European Journal of Technique

journal homepage: https://dergipark.org.tr/en/pub/ejt

Vol.15, No.1, 2025

Research Article

The Potential of Transformer-Based Models for Automated Lung Cancer Detection from CT Scans

Oguzhan Katar[1*], Tulin Ozturk[2], Ozal Yildirim[1]

[1*]Department of Software Engineering, Firat University, Elazig, Turkey. (e-mail: okatar@firat.edu.tr) (e-mail: ozalyildirim@firat.edu.tr)
[2] Department of Radiology, Fethi Sekin City Hospital, Elazig, Turkey. (e-mail: tulin58@hotmail.com).

ARTICLE INFO

Received: Nov., 09. 2024
Revised: Mar., 28. 2025
Accepted: May., 12. 2025

Keywords:
Lung cancer
Transformer models
CT imaging
Diagnosis
Deep learning

Corresponding author: *Oguzhan Katar*

ISSN: *2536-5010* | e-ISSN: *2536-5134*

DOI: https://doi.org/10.36222/ejt.1582121

ABSTRACT

Lung cancer is the most common type of cancer worldwide and the leading cause of cancer-related deaths. Early diagnosis and treatment can significantly increase the survival rate of this disease. Radiological methods used in the diagnosis of lung cancer, especially Computed Tomography (CT) imaging, allow tumors to be detected more precisely. However, manual analysis of these images is time consuming and error prone due to human factors. In this study, we compared the potential of three different transformer-based state-of-the-art models (ViT, DeiT and Swin Transformer) for automatic lung cancer detection. We collected 690 CT images including small cell lung cancer (SCLC), non-small cell lung cancer (NSCLC) and normal findings from a local hospital. Each image was carefully reviewed and labeled by our expert radiologist, and these labeled images were used to train the models. The ViT, DeiT and Swin Transformer models achieved accuracy rates of 91.3%, 84.1% and 80.4% respectively on the test samples. This study shows that the use of transformer-based models for lung cancer classification is promising in overcoming the difficulties in manual analysis.

1. INTRODUCTION

Lung cancer is the most common type of cancer and the leading cause of cancer-related deaths worldwide [1]. According to the GLOBOCAN database, approximately 2.09 million new cases were reported in 2018, with around 1.76 million deaths attributed to the disease [2]. In recent years, both incidence and mortality rates have risen sharply [3].

Lung cancer is classified into two main types: Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC) [4]. NSCLC accounts for 80%–85% of cases and includes three subtypes: adenocarcinoma, squamous cell carcinoma, and large cell carcinoma [5]. SCLC, on the other hand, represents about 10%–15% of cases. Although survival rates vary by clinical stage, the overall 5-year survival rate remains low at approximately 22% [6].

Radiological imaging methods are frequently used to detect lung cancer early [7]. Chest radiography is one of the most basic methods [8]. It is both low-cost and widely used. However, the tumor's size and location can cause issues. It might be missed or mistaken for other lung diseases [9]. Due to these disadvantages, detecting small lesions or early-stage tumors on chest radiography is very challenging [10].

Computed tomography (CT) is a more sensitive method for lung cancer diagnosis [11]. It allows comprehensive volumetric images to be captured in a single breath-hold [12]. By scanning the lungs in thin sections, CT can detect both small nodules and the extent of tumor spread [13]. Several studies have shown that low-dose CT detects more nodules and early-stage lung cancers compared to chest radiography [14, 15]. However, interpreting lung CT scans is a particularly intensive task. It requires extensive experience to assess the malignancy risk accurately [16]. Without such experience, the risk of misinterpretation can increase, affecting the accuracy of diagnosis.

To reduce these risks, deep learning-based computer-aided diagnosis (CAD) systems have been developed [17, 18]. These AI-powered systems enhance diagnostic efficiency, potentially reducing the workload on radiologists [19]. Li et al. introduced the Reconstruction-Assisted Feature Coding Network (RAFENet) model [20]. This model automatically classifies adenocarcinoma and squamous cell carcinoma in CT images. In their study, CT images from the Cancer Imaging Archive (TCIA) were utilized. Due to hardware limitations, each CT slice was cropped into a 128×128 pixel patch centered on the target structure. An early stopping function was used during training to stop the process if validation accuracy didn't improve within 10 epochs. RAFENet achieved a classification accuracy of 79.70% on the test set. Pang et al. developed a model based on densely connected convolutional neural networks (CNNs) to automatically classify adenocarcinoma, squamous cell

