A Comparative Study of DenseNet-201 and Swin Transformer for Malignant and Benign Skin Lesion Classification

Dahlan Hidayat [1)], Ahmad Musyafa [2)*)], Murni Handayani [3)]

[1)2)3)]Teknik Informatika S-2, Program Pascasarjana, Universitas Pamulang
[*)]**Correspondence author**: dosen00668@unpam.ac.id, Tangerang Selatan, Indonesia
DOI: https://doi.org/10.37012/jtik.v12i1.3265

Abstract

Skin cancer has a high global prevalence, underscoring the need for accurate and efficient early detection systems to support screening. This study presents a comparative analysis of DenseNet-201 and Swin Transformer for binary classification of malignant and benign skin lesions using the BCN20000 dataset, which contains 12,413 dermoscopic images. The proposed workflow includes image preprocessing and augmentation, transfer learning-based model training, and evaluation under a 5-fold stratified cross-validation protocol. Performance is assessed using Accuracy, Precision, Sensitivity (Recall), F1-score, and the area under the receiver operating characteristic curve (AUC-ROC). In addition, computational efficiency is examined in terms of parameter count, model size, and training time. Across five folds, DenseNet-201 achieved 88.05% Accuracy, 88.90% Precision, 89.48% Sensitivity, 89.17% F1-score, and 94.73% AUC, whereas Swin Transformer achieved 87.42% Accuracy, 89.77% Precision, 87.06% Sensitivity, 88.39% F1-score, and 94.33% AUC. A paired t-test at $\alpha = 0.05$ indicated no statistically significant performance difference between the two models. Model interpretability was investigated using Grad-CAM for DenseNet-201 and EigenCAM for Swin Transformer to verify that predictions were driven by lesion-relevant regions. Overall, the results suggest that both architectures are suitable candidates for dermoscopic image-based skin lesion screening support systems, including teledermatology applications.

Keywords: *Skin Cancer, Image Classification, DenseNet-201, Swin Transformer, BCN20000, Grad-CAM, EigenCAM*

Abstrak

Kanker kulit memiliki prevalensi global yang tinggi, yang menggarisbawahi kebutuhan akan sistem deteksi dini yang akurat dan efisien untuk mendukung skrining. Studi ini menyajikan analisis komparatif DenseNet-201 dan Swin Transformer untuk klasifikasi biner lesi kulit ganas dan jinak menggunakan dataset BCN20000, yang berisi 12.413 gambar dermoskopi. Alur kerja yang diusulkan mencakup pra-pemrosesan dan augmentasi gambar, pelatihan model berbasis transfer learning, dan evaluasi di bawah protokol validasi silang bertingkat 5-fold. Kinerja dinilai menggunakan Akurasi, Presisi, Sensitivitas (Recall), skor F1, dan area di bawah kurva karakteristik operasi penerima (AUC-ROC). Selain itu, efisiensi komputasi diperiksa dalam hal jumlah parameter, ukuran model, dan waktu pelatihan. Pada lima lipatan (folds), DenseNet-201 mencapai Akurasi 88,05%, Presisi 88,90%, Sensitivitas 89,48%, skor F1 89,17%, dan AUC 94,73%, sedangkan Swin Transformer mencapai Akurasi 87,42%, Presisi 89,77%, Sensitivitas 87,06%, skor F1 88,39%, dan AUC 94,33%. Uji t berpasangan pada $\alpha = 0,05$ menunjukkan tidak ada perbedaan kinerja yang signifikan secara statistik antara kedua model tersebut. Interpretasi model diselidiki menggunakan Grad-CAM untuk DenseNet-201 dan EigenCAM untuk Swin Transformer untuk memverifikasi bahwa prediksi didorong oleh wilayah yang relevan dengan lesi. Secara keseluruhan, hasilnya menunjukkan bahwa kedua arsitektur tersebut merupakan kandidat yang cocok untuk sistem pendukung skrining lesi kulit berbasis citra dermoskopi, termasuk aplikasi teledermatologi.

Kata Kunci: Kanker Kulit, Klasifikasi Gambar, DenseNet-201, Swin Transformer, BCN20000, Grad-CAM, EigenCAM

INTRODUCTION

Skin cancer is among the most prevalent malignancies worldwide and may progress to advanced disease if not detected at an early stage. Its development is influenced by multiple risk factors, including prolonged ultraviolet exposure, genetic predisposition, and heterogeneous clinical manifestations. In clinical practice, dermoscopy is commonly used to enhance visualization of lesion morphology and subsurface structures. However, dermoscopic image interpretation remains dependent on clinician expertise and may be subject to inter-observer variability.

