INTENSIF, Vol.9 No.2 August 2025
ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24918

Enhancing Vision Transformer Performance with Rotation Based Augmentation for Classifying Images of Colon Cancer Pathology

Received:
23 March 2025
Accepted:
8 June 2025
Published:
12 July 2025

[1*]Rudy Eko Prasetya, [2]M. Arief Soeleman, [3]Farrikh Al Zami, [4]Affandy, [5]Aris Marjuni, [6]Mohammad Iqbal Saryuddin Assaqty
[1-5]Magister Teknologi Informasi, Universitas Dian Nuswantoro Semarang,
[6] School of Computer Science and Engineering, South China University of Technology
E-mail:
[1]rudyekoprasetya@gmail.com ,[2]m.arief.soeleman@dsn.dinus.ac.id , [3]alzami@dsn.dinus.ac.id , [4]affandy@dsn.dinus.ac.id , [5]aris.marjuni@dsn.dinus.ac.id, [6]csiqbal@mail.scut.edu.cn
*Corresponding Author

Abstract— **Background**: In medical imaging, classifying images of colon cancer pathology is still an essential challenge, especially for facilitating early diagnosis and successful intervention. Recently, Vision Transformer (ViT) models have demonstrated great promise for a variety of computer vision tasks, including the classification of medical images. However, the lack of annotated medical datasets and the intrinsic unpredictability of histopathology pictures sometimes restrict their performance. **Objective**: This study aims to enhance the performance of ViT models in colon cancer pathology classification by introducing a targeted data augmentation strategy, with a particular focus on rotation-based augmentation. **Methods**: We proposed a data augmentation pipeline that uses controlled changes to improve the number and diversity of training data. Like Rotation, Flip and Geometry are emphasized to replicate the real-world tissue orientation variations that are frequently seen in colon pathology slides. 10,000 JPEG pictures of colon cancer pathology, each with a resolution of 768 x 768 pixels, are used to train the models. We use models trained with and without the suggested augmentation pipeline to compare ViT performance across accuracy, sensitivity, and specificity in order to assess the impact of augmentation. **Results**: According to study results, rotation-based augmentation enhances ViT performance, achieving up to 99.30% accuracy and 99.50% sensitivity while preserving training times. In real-world pathology settings, where slide orientation varies greatly and can affect categorization consistency, these enhancements are especially pertinent. **Conclusion**: The proposed rotation-centric data augmentation technique enhances the performance of the ViT model in the classification of images showing colon cancer pathology.
Keywords—Vision Transformer; Data Augmentation; Images Classification; Colon Cancer

This is an open access article under the CC BY-SA License.

Corresponding Author:
Rudy Eko Prasetya,
Magister Teknik Informatika,
Universitas Dian Nuswantoro,
Email: rudyekoprasetya@gmail.com
Orchid ID: http://orcid.org/0009-0001-5303-0890

INTENSIF, Vol.9 No.2 August 2025
ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24918

I. INTRODUCTION

Colon cancer is one of the most common types of cancer in the world. According to Global Cancer Statistics (GLOBOCAN), colon cancer is now the second leading cause of cancer deaths worldwide [1]. This cancer usually develops gradually, starting with benign polyps growing on the colon wall, and can develop into malignant tumors if not found early. Recent decades have seen a rise in colon cancer occurrences, highlighting the need for aggressive steps to enhance early detection and prevention. Cancer can be treated more quickly and effectively when detected early, before it spreads to other organs [2]. However, a significant barrier to early detection is that colon cancer frequently exhibits no symptoms in its early stages. Therefore, screening techniques including colonoscopy, sigmoidoscopy, and DNA-based stool tests have been developed to detect pre-cancerous polyps or aberrant cells.

However, low awareness and engagement present hurdles for early diagnosis of this condition. There is a stigma attached to dread of medical procedures because the majority of people with this condition have no family history and only receive rudimentary knowledge [3]. In addition to the expensive expense of the examination, pathology results are delayed [4]. This is because these facilities are not available in every hospital. Furthermore, a lack of understanding of integrated screening programs such as FOBT (*Fecal Occult Blood Test*) [5] impedes illness detection.

To solve these issues, screening methods and AI (Artificial Intelligence) are expected to improve the accuracy of early detection [6]. Several studies have been undertaken to increase the accuracy of early colon cancer detection in order to reduce mortality. Deep learning algorithms in histopathology image processing, for example, can identify cancer cells [7] more accurately [8][9] than the medical procedures mentioned above. Convolutional Neural Network (CNN) is one way for utilizing artificial intelligence (AI), particularly in the field of computer vision. CNN can detect by extracting key features from histopathological images [10]. Through CNN-based transfer learning, such as VGG, previously taught models can distinguish between healthy and malignant tissue [11]. However, a drawback of CNN-based models is that they require a significant amount of resources for both the training period and the computation [12][13], which are directly proportional to the performance of the model that will be developed.