carcinoma, and large cell carcinoma in CT images [21]. They used real patient data collected from Shandong Provincial Hospital for training and validation. Since the dataset was limited, they applied data augmentation techniques such as rotation, translation, and transformation to increase the variability in the training data. The model achieved an accuracy of 89.85% in detecting lung cancer. Han et al. employed the VGG-16 architecture for automatic classification of adenocarcinoma and squamous cell carcinoma [22]. Their model was trained using a dataset collected from Peking University Cancer Hospital. The dataset was split using a 10-fold cross-validation approach. The VGG-16 model reached an accuracy of 84.10% on the test set. Chaunzwa et al. also proposed a VGG-16 based model for classifying adenocarcinoma and squamous cell carcinoma from CT images [23]. The model was trained using a private dataset collected from 311 early-stage NSCLC patients treated at Massachusetts General Hospital. The model achieved an AUC of 0.71 (p = 0.018) in classifying these cancer types. Zhao et al. proposed a Vision Transformer-based (ViT) model for the classification of NSCLC subtypes [24]. Their model was trained on CT images obtained from the TCIA. To optimize performance, images were resized to 224×224 pixels before being fed into the network. A data augmentation strategy, including rotation and flipping, was applied to improve generalization. The model was trained using a cross-entropy loss function, and an adaptive learning rate scheduler was employed. Experimental results demonstrated that the ViT model achieved a classification accuracy of 86.00%. Venkatesh et al. proposed a hybrid deep learning model for lung cancer detection, combining patch processing with CNN-based classification [25]. Using CT images from the LIDC and Kaggle datasets, the model automatically distinguishes between benign and malignant lung nodules. By extracting relevant features through CNN, the approach achieved an impressive classification accuracy of 99.96%.

The experimental findings of these studies show that deep learning architectures hold great potential for automatically classifying NSCLC subtypes. However, these studies focus solely on NSCLC detection and exclude SCLC and normal findings. While NSCLC makes up about 85% of lung cancer cases, models that only detect this class are not enough. A reliable decision support system should also accurately classify SCLC and normal cases. However, we have not come across a publicly available lung CT dataset containing NSCLC, SCLC and normal images labelled for training deep learning models. In addition, when the studies are analyzed, it is seen that most of them use CNN-based architectures for lung cancer detection. However, CNNs heavily depend on local receptive fields and pooling operations, which limits their ability to capture long-range dependencies within an image. This limitation makes it harder to fully understand the input and identify complex relationships between different regions of the image. In contrast, transformer-based models, which use self-attention mechanisms to capture interactions between distant image regions, have the potential to enhance the accuracy of lung cancer classification.

In this study, we compared the performance of transformer-based architectures for automatic lung cancer classification from CT images. Specifically, we evaluated three models commonly used in image classification: ViT, data-efficient image transformers (DeiT), and Swin Transformer. For this study, we collected a private lung cancer dataset, which includes CT images of patients with SCLC, NSCLC and normal findings. Then we trained each model with equal hyper-parameters. Using the weights obtained as a result of training, we examined the computational and statistical performance of the models on the test samples.

The main contributions of this study are as follows:

- We provided a thorough comparison of three state-of-the-art transformer-based models for the task of lung cancer classification from CT images.
- We applied the transformer-based models to a real world lung cancer dataset.
- We compared the computational efficiency of the models and assessed their potential for integration into diagnostic processes.
- Our work contributes to the growing field of CAD by showcasing the potential of transformer-based models.

2. MATERIAL AND METHODS

In this study, we utilized three different deep learning models for lung cancer classification. Our approach includes a pre-processing stage where input CT images are divided into patches that these models can process. Specifically, we used 16×16 pixel patches for ViT and DeiT, while the Swin Transformer operates with 4×4 pixel patches. Each model leverages a transformer-based architecture to extract features from the images, and classifies them into one of three categories: SCLC, NSCLC, or normal. To evaluate and compare the performance of these models, we analyzed confusion matrices, training and validation curves, and epoch-based duration. The block diagram of our used framework is given in Figure 1.

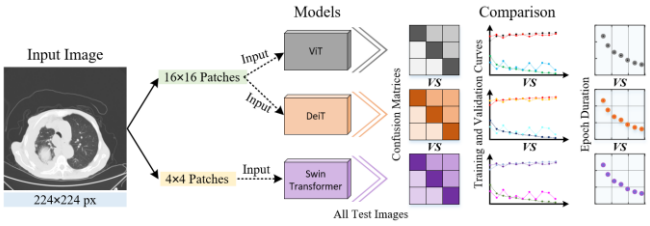


Figure 1. The block diagram of our framework.

2.1. Lung cancer CT dataset

In this study, we collected a private dataset with the approval of the non-interventional ethics committee from Firat University (Approval Number: 2024/13-38). The dataset contains 690 CT images, showing either lung cancer (SCLC or NSCLC) or normal findings. These scans were taken at Elazig Fethi Sekin City Hospital between 2020 and 2024. All CT scans were performed using a Philips Ingenuity-128 CT device. Our expert radiologist carefully reviewed and labeled each image. For images labeled as SCLC or NSCLC, a biopsy result confirmed the diagnosis.

TABLE I
DETAILS OF THE LUNG CANCER CT DATASET

Class	Label	Image Resolution	Number of Images	Percentage
Cancer	SCLC	768×768×3	125	54,35 %
	NSCLC	768×768×3	250	
Normal	Normal	768×768×3	315	45,65 %

CT images without finding of lung cancer were labeled as normal. The dataset comprises 125 images of SCLC, 250

images of NSCLC, and 315 images with no pathological findings. The SCLC and NSCLC images include diverse tumor sizes, locations, and densities to ensure clinical variability. Normal images include patients with no signs of nodules, masses, or other abnormalities. Further details about the dataset are given in Table 1.

