbalance, especially in medical datasets where minority classes are clinically significant but underrepresented [55]. The balanced and diverse synthetic images generated by GAN allow the VGG16 model to learn class-specific features more comprehensively [56]. As a result, the model exhibits greater generalization capability and reduced bias towards dominant classes [57]. These findings validate the strategic use of GAN as a powerful tool to enhance the reliability and fairness of deep learning models in healthcare applications [58].

In Table 7 (without GAN), the performance of VGG19 is already high, with an average accuracy of 95.94% and F1-score of 92.05%. However, there are differences in performance between classes. While the Normal (NL) and SC classes performed well, the LC class recorded the lowest precision (77.42%) and highest FPR (5.15%), indicating that the model struggled to distinguish this cancer type from other classes. The AC class also had a high FPR (4.90%) with a TPR of only 91.38%. This shows that VGG19, although architecturally robust, is still affected by data imbalance and limited visual variation in minority classes. After applying GAN (Table 8), the performance of VGG19 improved dramatically and evenly. All metrics are close to perfect, with an average precision, TPR, F1-score, and accuracy of 99.99%, and a very small FPR (0.001%). Even the LC class, which previously had low performance, is now on par with the other classes, showing that GAN successfully overcomes the bias towards minority classes and improves the generalisability of the model.

**Table 5.** Performance of the VGG16 in Lung Cancer detection (Lung Cancer CT-Scan without GAN)

| Class | Performance Evaluation (%) | | | | |
|---|---|---|---|---|---|
| | Precision | TPR | FPR | F1-score | Accuracy |
| NL | 100.00 | 96.67 | 0.00 | 98.31 | 99.38 |
| AC | 94.34 | 86.21 | 2.94 | 90.09 | 93.13 |
| SC | 97.73 | 89.58 | 0.89 | 93.48 | 96.25 |
| LC | 67.65 | 95.83 | 8.09 | 79.31 | 92.50 |
| **Average** | **89.93** | **92.07** | **2.98** | **90.30** | **95.31** |

**Table 6.** Performance of the VGG16 in Lung Cancer detection (Lung Cancer CT-Scan + GAN)

| Class | Performance Evaluation (%) | | | | |
|---|---|---|---|---|---|
| | Precision | TPR | FPR | F1-score | Accuracy |
| NL | 100.00 | 99.99 | 0.00 | 100.00 | 100.00 |
| AC | 99.89 | 100.00 | 0.04 | 99.94 | 99.97 |
| SC | 100.00 | 99.89 | 0.00 | 99.94 | 99.97 |
| LC | 99.99 | 100.00 | 0.00 | 100.00 | 100.00 |
| **Average** | **99.98** | **99.98** | **0.010** | **99.98** | **99.98** |

**Table 7.** Performance of the VGG19 in Lung Cancer detection (Lung Cancer CT-Scan without GAN)

| Class | Performance Evaluation (%) | | | | |
|---|---|---|---|---|---|
| | Precision | TPR | FPR | F1-score | Accuracy |
| NL | 100.00 | 96.67 | 0.00 | 98.31 | 99.38 |
| AC | 91.38 | 91.38 | 4.90 | 91.38 | 93.75 |
| SC | 97.62 | 85.42 | 0.89 | 91.11 | 95.00 |
| LC | 77.42 | 100.00 | 5.15 | 87.27 | 95.63 |
| Average | 91.60 | 93.37 | 2.74 | 92.02 | 95.94 |

When compared with Table 5 and Table 6 using VGG16, there are some important findings. Before applying GAN, VGG19 (Table 7) performed better than VGG16 (Table 5), with higher average accuracy (95.94% vs. 95.31%), and more stable F1-score across all classes. This shows that the deeper architecture of VGG19 is able to capture more complex features in CT scan images. However, after applying GAN, both VGG16 (Table 6) and VGG19 (Table 8) showed very high and almost identical performance, with evaluation metrics close to 100%. This proves that data augmentation plays a much more important role than architectural complexity in overcoming the data imbalance problem [59]. GAN is able to improve the performance of simple models like VGG16 to match and even surpass deeper models like VGG19 in the context of balanced data [60].