The Vision Transformers (ViT) model uses the transformer architecture to address this training and computational problem [14][15] by dividing the image into small patches [16] that are then processed as a series, similar to the text extraction process in NLP. This approach performs better in medical image segmentation tasks [17] because it allows for more flexibility in capturing intricate spatial relationships between image portions while simultaneously reducing the computational overhead. According to the above-mentioned properties of the ViT architecture,

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24918

this model requires a somewhat big dataset [18][19] for training, whereas pathology datasets are typically restricted [20][21] and frequently imbalanced [22][23], raising the danger of overfitting.

CNN showed good accuracy in colon cancer classification, but it requires a lot of training time and computational resources [24]. Since it needs a sizable dataset, Vision Transformers (ViT) has not yet reached ideal performance while being able to lower processing requirements [25][26]. With the aim of generating a more ideal and computationally efficient classification, this work suggests using data augmentation techniques to enhance ViT's performance in order to get over these drawbacks.

II. RESEARCH METHOD

The Lung and Colon Cancer Histopathological Images dataset from Kaggle (https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images) was used in this study based on the literature study that was done. Microscopic pictures of human tissue obtained from patients with lung and colon cancer make up this dataset. Each picture depicts one of the three primary types of cancer: lung adenocarcinoma, lung squamous cell carcinoma, and colon adenocarcinoma. The researchers examined a sample of 10,000 images from the Colon Cancer Histopathological dataset, which was divided into two classes or labels: Colon Adenocarcinoma (Colon Cancer) and Colon Benign Tissue (Benign Tumor). This dataset's photos are from tissue histology that was stained using the Hematoxylin and Eosin (H&E) staining method, which creates a visible contrast between the cell's cytoplasm and nucleus. Significant color variations in tissue images are produced by this staining, which is helpful for classifying medical images [27], [28]. The figure 1 below shows the distribution of the dataset.

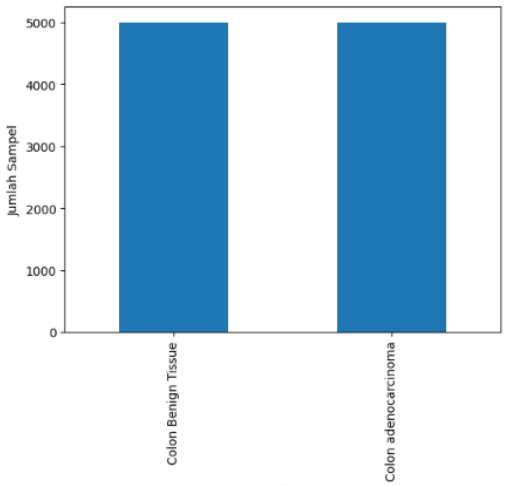

Fig 1. Dataset Distribution

INTENSIF, Vol.9 No.2 August 2025
ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24918

This dataset can be considered balanced since, as can be seen from the above graphic, each class or label has a sample size of 5,000 images.

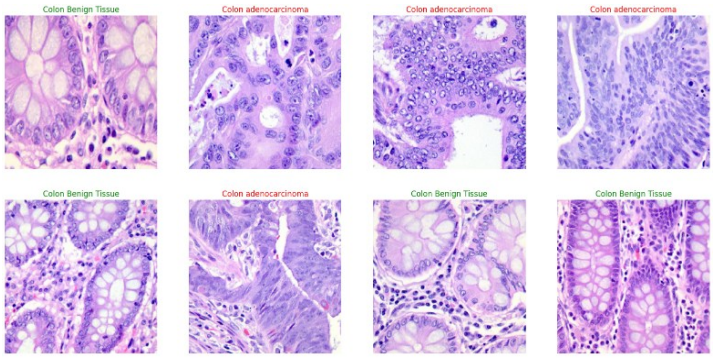

Fig 2. Sample Image Dataset

Then, as shown in Figure 2, the researcher presents eight randomly selected dataset samples with each class or label. For colon benign tissue, the label is green, and for colon adenocarcinoma, it is red. where each image is in the JPG format and has a dimension of 768 x 768 pixels. The study's proposed method, which includes enhancing Vision Transformer (ViT) through data augmentation, is compared with the CNN model in terms of classifying images of colon cancer pathology. The flow of the proposed method is shown in Figure 3 below.

