

An Attention-Guided GAN–DeiT Framework for Histopathological Lung Cancer Detection and Classification



Vishnu Vardhan Raju D.*^{ORCID}, Reddy Madhavi K.^{ORCID}

School of Computing, Mohan Babu University, Tirupati 517102, India

Corresponding Author Email: mailinvishnu@gmail.com

Copyright: ©2026 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.310106>

ABSTRACT

Received: 5 September 2025

Revised: 10 November 2025

Accepted: 18 January 2026

Available online: 31 January 2026

Keywords:

histopathological image analysis, lung cancer detection, generative adversarial networks, vision transformers, attention mechanisms, medical image classification, computer-aided diagnosis

Patients diagnosed with lung cancer are more likely to experience a poor prognosis and reduced survival due to metastasis. Deep learning can facilitate the non-invasive estimation of cancer tissue likelihood, providing doctors with important information that improves diagnostic precision and ultimately enhances patient survival and prognosis in cancer tissue detection and classification. In this study, a novel AGGAN–DeiT framework is proposed that combines an Attention-Guided Generative Adversarial Network (AGGAN) with a Data-efficient Image Transformer (DeiT) to improve the detection and classification of cancer tissue in lung cancer patients. The AGGAN model synthesizes high-resolution and realistic medical images that highlight key pathological traits. This process supplements training data and addresses class imbalance commonly observed in medical imaging datasets. The status of cancer tissue in histological images is then classified using the DeiT model, which is well known for capturing long-range dependencies and spatial hierarchies. Thus, the goal is to improve classification performance. Experimental results show that the proposed approach, integrating AGGAN and DeiT, achieves state-of-the-art performance in classifying cancer tissue compared with a standard Convolutional Neural Network (CNN). Regarding histopathological cancer tissue detection and classification, the proposed framework provides promising support for pathologists and cancer researchers in evaluating detection performance.

1. INTRODUCTION

Lung cancer remains one of the primary causes of morbidity and mortality because it is highly aggressive and behaves unpredictably in response to available treatments. The presence of mediastinal lymph nodes (MLNs) significantly influences the stage of the disease and the clinical outcome, as MLNs are considered the primary pathways for metastatic spread [1]. Therefore, evaluation of MLNs before surgery is important for appropriate treatment planning.

Various methods, such as ultrasound-guided fine-needle aspiration biopsy, mediastinoscopy, computed tomography (CT) scan, magnetic imaging reasoning (MRI) scan, and PET-CT scan, are collectively employed by physicians to assess metastasis in cancer. The conventional CT scan has low sensitivity (approximately 12.8%) but high specificity (approximately 99.6%) compared with contrast-enhanced CT scans [2]. The MRI scan has the potential to identify tumor areas based on morphology. It can also contribute to the overall assessment of the cancer condition through PET-CT imaging. Unfortunately, positron emission tomography (PET) scans may produce false positives due to tracer uptake in inflammatory and benign tissues. Mediastinoscopy and ultrasound-guided biopsies are invasive procedures and may lead to complications such as hemorrhage, nerve injury, and pneumothorax. These procedures are also technically

challenging. These limitations highlight the importance of a reliable non-invasive technique with high sensitivity for detecting cancer tissues.

Conventional diagnostic methods have difficulty accurately predicting malignancies in a systematic manner. Recently, the emergence of artificial intelligence, particularly deep learning, has provided promising solutions to address these challenges. Deep learning algorithms learn in an end-to-end manner and often outperform hand-crafted feature approaches in diagnostic accuracy. These algorithms have shown significant success in various medical image processing tasks, including diagnosis, detection, segmentation, and registration, as well as lung cancer detection, demonstrating their applicability [3].

However, deep learning algorithms rely on large annotated datasets [4]. Such datasets are often limited in medical research, creating a need for algorithms that can perform effectively with limited data. In this study, we propose an AGGAN–DeiT framework designed to detect and classify metastatic cancerous regions in lung cancer images. The Attention-Guided Generative Adversarial Network (AGGAN) refines cancerous region representations by emphasizing the regions most relevant to cancer, thereby capturing metastatic information more effectively. In addition, DeiT (Data-efficient Image Transformer) is designed to perform effectively even with limited data, making it suitable for medical image applications. By combining these two methods, the proposed

framework aims to enhance both the sensitivity and specificity of lung cancer metastasis detection while reducing dependence on large annotated datasets.

1.1 Motivation of this research

This work aims to enhance the detection and classification of cancerous regions related to lung cancer by combining AGGAN with DeiT. The proposed method leverages attention-based mechanisms to improve the reliability of metastasis detection in healthcare images. The integration of GANs and transformer-based models provides a new perspective for cancer-focused optimization in diagnostic approaches that face significant challenges in early-stage cancer intervention.

1.2 The main contribution of the research

- To improve histopathological cancer tissue detection and classification in lung cancer patients, a novel AGGAN–DeiT framework is proposed.
- The AGGAN synthesizes high-resolution and realistic medical images that emphasize important pathological features. This process efficiently augments the training data and addresses class imbalances commonly found in medical imaging datasets.
- Finally, the DeiT model, known for its strong ability to capture long-range relationships and spatial hierarchies, classifies cancer tissue states from histological images, thereby improving accuracy.

The remainder of this article is organized as follows: Section 2 offers a literature survey, Section 3 presents the proposed technique, Section 4 discusses the results and provides a discussion, and Section 5 concludes with the conclusion and suggestions for future study.

2. LITERATURE SURVEY

Research on the detection of cancer tissue in lung disease continually points out that early detection leads to much better outcomes. Recently, emphasis has been placed on the role of advances in imaging techniques, such as MRI and PET-CT, in providing more detailed information about cancer tissue. There is also an increasing trend of utilizing machine learning techniques on medical images to perform automated analysis and categorization of cancer tissue. Histologically, studies investigate tissue characteristics that make cancer distinct and enable targeted treatments. These advances demonstrate the importance of integrating newer technologies into the diagnosis and management of lung cancer metastases.

Ramesh et al. [5] introduced a multilayer CNN framework to identify different forms of lung cancer. Their approach demonstrated effective feature extraction for lung nodules with variations in size and shape using a multiscale layered architecture. The model was evaluated using the LC25000 dataset, comprising histological images of squamous cell carcinoma and adenocarcinoma. The results showed improved performance compared with traditional approaches, achieving a validation accuracy of 89% and a training accuracy of 64%.

In another attempt to develop a lung cancer prediction model, Mamun et al. [6] employed several ensemble learning methods, including XGBoost, LightGBM, Bagging, and

AdaBoost. These methods were tested on a Kaggle dataset containing 309 records, where factors such as age, smoking habits, and symptoms including chest pain, allergies, and fatigue were considered. Among the evaluated models, XGBoost achieved the best performance, reaching an accuracy of 94.42%.