**Table 8.** Performance of the VGG19 in Lung Cancer detection (Lung Cancer CT-Scan + GAN)

| Class | Performance Evaluation (%) | | | | |
|---|---|---|---|---|---|
| | Precision | TPR | FPR | F1-score | Accuracy |
| NL | 100.00 | 100.00 | 0.000 | 100.00 | 100.00 |
| AC | 99.99 | 100.00 | 0.002 | 99.99 | 99.99 |
| SC | 99.99 | 100.00 | 0.002 | 99.99 | 99.99 |
| LC | 100.00 | 99.99 | 0.000 | 99.99 | 99.99 |
| **Average** | **99.99** | **99.99** | **0.001** | **99.99** | **99.99** |

Overall, these results confirm that in the lung cancer detection task, model performance is highly dependent on the quality and balance of the training data [61]. Complex models such as VGG19 excel under initial conditions, but GAN augmentation is key in achieving optimal classification performance, regardless of the architecture used [62].

### 3.4. The Proposed VGG + GAN Approach Compared with Existing Methods

The results of VGG model testing on both the IQ-OTH/NCCD and Lung Cancer CT Scan datasets indicate that incorporating GANs for data augmentation effectively boosts classification performance, although the extent of improvement varies across different classes. Combining the VGG architecture with GAN-generated data notably strengthens the model's capability to identify various lung cancer types, especially in detecting underrepresented classes that were previously difficult to classify. The use of GANs plays a vital role in balancing and enriching the dataset, leading to a marked enhancement in the VGG model's classification accuracy. The next section provides a comparative analysis between this proposed method and other existing approaches in the field.

Table 9 shows the performance comparison between the proposed model (GAN+VGG) and various other methods that have been developed in CT image-based lung cancer detection. This table summarizes information from various previous studies, including the type of dataset used, model architecture, and accuracy values achieved. The compared models come from various approaches, ranging from conventional architectures such as CNN, to complex model combinations such as 3D-CNN + RNN, YOLOv8, Squeeze-Inception-ResNeXt, and Mask R-CNN. Most of the methods listed achieved high accuracy, ranging from 91.56% to 98.39%. For example, the MinClassNet approach on the LC25000 dataset by Talib et al. (2024) achieved 98.39% accuracy [65], while CNN + YOLOv8 by Elhassan et al. (2024) on the IQ-OTH/NCCD dataset achieved 97.67% accuracy [66]. Although this performance is considered excellent, there is still room for improvement, especially in recognizing minority classes and maintaining classification consistency across diverse datasets.

**Table 9.** Comparison of the performance of the proposed GAN + VGG with existing methods

| Research | Dataset | Models | Accuracy |
|---|---|---|---|
| Pandian et al. (2022) [5] | CT Images | CNN and Google Net | 98% |
| Wankhade et al. (2023) [63] | LUNA16 | 3D-CNN + RNN | 95% |
| Yan et al. (2023) [64] | IQ-OTH/NCCD | CNN | 96.58% |
| Talib et al. (2024) [65] | LC25000 | MinClassNet | 98.39% |
| Elhassan et al. (2024) [66] | IQ-OTH/NCCD | CNN + YOLOv8 | 97.67% |
| Basha et al. (2025) [67] | LIDC-IDRI + IQ-OTH/NCCD | PCA-F-SHCNNet | 91.56% |
| Lakshmi et al. (2025) [68] | CT Images | Squeeze-Inception-ResNeXt | 97.7% |
| Lavanya et al. (2025) [69] | LIDC | Locust Assisted CS based CNN | 96.33% |
| Akintola et al. (2025) [70] | IQ-OTH/NCCD | CNN + Mask R-CNN | 95.6% |
| Ansari et al. (2025) [71] | LIDC-IDRI | SVMVGGNet-16 | 97.22% |
| Proposed models | IQ-OTH/NCCD + Lung Cancer CT-Scan | GAN + VGG | 99.99% |

Moreover, the proposed GAN+VGG model shows outstanding results, with accuracy reaching 99.99% when tested on two large datasets combined: IQ-OTH/NCCD and Lung Cancer CT-Scan. This performance is consistently higher than all the methods in the table. This reflects the successful integration of GAN as an augmentation method to balance data distribution between classes and enrich

visual features, as well as the reliability of the VGG architecture in extracting spatial patterns from medical images in depth [72]-[74]. In other words, the combination of GAN and VGG effectively overcomes the limitations of unbalanced medical datasets, and improves the sensitivity and specificity of the model to all classes.

