Available online at http://scik.org

Commun. Math. Biol. Neurosci. 2025, 2025:136

https://doi.org/10.28919/cmbn/9602

ISSN: 2052-2541

LUNG CANCER CLASSIFICATION USING VISION TRANSFORMER: A CRISP-DM APPROACH WITH HISTOPATHOLOGICAL IMAGING

WAHYUDI SETIAWAN[1,*], WAHYUDI AGUSTIONO[1], YOGA DWITYA PRAMUDITA[2]

[1]Department of Information System, University of Trunojoyo Madura, Bangkalan 69162, Indonesia

[2]Department of Informatics, University of Trunojoyo Madura, Bangkalan 69162, Indonesia

Abstract: Lung cancer classification based on histopathological imaging plays a pivotal role in achieving early detection, accurate diagnosis, and effective treatment planning. Conventional diagnostic methods, including manual examination of histopathological slides and radiological imaging, are often subjective and time-consuming. Their limited ability to capture complex morphological patterns further constrains diagnostic accuracy. In response to these limitations, the present study employs the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework to systematically evaluate the application of the Vision Transformer (ViT) in lung cancer classification. The LC25000 dataset, comprising three histopathological categories: adenocarcinoma, squamous cell carcinoma, and benign lung tissue. It was utilized for model evaluation. All images were resized to 224 × 224 pixels, and data augmentation techniques were applied to enhance generalization capability. The ViT model was implemented using TensorFlow and trained with the Adam optimizer (learning rate = 0.0001, batch size = 16, epochs = 50), employing early stopping and learning rate scheduling to mitigate overfitting. The proposed model achieved an overall accuracy of 0.97, with precision, recall, and F1-scores consistently exceeding 0.97. Class-level analysis demonstrated exceptional performance in identifying benign tissue (precision = 0.999, recall = 1.000, F1 = 0.999) and robust classification of malignant subtypes, including adenocarcinoma (F1 = 0.957) and squamous cell carcinoma (F1 = 0.959). These results emphasize the ViT's strong capability in capturing global contextual features, surpassing conventional CNN-based methods that primarily rely on local feature extraction.

Keywords: lung cancer; vision transformer; histopathological image classification; deep learning; CRISP-DM.

2020 AMS Subject Classification: 68T07, 92C55.

*Corresponding author

E-mail address: wsetiawan@trunojoyo.ac.id

Received September 15, 2025

1. INTRODUCTION

Lung cancer remains one of the most pressing global health concerns, ranking as the second most frequently diagnosed cancer and the leading cause of cancer-related mortality worldwide. According to the World Health Organization (WHO), lung cancer accounted for nearly 1.8 million deaths in 2020, representing a significant burden on healthcare systems [1], [2]. In Indonesia, it ranks third in prevalence, with 34,783 reported cases (8.8% of all cancers), contributing to approximately 13.2% of cancer-related deaths [3], [4]. These data emphasize the need for reliable and affordable diagnostic methods to help improve clinical management and patient prognosis.

Traditional diagnostic techniques such as manual histopathological slide examination, chest radiography, and computed tomography (CT) scans, remain fundamental in clinical practice. However, these methods suffer from limitations including inter-observer variability, difficulty in detecting subtle morphological differences, and lengthy interpretation times [5], [6]. The rapid expansion of digital pathology and imaging data drives the need for computational methods that improve diagnostic accuracy and consistency.

Deep learning (DL) has revolutionized medical image analysis by enabling automated feature extraction and high-level pattern recognition. Convolutional Neural Networks (CNNs) have been widely used for histopathological image classification, achieving remarkable performance in various cancer detection tasks [7], [8]. Despite these successes, CNNs primarily rely on local receptive fields, which limits their ability to capture global spatial dependencies. This shortcoming can hinder their effectiveness when distinguishing between histologically similar tissue types, such as adenocarcinoma and squamous cell carcinoma [9], [10].

Recently, transformer-based architectures have emerged as a paradigm shift in computer vision. The Vision Transformer (ViT), in particular, leverages self-attention mechanisms to model long-range dependencies and capture global contextual information in images. Unlike CNNs, ViT processes image patches as sequences, allowing it to learn spatial relationships across the entire image [11], [12]. Several studies have demonstrated ViT's superiority in diverse visual recognition tasks, including medical imaging. However, its systematic application to lung cancer histopathology remains limited [12].

Recognizing this research gap, the present study explores the capability of the Vision Transformer (ViT) architecture in classifying lung adenocarcinoma, squamous cell carcinoma, and benign lung tissues using the LC25000 histopathological dataset [13]. The study adopts the Cross-Industry

Standard Process for Data Mining (CRISP-DM) framework to ensure methodological rigor and transparency. This approach enhances reproducibility across all analytical stages, from business understanding to deployment. The primary contributions of this research are summarized as follows:

- Demonstration of ViT applicability for histopathological lung cancer image classification.
- Implementation of a reproducible CRISP-DM workflow specifically adapted for medical image analysis.
- Establishment of benchmark results showcasing ViT's superiority over conventional CNN-based methods in capturing global contextual features.