The application of deep learning in medical image analysis, including dermoscopic imaging, has increased substantially due to its capacity for automated feature extraction and improved classification performance. Convolutional neural network (CNN) architectures such as DenseNet-201 have demonstrated strong results across a range of medical imaging tasks. DenseNet employs dense connectivity between layers, which facilitates stable gradient propagation and promotes efficient feature reuse throughout the network. In contrast, transformer-based vision models such as the Swin Transformer leverage self-attention mechanisms to capture broader feature dependencies, potentially improving robustness and adaptability across diverse imaging conditions.

Although numerous studies have investigated CNN and transformer-based approaches for skin lesion classification, direct comparisons between DenseNet-201 and Swin Transformer on real-world dermoscopic datasets such as BCN20000 remain limited. The BCN20000 dataset exhibits substantial image variability and more closely reflects clinical conditions encountered in practice. Consequently, comparative evaluations are required that extend beyond predictive performance to include computational efficiency and human-interpretable decision analysis.

Accordingly, this study makes three main contributions: (1) a systematic comparison of DenseNet-201 and Swin Transformer for benign versus malignant skin lesion classification on the BCN20000 dataset using a 5-fold stratified cross-validation protocol; (2) an assessment of computational efficiency, including model complexity and training characteristics, to support feasibility for practical deployment; and (3) an

interpretability analysis using Grad-CAM and EigenCAM to visualize lesion-relevant regions that drive model predictions.

RESEARCH METHOD

This study adopts a computational experimental design to compare two deep learning architectures, DenseNet-201 and Swin Transformer, for binary skin lesion classification using the BCN20000 dataset. The workflow comprises image preprocessing, data partitioning under a 5-fold stratified cross-validation protocol, model training, performance evaluation using standard classification metrics, statistical comparison via a paired t-test, and interpretability analysis using Grad-CAM and EigenCAM.

1. **Dataset**

   This study utilizes the BCN20000 dataset, which comprises dermoscopic images representing a wide range of skin lesion conditions. The dataset is formulated as a binary classification task with two categories: benign and malignant lesions.

2. **Research Stages**

   The study was carried out through sequential stages, including data collection, image preprocessing, dataset partitioning, model training, performance evaluation, and interpretability analysis. The overall workflow is summarized in Figure 1.

   As illustrated in Figure 1 below, the workflow begins with the acquisition of dermoscopic image data, followed by preprocessing to prepare the inputs for model development. The dataset is then partitioned into training, validation, and testing subsets. During training, data augmentation is applied to increase sample diversity and promote model generalization. Finally, classification is performed using DenseNet-201 and Swin Transformer, and the performance of both models is compared using the selected evaluation metrics.