## 4. Conclusion

The results of tests conducted on the VGG16 and VGG19 models in detecting lung cancer using two different datasets, namely IQ-OTH/NCCD and Lung Cancer CT Scan Dataset, show that the application of GAN has a very significant impact on improving model performance. Prior to data augmentation, both VGG16 and VGG19 performed quite well in general, but not consistently across classes. This was especially evident in minority classes such as Benign in the IQ-OTH/NCCD dataset and Large Cell Carcinoma (LC) in the Lung Cancer CT Scan dataset, where precision, TPR, and F1-score were much lower than the majority classes. The model tends to be biased towards classes that have more data, and the misclassification error (FPR) in the minority class is quite high, even reaching more than 8% in some cases. However, after data augmentation using GAN, all performance metrics improved drastically and evenly across all classes. The average precision, TPR, F1-score, and accuracy increased to close to or reach 100%, while the FPR decreased drastically to an average of only about 0.001%. This indicates that GAN contributes not only to balancing the number of samples across classes, but also to enhancing the diversity of visual features, which is essential for boosting the model's sensitivity and specificity in detecting cancer-related image patterns. Notably, following the augmentation process, the performance of the VGG16 and VGG19 models becomes nearly identical, implying that the quality of training data may influence model performance more significantly than the architectural complexity of the network.

The consistent performance improvement on two datasets with different label characteristics, IQ-OTH/NCCD with clinical classification (Benign, Malignant, Normal) and Lung Cancer CT-Scan with histopathological classification (AC, SC, LC, Normal), shows that this approach is robust and widely applicable. Thus, the application of GAN proved effective in overcoming data imbalance and improving the generalisation ability of the VGG model in lung cancer classification. These findings confirm that the combination of GAN and CNN architectures such as VGG is a very promising strategy for the development of artificial intelligence-based lung cancer diagnosis support systems, especially under conditions of limited and imbalanced medical data. For future research, it is recommended to explore the use of more advanced GAN variants, such as StyleGAN or CycleGAN, to further improve the realism of synthetic images. Additionally, evaluating this approach on larger and more diverse datasets, including multi-center or real-world clinical data, would enhance its practical relevance. Investigating its performance in multimodal frameworks, such as combining imaging with clinical or genomic data, could also extend its applicability.

**Author Contribution:** Y. Pamungkas: Conceptualization, funding acquisition, formal analysis, investigation, project administration, methodology, resources, supervision, software, and writing - review & editing. D. Kuswanto: Data curation, resources, investigation, software, visualization, validation, and writing - original draft. A. Syaifudin: Data curation, resources, investigation, software, visualization, validation, and writing - original draft. E. Triandini: Investigation, visualization, and writing - review & editing. D. P. Hapsari: Investigation, project administration, and writing - review & editing. K. Nakkliang: Investigation, visualization, and writing - review & editing. M. N. A. Uda: Investigation, project administration, and writing - review & editing. U. Hashim: Investigation, project administration, and writing - review & editing.

**Funding:** This research was provided by the Ministry of Higher Education, Science and Technology through the Indonesia Endowment Funds for Education (LPDP) in the Enhancing Quality Education for International University Recognition (EQUITY) program, hereafter referred to as the Higher Education Endowment Fund Program (DAPT), under the scheme of ITS' Inbound Researcher Mobility 2025 and the Universiti Malaysia Sabah (UMS) for the collaboration.

**Acknowledgment:** The authors would like to acknowledge the Department of Medical Technology, Institut Teknologi Sepuluh Nopember, for the facilities and support in this research. The authors also gratefully

acknowledge financial support from the Institut Teknologi Sepuluh Nopember for this work, under project scheme of the Publication Writing and IPR Incentive Program (PPHKI) 2025.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

[1] E. Piekarz-Porter and S. J. Kim, "State law at the intersection of lung cancer screening guidelines and social determinants of health," *Journal of Cancer Policy*, vol. 43, p. 100561, 2025, https://doi.org/10.1016/j.jcpo.2025.100561.

[2] G. Vicidomini, "Current Challenges and Future Advances in Lung Cancer: Genetics, Instrumental Diagnosis and Treatment," *Cancers*, vol. 15, no. 14, p. 3710, 2023, https://doi.org/10.3390/cancers15143710.