Table 1 presents a comparative summary of recent studies focusing on lung cancer image classification using deep learning approaches, particularly the ViT architectures and their variants. These studies reflect a growing shift from conventional convolutional neural networks (CNNs) toward transformer-based models, which demonstrate superior capability in capturing global contextual dependencies within medical images. The comparison includes the types of datasets utilized (ranging from CT scans and histopathology slides to chest X-ray images), the applied methods, and key evaluation metrics such as accuracy, AUC, precision, and F1-score. Overall, the results indicate that transformer-based approaches consistently outperform traditional CNN models in lung cancer image classification tasks.

Table 1. Comparison of Related Studies on Lung Cancer Image Classification

Ref.	Dataset	Method Applied	Accuracy/AUC
[14]	TCGA (Histopathology), LIDC-IDRI (CT), ImageNet (pretraining)	Vision Transformer (ViT), Swin Transformer, TransUNet, ConvNeXt	AUC ≈ 0.98, Accuracy > 95 % (Transformer) vs. 90 – 93 % (CNN)
[15]	CXR Images (19,003 images, 7 classes)	Vision Transformer (ViT-B/16, FastViT, CrossViT) + Optimizer Comparison (Adam, AdamW, NAdam, RAdam, SGDW)	FastViT + NAdam = 97.63 % accuracy, F1 = 97.64 %; ViT + RAdam = 95.87 %
[16]	LIDC-IDRI (CTScans, 3D nodules)	3D CMixNet + Gradient Boosting Machine (+ Clinical Biomarkers)	Accuracy = 94.17 %, AUC ≈ 1.0 (proposed); without biomarker = 91.13 %
[17]	LUNA16 & MSD (CT Scans 3D to 2D slices)	Swin Transformer (T, S, B224, B384)	Classification: 97.3 % Top-1 Acc; Segmentation: mIoU = 0.942
[18]	LC25000 (Histopathology Lung & Colon)	Hybrid Vision Transformer + Deformable CNN (ViT-DCNN) + Feature Fusion + SE Block	Accuracy = 94.24 %, Precision = 94.37 %, Recall = 94.24 %, F1 = 94.23

As shown in Table 1, transformer-based architectures demonstrate remarkable performance improvements across various imaging modalities and datasets. Studies employing ViT variants, such as Swin Transformer, FastViT, and hybrid models integrating CNN components, consistently report higher classification accuracy and robustness compared to conventional CNN models. These findings suggest that the self-attention mechanism within transformers enables more effective feature extraction from complex lung cancer images, capturing both local and global spatial relationships. Furthermore, the integration of multimodal data, optimization techniques, and hybrid feature fusion further enhances diagnostic precision. This trend highlights the growing potential of transformer-based models as a dominant paradigm in medical image analysis, particularly for early detection and classification of lung cancer.

2. METHOD

This study employs the Cross-Industry Standard Process for Data Mining (CRISP-DM) as the overarching methodological framework to ensure a structured, systematic, and reproducible crispresearch workflow. The CRISP-DM framework was selected for its well-established, domain-independent structure and proven adaptability to medical image analysis. It consists of six interrelated phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment [19], [20], [21]. These phases collectively guide the analytical process from problem formulation to model implementation. Each phase is described in the following subsections, and the overall research methodology is illustrated in Figure 1.

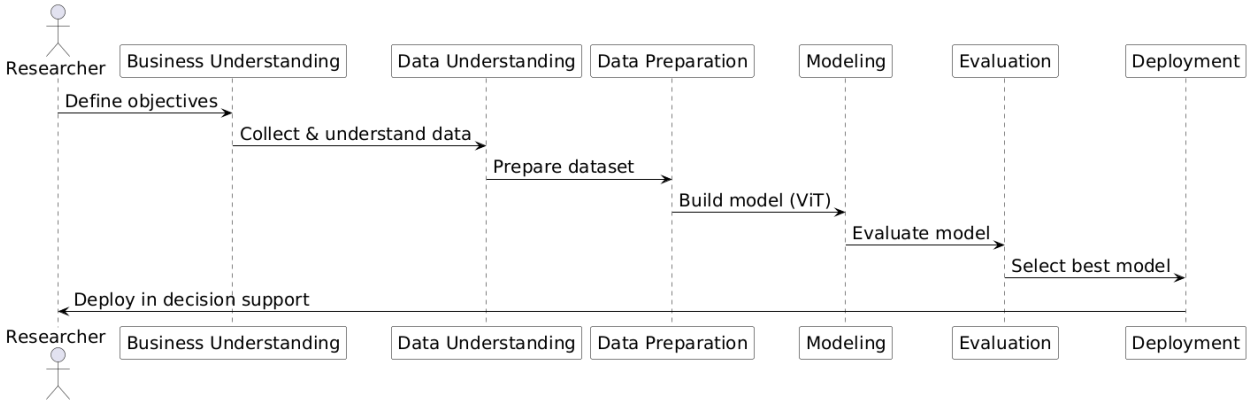

Figure 1. The methology research

2.1 Business Understanding

The primary objective of this study is to develop an accurate and efficient computer-aided diagnosis (CAD) system for classifying lung cancer subtypes from histopathological images. By minimizing diagnostic subjectivity and reducing interpretation time, the proposed system is expected to support pathologists in achieving more consistent and reliable diagnostic outcomes. Ultimately, this approach aims to enhance clinical decision-making and contribute to the advancement of precision medicine in oncology.