2.2. Transformer-based image classification models

In recent years, transformer-based models have gained significant popularity in computer vision studies [26, 27]. These models have the ability to capture long-distance dependencies thanks to their self-attention mechanisms [28]. This capability sets them apart from traditional CNNs. In this study, we employed three different transformer-based models. Each of these models can be summarized as follows.

In 2020, Dosovitskiy et al. introduced the ViT model, which marked a significant step in applying transformer-based architectures to image classification [29]. Unlike CNNs, ViT divides input images into fixed-size patches. These patches are treated like tokens in natural language processing tasks. Each patch is linearly embedded and passed through several layers of self-attention [30]. The model uses a specialized classification token (CLS token) to summarize the information gathered from the patches. The output of this token is converted into a class prediction through a small multilayer perceptron (MLP) using a tanh activation function in a single hidden layer.

The DeiT model, developed by Touvron et al. in 2021, aims to make transformer-based architectures for image classification more efficient [31]. Similar to ViT, DeiT divides the input images into fixed-size patches and treats these patches as tokens. These patches are placed linearly and then passed through self-attention layers. One of the most important features of DeiT is that it uses a teacher model to improve performance even with less data. This approach, called knowledge distillation, involves transferring knowledge from a larger and well-trained teacher model to a smaller student model [32]. The model trained on a larger dataset not only provides accurate predictions, but also valuable information about the relationships between different classes. This additional information allows DeiT to be trained more efficiently even with limited data.

In 2021, Liu et al. proposed the Swin Transformer model, which builds on transformer-based architectures for image classification [33]. This model addresses some of the challenges seen in ViT. While ViT processes an entire image at once, the Swin Transformer divides the image into fixed-size patches called windows. Self-attention is applied within each window, focusing on local regions, which reduces the computational load. To connect information between windows, the model uses a shifting window mechanism that shifts the windows at different layers, allowing the model to gather features from across the image [34]. Swin Transformer also uses a hierarchical structure, where the patch size increases as the network progresses, letting it capture both detailed and broader features. This multi-scale processing helps the model handle both fine and coarse information effectively.

2.3. Evaluation metrics

We used confusion matrix based metrics to evaluate the performance of the models. Confusion matrix is a simple table showing the relationship between the actual and predicted classes. This matrix contains the number of true predictions and false predictions. These situations are represented by 4 different elements. In a multi-class study, these elements are usually evaluated separately for each class. These elements can be summarized as follows.

- True Positive (TP): The number of images whose labels are correctly predicted from the samples belonging to the target class.
- True Negative (TN): The number of images whose labels are correctly predicted from samples of classes other than the target class.
- False Positive (FP): The number of images whose labels are predicted as the target class although their actual labels are different from the target class.
- False Negative (FN): The number of predicted images with labels different from the target class.

When measuring the classification performance of deep learning models, four main metrics are typically used. Accuracy (Acc) measures how correctly the model predicts across all test data. It is calculated using Equation 1.

$$Acc = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

Precision (Pre) assesses how accurate the model is in its positive classifications. It is calculated using Equation 2.

$$Pre = \frac{TP}{TP + FP} \tag{2}$$

Recall (Rec) evaluates how well the model identifies true positives. It is calculated using Equation 3.

$$Rec = \frac{TP}{TP + FN} \tag{3}$$

F-1 score provides a balance between precision and recall by calculating their harmonic mean. It is calculated using Equation 4.

$$F1 = \frac{2 \times Pre \times Rec}{Pre + Rec} \tag{4}$$

3. EXPERIMENTS

In this section, we present our experimental setup and results. First, we describe the training scenario and pre-processing steps. Next, we detail the training and testing processes of the models used in the study. Finally, we compare the performance of the models using various metrics.

3.1. Experimental setup

In this study, we evaluated the performance of transformer-based models for classifying lung cancer from CT images. Each model was initialized with pre-trained weights. We fine-tuned these models on our lung cancer dataset. Randomly selected samples from the classes in our dataset are given in Figure 2.

The models were trained to classify the CT images into one of three categories: SCLC, NSCLC, or normal. The collected CT images were first resized to a resolution of 224×224 pixels. The dataset samples are randomly divided as follows: 60% for training, 20% for validation, and 20% for

testing. The distribution of the dataset samples is given in Table 2.

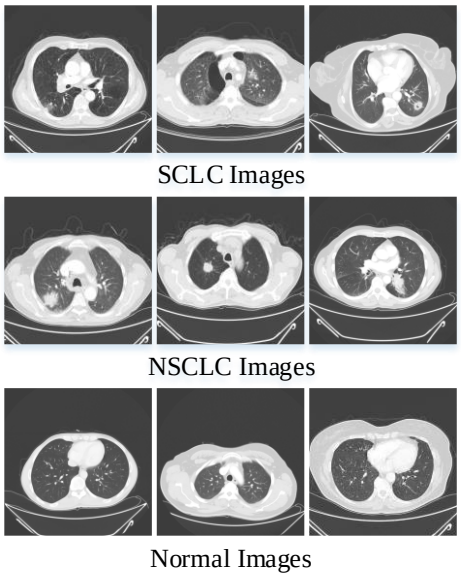


Figure 2. Randomly selected samples from our dataset.