Masud et al. [7] applied unsharp masking (UM) to enhance contrast along color boundaries as part of a lung cancer classification methodology. This technique enhances visual details by subtracting blurred regions from the original image. Two-dimensional wavelet and Fourier features were extracted for feature representation. After feature extraction, the collected features were classified using a convolutional neural network (CNN). Using the LC25000 dataset, the method achieved an accuracy of up to 96.33%.

Considering modifications to existing CNN-based pre-training procedures, such as increased data augmentation strategies, Garg and Garg [8] proposed that HPI could be beneficial for detecting lung and colon cancer. Eight pre-trained CNN models were trained using the LC25000 dataset, including Xception, MobileNet, InceptionResNetV2, VGG16, InceptionV3, DenseNet169, ResNet50, and NASNetMobile.

To detect lung and colon tumors using HPI, Adu et al. [9] developed DHS-CapsNet (dual horizontal squash capsule network). A new horizontal squash (HSquash) function was incorporated into DHS-CapsNet to improve encoder performance. HSquash enables effective vector squashing and introduces sparsity, which enhances capsule discrimination when extracting meaningful information from images with varied backgrounds.

Shafi et al. [10] developed a computer-aided design (CAD) model to identify pathological and physiological alterations in soft tissue cross-sections of lung cancer lesions. The model was initially trained to detect lung cancer by analyzing and comparing specific profile features in CT scans of patients and control subjects at diagnosis. CT images not used during training were then employed to test and validate the model. The study analyzed 888 annotated CT scans from the publicly available LIDC/IDRI dataset. The proposed SVM-based deep learning model achieved an accuracy of 94% in detecting pulmonary nodules that may indicate early-stage lung cancer.

Uddin [11] recommended using DenseNet for lung cancer diagnosis because it enables continuous transmission of learned features across layers. This property improves local feature learning, reduces the number of model parameters, and helps interpret the complex and uneven distribution observed in CT scans and histopathological cancer images. Additionally, DenseNet can be combined with an attention mechanism to form Attention-based DenseNet (ATT-DenseNet), allowing the model to focus more effectively on relevant regions of an image. With average improvements of 20%, 19.66%, and 24.33% in F1-score, accuracy, precision, and recall, respectively, ATT-DenseNet outperformed several alternative techniques.

Using deep learning techniques, Shin et al. [12] analyzed the characteristics of cell-derived exosomes and identified similarities with extracellular vesicles present in human plasma. Using exosome Surface-Enhanced Raman Scattering (SERS) data from lung cancer and normal cells, the deep learning classifier achieved a classification accuracy of 95%. According to this methodology, 90.7% of patient plasma exosomes in a sample of 43 patients with early-stage and stage II lung cancer were more similar to lung cancer cells than to healthy controls.

2.1 Research gap

Despite significant progress in the application of AI and ML for detecting and identifying lung diseases, several research gaps remain. State-of-the-art models often show limited accuracy and lack the ability to generalize due to small datasets, which may lack diversity in terms of tumor types, stages, or patient characteristics. Conventional methods are often unable to identify cancers at an early stage or effectively distinguish between benign and malignant nodes. Another limitation of traditional methods is their lack of interpretability, resulting in black-box AI systems that may reduce clinicians' confidence in the results. Integrating multiple data types, such as CT and PET scan imaging, together with clinical data, is highly promising but remains largely unexplored and could improve both accuracy and robustness.

3. PROPOSED SYSTEM

In this proposed section, a novel framework is presented that integrates an Attention-Guided Generative Adversarial Network (AGGAN) with a Data-efficient Image Transformer (DeiT) to improve the identification and classification of cancer tissue in patients with lung disease. High-resolution, realistic medical images that emphasize essential pathological traits are synthesized using the AGGAN model. This approach effectively augments training data and addresses class imbalances commonly found in medical imaging datasets. Figure 1 illustrates the architecture of the proposed AGGAN-DeiT framework.

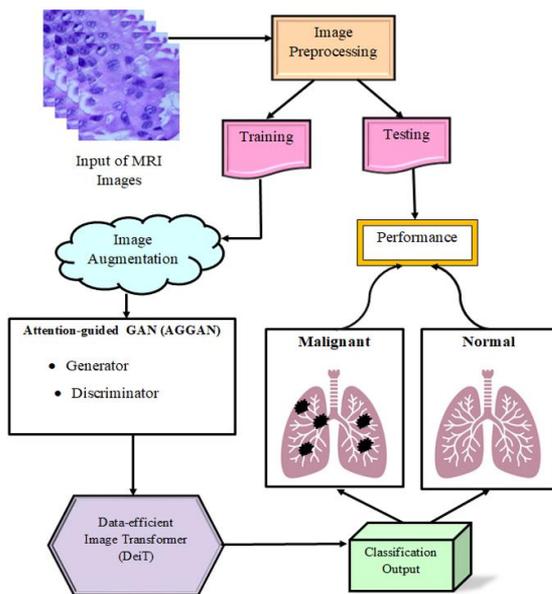


Figure 1. Proposed diagram of AGGAN-DeiT

3.1 Dataset

This collection consists of over 25,000 histopathological images across five categories from the LC25000 dataset. Individual images are stored in JPEG format with a resolution of 768×768 pixels [13]. The dataset contains 500 images of colon tissue (250 of which were benign and 250 of which were colon adenocarcinomas) and 750 pictures of lung tissue (250 of which were benign and 250 of which were lung adenocarcinomas, and 250 of which were lung squamous cell

carcinomas). These pictures came from a pilot group of sources that were checked out and followed HIPAA rules. Using the Augmentor software, these images were then enhanced to a total of 25,000. Figure 2 shows sample images from the dataset.

There are five classes in the dataset, each with 5,000 images, being:

- Lung squamous cell carcinoma
- Colon adenocarcinoma
- Colon benign tissue
- Lung adenocarcinoma
- Lung benign tissue

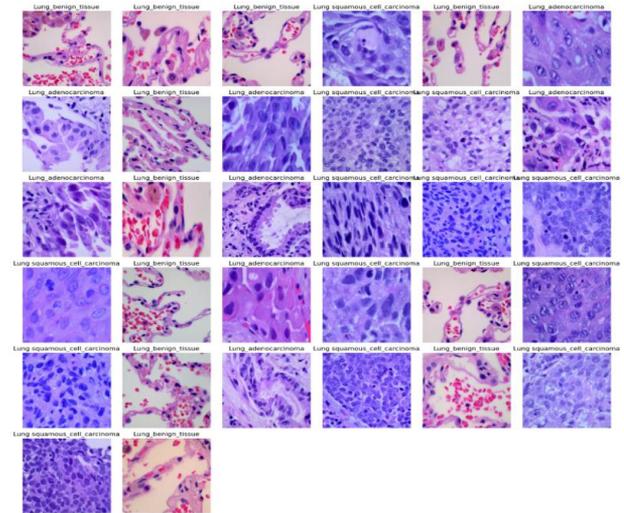


Figure 2. Sample images of dataset

3.2 Image pre-processing

Due to variations in staining techniques, lighting, tissue preparation, and scanning equipment, histopathological images exhibit a significant variation. The resilience and generalization capacity of deep learning models may be adversely affected by these differences. To improve discriminative tissue characteristics and normalize input images before model training, a specific preprocessing workflow was used.