2.2 Data Understanding

The dataset utilized in this study is the LC25000, a publicly available collection of histopathological lung tissue images. It contains a total of 15,000 samples distributed across three categories: lung adenocarcinoma (5,000 images), lung squamous cell carcinoma (5,000 images), and benign lung tissue (5,000 images). Although the dataset offers sufficient variability to support robust model training and evaluation, all images originate from a single institutional source. This limitation may influence the model's generalizability and restrict its applicability in diverse clinical settings.

2.3 Data Preparation

All images were resized to 224 × 224 pixels to comply with the input requirements of the Vision Transformer (ViT) architecture. The dataset was subsequently divided into training (80%) and testing (20%) subsets to facilitate unbiased model evaluation. A series of data augmentation techniques, including random rotation, horizontal and vertical flipping, zooming, and contrast adjustment, was applied to increase sample diversity and mitigate the risk of overfitting. These transformations simulate real-world histopathological variations, thereby enhancing the model's robustness and generalization capability.

2.4 Modeling

The Vision Transformer (ViT-B/16) architecture was implemented using the TensorFlow deep learning framework. In this configuration, each input image was partitioned into non-overlapping 16 × 16 patches, which were subsequently flattened and linearly embedded into vector representations. These patch embeddings were then processed through a Transformer Encoder comprising multi-head self-attention and feed-forward layers to capture global contextual relationships within the image. A classification head was appended to generate predictions across the three target classes. The complete experimental setup, including the hyperparameter

configuration, is summarized in Table 2.

Table 2. Experimental Setup and Hyperparameter Configuration

Parameter	Value/Method
Architecture	Vision Transformer (ViT-B/16)
Image Input Size	224 × 224 pixels
Patch Size	16 × 16
Optimizer	Adam
Initial Learning Rate	0.0001
Batch Size	16
Epochs	50
Regularization	Early stopping, learning rate scheduling
Data Augmentation	Rotation, flipping, zooming, contrast adjust.

2.5 Evaluation

The model's performance was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. In addition, a confusion matrix was generated to analyze class-wise prediction results and identify potential misclassifications among lung adenocarcinoma, lung squamous cell carcinoma, and benign lung tissue samples.

2.6 Deployment

During the deployment phase, the feasibility of integrating the trained ViT model into Clinical Decision Support Systems (CDSS) was evaluated. The model training and validation were performed using Google Colab Pro, equipped with an NVIDIA Tesla T4 GPU (16 GB RAM), which enabled efficient convergence within approximately one hour over 50 epochs.

3. MAIN RESULTS

3.1 Training and Validation Performance

The Vision Transformer (ViT-B/16) model was trained with a learning rate of 0.0001, a batch size of 16, and for 50 epochs. The training process demonstrated stable convergence throughout the optimization phase. The training accuracy increased steadily from 82% in the initial epoch to approximately 97% in the final epoch, while the validation accuracy followed a similar upward trend and marginally exceeded the training performance in several epochs. Both the training and validation loss curves showed consistent declines, stabilizing at approximately 0.08 and 0.07,

respectively. These results indicate the absence of overfitting and confirm that the model achieved effective generalization. Figure 2 presents the training and validation accuracy alongside the corresponding loss trajectories.