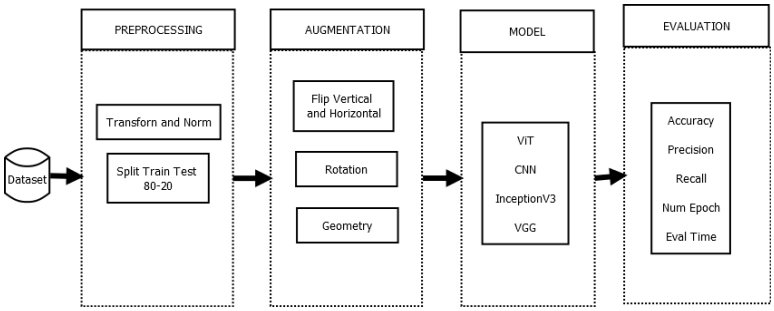

Fig 3. Proposed Method

The Figure 3 above depicts the workflow of picture categorization with Vision Transformer (ViT) used in this study. This is particularly true when it comes to classifying pathology images. To enhance the model's performance and stability, the raw data must first undergo preprocessing, which involves transformation and normalization. After then, the data is split in an 80%-20% ratio between a training set and a test set. The augmentation stage then starts, which includes rotation, flip, and geometry adjustments to reinforce the training data and enhance the model's generalization. The training set is only augmented to increase data diversity during training, whereas the test set is simply enlarged without augmentation to ensure evaluation integrity.

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24918

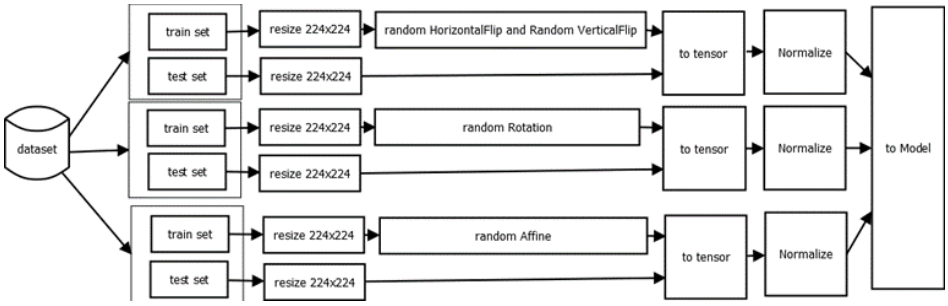

Fig 4. scenarios for data augmentation.

The data preliminary treatment figure 4 above for this study depicts three data augmentation scenarios applied to the dataset. The dataset is partitioned into training and test sets at an 80%-20% ratio and then processed using three separate augmentation methods: Random Horizontal and Vertical Flip, Random Rotation, and Random Affine. The photos for each scenario are first decreased to a standard resolution (224 x 224) to ensure that the input dimensions match the Vision Transformer model. In Flip, vertical and horizontal augmentation is carried out at random with a 50% chance. In principle, when this function is used on a picture, the algorithm will use the probability value p to randomly decide whether or not flipping takes place. If flipping takes place, the pixels in the image's top and bottom rows will switch places, as will the pixels in the remaining rows until they reach the image's center. If not, there's a $1–p = 0.5$ chance that the image will stay in its initial orientation.

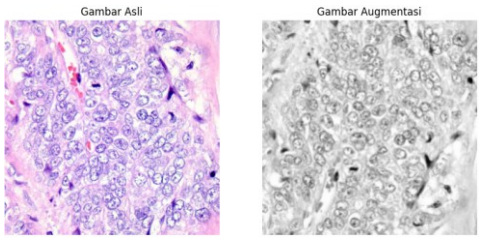

Fig 5. Image Result of Flip Augmentation.

The process of rotation augmentation involves randomly rotating the image at a specific angle. Using the setting *degree=15*, the image will be rotated at random within a range of 15 degrees. In technical terms, the method selects a random rotation angle from the specified range (-15 to +15 degrees) each time this function is applied to an image. After the image has been rotated in the middle, the pixel values in the remaining region outside the original image boundaries can be filled using either a specific color typically black or specific interpolation techniques, like bilinear interpolation. The augmented image result can be seen in figure 6

INTENSIF, Vol.9 No.2 August 2025
ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24918

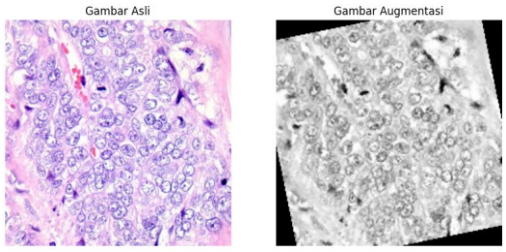

Fig 6. Image Result of Rotation Augmentation

Finally for geometric augmentation. Where it involves processes such as translation, rotation, scaling, and shearing, which ensure that parallel and collinear relationships remain in the image. This process involves three types of transformations. Rotation with Parameter *degrees=0* indicates that no rotation is applied to the image. If this value is greater than zero, the image will be rotated randomly within the specified range. Translation with Parameter *translate=(0.1, 0.1)* allows the image to move randomly up to 10% of the image size along the horizontal (x) and vertical (y) axes. For example, a 224×224 image can be shifted up to 22.4 pixels in the horizontal (right or left) or vertical (up or down) direction. By using this transformation, the model becomes more robust to shifts in object positions. Then scaling is applied with Parameter *Scale=(0.9, 1.1)* allows the image to scale randomly from 90% to 110% of its original size, which means the image can be enlarged or reduced randomly that seen in Figure 7.