- Image Resizing and Color Space Standardization: Every histopathological image from the LC25000 dataset was first reduced to a fixed spatial resolution of 224×224 pixels to ensure compliance with the Data-efficient Image Transformer (DeiT) design. Let, $I \in \mathbb{R}^{H \times W \times 3}$ represent an input RGB image where H and W denote the height and width of the image, and 3 represents the RGB channels. The resized image I_r is obtained as:

$$I_r = R(I, 224, 224) \quad (1)$$

where, $R(\cdot)$ represents the resizing operation.

- Stain Normalization: A staining pattern for each image was matched with a reference template via color normalization to make the slides more uniform. So, variations within hematoxylin and hematoxylin and eosin parts would be avoided. Here, given the resized image I_r , stain normalization is described as follows:

$$I_s = \mathcal{N}(I_r, I_{ref}) \quad (2)$$

where, I_{ref} is a reference image.

- **Noise Reduction:** Histopathological images may contain high-frequency noise introduced during digitization. A Gaussian smoothing filter was applied to minimize noise while preserving essential tissue characteristics:

$$I_d(x, y) = \sum_{i=-k}^k \sum_{j=-k}^k I_s(x+i, y+j)G(i, j) \quad (3)$$

where, $G(i, j)$ is a Gaussian kernel with standard deviation σ .

- **Contrast Enhancement:** To enhance tissue form and cellular borders, contrast-limited adaptive histogram equalization (CLAHE) was applied. The enhanced image I_c can be described as follows:

$$I_c = C(I_d) \quad (4)$$

where, $C(I_d)$ represents the CLAHE operation applied independently to each color channel.

- **Intensity Normalization:** To stabilize gradient updates during training, pixel intensities were finally standardized. The values of each pixel were scaled to the interval $[0,1]$ as follows:

$$I_n = \frac{I_c - \min(I_c)}{\max(I_c) - \min(I_c)} \quad (5)$$

The normalized image I_n that results is the last input for both the DeiT model for cancer tissue categorization and the Attention-Guided GAN for data augmentation.

3.3 Image augmentation

The author employed image augmentation techniques to enhance model generalization, reduce overfitting, and expand the dataset size. Image dimensions were modified through geometric transformations. To simulate variations in patient orientation, images were rotated by 30, 60, 90, and 120 degrees. To create anatomical differences, images were inverted. Tissue location and shape variations were introduced through elastic deformations. The model's capacity to classify various forms of lung cancer (LC) was evaluated by adding synthetic lesions to mimic disease states using the superimposition method. Using a system of two neural networks a discriminator and a generator that contend with one another, GANs made it easier to produce high-quality synthetic images. Without further training, the author was able to create PET/CT images that resembled actual PET/CT scans thanks to the pre-trained GANs. Moreover, images were generated at multiple locations through the latent space in order to achieve semantic interpolation. The mathematical technique for creating images with GANs is given by Eq. (6).

$$I_i = GANs(I_i), i = 1, \dots, n \quad (6)$$

3.4 Attention-Guided Generative Adversarial Network

The AGGAN is an advanced framework considered to enhance the detection and classification of cancer tissue in lung cancer. By incorporating attention mechanisms into a

Generative Adversarial Network (GAN) architecture, AGGAN efficiently focuses on key features in medical imaging, enhancing the model's ability to classify subtle patterns associated with metastasis.

1. Generative Adversarial Network (GAN):

The discriminator D and generator G , two neural networks trained simultaneously, make up a GAN [14]. Eq. (7) shows how the discriminator determines if the bogus samples provided by the generator of samples are genuine.

$$\text{Objective of GAN: } \min_G \max_D V(D, G) = E_{p_{data}(x)}[\log D(x)] + E_{p(z)}[\log(1 - D(G(z)))] \quad (7)$$

2. Attention Mechanism:

The attention mechanism directs the simulation's focus to significant regions in the images, enhancing feature representation for classification. This can be formulated as Eq. (8):

$$A = \text{soft max}(WF) \quad (8)$$

where, W represents learnable parameters and F is the feature map derived from the convolutional layers of the generator.

3. Loss Function:

The loss function can include both adversarial loss and attention loss to make sure that the generated samples not only look realistic but also continue attention on relevant features, as indicated in Eq. (9).

$$L_{AGGAN} = L_{adv} + \lambda L_{att} \quad (9)$$

where, L_{adv} is the adversarial loss, L_{att} is the attention loss, and λ is a hyperparameter that balances the two loss components.

4. Attention Loss:

Attention loss encourages the model to improve its focus on specific regions of interest (e.g., lymph nodes), as defined in Eq. (10).

$$L_{att} = \|A - A^*\|^2 \quad (10)$$

where, A is the anticipated attention map, and A^* shows the areas of cancerous tissue on the ground truth attention map.

5. Classification Loss:

Eq. (11) illustrates how a softmax function for the classification problem can be used to calculate the classification loss:

$$L_{class} = - \sum_{i=1}^C y_i \log(\hat{y}_i) \quad (11)$$

where, C is the number of classes is, y_i is the true label, and \hat{y}_i is the probability possibility.

The AGGAN technique makes use of the attention mechanism in being able to recognize and classify the cancerous region of lung cancer more effectively using the frameworks developed in their approach. This technique could greatly enhance the accuracy of diagnosis in medical images.

3.5 Data-efficient Image Transformer

DeiT aims to enhance the performance of vision transformers when labeled data is limited, which is typically the case in medical imaging applications. The model approach used in DeiT in lung cancer tissue detection and classification relies on a transformer framework that checks images by the mechanism of attention, highlighting the most important features that help pick up the subtle visual signals so often encountered in medical images [15].

DeiT is trained on a small dataset of annotated medical images using methods such as knowledge distillation, where a smaller student learns from a larger teacher perfectly, thereby enhancing its ability to generalize from limited examples. The resulting model can accurately classify the presence of histopathological cancer tissue classification, contributing to early diagnosis and treatment planning.

1. Self-Attention Mechanism: The essential of the changer architecture involves self-attention, which computes the output as shown in Eq. (12).

$$Attention(Q, K, V) = \text{soft max} \left(\frac{QK^T}{\sqrt{d_k}} \right) V \quad (12)$$

where:

- Q : Query matrix
- K : Key matrix
- V : Value matrix
- d_k : Dimension of the keys

2. Multi-Head Attention: To capture changes in relationships within the data, the MHA mechanism is defined as shown in Eq. (13):

$$MultiHead(Q, K, V) = \text{Concat}(head_1, \dots, head_h)W^O \quad (13)$$

where, to each head is definite as:

$$head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \quad (14)$$

with, W_i^Q, W_i^K, W_i^V, W^O being learned projection matrices.