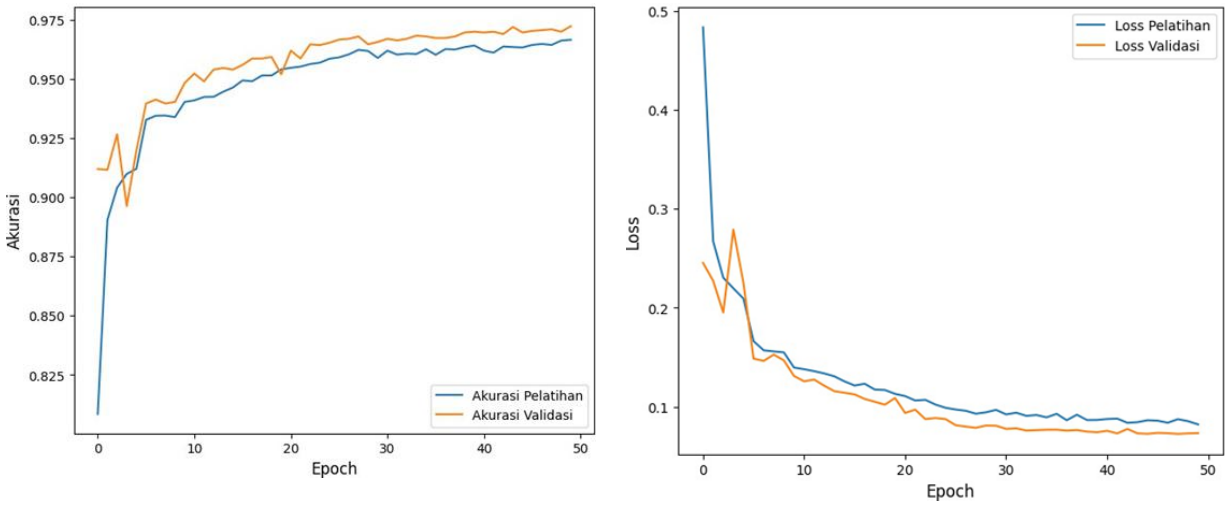

Figure 2. Training and validation of accuracy (left) and loss (right)

3.2 Classification Performance

Table 3 presents the performance metrics of the proposed ViT model across the three lung tissue classes.

Table 3. Performance Metrics of the ViT Model

Class	Precision	Recall	F1-score
Adenocarcinoma	0.980	0.936	0.957
Squamous Cell Carcinoma	0.939	0.981	0.959
Benign Tissue	0.999	1.000	0.999
Average (Macro)	0.973	0.972	0.972

The proposed model achieved an overall accuracy of 97%, with macro-averaged precision, recall, and F1-score values consistently exceeding 0.97. These findings highlight the strong discriminative ability of the ViT architecture in accurately distinguishing between malignant and benign lung tissue samples.

3.3 Confusion Matrix Analysis

The confusion matrix shown in Figure 3 illustrates the class-specific prediction outcomes and

associated misclassifications. Most classification errors occurred between adenocarcinoma and squamous cell carcinoma, which is consistent with their well-documented morphological similarities in histopathological images. In contrast, benign lung tissue was identified with near-perfect accuracy, underscoring the robustness of the ViT model in effectively discriminating between malignant and non-malignant samples.

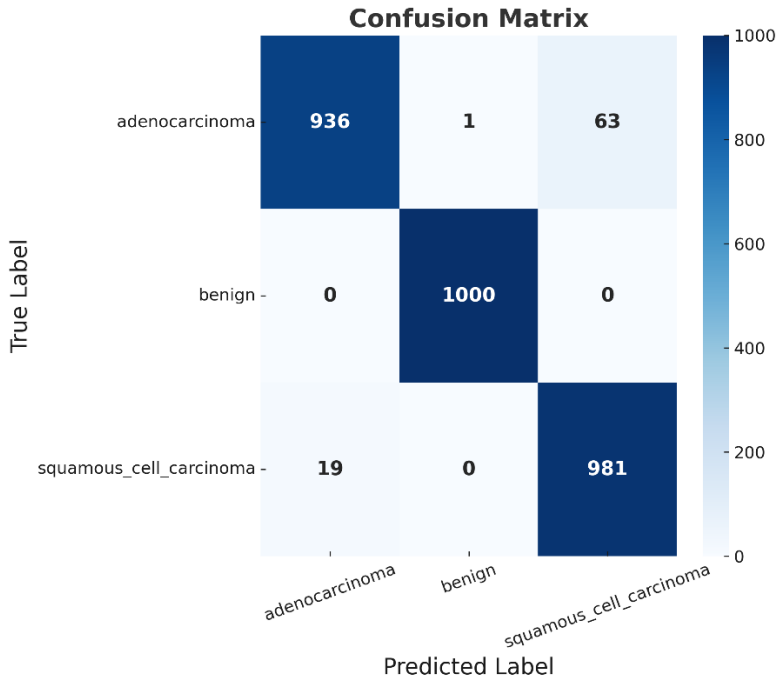

Figure 3. Confusion matrix of the ViT model on the LC25000 dataset

The confusion matrix presented in Figure 3 provides a comprehensive overview of the model's classification performance across three lung tissue categories: adenocarcinoma, benign, and squamous cell carcinoma. The model achieved high predictive accuracy, correctly identifying 936 adenocarcinoma, 1,000 benign, and 981 squamous cell carcinoma samples. Only a small number of misclassifications were observed, 63 adenocarcinoma samples were incorrectly predicted as squamous cell carcinoma, while 19 squamous cell carcinoma samples were misclassified as adenocarcinoma. Remarkably, only one benign case was erroneously classified as adenocarcinoma, demonstrating the model's strong discriminative capability in distinguishing benign from malignant tissues. Overall, the confusion matrix underscores the robustness and reliability of the Vision Transformer (ViT) model in differentiating cancerous from non-cancerous

histopathological images with minimal error.

Figure 4. Representative histopathological image samples illustrating both correct and incorrect predictions across three classes: adenocarcinoma, squamous cell carcinoma, and benign tissues. This figure presents several examples of histopathological image patches used in the classification experiment. Each sub-image is annotated with the true and predicted labels, along with an indication of correctness ("correct") or misclassification ("wrong"). The figure is organized into multiple rows to provide a balanced visual representation of all classes.