[3] N. Kenaan *et al*., "Advances in early detection of non-small cell lung cancer: A comprehensive review," *Cancer Medicine*, vol. 13, no. 18, p. e70156, 2024, https://doi.org/10.1002/cam4.70156.

[4] A. H. Nielsen and U. Fredberg, "Earlier diagnosis of lung cancer," *Cancer Treatment and Research Communications*, vol. 31, p. 100561, 2022, https://doi.org/10.1016/j.ctarc.2022.100561.

[5] R. Pandian, V. Vedanarayanan, D. N. S. R. Kumar, and R. Rajakumar, "Detection and classification of lung cancer using CNN and Google net," *Measurement: Sensors*, vol. 24, p. 100588, 2022, https://doi.org/10.1016/j.measen.2022.100588.

[6] A. R. W. Sait, "Lung Cancer Detection Model Using Deep Learning Technique," *Applied sciences*, vol. 13, no. 22, p. 12510, 2023, https://doi.org/10.3390/app132212510.

[7] R. Klangbunrueang, P. Pookduang, W. Chansanam, and T. Lunrasri, "AI-Powered Lung Cancer Detection: Assessing VGG16 and CNN Architectures for CT Scan Image Classification," *Informatics*, vol. 12, no. 1, p. 18, 2025, https://doi.org/10.3390/informatics12010018.

[8] T. I. A. Mohamed, O. N. Oyelade, and A. E. Ezugwu, "Automatic detection and classification of lung cancer CT scans based on deep learning and ebola optimization search algorithm," *PLOS ONE*, vol. 18, no. 8, p. e0285796, 2023, https://doi.org/10.1371/journal.pone.0285796.

[9] A. A. Shah, H. A. M. Malik, A. Muhammad, A. Alourani, and Z. A. Butt, "Deep learning ensemble 2D CNN approach towards the detection of lung cancer," *Scientific Reports*, vol. 13, no. 1, p. 2987, 2023, https://doi.org/10.1038/s41598-023-29656-z.

[10] M. Thangamani, M. S. Koti, B. A. Nagashree, V. Geetha, K. P. Shreyas, S. K. Mathivanan, and G. T. Dalu, "Lung cancer diagnosis based on weighted convolutional neural network using gene data expression," *Scientific Reports*, vol. 14, no. 1, p. 3656, 2024, https://doi.org/10.1038/s41598-024-54124-7.

[11] H. Alyasriy and M. Al-Huseiny, "The IQ-OTH/NCCD lung cancer dataset," *Mendeley Data*, 2023, https://doi.org/10.17632/bhmdr45bh2.4.

[12] Kaggle, "Chest CT-Scan images Dataset," *Kaggle*, 2020, https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images.

[13] V. Shariff, C. Paritala, and K. M. Ankala, "Optimizing non small cell lung cancer detection with convolutional neural networks and differential augmentation," *Scientific Reports*, vol. 15, no. 1, p. 15640, 2025, https://doi.org/10.1038/s41598-025-98731-4.

[14] M. Ibrahim *et al*., "Generative AI for synthetic data across multiple medical modalities: A systematic review of recent developments and challenges," *Computers in Biology and Medicine*, vol. 189, p. 109834, 2025, https://doi.org/10.1016/j.compbiomed.2025.109834.

[15] I. Kajo, M. Kas, A. Chahi, and Y. Ruichek, "Learning by competing: Competitive multi-generator based adversarial learning," *Applied Soft Computing*, vol. 146, p. 110698, 2023, https://doi.org/10.1016/j.asoc.2023.110698.

[16] A. Luschi, L. Tognetti, A. Cartocci, G. Cevenini, P. Rubegni, and E. Iadanza, "Advancing synthetic data for dermatology: GAN comparison with multi-metric and expert validation approach," *Health and Technology*, vol. 15, pp. 553-562, 2025, https://doi.org/10.1007/s12553-025-00971-x.

[17] R. Sauber-Cole and T. M. Khoshgoftaar, "The use of generative adversarial networks to alleviate class imbalance in tabular data: a survey," *Journal of Big Data*, vol. 9, no. 1, p. 98, 2022, https://doi.org/10.1186/s40537-022-00648-6.