3. Position-wise Feed-Forward Networks: The data is then routed through a feed-forward network following the attention process, as shown in Eq. (15).

$$FFN(x) = L U(xW_1 + b_1)W_2 + b_2 \quad (15)$$

where, W_1, W_2 are weights and b_1, b_2 are biases.

4. Loss Function: For classificational tasks, the cross-entropy loss function is generally used in Eq. (16).

$$L = - \sum_{i=1}^N y_i \log(\hat{y}_i) \quad (16)$$

where:

- y_i : True label
- \hat{y}_i : Anticipated probability for class i
- N : Total number of classes

5. Knowledge Distillation Loss: When using a teacher-student framework, the loss can embrace a distillation term, as signified in Eq. (17).

$$L_{distill} = \alpha L_{CE} + (1 - \alpha) \cdot T^2 \cdot L_{KL} \quad (17)$$

where:

- L_{CE} : Cross-entropy loss for the student model
- T : Temperature parameter to soften the probabilities
- L_{KL} : Kullback-Leibler difference between the teacher's and student's output distributions
- α : Weighting factor to stabilize the two loss terms

DeiT's transformer-based architecture, combined with methods such as knowledge distillation, makes it a powerful tool for detecting and categorizing histopathological cancer tissue classification in lung cancer using limited labeled data, thereby providing effective and accurate diagnostic support in medical imaging.

3.6 Implementation

The implementation of an Attention-guided GAN (AGGAN) collective with a Vision Transformer (DeiT) for detecting and classifying cancer tissue in lung cancer includes a multi-step method. For increasing the diversity in the dataset and for dealing with the drawback associated with less annotated data, a GAN architecture is employed for the purpose of generating medicinal images. With the implementation of an attention mechanism, which pays special attention to the most important information present in the cancerous regions, the generated images are enhanced in quality. The DeiT model is pre-trained on a massive image dataset and then fine-tuned for better abstract representations and global dependencies. The assessment of the model is done by metrics such as accuracy, precision, and recall to classify and detect histopathological cancerous regions effectively. Ultimately, a final evaluation of the performance of the model is performed to verify its practical effectiveness.

3.7 Summary of the proposed model

Attention-guided GANs and DeiT come together for lung cancer tissue classification in histopathology, as explored by Lung Cancer Discovery and Organization. Such a method marries GANs with DeiT for the attention mechanism that highlights informative features of medical images to assist in more precise identification of cancerous histopathology. GANs help generate high-quality synthetic images to augment the training set, while increasing classifier robustness. DeiT further enhances this because of its good data efficiency and attention mechanism-driven processes for making decisions about the status of cancerous tissues. In other words, this novel method is likely to improve early detection and prognosis among lung cancer patients by effectively integrating state-of-the-art deep learning techniques tailored for medical image analysis.

The AGGAN model incorporates attention mechanisms into the generator and discriminator of a classic GAN architecture. It allows the attention component to help guide the network toward the regions in the histopathological images where the key structures lie that distinguish cancerous tissue from normal tissue. Although a related Attention-GAN model exists for the generation and augmentation of histopathology images, this AGGAN will be constructed from scratch with a focus on handling cancer tissue augmentation for histopathology using the LC25000 dataset, which can capture morphological features in great detail and is able to balance the collection of LC25000. In practice, the generator would use attention to refine features in the most important areas of tissues, while the

discriminator would utilize these attention masks for evaluating not only coherence but also the accuracy of tissues. This in turn pushes the resolution of the images higher and thus improves classification. At the system level, it combines the advantages of AGGAN and DeiT by first generating realistic images using the generative approach, and then later uses them to train the transformer:

- AGGAN addresses class imbalance and small sample sizes while enhancing data diversity and feature quality.
- By using self-attention to learn subtle morphological traits and long-range spatial dependencies, DeiT improves classification accuracy.

3.8 Advantages of proposed method

- The integration of attention mechanisms enables a focused emphasis on critical features, improving the sensitivity of metastasis detection.
- The GAN framework can produce high-quality synthetic images, which augment the training dataset and enhance model performance.
- The use of Data-efficient Image Transformer (DeiT) contributes to greater robustness against differences in image acquisition, such as noise and dissimilar imaging modalities.
- This combined method can lead to developing classification performance by leveraging the strengths of both GANs and transformers, allowing precise differentiation among metastatic and non-metastatic lymph nodes.

3.9 Key innovations and contributions

Attention is woven into the how of data generation and decision-making in the approach employed by the AGGAN-DeiT framework for the reimagining of histopathology analysis.

- AGGAN is designed to identify the most diagnostically relevant tissue regions, rather than simply generating images. It can train the generator and discriminator to focus on such zones with more realism in histopathology visuals, which helps downstream models learn better.
- Synthetic images are used after augmentation, and a classifier is trained using the framework DeiT, which marries focus-based data augmentation with transformer-powered feature extraction.
- While transformers for feature extraction are not new in medical image classification, this model has access to a broader set of augmented data sources complemented by sound knowledge regarding clinically meaningful tissue properties.
- Fusion of synthetic data with the data-efficient DeiT will give outstanding performance in classification, even on small annotated datasets, which is an important barrier in medical image analysis. Attention modules of both AGGAN and DeiT preferentially handle their value matrices toward regions of interest in the morphological, rather than every pixel, as attention mechanisms frequently focus on features that training

emphasizes. This enriches the interpretability and predictability of results.

In sum, AGGAN-DeiT framework integrates attention-driven generative modeling (AGGAN) with transformer-driven classification (DeiT) into a comprehensive, reinforcing learning pipeline for histopathological cancer tissue analysis. This also differentiates it from the predecessors, in which either only GANs or only attentional mechanisms are employed. The key novelties are the emphasis on higher-quality training data and superior classification performance.

4. RESULT AND DISCUSSION

4.1 Experimental setup

An Intel Core i3 6006U CPU (2.00 GHz, 2 cores, and 4 logical processors), 16 GB of RAM, a 120 GB SSD, and a 1 TB HDD were all included in the Windows 10 Pro PC used for the investigations. The system ran a Jupyter Notebook environment via Anaconda Navigator. Python 3.7.5 was used, along with libraries like NumPy (1.18.5), Pandas (1.3.5), TensorFlow (2.3.0), and Scikit-learn (1.0.2). A high-spec laptop with GPU acceleration was employed for lung cancer image classification, enhancing deep learning model performance and reducing training time. Table 1 presents the performance metrics.