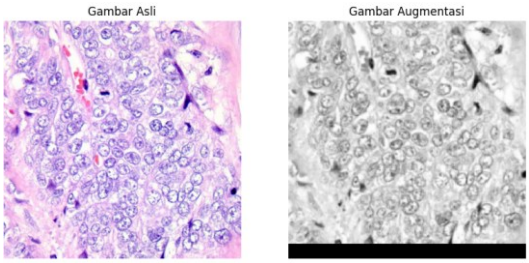

Fig 7. Image Result of Geometry Augmentation.

Next, by converting the data to tensor format, the image may be transformed into a tensor, a multidimensional numeric representation that PyTorch can utilize. The data used in this investigation was transformed into uint8 format. The image's pixel values are then modified to a specific range, often the dataset's mean (0.5, 0.5, 0.5) and standard deviation (0.5, 0.5, 0.5) for each RGB color channel. By ensuring that the input values are on the appropriate scale, normalization aids in stabilizing the model training process. By scaling the pixel values to the interval [−1,1][-1,1][−1,1], this normalization preserves numerical stability and speeds up convergence during model training.

To improve the accuracy and efficacy of colon pathology image classification with Vision Transformer (ViT), the model was adjusted employing a variety of methodologies. First, the

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24918

Adam optimization technique was employed to provide efficient convergence even on complicated datasets, using a *3e-4* learning rate to strike a balance between convergence speed and stability. Model training was optimized with PyTorch and cuDNN using the *torch.backends.cudnn.benchmark = True* method. This function lets you select the most efficient GPU convolution algorithm.

The Mixed Precision Training technique uses *GradScaler()* to optimize training and reduce memory usage with higher precision data types (float16). In addition, the early stopping method with a target training accuracy of at least 99% is used to ensure that treatment is completed automatically when the goal is accomplished, to avoid overfitting, and to reduce computing power consumption. This advancement makes the treatment process more effective, especially when working with a large model such as ViT on a ca colon dataset. The training results are then evaluated using several kinds of performance parameters, including accuracy, precision, specificity, and sensitivity, as well as evaluation time and epoch count. This gives an entire overview of the model's ability to effectively classify images.

III. RESULT AND DISCUSSION

In this study, the researchers used Google Colabs using Runtime T4 (GPU). The study then employed many comparisons of augmentation findings and hyperparameter adjustment of the ViT model to classify colon pathology photos. The hyperparameters employed here are training-related, such as the number of epoch iterations, learning rate, and early termination to preserve computer resources. First, the ViT Model was tested without any hyperparameter adjustment or data augmentation. With the default architecture and no early halting, 10 epochs took 1955.69 seconds, or approximately 33 minutes of training time. The training accuracy and loss process for each epoch are shown in Table 1.

Table 1. Loss and Training Accuracy Model ViT without Tuning and Augmentation

Epoch	Loss	Training Accuracy
1	0.3369	0.8510
2	0.1725	0.9315
3	0.1231	`0.9521
4	0.0565	0.9791
5	0.0818	0.9662
6	0.0592	0.9766
7	0.0584	0.9775
8	0.0483	0.9819
9	0.0399	0.9861
10	0.0351	0.9869

INTENSIF, Vol.9 No.2 August 2025
ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24918

Authors assessed the ViT Model with hyperparameter adjustment but no data augmentation. Figure 5 displays training accuracy and loss values for each epoch.

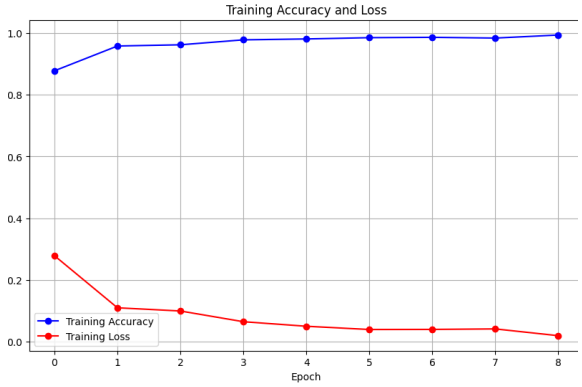

Fig 5. Training Accuracy and Loss Proposed ViT Model with hyperparameter

In the iteration above, beginning with epoch 1, the loss is 27.88% with a training accuracy of 87.74% until the 9th epoch, when the training accuracy value has reached 99.32% because it is greater than 99%, therefore early halting terminates the process. Model training in this trial took 626.94 seconds, or approximately 10 minutes. For the second trial, the ViT Model underwent hyperparameter tweaking as well as the Flip Vertical and Horizontal data augmentation method. The training accuracy and loss values for each epoch are as shown in Figure 6.