TABLE II
THE DISTRIBUTION OF THE DATASET SAMPLES

Image Type	Resolution	Number of Train Images	Number of Validation Images	Number of Test Images
PNG	224×224×3	SCLC=75 NSCLC=150 Normal=189	SCLC=25 NSCLC=50 Normal=63	SCLC=25 NSCLC=50 Normal=63

All models were implemented using the PyTorch framework (version 2.2.1). The training pipeline was managed using timm. We used the AdamW optimizer with an initial learning rate of 0.00002. A batch size of 16 was used for training. Each model was trained for a maximum of 100 epochs. Early stopping function was used to prevent overfitting. The patience value was set to five, and validation loss was monitored at the end of each epoch. The performance of each model was evaluated using confusion matrix-based metrics. All experimental studies were carried out on RTX 4090 24 GB GPU.

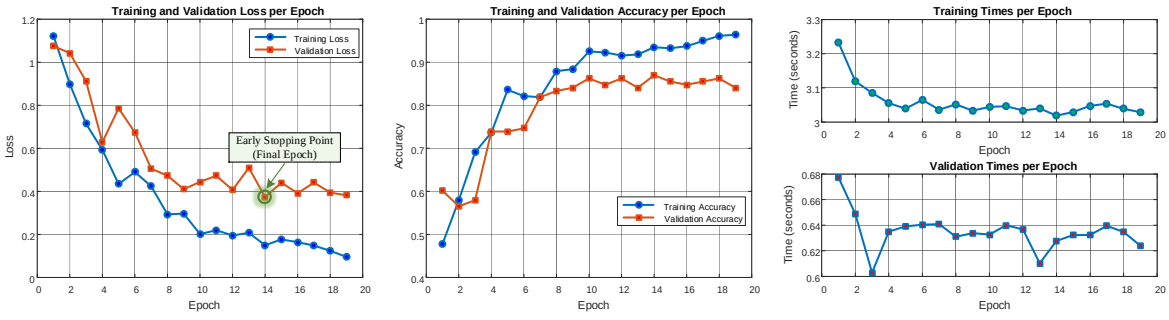
3.2. Results

This section presents the experimental results of transformer-based models for lung cancer classification. The evaluation includes accuracy, precision, recall, F1-score, and computational efficiency. A summary of the model-specific training processes is provided below.

The training loss of the ViT model decreased steadily. It dropped from 1.12 in the first epoch to 0.14 in the final epoch (14th epoch). Training accuracy improved significantly, rising from 47.8% at the beginning to 93.4% in the final epoch. This shows that the model adapted well to the training data, and the learning process was successful. The continuous decrease in training loss correlated with an increase in accuracy. Accuracy, which started at lower levels, increased rapidly as the loss decreased. Validation loss followed a similar trend. It dropped from 1.07 at the beginning to 0.37 in the final epoch. Validation accuracy improved from 60.1% to 86.9%. Training times ranged between 3.03 and 3.23 seconds per epoch, while validation times were approximately 0.60 to 0.68 seconds. The training and validation curves, along with the epoch-based time plot for the ViT model, are shown in Figure 3.

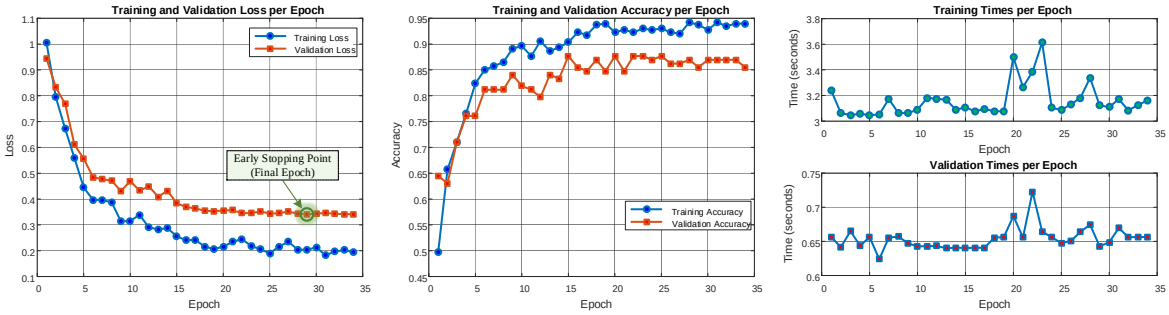
The DeiT model also showed good results. Training loss decreased from 1.00 in the first epoch to 0.20 by the final epoch (29th epoch). Training accuracy rose from 49.8% to 93.7%. The decrease in training loss was in line with the increase in accuracy. Although accuracy was low at first, it improved as the losses decreased. Validation loss showed a similar pattern. It dropped from 0.94 at the beginning to 0.34 by the final epoch. Validation accuracy increased from 64.5% in the first epoch to 85.5% by the end of training. This indicates that the model performed well on unseen data. Training times ranged from 3.05 to 3.61 seconds per epoch, while validation times were approximately 0.64 to 0.72 seconds. The training and validation curves, along with the epoch-based time plot for the DeiT model, are shown in Figure 4.