Table 1. Performance metrics

Performance Metrics	Equation	Significance
Accuracy (%)	$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$	The average ratio of positive to negative examples.
Kappa	$\text{Kappa} = \frac{Pr(a)-Pr(e)}{1-Pr(e)}$	Inter-rater agreement is a criterion used to assess the level of agreement between two approaches of classifying cancer cases. Combining specificity and sensitivity into a single value that achieves both objects.
g-mean(%)	$\text{g-mean} = \sqrt{\frac{TP*TN}{TP+FN+TN+FP}}$	The expected true positives outnumber the actual positives, whether genuine or predicted.
JaccardIndex(%)	$\text{Jaccard} = \frac{TP}{TP+FP+FN}$	Execution time is the quantity of time it requests for a computer program or procedure to complete its tasks. It is an important metric for determining the performance and efficiency of software applications.
Execution Time (ms)	$T = T_{start} - T_{end}$	

4.2 Performance metrics

4.2.1 Accuracy analysis

Table 2 and Figure 3 present a side-by-side comparison of how various machine learning and deep learning models perform binary classification of histopathological cancerous tissue (i.e., distinguishing malignant from normal) using the LC25000 data. The traditional CNN-based methods, such as ML-CNN and CNN, fall in the middle, with accuracies ranging from about 80.38% up to 88.78%. They do a

reasonably good job of extracting local spatial features. Still, because of their limited reach, they struggle to capture long-range patterns that might help differentiate complex tissue structures. The DL-SVM combo ups the ante by combining deep feature extraction with a strong classifier, reaching 86.87% for malignant tissue and 94.39% for normal tissue. This increases accuracy, as ATT-DenseNet leverages attention to focus on informative regions in images, achieving 89.13% for malignant tissue and 95.16% for normal tissue. This proves the benefits that attention mechanisms bring about in image analysis tasks in the context of pathology. The highest performers in identifying malignant vs. normal tissue images are from the AGGAN-DeiT approach itself, at 96.72% for malignant tissue images and at 97.46% for normal tissue images. The two main reasons for which the proposed approach attains superiority over existing CNN-based solutions are: (i) Attention-Guided GAN for the augmentation of the training dataset with better-quality images, and (ii) DeiT architecture for pursuing global knowledge about the image with the help of self-attention mechanisms in addition to the tissue detail in Cancer Histopathological images. This approach provides a more discerning approach than the existing CNN-based solutions in identifying histopathological Cancer tissue images.

Table 2. Accuracy analysis for AGGAN-DeiT method

Methods	Accuracy	
	Malignant	Normal
ML-CNN	80.384	85.201
CNN	82.281	88.781
DL-SVM	86.873	94.387
ATT-DenseNet	89.132	95.162
AGGAN-DeiT	96.721	97.462

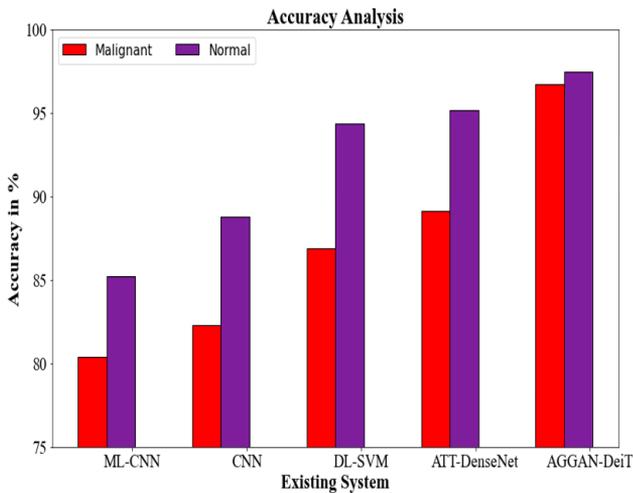


Figure 3. Accuracy analysis for AGGAN-DeiT method

4.2.2 Execution time analysis

Table 3 and Figure 4 present the time consumption of different machine learning models in detecting malignant and normal brain MRI images. ML-CNN takes longer for the malignant images at 14.384 seconds and relatively quicker for normal images at 12.201 seconds. The CNN model holds back the shortest time for malignant image detection at 11.281 seconds, but marginally longer for normal images at 14.781 seconds. The DL-SVM maintains a relatively consistent performance for both images, and ATT-DenseNet follows the same trend with a mere whisper quicker response times—at

11.132 seconds for malignant and 11.162 seconds for normal images. Lastly, the AGGAN-DeiT has been observed as the quickest among all models with no time difference between the detection of both malignant (10.721 seconds) and normal (10.462 seconds) images, thereby hinting at the most optimized machine learning model among the lot.

Table 3. Execution time analysis for AGGAN-DeiT method

Methods	Execution Time	
	Malignant	Normal
ML-CNN	14.384	12.201
CNN	11.281	14.781
DL-SVM	13.873	12.387
ATT-DenseNet	11.132	11.162
AGGAN-DeiT	10.721	10.462

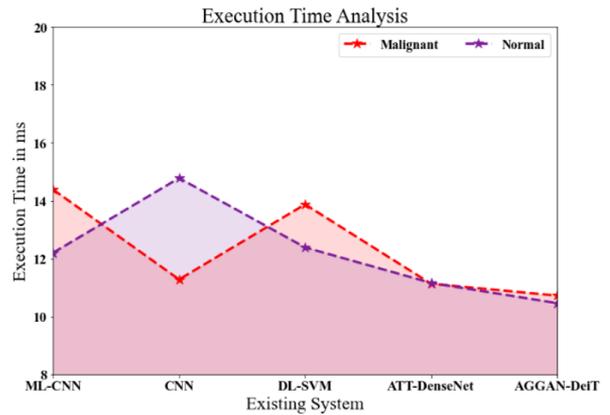


Figure 4. Execution time analysis for AGGAN-DeiT method

Table 4. Cohen’s Kappa Score analysis for AGGAN-DeiT method

Methods	Cohen’s Kappa Score	
	Malignant	Normal
ML-CNN	0.52	0.78
CNN	0.69	0.82
DL-SVM	0.57	0.86
ATT-DenseNet	0.74	0.84
AGGAN-DeiT	0.98	0.90

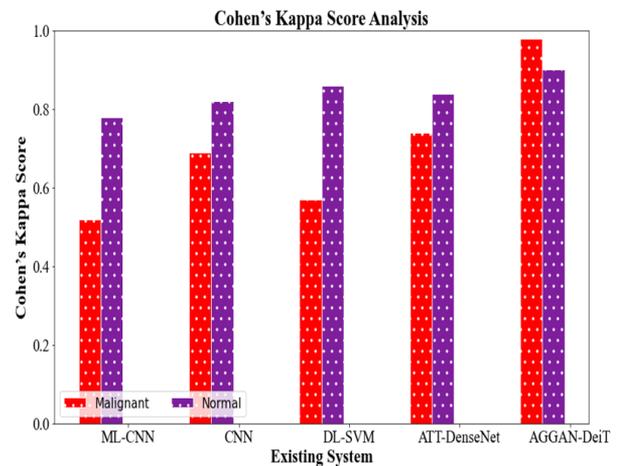


Figure 5. Cohen’s Kappa Score analysis for AGGAN-DeiT method

4.2.3 Cohen’s Kappa Score analysis

Table 4 and Figure 5 present the Cohen’s Kappa Score analysis for various machine learning models, comparing their

performance in classifying malignant and normal cases, with a focus on the AGGAN-DeiT method. The Cohen’s Kappa Score, which measures inter-rater agreement, indicates the reliability of each model. The ML-CNN, CNN, DL-SVM, and ATT-DenseNet methods show moderate to high agreement scores across malignant and normal classifications. Notably, the AGGAN-DeiT method outperforms all others, achieving a high agreement score of 0.98 for malignant and 0.90 for normal cases, demonstrating superior consistency and reliability in classification.