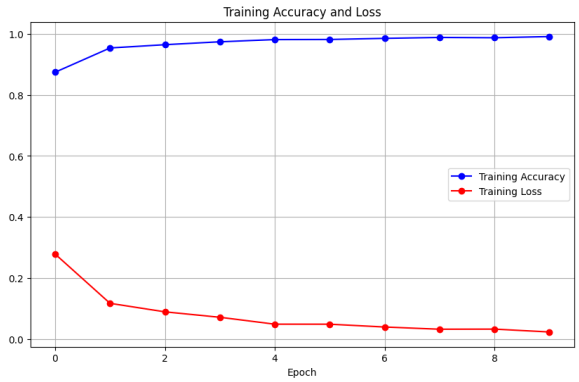

Fig 6. Training Accuracy and Loss Proposed Method 1

Starting with epoch 1, the loss was 27.88% with a training accuracy of 87.44%, and by the 10th epoch, the training accuracy had increased to 99.12% with a loss of 0.23%. In this trial, the model was trained in 696.34 seconds (about 11 minutes). The third trial contained hyperparameter adjustment and Rotation data augmentation for the ViT Model. Below is a chart of the training process at each loss and training accuracy value, as shown in Figure 7.

INTENSIF, Vol.9 No.2 August 2025
ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24918

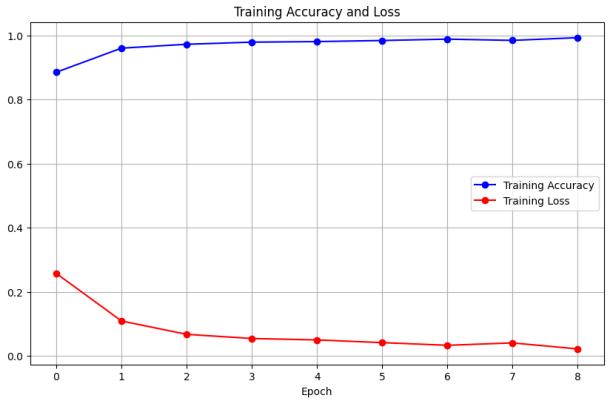

Fig 7. Training Accuracy and Loss Proposed Method 2

In the iteration above, starting with epoch 1, the loss is 25.77% with a training accuracy of 88.54% until the 9th epoch, when the training accuracy value has reached 99.32% because it is greater than 99%, therefore early stopping terminates the process. Model training in this trial took 626.20 seconds, or about 10 minutes. The ViT Model with geometric augmentation trained in this trial in 709.32 seconds, or roughly 12 minutes. Figure 8 shows a graph of the training process, with loss and accuracy values per each iteration.

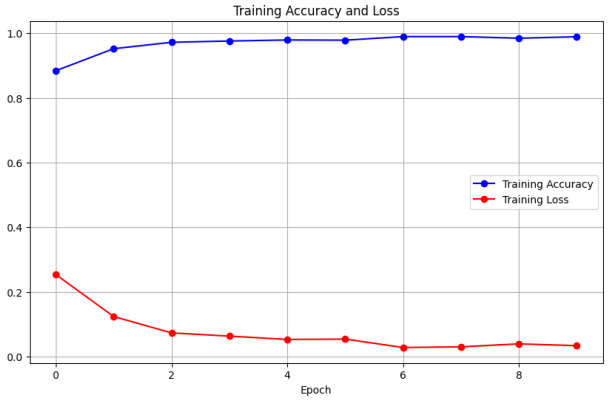

Fig 8. Training Accuracy and Loss Proposed Method 3

In the iteration above, beginning with epoch 1, the loss is 25.41% with a training accuracy of 88.39%, and by the tenth epoch, the training accuracy has reached 98.90% with a loss of 0.33%. Table 2 compares each method's performance with and without hyperparameter adjustment and augmentation.

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24918

Table 2. Comparison of Vit toward Various Proposed Methods

No	Model	Accuracy	Precision	Specificity	Sensitivity	Num Epoch	Time Eval (s)
1	ViT without tunning hyperparameter and augmented	99.72	99.77	99.77	99.67	10	94.50
2	ViT + tunning hyperparameter without augmented	98.20	96.53	96.40	100	9	18.01
3	ViT + tunning hyperparameter + Flip augmented	97.25	95.83	95.70	98.80	10	16.60
4	**ViT + tunning hyperparameter + Rotation augmented**	**99.30**	**99.10**	**99.10**	**99.50**	**9**	**16.26**
5	ViT + tunning hyperparameter + Geometry augmented	98.85	98.04	98.00	99.70	10	17.23

Before adjusting hyperparameters or adding augmentation, the Vision Transformer (ViT) model worked well. However, the assessment time was extremely long (94.50 seconds), and lowering the hyperparameter settings greatly reduced the evaluation time from 94.50 seconds to 18.01 seconds. However, as compared to the model without tuning, the accuracy and other metrics were slightly lower. Rotation augmentation enhanced model performance, which was comparable to test results obtained without augmentation and tweaking. The addition of Flip augmentation reduced performance compared to the model without augmentation (Proposed 1), but the sensitivity remained extremely high (98.80%). Geometry augmentation (Proposed 3) performed well, with a very high sensitivity (99.70%) and good specificity. Rotation augmentation (Proposed 2) improved model performance the most, with accuracy (99.30%), precision (99.10%), specificity (99.10%), and sensitivity (99.50%), while reducing evaluation time to a minimum of 16.26 seconds. In addition, the authors compared their results to the CNN model in previous study approaches. Here's a performance comparison table 3.