[18] L. J. Crasta, R. Neema, and A. R. Pais, "A novel Deep Learning architecture for lung cancer detection and diagnosis from Computed Tomography image analysis," *Healthcare analytics*, vol. 5, p. 100316, 2024, https://doi.org/10.1016/j.health.2024.100316.

[19] P. Purwono, E. Wulandari, A. Ma'arif, and W. A. Salah, "Understanding Generative Adversarial Networks (GANs): A Review," *Control Systems and Optimization Letters*, vol. 3, no. 1, pp. 36-45, 2025, https://doi.org/10.59247/csol.v3i1.170.

[20] W. Serrano, "The Deep Learning Generative Adversarial Random Neural Network in data marketplaces: The digital creative," *Neural Networks*, vol. 165, pp. 420-434, 2023, https://doi.org/10.1016/j.neunet.2023.05.028.

[21] M. Goyal and Q. H. Mahmoud, "A Systematic Review of Synthetic Data Generation Techniques Using Generative AI," *Electronics*, vol. 13, no. 17, p. 3509, 2024, https://doi.org/10.3390/electronics13173509.

[22] A. Dash, J. Ye and G. Wang, "A Review of Generative Adversarial Networks (GANs) and Its Applications in a Wide Variety of Disciplines: From Medical to Remote Sensing," *IEEE Access*, vol. 12, pp. 18330-18357, 2024, https://doi.org/10.1109/ACCESS.2023.3346273.

[23] J. Jheelan and S. Pudaruth, "Using Deep Learning to Identify Deepfakes Created Using Generative Adversarial Networks," *Computers*, vol. 14, no. 2, p. 60, 2025, https://doi.org/10.3390/computers14020060.

[24] L. Alhoraibi, D. Alghazzawi, and R. Alhebshi, "Generative Adversarial Network-Based Data Augmentation for Enhancing Wireless Physical Layer Authentication," *Sensors*, vol. 24, no. 2, p. 641, 2024, https://doi.org/10.3390/s24020641.

[25] J. Wei, Q. Wang, and Z. Zhao, "Generative adversarial network based on Poincaré distance similarity constraint: Focusing on overfitting problem caused by finite training data," *Applied Soft Computing*, vol. 151, p. 111147, 2024, https://doi.org/10.1016/j.asoc.2023.111147.

[26] S.-L. Jeng, "Generative Adversarial Network for Synthesizing Multivariate Time-Series Data in Electric Vehicle Driving Scenarios," *Sensors*, vol. 25, no. 3, p. 749, 2025, https://doi.org/10.3390/s25030749.

[27] S. S. Sengar, A. B. Hasan, S. Kumar, and F. Carroll, "Generative artificial intelligence: a systematic review and applications," *Multimedia Tools and Applications*, 2024, https://doi.org/10.1007/s11042-024-20016-1.

[28] H. Ding, N. Huang, Y. Wu, and X. Cui, "Improving imbalanced medical image classification through GAN-based data augmentation methods," *Pattern Recognition*, vol. 166, p. 111680, 2025, https://doi.org/10.1016/j.patcog.2025.111680.

[29] H. Li, Z. Wang, X. Yue, W. Wang, H. Tomiyama, and L. Meng, "An architecture-level analysis on deep learning models for low-impact computations," *Artificial Intelligence Review*, vol. 56, no. 3, pp. 1971-2010, 2022, https://doi.org/10.1007/s10462-022-10221-5.

[30] S. U. Krishna, A.N Barath Lakshman, T. Archana, K. Raja, and M. Ayyadurai, "Lung Cancer Prediction and Classification Using Decision Tree and VGG16 Convolutional Neural Networks," *The Open Biomedical Engineering Journal*, vol. 18, no. 1, 2024, https://doi.org/10.2174/0118741207290271240322061032.

[31] F. Behrad and M. S. Abadeh, "An overview of deep learning methods for multimodal medical data mining," *Expert Systems with Applications*, vol. 200, p. 117006, 2022, https://doi.org/10.1016/j.eswa.2022.117006.

[32] W. Bakasa and S. Viriri, "VGG16 Feature Extractor with Extreme Gradient Boost Classifier for Pancreas Cancer Prediction," *Journal of Imaging*, vol. 9, no. 7, p. 138, 2023, https://doi.org/10.3390/jimaging9070138.