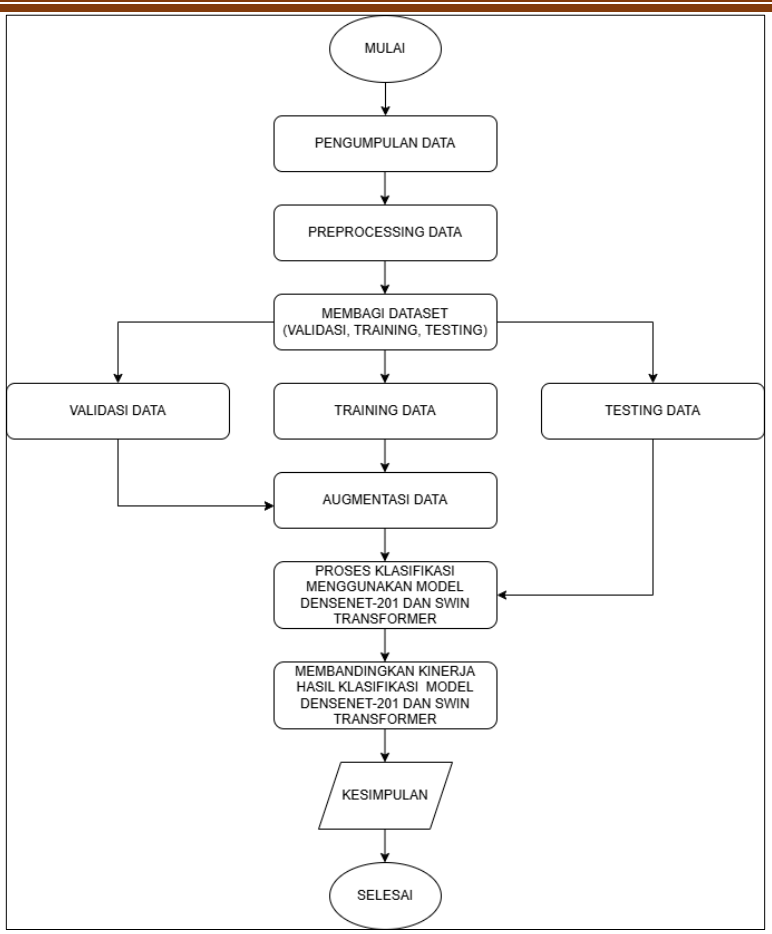

Figure 1. Research Design

3. Preprocessing and Augmentasi

Preprocessing is performed to ensure that all images conform to the input specifications required by the models. This stage includes image resizing, normalization, and other necessary transformations. To increase training data diversity and mitigate overfitting, data augmentation techniques such as rotation, flipping, and other relevant transformations are applied.

4. Model Architectures

This study compares two deep learning architectures:

1) DenseNet-201, a convolutional neural network (CNN) that employs dense inter-layer connectivity to facilitate efficient feature reuse and stable gradient propagation.

2) Swin Transformer, a shifted window-based vision transformer that applies hierarchical self-attention to capture both local and global feature relationships.

5. Evaluation Protocol and Metrics

Model evaluation was performed using a 5-fold stratified cross-validation protocol to preserve class proportions within each fold. Performance was assessed using the following metrics:

- Accuracy
- Precision
- Sensitivity (Recall)
- F1-score
- AUC (Area Under Curve)

In addition, a paired-sample t-test at $\alpha = 0.05$ was conducted to determine whether performance differences between the two models were statistically significant.

6. Model Interpretability

To enhance the transparency of model predictions, interpretability analyses were conducted using two visualization techniques:

- Grad-CAM was applied to DenseNet-201 to highlight image regions that contributed most strongly to the predicted class.
- EigenCAM was applied to the Swin Transformer to generate attention maps derived from feature representations, providing insight into regions that influenced the model's decision-making process.

RESULTS AND DISCUSSION

This section presents the evaluation results of DenseNet-201 and Swin Transformer for binary skin lesion classification using the BCN20000 dataset. All experiments were conducted under a consistent 5-fold stratified cross-validation protocol to ensure a fair and objective comparison between the two models.

Classification Performance Comparison (Mean 5-Fold)

A summary of the mean 5-fold results for both models is shown in Table 1.

Table 1. Summary of Comparison Results for DenseNet-201 and Swin Transformer (Mean 5-Fold)

Model	Accuracy	Precision	Sensitivity	F1-score	AUC
DenseNet-201	88,05%	88,90%	89,48%	89,17%	94,73%
Swin Transformer	87,42%	89,77%	87,06%	88,39%	94,33%

Based on Table 1, DenseNet-201 achieves higher Accuracy, Sensitivity, F1-score, and AUC, whereas Swin Transformer attains higher Precision. In clinical screening applications, Sensitivity is particularly critical because false-negative predictions may result in missed malignant cases and delayed treatment. Although the overall performance differences are modest, the higher Sensitivity of DenseNet-201 suggests that it may be more suitable for screening scenarios that prioritize the detection of malignant lesions.

Computational Efficiency

Beyond predictive performance, computational efficiency was assessed to evaluate the feasibility of deploying each model in practical settings. A comparison of computational efficiency between DenseNet-201 and Swin Transformer is presented in Table 2.

Table 2. Comparison of Computational Efficiency between DenseNet-201 and Swin Transformer

NO	Efficiency Metric	DenseNet-201	Swin Transformer	Advantages
1	Number of Parameters	±20 million (20M)	±28 million (28M)	DenseNet (40% lighter)
2	Model File Size	81.1 MB	114 MB	DenseNet (29% smaller)
3	Total 5-Fold Training Time	7 hours 20 minutes	6 hours 25 minutes	Swin Transformer (12.5% faster)
4	Average per Fold	88 minutes	77 minutes	Swin Transformer (12.5% faster)

Table 2 indicates a clear trade-off between the two architectures. DenseNet-201 has fewer parameters and a smaller model size, which reduces memory requirements and

supports deployment on resource-constrained devices. In contrast, Swin Transformer exhibits shorter training time, making it more appropriate for server-based environments where retraining efficiency is a priority.