4.2.4 G-mean analysis

The geometric mean (g-mean) is used to balance sensitivity and specificity in binary cancer tissue classification tasks, where class-wise performance must be fairly assessed.

Table 5. G-mean analysis for AGGAN-DeiT method

Methods	G-mean	
	Malignant	Normal
ML-CNN	0.47	0.63
CNN	0.12	0.57
DL-SVM	0.48	0.92
ATT-DenseNet	0.73	0.94
AGGAN-DeiT	0.91	0.96

Table 5 and Figure 6 illustrate how different models result in different g-means for malignant and normal tissues while utilizing tumor classification and detection methods. For malignant and normal tissues, the ML-CNN model results in g-means of 0.47 and 0.63, respectively. However, the DL-SVM model performs better, producing g-means of 0.48 for malignant and 0.92 for normal tissues, while the traditional CNN model performs worst, producing g-means of 0.57 for malignant and 0.12 for normal tissues. The ATT-DenseNet model performs best for normal tissues, producing a g-mean of 0.94 for normal and 0.73 for malignant tissues. Looking at balance for sensitivity and specificity, CNN and ML-CNN models are closer to the bottom, which indicates poor performance. DL-SVM performs much better, and this is for normal tissues. However, the best result is from the AGGAN-DeiT model, which produces the largest g-means of 0.91 for malignant and 0.96 for normal tissues, respectively, which clearly indicates better balance. It is quite effective for classification of histopathological tissues related to cancers, utilizing features and data-driven augmentation strategies.

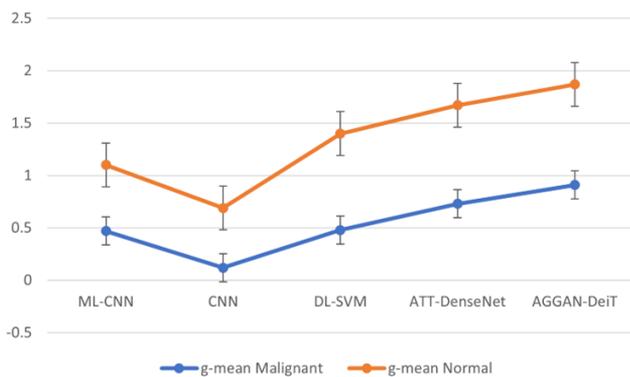


Figure 6. G-mean analysis for AGGAN-DeiT method

4.2.5 Jaccard index analysis

When comparing the performance of several simulation models with the Jaccard Index between malignant and normal

tissue, Table 6 and Figure 7 clearly show that there have been significant modifications. The Jaccard Index for the ML-CNN model is 0.73 for malignant tissue and 0.46 for normal tissue, indicating fair performance in distinguishing between the two. The CNN advantage appears slightly greater for malignant tissue, where the Jaccard Index is 0.79, but poor for normal tissue, which drops to 0.12. The performance of the DL-SVM also dipped, to 0.21 for malignant tissue and 0.30 for normal tissue. The ATT-DenseNet model, however, performed poorly compared with other models, with Jaccard Indexes of 0.83 for malignant tissue and 0.38 for normal tissue. In contrast, the proposed new model, denoted the AGGAN-DeiT model, achieved the highest performance, with indices of 0.88 for malignant and 0.91 for normal, indicating a strong mapping between the predicted outputs and ground truths, compared with other models and within their own definitions.

Table 6. Jaccard index analysis for AGGAN-DeiT method

Methods	Jaccard Index	
	Malignant	Normal
ML-CNN	0.73	0.46
CNN	0.79	0.12
DL-SVM	0.21	0.30
ATT-DenseNet	0.83	0.38
AGGAN-DeiT	0.88	0.91

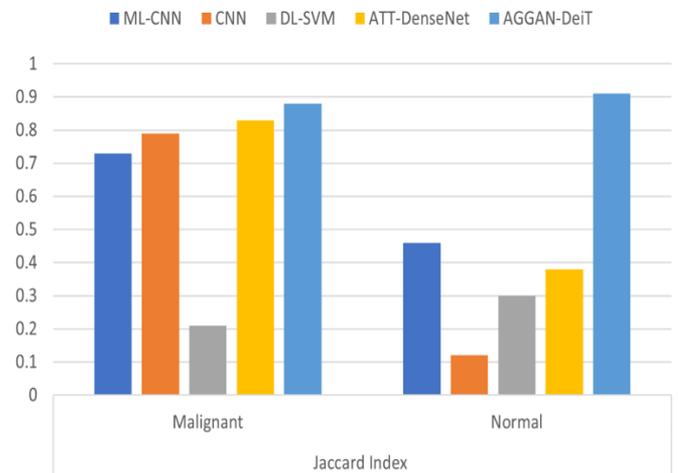


Figure 7. Jaccard Index for AGGAN-DeiT method

4.3 Training and testing validation

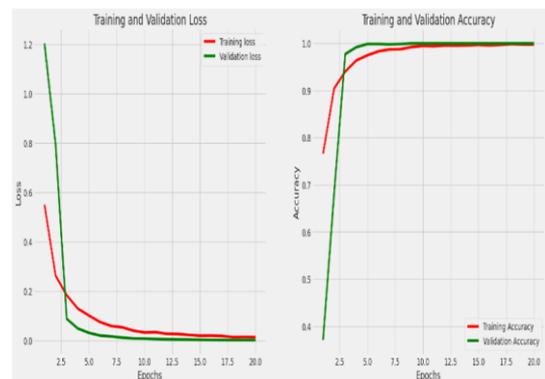


Figure 8. Training and validation loss and accuracy

Training validation loss and accuracy are critical parameters for estimating the success of a machine learning model during

the training procedure, as shown in Figure 8. Training validation loss procedures how well the model predicts the training data, with lower values representing better performance and reduced error. Training validation accuracy, on the other hand, measures the percentage of properly predicted examples in the validation dataset, demonstrating how well the model generalizes. Monitoring these metrics helps classify issues such as overfitting or underfitting, enabling adjustments to the training process to optimize model presentation.

4.4 Comparative methods

Multi-level CNN (ML-CNN) [6]: The model was validated using the LC25000 dataset, which includes histological images of carcinoma of squamous cells and adenocarcinoma.