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24918

Table 3 Comparison of the Proposed Method to the CNN Model and Several Previous The study Models

No	Model	Accuracy	Effect Size Baseline (ViT Default)	Effect Size vs. ViT+Rotation
1	CNN [29]	74.80	-24.92	-24.50
2	RCG-NET [19]	90.62	-9.10	-8.68
3	CViTS-NET [25]	99.61	-0.11	+0.31
4	VGG [30]	95.15	-4.57	-4.15
5	CellViTs-NET [22]	84.00	-15.72	-15.30
2	**ViT without tunning hyperparameter and augmented (Default)**	**99.72**	**Baseline**	**+0.42**
3	ViT + tunning hyperparameter without augmented	98.20	-1.52	-1.10
4	ViT + tunning hyperparameter + Flip augmented (proposed method 1)	97.25	-2.47	-2.05
5	**ViT + tunning hyperparameter + Rotation augmented (proposed method 2)**	**99.30**	**-0.42**	**Baseline**
6	ViT + tunning hyperparameter + Geometry augmented (proposed method 3)	98.85	-0.87	-0.45

The proposed model accuracy comparison table demonstrates that the Vision Transformer (ViT) model outperforms traditional Convolutional Neural Network (CNN) models such as CNN [29], RCG-NET [19], CVITS-NET[25], and VGG[30] in prior studies. The ViT model trained without hyperparameter adjustment and data augmentation (default) achieves 99.72% accuracy. In addition, the results suggest that hyperparameter adjustment and data augmentation in the ViT model can increase classification performance. The rotation data augmentation approach achieves the best results with an accuracy of 99.30%, improves performance by +0.42% when compared to the ViT baseline without tuning, and approaches the performance of the optimal ViT model without parameter adjustments, even though not all data augmentation strategies provide significant benefits. demonstrating that data variety through rotation can improve the ViT model's generalization and lower a risk of overfitting. Overall, the findings of this study show that the ViT model has tremendous promise for use in the problem of colon pathology image classification. The results also suggest that using the appropriate fine-tuning technique can considerably increase model accuracy.

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24918

IV. CONCLUSION

The proposed method, ViT with Hyperparameter Tuning and Rotation Augmentation (proposed method 2), outperforms existing models in terms of performance and efficiency. This model provides the optimum balance of accuracy, sensitivity, and training time, particularly for hyperparameter adjustment and rotation augmentation. Proposed method 2 data augmentation increases ViT performance with optimal results, boosting accuracy to 99.30% and sensitivity to 99.50% while preserving training time efficiency. This augmentation is more efficient than Flip and Geometry.

The ViT-based colon cancer classification approach with rotation augmentation is indicated for early diagnosis of high sensitivity (recall) [31]. This technique strikes a compromise between accuracy and efficiency, making it ideal for clinical applications. To increase the model's generalizability, a larger and more diversified dataset should be used, allowing for a more comprehensive performance evaluation on a wider range of data.

Integrating the ViT model into an AI-based classification system in a medical setting necessitates changes to training time, evaluation, and extra validation with real-world data. This procedure is required to validate the model's reliability before clinical application. Additional changes to the clinical process are also required for the system to be implemented properly. More research might look into how Vision Transformer's (ViT) performance in medical picture classification is enhanced when rotation augmentation is combined with other augmentation methods like Color Jitter, Elastic Transformation, or CutMix.

Author Contributions: *Rudy Eko Prasetya:* was responsible for writing the original draft, data curation, formal analysis, and conducting the experiments. *M Arief Soeleman, Farrikh Al Zami, Affandy, Aris Marjuni, Mohammad Iqbal Saryuddin Assaqty*: contributed to the conceptualization, methodology, and review for editing of the manuscript.

All authors have read and agreed to the published version of the manuscript.

Funding: this research received no specific grant from any funding agency.

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

Data Availability: The source code is publicly available at https://github.com/rudyekoprasetya/enhance_vit_ca_colon.git and the dataset is available at https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images

Informed Consent: There were no human subjects.

Animal Subjects: There were no animal subjects

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24918

ORCID:
Rudy Eko Prasetya: https://orcid.org/0009-0001-5303-0890
M Arief Soeleman: https://orcid.org/0000-0001-6099-7023
Farrikh Al Zami: https://orcid.org/0000-0003-2669-3864
Affandy: https://orcid.org/0000-0003-1897-8261
Aris Marjuni: https://orcid.org/0000-0002-4072-3081
Mohammad Iqbal Saryuddin Assaqty*:* https://orcid.org/0000-0001-7274-6299

REFERENCES

[1] H. Sung *et al.*, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA. Cancer J. Clin.*, vol. 71, no. 3, pp. 209–249, 2021, **doi:** 10.3322/caac.21660.