The visual examples demonstrate the model's capability in differentiating malignant from non-malignant lung tissues. Correctly classified instances predominate across all categories, suggesting that the model effectively learns and recognizes morphological patterns characteristic of each tissue type. For instance, benign tissue samples (second and third rows) display clear alveolar structures and uniform nuclei, which are accurately identified by the model.

In contrast, adenocarcinoma and squamous cell carcinoma images exhibit more complex morphological patterns characterized by higher nuclear atypia and increased cellular density. The model demonstrates robust feature extraction for these malignant tissues, as evidenced by the high number of correct predictions. Nevertheless, a few misclassifications are observed (highlighted with red labels), including instances where squamous cell carcinoma was incorrectly predicted as adenocarcinoma. These errors likely stem from overlapping histopathological features between regions of adenocarcinoma and squamous differentiation, particularly in transitional or poorly differentiated tumor areas. Such morphological overlap is a well-documented challenge even for expert pathologists and underscores the need for further refinement of feature extraction strategies, potentially through the incorporation of multi-scale attention mechanisms within the model.

The overall visual analysis confirms that the model not only attains strong quantitative performance, as evidenced by the confusion matrix, but also demonstrates consistent spatial and morphological recognition at the image-patch level. The few misclassified instances provide meaningful insights into regions exhibiting inter-class morphological ambiguity. Addressing these challenging cases through targeted data augmentation, fine-grained feature extraction, or the incorporation of hybrid CNN–Transformer architectures could further improve the diagnostic robustness and reliability of automated lung cancer classification systems.

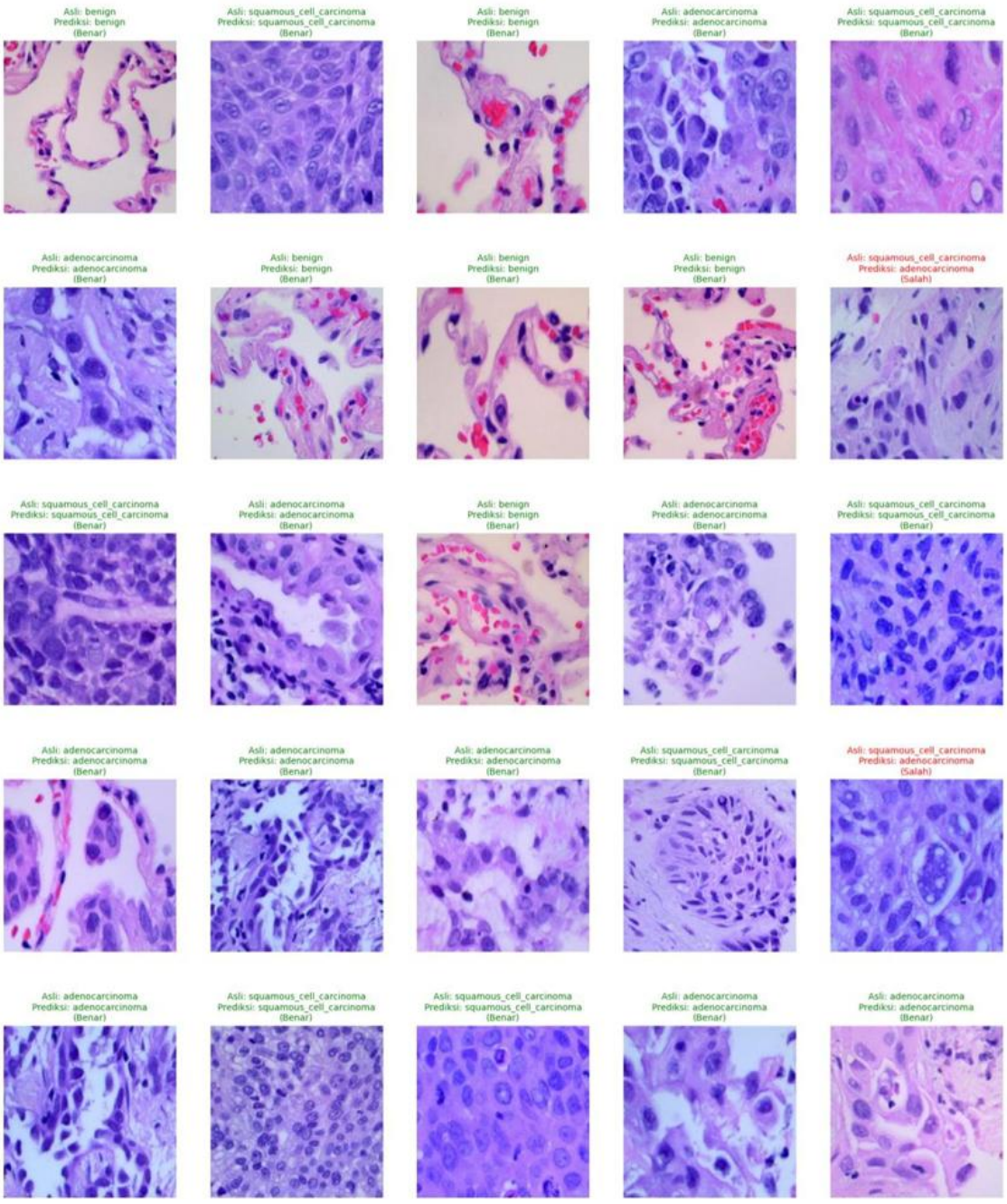

Figure 4. Correct recognition results (green text) and incorrect recognition results (red text).