Convolutional Neural Network (CNN) [9]: Enhanced augmentation approaches are an essential component of the current pre-training CNN-based strategy for detecting lung and cancers of the colon using HPI.

Deep Learning- Support Vector Machine (DL-SVM) [11]: It outperforms previous methods for detecting nodules on lung CT scans, including compound DL, simple ML, and hybrid algorithms.

Attention-based DenseNet (ATT-DenseNet) [12]: Furthermore, the model can focus on specific sections of an image by combining DenseNet with an attention mechanism (ATT-DenseNet), giving higher weight to relevant parts.

4.5 Discussion

This work explores the detection and classification of histopathological cancer tissue images related to lung cancer using an Attention-Guided GAN and a Data-efficient Image Transformer (DeiT). By generating high-quality synthetic images with detailed features, AGGAN improves the training process and increases model accuracy. DeiT efficiently handles image data. Therefore, the combination of AGGAN and DeiT enhances the detection accuracy of cancerous tissues

and demonstrates strong potential for differentiating malignant from normal tissues, thereby supporting the automatic diagnosis of lung cancer. The accuracy of the proposed methods for distinguishing between malignant and normal tissues is ranked as follows: the proposed AGGAN-DeiT framework achieved results of 96.721% and 97.462%, respectively. ATT-DenseNet ranked second with 89.132% for malignant tissues and 95.162% for normal tissues, while the DL-SVM approach showed moderate performance with results of 86.873% and 94.387%, respectively. The traditional CNN recorded slightly lower performance of 82.281% and 88.781% for malignant and normal tissues, respectively, whereas the lowest performance was observed with ML-CNN, achieving 80.384% and 85.201% for malignant and normal tissues, respectively.

4.6 Ablation study

In the suggested model, each module plays a crucial part. In this section, we estimate the proposed AGGAN-DeiT model versus current models like ML-CNN, CNN, DL-SVM, and ATT-DenseNet using a series of ablation tests on the Lung dataset to evaluate performance gains and elucidate the reasons behind AGGAN-DeiT.

i) Influence of the DeiT:

The Data-efficient Image Transformer (DeiT) has been a stealth leader in computer vision in the sense that it has been promoting the use of the transformer and showing the way for other models in the way it can offer good results with less labeled data in computer vision. Unlike other convolutional nets that require massive labeled datasets, the DeiT relies instead on the use of knowledge distillation in the way it trains by learning from the teacher models it is trained with. The vision transformer can allow for the understanding and manipulation of global views and enhanced features in images. The Data-efficient Image Transformer can mark the beginning of something big in using the transformer in image processing tasks.

Table 7. 10-fold cross validation of AGGAN-DeiT analysis

K-folds	AGGAN-DeiT for Malignant Accuracy	AGGAN-DeiT for Normal Accuracy
1-Fold	0.94	0.98
2-Fold	0.99	0.97
3-Fold	0.97	0.95
4-Fold	0.95	0.96
Fold-5	0.96	0.97
Fold-6	0.99	0.97
Fold-7	0.95	0.97
Fold-8	0.96	0.97
Fold-9	0.94	0.98
Fold-10	0.95	0.99
10-Fold Mean	0.96	0.97

Table 8. Comparison of the proposed method with the other models

S No	Author	Dataset Used	Models	Accuracy
1.	Dabass et al. [16]	CRAG Dataset	Atrous Convolved Hybrid Seg-Net Architecture	87.63%
2.	Setiawan et al. [17]	LC25000 Database	CNN	87.16%
3.	Ahmed. et al. [18]	LUNA16 Database	3D CNN	80%
4.	Mastouri. et al. [19]	LUNA16 Database	BCNN	91.99%
5.	Our Models	LC25000 Lung and colon histopathological image dataset	AGGAN-DeiT	97.091%

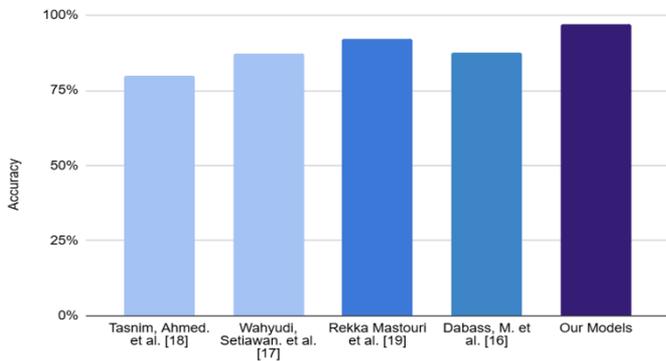


Figure 9. Comparison of the proposed method with the other models

In addition, when combined with the AGGAN model, DeiT attains an accuracy of 96.721% for malignant and 97.462% for normal images within our dataset. Conversely, current approaches such as ML-CNN, CNN, DL-SVM, and ATT-DenseNet yield accuracy of 80.384% and 85.201% (malignant/normal) for ML-CNN, 82.281% and 88.781% for CNN, 86.873%.

ii) Influence of the K-fold cross validation:

The robustness and dependability of the AGGAN and DeiT models for histopathological cancer tissue classification detection in lung cancer research are greatly improved by ten-fold cross-validation. Cross-validation is an arithmetical approach in which the dataset is divided into ten portions, or subsets, with nine utilized for training and another for validation. Using a 10-fold cross-validation procedure, the suggested AGGAN-DeiT model achieved a performance accuracy of 96.721% for our input data. Ultimately, both the AGGAN and DeiT models demonstrated superior performance, achieving accuracies of 96.721% for malignant data and 97.462% for normal data. In comparison, existing models such as ML-CNN, CNN, DL-SVM, and ATT-DenseNet showed accuracies for malignant and normal data of 80.384% and 85.201%, 82.281% and 88.781%, 86.873% and 94.387%, and 89.132% and 95.162%, respectively, as shown in Table 7.

4.7 Comparative analysis

Table 8 and Figure 9 compare the proposed AGGAN-DeiT framework with other existing lung cancer detection models. Tasnim Ahmed et al. [18] used LUNA16 data with a 3D CNN model, achieving an accuracy of 80%. Wahyudi Setiawan et al. [17] used a CNN-based model on the LC25000 dataset with an error of 87.16%. Mastouri et al. [19] introduced a BCNN composed of VGG16 and VGG19, with SVM, on LUNA16, achieving 91.99% accuracy. On the same note, Dabass et al. [16] presented an architecture, Atrous Convolved Hybrid Seg-Net, for gland detection in the CRAG dataset, achieving 87.63% accuracy. Compared with the suggested AGGAN-DeiT model using LC25000 histopathological data, the model achieves higher accuracy (97.091), making it effective for cancer tissue classification.