Paired t-test Statistical Analysis

To determine whether the observed performance differences between the two models were statistically significant, a paired-sample t-test was conducted at a significance level of $\alpha = 0.05$. The results are reported in Table 3.

Table 3. Results of Paired t-test Comparison between DenseNet-201 and Swin Transformer ($\alpha = 0.05$)

Metric	t-statistic	p-value	α	Decision	Conclusion
Accuracy	0,5056	0,6397	0,05	p > α	Not significant
Precision	-1,1477	0,3151	0,05	p > α	Not significant
Sensitivity	1,0897	0,3371	0,05	p > α	Not significant
F1-score	0,6235	0,5668	0,05	p > α	Not significant
AUC	0,4516	0,6750	0,05	p > α	Not significant

As shown in Table 3, all metrics yield p-values greater than 0.05; therefore, the null hypothesis (H0) cannot be rejected. This finding indicates that the performance differences between DenseNet-201 and Swin Transformer are not statistically significant. Consequently, both models can be considered to exhibit comparable performance on the BCN20000 dataset under the 5-fold stratified cross-validation protocol.

DenseNet-201 Evaluation (Best Fold)

1. **DenseNet-201 Confusion Matrix**

   To examine classification errors in greater detail, a confusion matrix is used to summarize the distribution of correct and incorrect predictions across classes. The confusion matrix for DenseNet-201 in the best-performing fold is presented in Figure 2.

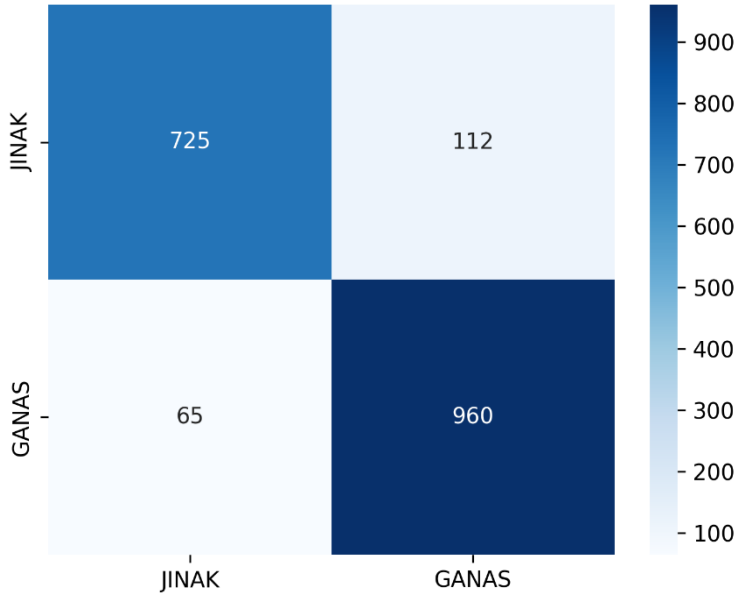

Figure 2. DenseNet-201 Confusion Matrix (Best-Performing Fold)

As shown in Figure 2, DenseNet-201 correctly classified 725 benign samples and 960 malignant samples. Misclassifications included 112 benign samples predicted as malignant (false positives) and 65 malignant samples predicted as benign (false negatives). In clinical screening, false-negative errors are particularly critical because they may result in missed malignant cases and delayed intervention. Nevertheless, the false-negative count in the best-performing fold is lower than the false-positive count, indicating that DenseNet-201 maintains adequate sensitivity for malignant lesion detection.

2. DenseNet-201 ROC Curve

In addition to the confusion matrix, model performance was further evaluated using the receiver operating characteristic (ROC) curve, which illustrates the trade-off between the true positive rate (TPR) and false positive rate (FPR) across different classification thresholds. The ROC curve for DenseNet-201 in the best-performing fold is shown in Figure 3.