[33] H. M. Ahmed and M. Y. Kashmola, "Performance Improvement of Convolutional Neural Network Architectures for Skin Disease Detection," *International journal of computing and digital system/International Journal of Computing and Digital Systems*, vol. 13, no. 1, pp. 657-669, 2023, https://doi.org/10.12785/ijcds/130152.

[34] X.-X. Yin, L. Sun, Y. Fu, R. Lu, and Y. Zhang, "U-Net-Based Medical Image Segmentation," *Journal of Healthcare Engineering*, vol. 2022, p. 4189781, 2022, https://doi.org/10.1155/2022/4189781.

[35] A. A. Adegun, S. Viriri, and J.-R. Tapamo, "Review of deep learning methods for remote sensing satellite images classification: experimental survey and comparative analysis," *Journal of Big Data*, vol. 10, no. 1, 2023, https://doi.org/10.1186/s40537-023-00772-x.

[36] X. Jiang, Z. Hu, S. Wang, and Y. Zhang, "Deep Learning for Medical Image-Based Cancer Diagnosis," *Cancers*, vol. 15, no. 14, p. 3608, 2023, https://doi.org/10.3390/cancers15143608.

[37] Y. Xu, M. Xia, K. Hu, S. Zhou, and L. Weng, "Style Transfer Review: Traditional Machine Learning to Deep Learning," *Information*, vol. 16, no. 2, p. 157, 2025, https://doi.org/10.3390/info16020157.

[38] B. Qawasmeh, J.-S. Oh, and Valerian Kwigizile, "Comparative Analysis of AlexNet, ResNet-50, and VGG-19 Performance for Automated Feature Recognition in Pedestrian Crash Diagrams," *Applied Sciences*, vol. 15, no. 6, p. 2928, 2025, https://doi.org/10.3390/app15062928.

[39] H. Xu, Y. Yu, J. Chang, X. Hu, Z. Tian, and O. Li, "Precision lung cancer screening from CT scans using a VGG16-based convolutional neural network," *Frontiers in Oncology*, vol. 14, 2024, https://doi.org/10.3389/fonc.2024.1424546.

[40] P. K. Mall *et al*., "A comprehensive review of deep neural networks for medical image processing: Recent developments and future opportunities," *Healthcare Analytics*, vol. 4, p. 100216, 2023, https://doi.org/10.1016/j.health.2023.100216.

[41] J. Li *et al*., "Quantum k-fold cross-validation for nearest neighbor classification algorithm," *Physica A: Statistical Mechanics and its Applications*, vol. 611, p. 128435, 2023, https://doi.org/10.1016/j.physa.2022.128435.

[42] V. H. Kamble and M. P. Dale, "Machine learning approach for longitudinal face recognition of children," *Machine Learning for Biometrics*, pp. 1-27, 2022, https://doi.org/10.1016/b978-0-323-85209-8.00011-0.

[43] F. Maleki, N. Muthukrishnan, K. Ovens, C. Reinhold, and R. Forghani, "Machine Learning Algorithm Validation," *Neuroimaging Clinics of North America*, vol. 30, no. 4, pp. 433-445, 2020, https://doi.org/10.1016/j.nic.2020.08.004.

[44] M. Khalsan, M. Mu, A.-S. E. Salih, L. Machado, S. Ajit, and M. O. Agyeman, "Fuzzy Gene Selection and Cancer Classification Based on Deep Learning Model," *arXiv*, 2023, https://doi.org/10.48550/arxiv.2305.04883.

[45] J. White and S. D. Power, "k-Fold Cross-Validation Can Significantly Over-Estimate True Classification Accuracy in Common EEG-Based Passive BCI Experimental Designs: An Empirical Investigation," *Sensors*, vol. 23, no. 13, p. 6077, 2023, https://doi.org/10.3390/s23136077.

[46] T. Pahikkala, H. Suominen, and J. Boberg, "Efficient cross-validation for kernelized least-squares regression with sparse basis expansions," *Machine Learning*, vol. 87, no. 3, pp. 381-407, 2012, https://doi.org/10.1007/s10994-012-5287-6.

[47] G. Nicora, M. Rios, A. Abu-Hanna, and R. Bellazzi, "Evaluating pointwise reliability of machine learning prediction," *Journal of Biomedical Informatics*, vol. 127, p. 103996, 2022, https://doi.org/10.1016/j.jbi.2022.103996.