The Swin Transformer model also demonstrated effective learning. Training loss decreased from 1.02 in the first epoch to 0.38 by the final epoch (21st epoch). Training accuracy improved from 46.9% to 84.7%. Validation loss followed a similar pattern. Validation accuracy increased from 56.5% in the first epoch to 85.5% by the final epoch. Training times averaged between 2.35 and 2.76 seconds per epoch, while validation times were approximately 0.53 to 0.58 seconds. The training and validation curves, along with the epoch-based time plot for the Swin Transformer model, are shown in Figure 5.



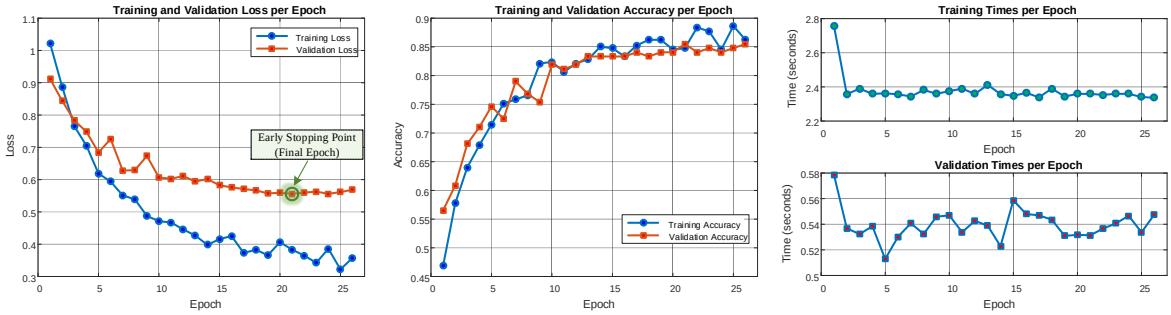
(a) Loss Graph (b) Accuracy Graph (c) Time Graph

Figure 3. The performance of ViT Model: (a) Loss Graph, (b) Accuracy Graph and (c) Time Graph



(a) Loss Graph (b) Accuracy Graph (c) Time Graph

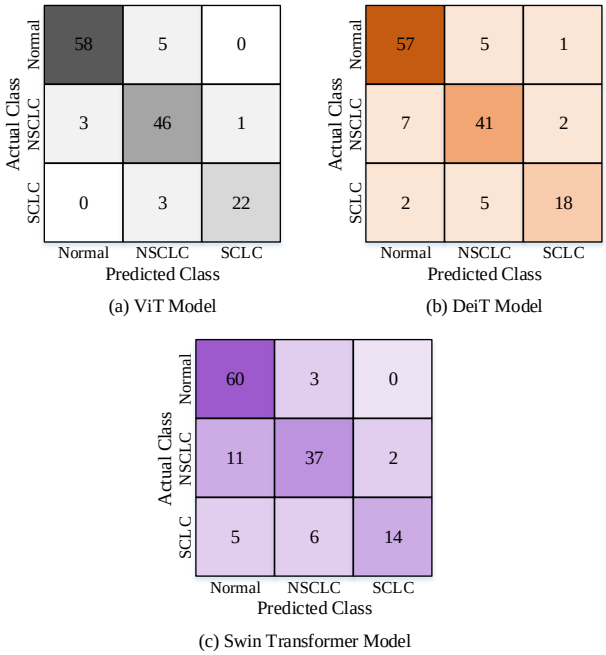
Figure 4. The performance of DeiT Model: (a) Loss Graph, (b) Accuracy Graph and (c) Time Graph



(a) Loss Graph (b) Accuracy Graph (c) Time Graph

Figure 5. The performance of Swin Transformer Model: (a) Loss Graph, (b) Accuracy Graph and (c) Time Graph

The performance of all models during the training and validation processes was generally successful. After completing these processes, each model was evaluated using the test images. The confusion matrices, generated from the predictions of each model on the test samples, are shown in Figure 6.



(a) ViT Model (b) DeiT Model

(c) Swin Transformer Model

Figure 6. Lung cancer classification results: (a) ViT Model, (b) DeiT Model and (c) Swin Transformer Model.

The ViT model achieved a high prediction rate of 92.1% (58/63) in the Normal class. It correctly predicted 92% (46/50) in the NSCLC class and 88% (22/25) in the SCLC class. The model's biggest challenge was misclassifying some Normal samples as NSCLC. Additionally, several SCLC samples were predicted as NSCLC.