4.8 Challenges and limitations

The requirement for large, high-quality labeled training datasets is one of the challenges associated with using an Attention-Guided Generative Adversarial Network (AGGAN) and a Data-efficient Image Transformer (DeiT) for

histopathological cancer tissue classification in lung cancer. Although AGGAN-DeiT framework is designed to perform well even with limited data, cancer tissues often contain complex patterns. If the datasets do not provide sufficient variability or contain imaging artifacts, the precision of these models may be affected. In addition, AGGAN's attention mechanism requires intensive computation, especially when high-resolution imaging is used, which can increase training time. Furthermore, despite its efficiency, issues related to generalization may arise when smaller datasets are used or when there are variations in imaging sources [20].

5. CONCLUSION

Thus, the combination of AGGAN and the Data-efficient Image Transformer (DeiT) provides a promising direction for the detection and classification of histopathological cancer tissue in lung cancer. The complementary use of these approaches improves image quality as well as the accuracy of the classification process. The capability of AGGAN to focus on key structures within images helps capture important characteristics of cancer tissues, while the use of DeiT ensures effective performance even in situations with limited data. This technique not only contributes to advances in cancer diagnosis but also highlights the potential of combining generative models with transformer-based models in medical imaging for more efficient cancer analysis. Future research will focus on further optimizing this framework and exploring its applicability across other oncological domains.

REFERENCES

- [1] Zhang, S., Wang, H., Xu, Z., Bai, Y., Xu, L. (2020). Lymphatic metastasis of NSCLC involves chemotaxis effects of lymphatic endothelial cells through the CCR7–CCL21 axis modulated by TNF- α . *Genes*, 11(11): 1309. <https://doi.org/10.3390/genes11111309>
- [2] Fang, C., Xiang, Y., Han, W. (2022). Preoperative risk factors of lymph node metastasis in clinical N0 lung adenocarcinoma of 3 cm or less in diameter. *BMC Surgery*, 22(1): 153. <https://doi.org/10.1186/s12893-022-01605-z>
- [3] Chao, H., Shan, H., Homayounieh, F., Singh, R., et al. (2021). Deep learning predicts cardiovascular disease risks from lung cancer screening low dose computed tomography. *Nature Communications*, 12(1): 2963. <https://doi.org/10.1038/s41467-021-23235-4>
- [4] Lotfollahi, M., Naghipourfar, M., Luecken, M.D., Khajavi, M., et al. (2022). Mapping single-cell data to reference atlases by transfer learning. *Nature Biotechnology*, 40(1): 121-130. <https://doi.org/10.1038/s41587-021-01001-7>
- [5] Ramesh, M., Maheswaran, S., Theivanayaki, S., Kodeeswari, K., Sriram, N. (2023). Efficient lung cancer classification on multi level convolution neural network using histopathological images. In 2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT), Delhi, India, pp. 1-7. <https://doi.org/10.1109/ICCCNT56998.2023.10307852>
- [6] Mamun, M., Farjana, A., Al Mamun, M., Ahammed, M.S. (2022). Lung cancer prediction model using ensemble

- learning techniques and a systematic review analysis. In 2022 IEEE world AI IoT congress (AIoT), Seattle, WA, USA, pp. 187-193. <https://doi.org/10.1109/AIoT54504.2022.9817326>
- [7] Masud, M., Sikder, N., Nahid, A.A., Bairagi, A.K., AlZain, M.A. (2021). A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework. *Sensors*, 21(3): 748. <https://doi.org/10.3390/s21030748>
- [8] Garg, S., Garg, S. (2020). Prediction of lung and colon cancer through analysis of histopathological images by utilizing Pre-trained CNN models with visualization of class activation and saliency maps. In *Proceedings of the 2020 3rd Artificial Intelligence and Cloud Computing Conference*, Kyoto, Japan, pp. 38-45. <https://doi.org/10.1145/3442536.3442543>
- [9] Adu, K., Yu, Y., Cai, J., Owusu-Agyemang, K., Twumasi, B.A., Wang, X. (2021). DHS-CapsNet: Dual horizontal squash capsule networks for lung and colon cancer classification from whole slide histopathological images. *International Journal of Imaging Systems and Technology*, 31(4): 2075-2092. <https://doi.org/10.1002/ima.22569>
- [10] Shafi, I., Din, S., Khan, A., Díez, I.D.L.T., Casanova, R.D.J.P., Pifarre, K.T., Ashraf, I. (2022). An effective method for lung cancer diagnosis from CT scan using deep learning-based support vector network. *Cancers*, 14(21): 5457. <https://doi.org/10.3390/cancers14215457>
- [11] Uddin, J. (2024). Attention-based densenet for lung cancer classification using CT scan and histopathological images. *Designs*, 8(2): 27. <https://doi.org/10.3390/designs8020027>
- [12] Shin, H., Oh, S., Hong, S., Kang, M., et al. (2020). Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes. *ACS Nano*, 14(5): 5435-5444. <https://doi.org/10.1021/acsnano.9b09119>
- [13] Lung and Colon Cancer Histopathological Images. <https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images/data>
- [14] Tang, H., Xu, D., Sebe, N., Yan, Y. (2019). Attention-guided generative adversarial networks for unsupervised image-to-image translation. In *2019 International Joint Conference on Neural Networks (IJCNN)*, Budapest, Hungary, pp. 1-8. <https://doi.org/10.1109/IJCNN.2019.8851881>
- [15] Anzum, H., Sammo, M.N.S., Akhter, S. (2024). Leveraging data efficient image transformer (DeIT) for road crack detection and classification. In *2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS)*, Dhaka, Bangladesh, pp. 1-6. <https://doi.org/10.1109/iCACCESS61735.2024.10499539>
- [16] Dabass, M., Dabass, J. (2023). An Atrous Convolved Hybrid Seg-Net Model with residual and attention mechanism for gland detection and segmentation in histopathological images. *Computers in Biology and Medicine*, 155: 106690. <https://doi.org/10.1016/j.combiomed.2023.106690>
- [17] Setiawan, W., Suhadi, M.M., Pramudita, Y.D. (2022). Histopathology of lung cancer classification using convolutional neural network with gamma correction. *Communications in Mathematical Biology and Neuroscience*, 2022: 7611. <https://doi.org/10.28919/cmbn/7611>
- [18] Ahmed, T., Parvin, M.S., Haque, M.R., Uddin, M.S. (2020). Lung cancer detection using CT image based on 3D convolutional neural network. *Journal of Computer and Communications*, 8(3): 35. <https://doi.org/10.4236/jcc.2020.83004>
- [19] Mastouri, R., Khelifa, N., Neji, H., Hantous-Zannad, S. (2021). A bilinear convolutional neural network for lung nodules classification on CT images. *International Journal of Computer Assisted Radiology and Surgery*, 16(1): 91-101. <https://doi.org/10.1007/s11548-020-02283-z>
- [20] Ijjina, E.P., Chalavadi, K.M. (2016). Human action recognition using genetic algorithms and convolutional neural networks. *Pattern Recognition*, 59: 199-212. <https://doi.org/10.1016/j.patcog.2016.01.012>