[48] M. O.-Adjei, J. B. H.-Acquah, F. Twum, and G. A.-Salaam, "Imbalanced class distribution and performance evaluation metrics: A systematic review of prediction accuracy for determining model performance in healthcare systems," *PLOS digital health*, vol. 2, no. 11, p. e0000290, 2023, https://doi.org/10.1371/journal.pdig.0000290.

[49] I. N. M. Adiputra, P.-C. Lin, and P. Wanchai, "The Effectiveness of Generative Adversarial Network-Based Oversampling Methods for Imbalanced Multi-Class Credit Score Classification," *Electronics*, vol. 14, no. 4, p. 697, 2025, https://doi.org/10.3390/electronics14040697.

[50] J. Hussain, M. Båth, and J. Ivarsson, "Generative adversarial networks in medical image reconstruction: A systematic literature review," *Computers in Biology and Medicine*, vol. 191, p. 110094, 2025, https://doi.org/10.1016/j.compbiomed.2025.110094.

[51] C. Xu, J. Wu, F. Zhang, J. Freer, Z. Zhang, and Y. Cheng, "A deep image classification model based on prior feature knowledge embedding and application in medical diagnosis," *Scientific Reports*, vol. 14, no. 1, p. 13244, 2024, https://doi.org/10.1038/s41598-024-63818-x.

[52] H. Javed, S. El-Sappagh, and T. Abuhmed, "Robustness in deep learning models for medical diagnostics: security and adversarial challenges towards robust AI applications," *Artificial Intelligence Review*, vol. 58, no. 1, p. 12, 2024, https://doi.org/10.1007/s10462-024-11005-9.

[53] G. Iglesias, E. Talavera, and A. Díaz-Álvarez, "A survey on GANs for computer vision: Recent research, analysis and taxonomy," *Computer Science Review*, vol. 48, p. 100553, 2023, https://doi.org/10.1016/j.cosrev.2023.100553.

[54] Y. J.-Gaona, D. C.-Figueroa, V. Lakshminarayanan, and M. J. R.-Álvarez, "Gan-based data augmentation to improve breast ultrasound and mammography mass classification," *Biomedical Signal Processing and Control*, vol. 94, p. 106255, 2024, https://doi.org/10.1016/j.bspc.2024.106255.

[55] M. Salmi, D. Atif, D. Oliva, A. Abraham, and S. Ventura, "Handling imbalanced medical datasets: review of a decade of research," *Artificial Intelligence Review*, vol. 57, no. 10, p. 273, 2024, https://doi.org/10.1007/s10462-024-10884-2.

[56] Q. Su, H. N. A. Hamed, M. A. Isa, X. Hao and X. Dai, "A GAN-Based Data Augmentation Method for Imbalanced Multi-Class Skin Lesion Classification," *IEEE Access*, vol. 12, pp. 16498-16513, 2024, https://doi.org/10.1109/access.2024.3360215.

[57] M. N. Razali, N. Arbaiy, P.-C. Lin, and S. Ismail, "Optimizing Multiclass Classification Using Convolutional Neural Networks with Class Weights and Early Stopping for Imbalanced Datasets," *Electronics*, vol. 14, no. 4, p. 705, 2025, https://doi.org/10.3390/electronics14040705.

[58] Y. Chen and P. Esmaeilzadeh, "Generative AI in Medical Practice: In-Depth Exploration of Privacy and Security Challenges," *Journal of Medical Internet Research*, vol. 26, p. e53008, 2024, https://doi.org/10.2196/53008.

[59] A. Mumuni and F. Mumuni, "Data augmentation: A comprehensive survey of modern approaches," *Array*, vol. 16, p. 100258, 2022, https://doi.org/10.1016/j.array.2022.100258.

[60] M. Hussain, T. Thaher, M. B. Almourad, and M. Mafarja, "Optimizing VGG16 deep learning model with enhanced hunger games search for logo classification," *Scientific Reports*, vol. 14, no. 1, p. 31759, 2024, https://doi.org/10.1038/s41598-024-82022-5.

[61] R. Durgam, B. Panduri, V. Balaji, A. O. Khadidos, and S. Selvarajan, "Enhancing lung cancer detection through integrated deep learning and transformer models," *Scientific Reports*, vol. 15, no. 1, p. 15614, 2025, https://doi.org/10.1038/s41598-025-00516-2.