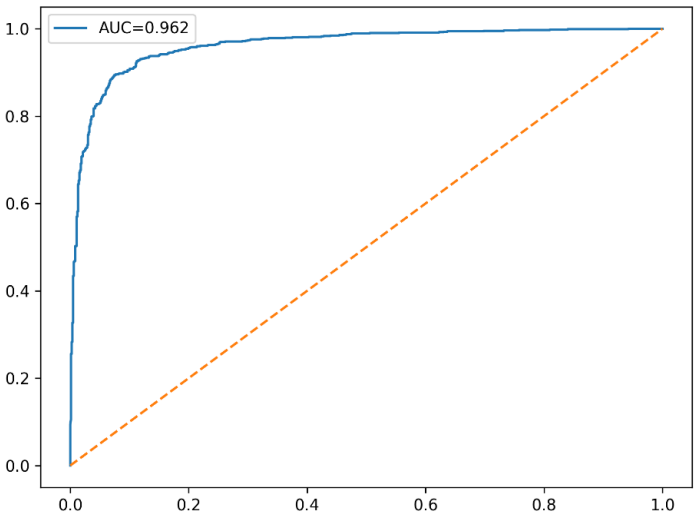

Figure 3. DenseNet-201 ROC Curve (Best-Performing Fold)

As shown in Figure 3, DenseNet-201 achieves an AUC of 0.962, indicating strong discriminative performance in distinguishing malignant from benign skin lesions. The ROC curve remains well above the diagonal baseline, demonstrating performance that exceeds random classification.

3. DenseNet-201 Grad-CAM Visualization

To improve the transparency of model predictions, an interpretability analysis was conducted using Grad-CAM on DenseNet-201. This technique highlights image regions that contribute most strongly to the model's classification decisions.

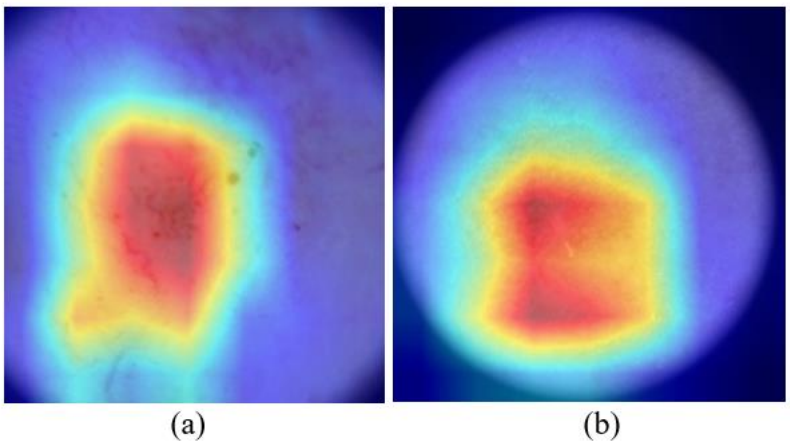

Figure 4. DenseNet-201 Grad-CAM Visualization for Malignant Skin Lesions: (a) Example 1, (b) Example 2

As shown in Figure 4, regions with the highest activation (red to yellow) are concentrated within the lesion core, whereas areas outside the lesion exhibit lower activation. This pattern suggests that DenseNet-201 primarily relies on textural and structural features within the lesion region when predicting the malignant class.

In addition, Grad-CAM was applied to benign lesion images to verify that DenseNet-201 consistently attended to lesion-relevant regions rather than image artifacts or background features.

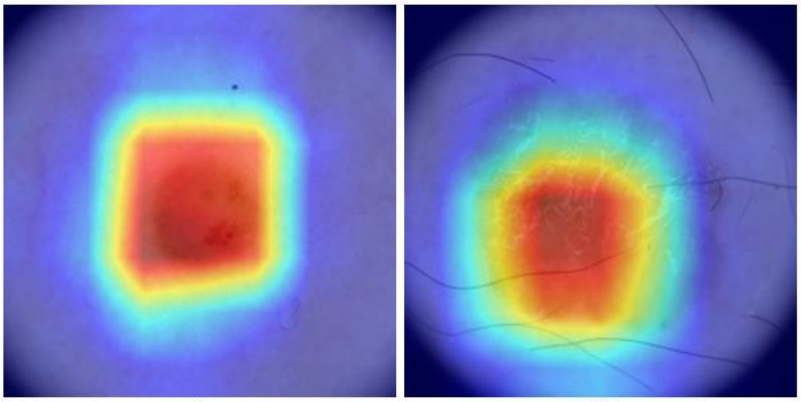

Figure 5. DenseNet-201 Grad-CAM Visualization for Benign Skin Lesions: (a) Example 1, (b) Example 2

As shown in Figure 5, the highest activation regions are concentrated within the lesion area, whereas surrounding regions exhibit lower activation. This consistent localization indicates that DenseNet-201 primarily focuses on lesion-relevant regions when generating benign predictions.