3.4 Comparison with Related Studies

The performance of the proposed ViT model was compared with previous CNN-based approaches (Table 4).

Table 4. Comparison with Related Studies on Lung Cancer Classification

Reference	Dataset	Method	Accuracy / AUC
[22]	TCGA	Inception-v3 (CNN)	97% AUC
[23]	LC25000 Binary class	CNN	96-100% Accuracy
[24]	LC25000	CNN	87.16% accuracy
[25]	LC25000	Inception-V3	93% accuracy
Proposed	LC25000	Vision Transformer	97% Accuracy

Table 4 presents a comparative analysis of the proposed model's performance against previously published studies on lung cancer classification. As shown, [22] achieved a high diagnostic performance using the TCGA dataset and the Inception-v3 architecture, reporting an AUC of 97%. [23] utilized the LC25000 dataset for binary classification tasks and obtained accuracy values ranging from 96% to 100% using various convolutional neural network (CNN) models. In contrast, studies [24] and [25] employed CNN and Inception-v3 architectures on the same LC25000 dataset, achieving lower accuracies of 87.16% and 93%, respectively.

Compared with these prior works, the proposed ViT model attained an accuracy of 97%, demonstrating competitive or superior performance relative to conventional CNN-based approaches. This improvement highlights the ViT model's enhanced capability to capture global contextual and morphological features from histopathological images, thereby improving classification reliability in automated lung cancer diagnosis.

The comparative analysis indicates that, although conventional Convolutional Neural Networks (CNNs) deliver competitive performance in histopathological image classification, their inherent focus on local feature extraction constrains their ability to distinguish between morphologically similar cancer subtypes. In contrast, the ViT, leveraging its self-attention mechanism, effectively captures global contextual dependencies across the entire image. This capability enables the ViT to surpass CNN-based architectures, confirming its superior suitability for complex histopathological image analysis tasks.

3.5 Clinical Implications

The ViT model demonstrates a consistent capability to differentiate benign from malignant tissues, representing a notable advancement in computational pathology. This capability holds promise for reducing unnecessary biopsies, accelerating treatment planning, and minimizing diagnostic inconsistencies caused by inter-observer variability. The misclassifications observed between

adenocarcinoma and squamous cell carcinoma highlight the importance of employing larger, multi-institutional datasets and implementing preprocessing techniques such as stain normalization to improve model robustness and generalizability.

In addition, the relatively short training time, which was less than one hour using Google Colab Pro with GPU acceleration, illustrates the practical feasibility of integrating the ViT model into Clinical Decision Support Systems (CDSS). With further optimization, the model could be deployed on cloud-based infrastructures for remote diagnostic applications or adapted for edge computing environments, enabling real-time histopathological analysis within clinical laboratory workflows.

3.6 Discussion Summary

The results of this study demonstrate that the ViT architecture offers a significant advancement in lung cancer histopathology classification. With an overall accuracy of 97% and F1-scores exceeding 0.95 across all classes, the ViT model effectively captures both global and local contextual relationships within histopathological imagery. This performance surpasses conventional CNN-based methods, which often struggle to differentiate between morphologically similar tissue types due to their reliance on localized receptive fields. The ViT's self-attention mechanism enables the modeling of long-range dependencies, thereby improving its capacity to discern subtle morphological variations that are critical in distinguishing adenocarcinoma from squamous cell carcinoma.

The integration of the CRISP-DM framework provided a structured, transparent, and reproducible process encompassing all stages, from data understanding to deployment. This systematic approach enhances the interpretability of each phase, particularly the data preparation and modeling stages, which are critical for ensuring model generalization. Compared with ad hoc deep learning studies, the adoption of the CRISP-DM methodology reinforces methodological rigor and facilitates reproducibility across diverse medical imaging domains.

The confusion matrix and visual inspection of misclassified samples reveal clinically relevant insights. Most classification errors occurred between adenocarcinoma and squamous cell carcinoma, reflecting the intrinsic morphological overlap between these subtypes. This finding aligns with prior studies indicating that inter-class ambiguity remains a persistent challenge, even for expert pathologists. The use of stain normalization, domain adaptation, or hybrid ViT–CNN configurations could further mitigate these issues by improving feature consistency and capturing

fine-grained textural cues. Moreover, incorporating multi-scale attention layers may help the model integrate global and local patterns more effectively.

In terms of computational performance, the ViT achieved rapid convergence within approximately one hour on GPU-supported Google Colab Pro. This efficiency underscores its practicality for clinical deployment, especially when compared with other transformer variants that require extensive computational resources. The feasibility of integrating the trained model into Clinical Decision Support Systems (CDSS) or cloud-based diagnostic platforms demonstrates the scalability of this approach. Such integration could facilitate remote diagnosis, assist in triaging histopathological cases, and enhance decision-making in under-resourced medical settings.