The DeiT model performed well in the Normal class with 90.5% (57/63) prediction rate. However, it achieved 82% prediction rate (41/50) in the NSCLC class, with slightly more errors in this category. In the SCLC class, it performed worse, with a prediction rate of 72% (18/25). The NSCLC class was the most difficult for this model, as some samples were misclassified as Normal.

The Swin Transformer model had the highest prediction rate in the Normal class, achieving 95.2% (60/63). However, its performance was lower in the NSCLC class, with 74% prediction rate (37/50), and even lower in the SCLC class, at 56% (14/25). Misclassifications were particularly notable in the NSCLC class, with many samples predicted as Normal. The prediction rate in the SCLC class was also the lowest among the three models.

Table 3 shows in detail the performance of each model on the test samples.

TABLE III
DETAIL THE PERFORMANCE OF EACH MODEL

Model	Class	Pre (%)	Rec (%)	F-1 (%)	Acc (%)
ViT	Normal	95.08%	92.06%	93.55%	91.30%
	NSCLC	85.19%	92.00%	88.46%	
	SCLC	95.65%	88.00%	91.67%	
DeiT	Normal	86.36%	90.48%	88.37%	84.06%
	NSCLC	80.39%	82.00%	81.19%	
	SCLC	85.71%	72.00%	78.26%	
Swin Transformer	Normal	78.95%	95.24%	86.33%	80.43%
	NSCLC	80.43%	74.00%	77.08%	
	SCLC	87.50%	56.00%	68.29%	

4. DISCUSSION

The experimental results highlight the impressive ability of transformer-based models to effectively classify lung cancer from CT images. Among these models, ViT stood out, delivering the highest overall performance, especially in classifying NSCLC and SCLC cases. Its balanced accuracy across all three classes reflects its strong generalization

capability. However, minor misclassifications were observed, particularly between NSCLC and SCLC. The DeiT model, while performing well, showed slightly lower accuracy than ViT, particularly in distinguishing NSCLC from SCLC and normal cases. The Swin Transformer, on the other hand, achieved the best performance in classifying normal cases but faced challenges in accurately differentiating between NSCLC and SCLC.

These differences in performance can be attributed to several factors. From an architectural standpoint, ViT uses global self-attention mechanisms that allow it to capture long-range dependencies across the image, which is particularly effective for capturing complex patterns in medical imaging. DeiT, uses distillation-based training that can slightly limit its capacity in fine-grained classification. Swin Transformer adopts a hierarchical structure with shifted windows, which is more efficient but may lose some global contextual information potentially affecting performance on subtle differences between cancer subtypes.

While the results are promising, several challenges remain. Dataset size and class imbalance continue to impact the models' performance. Transformer-based architectures require larger, more balanced datasets to improve robustness and prevent overfitting. Integrating widely-used public datasets, such as LIDC-IDRI, could further enhance the models' performance.

As for the practical application of these models in clinical settings, we emphasize their potential to significantly reduce the workload of radiologists and assist in early diagnosis. The ViT model, with its strong generalization ability, offers a promising path toward faster, more accurate diagnoses. By automating the classification process, these models not only reduce the time and effort needed for manual analysis but also enable earlier detection, ultimately facilitating quicker treatment decisions. However, while these models show strong performance, their computational demands present a challenge for seamless integration into clinical workflows. The memory and processing power required by transformer-based models, particularly ViT, could limit their real-time application. Moreover, the need for thorough validation, regulatory approval, and interpretability further complicates clinical integration. Despite this, the continued advancement of hardware capabilities and optimization techniques such as model pruning, quantization, and distributed computing suggests that these models can be adapted for practical use in clinical environments. In future work, we aim to explore various model optimization strategies to improve inference times and reduce memory consumption, making these transformer-based models even more suitable for integration into clinical practice. Furthermore, explainability methods will be investigated to ensure that predictions are interpretable and clinically trustworthy.

5. CONCLUSION

This study demonstrated the potential of transformer-based models, specifically ViT, DeiT, and Swin Transformer, for lung cancer classification from CT images. These models were evaluated on a private dataset with images of NSCLC, SCLC, and normal lung. Their performance was compared using accuracy, precision, recall, and F1-score. The results showed that all three models were effective in detecting lung cancer. Each model excelled in different aspects. The ViT model achieved the highest overall accuracy. It showed strong performance across all categories, particularly in the Normal and NSCLC classes. The DeiT model also performed well, but it struggled more in the SCLC class. The Swin Transformer achieved the highest prediction rate for the Normal class. However, it had weaker results in distinguishing between NSCLC and SCLC. Despite their success, there is room for improvement, particularly in increasing the classification accuracy for more aggressive cancer types like SCLC. Future research could focus on improving the generalization of these models by incorporating larger and more diverse datasets. Additionally, hybrid approaches that combine the strengths of CNNs and transformers could be explored to enhance performance further.