[2] E. Dekker, P. J. Tanis, J. L. A. Vleugels, P. M. Kasi, and M. B. Wallace, "Colorectal cancer," *Lancet (London, England)*, vol. 394, no. 10207, pp. 1467–1480, Oct. 2019, **doi:** 10.1016/S0140-6736(19)32319-0.

[3] Asmaul Husnah, Andi Kartini Eka Yanti, Arina Fathiyyah Arifin, Berry Erida Hasbi, and Dzul Ikram, "Karakteristik Penderita Kanker Kolorektal Di Rumah Sakit Pendidikan Ibnu Sina Makassar Tahun 2022," *Fakumi Med. J. J. Mhs. Kedokt.*, vol. 4, no. 1, pp. 19–28, 2024, **doi:** 10.33096/fmj.v4i1.435.

[4] N. P. V. Primatama, A. Siswandi, T. Triwahyuni, and E. Purnanto, "Gambaran Faktor Resiko Kejadian Kanker Kolorektal di RSUD Dr. H. Abdi; Moeloek," *J. Ilmu Kedokt. dan Kesehat.*, vol. 10, no. 7, pp. 2461–2467, 2023, **doi:** 10.33024/jikk.v10i7.10808.

[5] L. W. Pratika Yuhyi Hernanda, Novina Aryanti, Maria Widijanti Sugeng, Febtarini Rahmawati,AjengTribawati, "Pemberdayaan Posyandu Lansia untuk Deteksi Dini Kanker Kolorektal dengan Tes Darah Samar Feses (FOBT)," in *Seminar Nasional Kusuma III Kualitas Sumberdaya Manusia*, 2024, pp. 10–19.

[6] I. Pacal, D. Karaboga, A. Basturk, B. Akay, and U. Nalbantoglu, "A comprehensive review of deep learning in colon cancer," *Comput. Biol. Med.*, vol. 126, no. August, p. 104003, 2020, **doi:** 10.1016/j.compbiomed.2020.104003.

[7] S. Sharmin, T. Ahammad, A. Talukder, and P. Ghose, "A Hybrid Dependable Deep Feature Extraction and Ensemble-Based Machine Learning Approach for Breast Cancer Detection," *IEEE Access*, vol. 11, pp. 87694–87708, Jan. 2023, **doi:** 10.1109/access.2023.3304628.

[8] S. L. Verghese, I. Y. Liao, T. H. Maul, and S. Y. Chong, "An Empirical Study of Several Information Theoretic Based Feature Extraction Methods for Classifying High Dimensional Low Sample Size Data," *IEEE Access*, vol. 9, pp. 69157–69172, Jan. 2021, **doi:** 10.1109/access.2021.3077958.

[9] S. Poudel, Y. J. Kim, D. M. Vo, and S.-W. Lee, "Colorectal Disease Classification Using Efficiently Scaled Dilation in Convolutional Neural Network," *IEEE Access*, vol. 8, pp. 99227–99238, Jan. 2020, **doi:** 10.1109/access.2020.2996770.

[10] F. J. P. Montalbo, "Diagnosing gastrointestinal diseases from endoscopy images through a multi-fused CNN with auxiliary layers, alpha dropouts, and a fusion residual block," *Biomed. Signal Process. Control*, vol. 76, p. 103683, Jul. 2022, **doi:** 10.1016/j.bspc.2022.103683.

[11] A. Bechar, Y. Elmir, R. Medjoudj, Y. Himeur, and A. Amira, "Transfer Learning for Cancer Detection based on Images Analysis," *Procedia Comput. Sci.*, vol. 239, pp. 1903–1910, 2024, **doi:** 10.1016/j.procs.2024.06.373.

[12] P. Haldar *et al.*, "XGBoosted Binary CNNs for Multi-Class Classification of Colorectal Polyp Size," *IEEE Access*, vol. 11, pp. 128461–128472, 2023, **doi:** 10.1109/ACCESS.2023.3332826.

[13] J. Lee, C. Han, K. Kim, G. H. Park, and J. T. Kwak, "CaMeL-Net: Centroid-aware metric

learning for efficient multi-class cancer classification in pathology images," *Comput. Methods Programs Biomed.*, vol. 241, p. 107749, Nov. 2023, **doi:** 10.1016/J.CMPB.2023.107749.

[14] A. Vaswani *et al.*, "Attention is all you need," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.

[15] N. Parmar *et al.*, "Image transformer," *35th Int. Conf. Mach. Learn. ICML 2018*, vol. 9, pp. 6453–6462, 2018.