Clinically, the ability of the ViT model to consistently distinguish benign from malignant tissues holds substantial implications. By improving diagnostic accuracy and reducing inter-observer variability, the model can support early cancer detection, streamline treatment planning, and decrease the likelihood of unnecessary invasive procedures. These outcomes align with the goals of precision oncology, where data-driven decision-making complements pathologists' expertise to achieve more individualized patient care.

Despite these promising findings, several limitations must be acknowledged. The LC25000 dataset, while balanced and widely used, originates from a single institution, which may constrain the model's generalizability to diverse clinical environments. Future research should prioritize the inclusion of multi-institutional datasets with heterogeneous staining protocols and imaging conditions. Additionally, the current study focused on three lung tissue categories; extending the analysis to encompass additional subtypes or integrating multimodal inputs such as genomic or radiological data, could yield more comprehensive diagnostic models. The incorporation of Explainable techniques is also essential to enhance interpretability, enabling clinicians to visualize attention maps and validate the model's decision rationale.

Overall, the findings confirm that Vision Transformers represent a transformative step in computational pathology. By combining the structured methodology of CRISP-DM with the representational power of transformer-based architectures, this study establishes a benchmark for future research in automated histopathological classification. Continued exploration in model explainability, dataset diversity, and clinical integration will be pivotal to realizing the full potential of ViT-based systems in precision medicine and oncological diagnostics.

4. CONCLUSION

This study demonstrated the effectiveness of the ViT architecture for lung cancer histopathology image classification using the LC25000 dataset. By employing the CRISP-DM framework, the research established a structured, transparent, and reproducible workflow encompassing all analytical stages. The ViT achieved an overall accuracy of 97%, along with high precision, recall, and F1-scores across adenocarcinoma, squamous cell carcinoma, and benign tissues, confirming its robustness in distinguishing malignant from non-malignant cases.

The key contributions of this study include: (i) validating the applicability of transformer-based architectures for histopathological image analysis, (ii) introducing a reproducible CRISP-DM framework tailored for medical imaging applications, and (iii) establishing benchmark performance that underscores the ViT's superiority in capturing global contextual features compared with conventional CNNs.

Future research should focus on expanding multi-institutional datasets, incorporating advanced preprocessing techniques such as stain normalization, and exploring hybrid ViT–CNN or transfer learning approaches. Furthermore, integrating explainable methods will be essential to enhance interpretability and foster clinical trust. Overall, the findings highlight the ViT's potential as a foundation for next-generation computational pathology and precision oncology systems.

ACKNOWLEDGEMENT

This research was supported by the Fundamental Research Scheme, Ministry of Higher Education, Science, and Technology of the Republic of Indonesia, under Grant Numbers 120/C3/DT.05.00/PL/2025 and B/045/UN46.1/PT.01.03/BIMA/PL/2025.

CONFLICT OF INTERESTS

The authors declare that there is no conflict of interests.

REFERENCES

[1] J. Zhou, Y. Xu, J. Liu, L. Feng, J. Yu, et al., Global Burden of Lung Cancer in 2022 and Projections to 2050: Incidence and Mortality Estimates from GLOBOCAN, Cancer Epidemiol. 93 (2024), 102693. https://doi.org/10.1016/j.canep.2024.102693.

[2] K. Chaitanya Thandra, A. Barsouk, K. Saginala, J. Sukumar Aluru, A. Barsouk, Epidemiology of Lung Cancer, Współczesna Onkol. 25 (2021), 45-52. https://doi.org/10.5114/wo.2021.103829.

[3] O.D. Asmara, E.D. Tenda, G. Singh, C.W. Pitoyo, C.M. Rumende, et al., Lung Cancer in Indonesia, J. Thorac.

Oncol. 18 (2023), 1134-1145. https://doi.org/10.1016/j.jtho.2023.06.010.

[4] GLOBOCAN, The Global Cancer Observatory: Indonesia, 2020. https://gco.iarc.who.int/media/globocan/factsheets/populations/360-indonesia-fact-sheet.pdf.

[5] R. Butter, L.M. Hondelink, L. van Elswijk, J.L. Blaauwgeers, E. Bloemena, et al., The Impact of a Pathologist's Personality on the Interobserver Variability and Diagnostic Accuracy of Predictive PD-L1 Immunohistochemistry in Lung Cancer, Lung Cancer 166 (2022), 143-149. https://doi.org/10.1016/j.lungcan.2022.03.002.

[6] A. Snoeckx, J. Cant, C. Franck, E. Luyckx, K. Carpentier, et al., Lesion Measurement on a Combined "All-in-One" Window for Chest CT: Effect on Intra- and Interobserver Variability, Cancer Imaging 19 (2019), 78. https://doi.org/10.1186/s40644-019-0262-0.

[7] M. Unger, J.N. Kather, Deep Learning in Cancer Genomics and Histopathology, Genome Med. 16 (2024), 44. https://doi.org/10.1186/s13073-024-01315-6.