Swin Transformer Evaluation (Best Fold)

1. Swin Transformer Confusion Matrix

To analyze the prediction error patterns of Swin Transformer, a confusion matrix is used to summarize the number of correct and incorrect predictions for each class. The confusion matrix for Swin Transformer in the best-performing fold is presented in Figure 6.

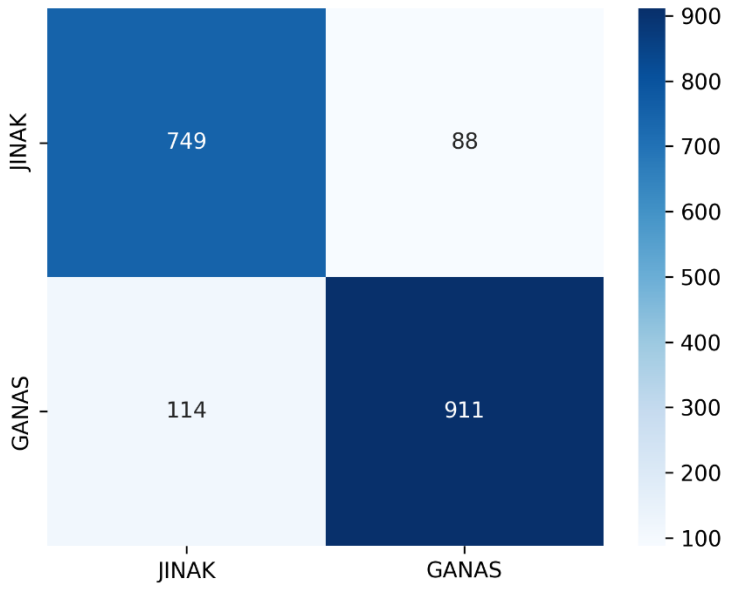

Figure 6. Swin Transformer Confusion Matrix (Best-Performing Fold)

As shown in Figure 6, Swin Transformer correctly classified 749 benign samples and 911 malignant samples. Misclassifications included 88 benign samples predicted as malignant (false positives) and 114 malignant samples predicted as benign (false negatives). Compared with DenseNet-201 in the best-performing fold, Swin Transformer produced fewer false positives but a higher number of false negatives.

2. Swin Transformer ROC Curve

The receiver operating characteristic (ROC) curve is used to illustrate the trade-off between the true positive rate (TPR) and false positive rate (FPR) across different classification thresholds. The ROC curve for Swin Transformer in the best-performing fold is shown in Figure 7.

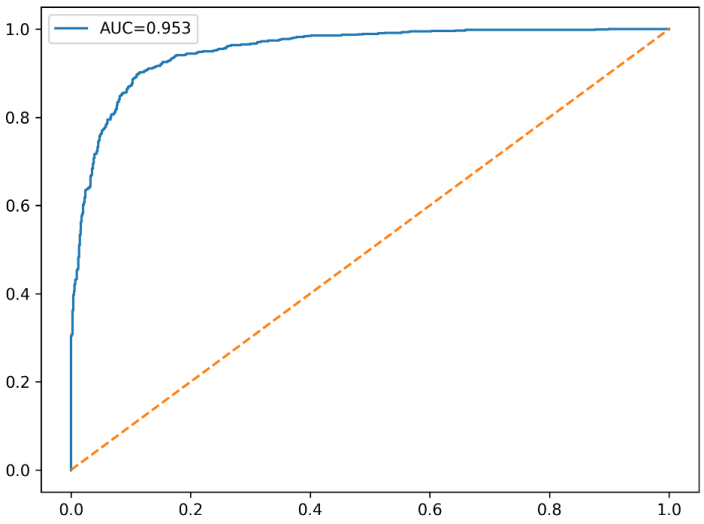

Figure 7. Swin Transformer ROC Curve (Best-Performing Fold)

As shown in Figure 7, Swin Transformer achieves an AUC of 0.953, indicating strong discriminative performance. The ROC curve remains well above the diagonal baseline, demonstrating performance that exceeds random classification.

3. Swin Transformer EigenCAM Visualization

The interpretability of Swin Transformer was examined using EigenCAM to generate attention maps that highlight image regions contributing to the model's predictions.