ACKNOWLEDGEMENT

This study was supported by Scientific Research Projects Unit of Firat University (FUBAP) under the Grant Number TEKF.24.42. The authors thank to FUBAP for their supports

REFERENCES

[1] A. Leiter, R. R. Veluswamy, and J. P. Wisnivesky, "The global burden of lung cancer: current status and future trends," *Nat. Rev. Clin. Oncol.*, vol. 20, no. 9, pp. 624–639, Sep. 2023, doi: 10.1038/s41571-023-00798-3.

[2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA. Cancer J. Clin., vol. 68, no. 6, pp. 394–424, Nov. 2018, doi: 10.3322/caac.21492.

[3] Y. Fang et al., "Burden of lung cancer along with attributable risk factors in China from 1990 to 2019, and projections until 2030," J. Cancer Res. Clin. Oncol., vol. 149, no. 7, pp. 3209–3218, Jul. 2023, doi: 10.1007/s00432-022-04217-5.

[4] M. Kriegsmann et al., "Deep Learning for the Classification of Small-Cell and Non-Small-Cell Lung Cancer," Cancers (Basel)., vol. 12, no. 6, p. 1604, Jun. 2020, doi: 10.3390/cancers12061604.

[5] L. E. L. Hendriks et al., "Non-small-cell lung cancer," Nat. Rev. Dis. Prim., vol. 10, no. 1, p. 71, Sep. 2024, doi: 10.1038/s41572-024-00551-9.

[6] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, "Cancer statistics, 2022," CA. Cancer J. Clin., vol. 72, no. 1, pp. 7–33, Jan. 2022, doi: 10.3322/caac.21708.

[7] S. J. Adams, E. Stone, D. R. Baldwin, R. Vliegenthart, P. Lee, and F. J. Fintelmann, "Lung cancer screening," Lancet, vol. 401, no. 10374, pp. 390–408, Feb. 2023, doi: 10.1016/S0140-6736(22)01694-4.

[8] R. Nooreldeen and H. Bach, "Current and Future Development in Lung Cancer Diagnosis," Int. J. Mol. Sci., vol. 22, no. 16, p. 8661, Aug. 2021, doi: 10.3390/ijms22168661.

[9] J. Kim and K. H. Kim, "Role of chest radiographs in early lung cancer detection," Transl. Lung Cancer Res., vol. 9, no. 3, pp. 522–531, Jun. 2020, doi: 10.21037/tlcr.2020.04.02.

[10] B. Philip et al., "Current investigative modalities for detecting and staging lung cancers: a comprehensive summary," Indian J. Thorac. Cardiovasc. Surg., vol. 39, no. 1, pp. 42–52, Jan. 2023, doi: 10.1007/s12055-022-01430-2.

[11] H. L. Lancaster, M. A. Heuvelmans, and M. Oudkerk, "Low-dose computed tomography lung cancer screening: Clinical evidence and implementation research," J. Intern. Med., vol. 292, no. 1, pp. 68–80, Jul. 2022, doi: 10.1111/joim.13480.

[12] M. A. Thanoon, M. A. Zulkifley, M. A. A. Mohd Zainuri, and S. R. Abdani, "A Review of Deep Learning Techniques for Lung Cancer Screening and Diagnosis Based on CT Images," Diagnostics, vol. 13, no. 16, p. 2617, Aug. 2023, doi: 10.3390/diagnostics13162617.

[13] E. W. Zhang et al., "Characteristics and Outcomes of Lung Cancers Detected on Low-Dose Lung Cancer Screening CT," Cancer Epidemiol. Biomarkers Prev., vol. 30, no. 8, pp. 1472–1479, Aug. 2021, doi: 10.1158/1055-9965.EPI-20-1847.

[14] K.-L. Huang, S.-Y. Wang, W.-C. Lu, Y.-H. Chang, J. Su, and Y.-T. Lu, "Effects of low-dose computed tomography on lung cancer screening: a systematic review, meta-analysis, and trial sequential analysis," BMC

Pulm. Med., vol. 19, no. 1, p. 126, Dec. 2019, doi: 10.1186/s12890-019-0883-x.

[15] E. F. Patz, E. Greco, C. Gatsonis, P. Pinsky, B. S. Kramer, and D. R. Aberle, "Lung cancer incidence and mortality in National Lung Screening Trial participants who underwent low-dose CT prevalence screening: a retrospective cohort analysis of a randomised, multicentre, diagnostic screening trial," Lancet Oncol., vol. 17, no. 5, pp. 590–599, May 2016, doi: 10.1016/S1470-2045(15)00621-X.

[16] S. J. van Riel et al., "Observer variability for Lung-RADS categorisation of lung cancer screening CTs: impact on patient management," Eur. Radiol., vol. 29, no. 2, pp. 924–931, Feb. 2019, doi: 10.1007/s00330-018-5599-4.

[17] S. H. Hosseini, R. Monsefi, and S. Shadroo, "Deep learning applications for lung cancer diagnosis: A systematic review," Multimed. Tools Appl., vol. 83, no. 5, pp. 14305–14335, Jul. 2023, doi: 10.1007/s11042-023-16046-w.