[16] A. Dosovitskiy *et al.*, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," *ICLR 2021 - 9th Int. Conf. Learn. Represent.*, Oct. 2020, Accessed: Dec. 12, 2023. [Online]. Available: https://arxiv.org/abs/2010.11929v2

[17] N. T. Duc, N. T. Oanh, N. T. Thuy, T. M. Triet, and V. S. Dinh, "ColonFormer: An Efficient Transformer Based Method for Colon Polyp Segmentation," *IEEE Access*, vol. 10, pp. 80575–80586, Jan. 2022, **doi:** 10.1109/access.2022.3195241.

[18] A. Bechar, Y. Elmir, R. Medjoudj, Y. Himeur, and A. Amira, "Transfer Learning for Cancer Detection based on Images Analysis," *Procedia Comput. Sci.*, vol. 239, pp. 1903–1910, Jan. 2024, **doi:** 10.1016/J.PROCS.2024.06.373.

[19] T. Mahmood, A. Wahid, J. S. Hong, S. G. Kim, and K. R. Park, "A novel convolution transformer-based network for histopathology-image classification using adaptive convolution and dynamic attention," *Eng. Appl. Artif. Intell.*, vol. 135, p. 108824, 2024, **doi:** https://doi.org/10.1016/j.engappai.2024.108824.

[20] A. I. Saad, F. A. Maghraby, and O. M. Badawy, "PolyDSS: computer-aided decision support system for multiclass polyp segmentation and classification using deep learning," *Neural Comput. Appl.*, vol. 36, no. 9, pp. 5031–5057, Mar. 2024, **doi:** 10.1007/S00521-023-09358-3/FIGURES/16.

[21] T. Aitazaz, A. Tubaishat, F. Al-Obeidat, B. Shah, T. Zia, and A. Tariq, "Transfer learning for histopathology images: an empirical study," *Neural Comput. Appl. 2022 3511*, vol. 35, no. 11, pp. 7963–7974, Jul. 2022, **doi:** 10.1007/S00521-022-07516-7.

[22] F. Hörst *et al.*, "CellViT: Vision Transformers for precise cell segmentation and classification," *Med. Image Anal.*, vol. 94, p. 103143, May 2024, **doi:** 10.1016/J.MEDIA.2024.103143.

[23] N. Marini *et al.*, "Multimodal representations of biomedical knowledge from limited training whole slide images and reports using deep learning," *Med. Image Anal.*, vol. 97, p. 103303, 2024, **doi:** https://doi.org/10.1016/j.media.2024.103303.

[24] S. V. Mahadevkar *et al.*, "A Review on Machine Learning Styles in Computer Vision - Techniques and Future Directions," *IEEE Access*, vol. 10, no. October, pp. 107293–107329, 2022, **doi:** 10.1109/ACCESS.2022.3209825.

[25] A. Kanadath, J. Angel Arul Jothi, and S. Urolagin, "CViTS-Net: A CNN-ViT Network With Skip Connections for Histopathology Image Classification," *IEEE Access*, vol. 12, pp. 117627–117649, 2024, **doi:** 10.1109/ACCESS.2024.3448302.

[26] C. H. J. Kusters *et al.*, "Will Transformers change gastrointestinal endoscopic image analysis? A comparative analysis between CNNs and Transformers, in terms of performance, robustness and generalization," *Med. Image Anal.*, vol. 99, p. 103348, Jan. 2025, **doi:** 10.1016/J.MEDIA.2024.103348.

[27] C.-M. Liu, Z. Niu, and K.-T. Liao, "Mechanisms to improve clustering uncertain data with UKmeans," *Data Knowl. Eng.*, vol. 116, pp. 61–79, Jul. 2018, **doi:** 10.1016/j.datak.2018.05.004.

[28] M. Al-Jabbar, M. Alshahrani, E. M. Senan, and I. A. Ahmed, "Histopathological Analysis for Detecting Lung and Colon Cancer Malignancies Using Hybrid Systems with Fused Features," *Bioengineering*, vol. 10, no. 3, 2023, **doi:** 10.3390/bioengineering10030383.

[29] M. Mahanty, D. Bhattacharyya, D. Midhunchakkaravarthy, and T. H. Kim, "Detection of colorectal cancer by deep learning: An extensive review," *Int. J. Curr. Res. Rev.*, vol. 12, no. 22, pp. 150–157, 2020, **doi:** 10.31782/IJCRR.2020.122234.

ISSN: 2580-409X (Print) / 2549-6824 (Online)
DOI: https://doi.org/10.29407/intensif.v9i2.24918

[30] A. Ben Hamida *et al.*, "Deep learning for colon cancer histopathological images analysis," *Comput. Biol. Med.*, vol. 136, no. August, 2021, **doi:** 10.1016/j.compbiomed.2021.104730.

[31] A. Gupta, A. Anand, and Y. Hasija, "Recall-based Machine Learning approach for early detection of Cervical Cancer," *2021 6th Int. Conf. Converg. Technol. I2CT 2021*, pp. 1–5, 2021, **doi:** 10.1109/I2CT51068.2021.9418099.