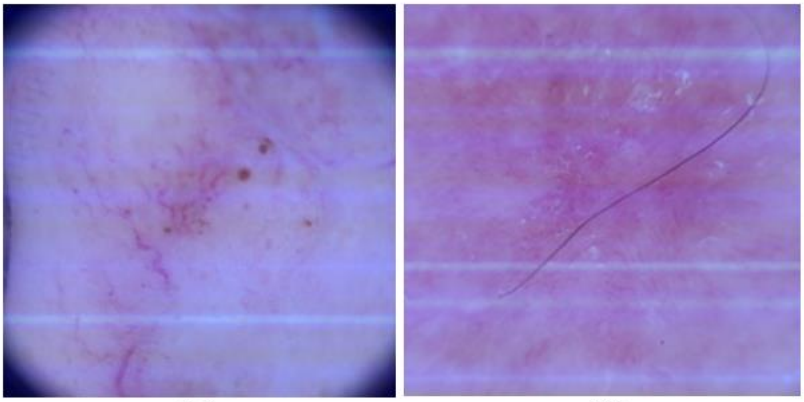

Figure 8. Swin Transformer EigenCAM Visualization for Malignant Skin Lesions: (a) Example 1, (b) Example 2

As shown in Figure 8, the EigenCAM activation maps indicate that the model attends to the lesion region as well as surrounding structures. The broader activation pattern is consistent with the self-attention mechanism of Swin Transformer, which captures feature relationships over a wider spatial context.

In addition, EigenCAM was applied to benign lesion images to verify that Swin Transformer consistently focused on lesion-relevant regions rather than background features.

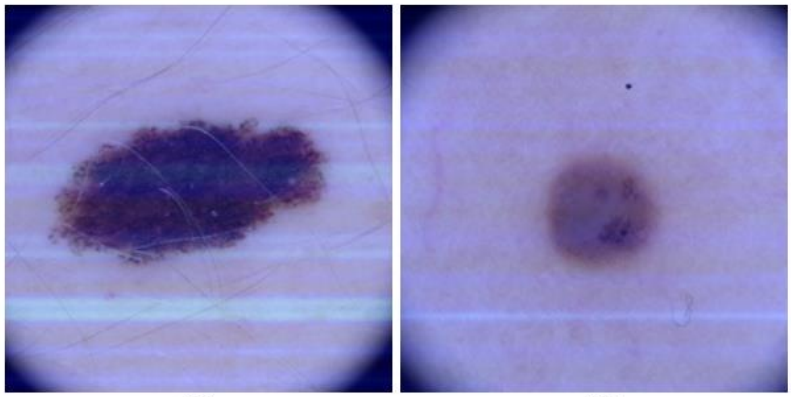

Figure 9. Swin Transformer EigenCAM Visualization for Benign Skin Lesions: (a) Example 1, (b) Example 2

As shown in Figure 9, model attention is primarily directed toward lesion regions that exhibit contrast relative to surrounding normal skin. This pattern suggests that Swin Transformer leverages both local lesion characteristics and broader contextual information to generate benign class predictions.

DenseNet-201 and Swin Transformer Comparison

1. Interpretability Comparison

Overall, Grad-CAM visualizations for DenseNet-201 (Figures 4 and 5) exhibit more localized activation concentrated within the lesion core, whereas EigenCAM visualizations for Swin Transformer (Figures 8 and 9) tend to display broader attention that encompasses both the lesion region and surrounding context. This difference is consistent with architectural characteristics: CNNs primarily emphasize local feature

extraction through convolutional operations, while transformer-based models employ self-attention mechanisms to capture feature dependencies over a wider spatial extent.

2. **ROC Comparison (Best-Performing Fold)**

   DenseNet-201 achieved an AUC of 0.962 in the best-performing fold (Figure 3), whereas Swin Transformer achieved an AUC of 0.953 (Figure 7). This difference indicates that DenseNet-201 provides slightly higher discriminative capability in the best fold, although both models demonstrate high overall performance.

3. **Error Pattern Analysis**

   Based on the confusion matrices from the best-performing folds, DenseNet-201 produced 65 false negatives (FN) and 112 false positives (FP), while Swin Transformer produced 114 FN and 88 FP. DenseNet-201 yields fewer false negatives, which is advantageous for screening settings that prioritize minimizing missed malignant cases. In contrast, Swin Transformer produces fewer false positives, suggesting greater selectivity and potential benefits in reducing false alarms and unnecessary referrals.