[18] R. Javed, T. Abbas, A. H. Khan, A. Daud, A. Bukhari, and R. Alharbey, "Deep learning for lungs cancer detection: a review," Artif. Intell. Rev., vol. 57, no. 8, p. 197, Jul. 2024, doi: 10.1007/s10462-024-10807-1.

[19] J. V. Naga Ramesh, R. Agarwal, P. Deekshita, S. A. Elahi, S. H. Surya Bindu, and J. S. Pavani, "Application of Several Transfer Learning Approach for Early Classification of Lung Cancer," EAI Endorsed Trans. Pervasive Heal. Technol., vol. 10, Mar. 2024, doi: 10.4108/eetpht.10.5434.

[20] H. Li, Q. Song, D. Gui, M. Wang, X. Min, and A. Li, "Reconstruction-Assisted Feature Encoding Network for Histologic Subtype Classification of Non-Small Cell Lung Cancer," IEEE J. Biomed. Heal. Informatics, vol. 26, no. 9, pp. 4563–4574, Sep. 2022, doi: 10.1109/JBHI.2022.3192010.

[21] S. Pang, Y. Zhang, M. Ding, X. Wang, and X. Xie, "A Deep Model for Lung Cancer Type Identification by Densely Connected Convolutional Networks and Adaptive Boosting," IEEE Access, vol. 8, pp. 4799–4805, 2020, doi: 10.1109/ACCESS.2019.2962862.

[22] Y. Han et al., "Histologic subtype classification of non-small cell lung cancer using PET/CT images," Eur. J. Nucl. Med. Mol. Imaging, vol. 48, no. 2, pp. 350–360, Feb. 2021, doi: 10.1007/s00259-020-04771-5.

[23] T. L. Chaunzwa et al., "Deep learning classification of lung cancer histology using CT images," Sci. Rep., vol. 11, no. 1, p. 5471, Mar. 2021, doi: 10.1038/s41598-021-84630-x.

[24] A. Fanizzi et al., "Comparison between vision transformers and convolutional neural networks to predict non-small lung cancer recurrence," Sci. Rep., vol. 13, no. 1, p. 20605, 2023, doi: 10.1038/s41598-023-48004-9

[25] C. Venkatesh et al., "A hybrid model for lung cancer prediction using patch processing and deeplearning on CT images," Multimedia Tools and Applications, vol. 83, no. 15, p. 43931-43952, 2024, doi: 10.1007/s11042-024-19208-6

[26] R. Azad et al., "Advances in medical image analysis with vision Transformers: A comprehensive review," Med. Image Anal., vol. 91, p. 103000, Jan. 2024, doi: 10.1016/j.media.2023.103000.

[27] A. Parvaiz, M. A. Khalid, R. Zafar, H. Ameer, M. Ali, and M. M. Fraz, "Vision Transformers in medical computer vision—A contemplative retrospection," Eng. Appl. Artif. Intell., vol. 122, p. 106126, Jun. 2023, doi: 10.1016/j.engappai.2023.106126.

[28] A. M. Hafiz, S. A. Parah, and R. U. A. Bhat, "Attention mechanisms and deep learning for machine vision: A survey of the state of the art," Jun. 2021.

[29] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," Oct. 2020.

[30] O. Katar and O. Yildirim, "An Explainable Vision Transformer Model Based White Blood Cells Classification and Localization," Diagnostics, vol. 13, no. 14, p. 2459, Jul. 2023, doi: 10.3390/diagnostics13142459.

[31] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jegou, "Training data-efficient image transformers & distillation through attention," in Proceedings of the 38th International Conference on Machine Learning, M. Meila and T. Zhang, Eds., in Proceedings of Machine Learning Research, vol. 139. PMLR, Oct. 2021, pp. 10347–10357. [Online]. Available: https://proceedings.mlr.press/v139/touvron21a.html

[32] G. Habib, T. J. Saleem, and B. Lall, "Knowledge Distillation in Vision Transformers: A Critical Review," Feb. 2023.

[33] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 10012–10022.

[34] J. Huang et al., "Swin transformer for fast MRI," Neurocomputing, vol. 493, pp. 281–304, Jul. 2022, doi: 10.1016/j.neucom.2022.04.051.

BIOGRAPHIES

Oguzhan Katar was born in Elazığ, Türkiye, in 1997. He received the B.Sc. and M.Sc. degrees in computer engineering from Firat University, Elazığ, in 2019 and 2022, respectively, where he is currently pursuing the Ph.D. degree in software engineering. His research interests include machine learning, deep learning, and image processing.

Tulin Ozturk graduated from Firat University Faculty of Medicine. She received radiology training at Firat University Radiology Clinic. She currently works as an associate professor at the Radiology clinic of the University of Health Sciences. She has published over 30 papers in journals. She main research interests include abdominal radiology and oncological radiology.

Ozal Yildirim received the Ph.D. degree in electrical and electronic engineering from Firat University, Türkiye. He is currently an Associate Professor in software engineering with Firat University. He has published over 60 papers in international refereed journals and conference proceedings. His main research interests include deep learning and medical signal and image processing.