[62] L. C. Ribas, W. Casaca, and R. T. Fares, "Conditional Generative Adversarial Networks and Deep Learning Data Augmentation: A Multi-Perspective Data-Driven Survey Across Multiple Application Fields and Classification Architectures," *AI*, vol. 6, no. 2, p. 32, 2025, https://doi.org/10.3390/ai6020032.

[63] S. Wankhade and S. Vigneshwari, "A novel hybrid deep learning method for early detection of lung cancer using neural networks," *Healthcare Analytics*, vol. 3, p. 100195, 2023, https://doi.org/10.1016/j.health.2023.100195.

[64] C. Yan and N. Razmjooy, "Optimal lung cancer detection based on CNN optimized and improved Snake optimization algorithm," *Biomedical Signal Processing and Control*, vol. 86, p. 105319, 2023, https://doi.org/10.1016/j.bspc.2023.105319.

[65] L. F. Talib, J. Amin, M. Sharif, and M. Raza, "Transformer-based semantic segmentation and CNN network for detection of histopathological lung cancer," *Biomedical Signal Processing and Control*, vol. 92, p. 106106, 2024, https://doi.org/10.1016/j.bspc.2024.106106.

[66] S. M. Elhassan, S. M. Darwish, and S. M. Elkaffas, "An Enhanced Lung Cancer Detection Approach Using Dual-Model Deep Learning Technique," *CMES - Computer Modeling in Engineering and Sciences*, vol. 142, no. 1, pp. 835-867, 2024, https://doi.org/10.32604/cmes.2024.058770.

[67] S. A. H. Basha, P. R. Kshirsagar, P. S. Rao, T. K. Tak, and B. Sivaneasan, "PCA-F-SHCNNet: Principal Component Analysis-Fused-Shepard Convolutional Neural Networks for lung cancer detection and severity level classification," *Biomedical Signal Processing and Control*, vol. 107, p. 107843, 2025, https://doi.org/10.1016/j.bspc.2025.107843.

[68] G. L. G and P. Nagaraj, "Lung cancer detection and classification using Optimized CNN features and Squeeze-Inception-ResNeXt model," *Computational Biology and Chemistry*, vol. 117, p. 108437, 2025, https://doi.org/10.1016/j.compbiolchem.2025.108437.

[69] P. Lavanya and K. Vidhya, "A novel lung cancer detection adopting Radiomic feature extraction with Locust assisted CS based CNN classifier," *Biomedical Signal Processing and Control*, vol. 100, p. 107139, 2025, https://doi.org/10.1016/j.bspc.2024.107139.

[70] A. G. Akintola, K. Y. Obiwusi, Y. O. Olatunde, M. Usman, F. G. Aberuagba, M. A. Adebisi , and S. A. Adebayo, "Integrated Deep Learning Paradigm for Comprehensive Lung Cancer Segmentation and Classification Using Mask R-CNN and CNN Models," *Franklin Open*, p. 100278, 2025, https://doi.org/10.1016/j.fraope.2025.100278.

[71] M. M. Ansari *et al*., "SVMVGGNet-16: A Novel Machine and Deep Learning Based Approaches for Lung Cancer Detection Using Combined SVM and VGGNet-16," *Current Medical Imaging Formerly Current Medical Imaging Reviews*, vol. 21, no. 1, p. e15734056348824, 2025, https://doi.org/10.2174/0115734056348824241224100809.

[72] H. M. Rai, J. Yoo, S. Agarwal, and N. Agarwal, "LightweightUNet: Multimodal Deep Learning with GAN-Augmented Imaging Data for Efficient Breast Cancer Detection," *Bioengineering*, vol. 12, no. 1, p. 73, 2025, https://doi.org/10.3390/bioengineering12010073.

[73] S. Oraby, A. Emran, B. El-Saghir, and S. Mohsen, "Hybrid of DSR-GAN and CNN for Alzheimer disease detection based on MRI images," *Scientific Reports*, vol. 15, no. 1, p. 12727, 2025, https://doi.org/10.1038/s41598-025-94677-9.

[74] A. Dash and T. Swarnkar, "Data-GAN augmentation techniques in medical image analysis: A deep survey," *SN Computer Science*, vol. 6, no. 4, pp. 1-21, 2025, https://doi.org/10.1007/s42979-025-03867-9.