[8] T. Kim, H. Chang, B. Kim, et al., Deep Learning-Based Diagnosis of Lung Cancer Using a Nationwide Respiratory Cytology Image Set: Improving Accuracy and Inter-Observer Variability, Am. J. Cancer Res. 13 (2023), 5493-5503.

[9] M.K. Faizi, Y. Qiang, Y. Wei, Y. Qiao, J. Zhao, et al., Deep Learning-Based Lung Cancer Classification of CT Images, BMC Cancer 25 (2025), 1056. https://doi.org/10.1186/s12885-025-14320-8.

[10] P. Gu, Y. Zhang, C. Wang, D.Z. Chen, ConvFormer: Combining CNN and Transformer for Medical Image Segmentation, arXiv:2211.08564 (2022). http://arxiv.org/abs/2211.08564v1.

[11] K. Al-hammuri, F. Gebali, A. Kanan, I.T. Chelvan, Vision Transformer Architecture and Applications in Digital Health: A Tutorial and Survey, Vis. Comput. Ind. Biomed. Art 6 (2023), 14. https://doi.org/10.1186/s42492-023-00140-9.

[12] S. Takahashi, Y. Sakaguchi, N. Kouno, K. Takasawa, K. Ishizu, et al., Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review, J. Med. Syst. 48 (2024), 84. https://doi.org/10.1007/s10916-024-02105-8.

[13] A.A. Borkowski, M.M. Bui, L.B. Thomas, C.P. Wilson, L.A. DeLand, et al., Lung and Colon Cancer Histopathological Image Dataset (LC25000), arXiv:1912.12142 (2019). http://arxiv.org/abs/1912.12142v1

[14] J. Li, J. Chen, Y. Tang, C. Wang, B.A. Landman, et al., Transforming Medical Imaging with Transformers? A Comparative Review of Key Properties, Current Progresses, and Future Perspectives, Med. Image Anal. 85 (2023), 102762. https://doi.org/10.1016/j.media.2023.102762.

[15] J. Ko, S. Park, H.G. Woo, Optimization of Vision Transformer-Based Detection of Lung Diseases from Chest X-Ray Images, BMC Med. Inform. Decis. Mak. 24 (2024), 191. https://doi.org/10.1186/s12911-024-02591-3.

[16] N. Nasrullah, J. Sang, M.S. Alam, M. Mateen, B. Cai, et al., Automated Lung Nodule Detection and Classification Using Deep Learning Combined with Multiple Strategies, Sensors 19 (2019), 3722. https://doi.org/10.3390/s19173722.

[17] R. Sun, Y. Pang, W. Li, Efficient Lung Cancer Image Classification and Segmentation Algorithm Based on an Improved Swin Transformer, Electronics 12 (2023), 1024. https://doi.org/10.3390/electronics12041024.

[18] A. Pal, H.M. Rai, J. Yoo, S. Lee, Y. Park, Vit-Dcnn: Vision Transformer with Deformable CNN Model for Lung

and Colon Cancer Detection, Cancers 17 (2025), 3005. https://doi.org/10.3390/cancers17183005.

[19] M. Elkabalawy, A. Al-Sakkaf, E. Mohammed Abdelkader, G. Alfalah, Crisp-Dm-Based Data-Driven Approach for Building Energy Prediction Utilizing Indoor and Environmental Factors, Sustainability 16 (2024), 7249. https://doi.org/10.3390/su16177249.

[20] W.Y. Ayele, Adapting Crisp-Dm for Idea Mining, Int. J. Adv. Comput. Sci. Appl. 11 (2020), 20-32. https://doi.org/10.14569/ijacsa.2020.0110603.

[21] C. Schröer, F. Kruse, J.M. Gómez, A Systematic Literature Review on Applying Crisp-Dm Process Model, Procedia Comput. Sci. 181 (2021), 526-534. https://doi.org/10.1016/j.procs.2021.01.199.

[22] N. Coudray, P.S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, et al., Classification and Mutation Prediction from Non–Small Cell Lung Cancer Histopathology Images Using Deep Learning, Nat. Med. 24 (2018), 1559-1567. https://doi.org/10.1038/s41591-018-0177-5.

[23] S. Garg, S. Garg, 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: 2020 3rd Artificial Intelligence and Cloud Computing Conference, ACM, New York, 2020, pp. 38-45. https://doi.org/10.1145/3442536.3442543.

[24] W. Setiawan, M.M. Suhadi, Husni, Y.D. Pramudita, Histopathology of Lung Cancer Classification Using Convolutional Neural Network with Gamma Correction, Commun. Math. Biol. Neurosci. 2022 (2022), 81. https://doi.org/10.28919/cmbn/7611.

[25] W. Setiawan, M.M. Suhadi, Y.D. Pramudita, Mulaab, Inception-V3 with Reduce Learning Rate for Optimization of Lung Cancer Histopathology Classification, Ingén. Syst. Inf. 29 (2024), 561-570. https://doi.org/10.18280/isi.290217.