CONCLUSIONS AND RECOMMENDATIONS

This study compares DenseNet-201 and Swin Transformer for malignant versus benign skin lesion classification using the BCN20000 dataset under a 5-fold stratified cross-validation protocol. DenseNet-201 achieved 88.05% Accuracy, 88.90% Precision, 89.48% Sensitivity, 89.17% F1-score, and 94.73% AUC, whereas Swin Transformer achieved 87.42% Accuracy, 89.77% Precision, 87.06% Sensitivity, 88.39% F1-score, and 94.33% AUC. Computational efficiency analysis indicates that DenseNet-201 is more suitable for deployment due to its smaller model size, while Swin Transformer requires less training time. A paired t-test at $\alpha = 0.05$ shows that performance differences between the two models are not statistically significant, indicating that both architectures are viable candidates for skin lesion screening support systems. Error pattern analysis on the best-performing fold further shows that DenseNet-201 produces fewer false negatives, supporting screening scenarios that prioritize minimizing missed malignant cases, whereas Swin Transformer produces fewer false positives, which may reduce false alarms. Future work should include external validation on additional datasets, the application of more

adaptive class imbalance handling strategies, and the exploration of ensemble or hybrid CNN–Transformer approaches to improve performance stability and generalization.

REFERENCES

Brutti, F., La Rosa, F., Lazzeri, L., Benvenuti, C., Bagnoni, G., Massi, D., & Laurino, M. (2023). Artificial intelligence algorithms for benign vs. malignant dermoscopic skin lesion image classification. *Bioengineering*, *10*(11), 1322.

Bhattacharjee, T (2025). *GoogLeNet/DenseNet-201 to classify near-infrared (NIR) spectrum graphs for cancer diagnosis–using pretrained image networks for medical spectroscopy.*, researchsquare.com, https://www.researchsquare.com/article/rs-6562812/latest

Das, S, & Sahoo, BK (2025). Brain tumor identification using DenseNet-201 deep learning model. *Design Optimization Using Artificial …*, taylorfrancis.com, https://doi.org/10.1201/9781003589716-13/brain-tumor-identification-using-densenet-201-deep-learning-model-soubhagya-das-bidush-kumar-sahoo

Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, *542*(7639), 115–118.

Fortarezza, F., Cazzato, G., & Ingravallo, G. (2024). The 2023 WHO updates on skin tumors: Advances since the 2018 edition. *Pathologica*, *116*(4), 193–206.

Hernández Pérez, C., Combalia, M., Podlipnik, S., Codella, N. C. F., Rotemberg, V., Halpern, A. C., Reiter, O., Carrera, C., Barreiro, A., Helba, B., Puig, S., Vilaplana, V., & Malvehy, J. (2024). BCN20000: Dermoscopic lesions in the wild. *Scientific Data*, *11*(1), Article 641.

Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings*

of the IEEE/CVF International Conference on Computer Vision (pp. 10012–10022).

Lu, T, Han, B, Chen, L, Yu, F, & Xue, C (2021). Author Correction: A generic intelligent tomato classification system for practical applications using DenseNet-201 with transfer learning. *Scientific Reports*, pmc.ncbi.nlm.nih.gov, https://pmc.ncbi.nlm.nih.gov/articles/PMC8455630/

Mahbod, A., Schaefer, G., Ellinger, I., et al. (2023). Swin transformer for multi-label skin lesion classification using ISIC datasets. *Computers in Biology and Medicine, 162*, 107075.

Pacal, I., Alaftekin, M., & Devrim, F. (2024). Enhancing skin cancer diagnosis using Swin transformer with hybrid shifted window-based multi-head self-attention and SwiGLU-based MLP. *Journal of Imaging Informatics in Medicine, 37*(6), 3174–3192. https://doi.org/10.1007/s10278-024-01140-8

Salim, F, Saeed, F, Basurra, S, Qasem, SN, & ... (2023). *DenseNet-201 and Xception Pre-Trained Deep Learning Models for Fruit Recognition. Electronics 2023, 12, 3132*., academia.edu, https://www.academia.edu/download/107856186/electronics-12-03132.pdf

Sung, H., Ferlay, J., Siegel, R. L., Laversanne, M., Soerjomataram, I., Jemal, A., & Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians, 71*(3), 209–249.

Wang, SH, & Zhang, YD (2020). DenseNet-201-based deep neural network with composite learning factor and precomputation for multiple sclerosis classification. *ACM Transactions on Multimedia Computing …*, dl.acm.org, https://doi.org/10.1145/3341095

Whiteman, D. C., et al. (2022). Mortality and metastasis in melanoma and non-melanoma skin cancers: An updated global overview. *The Lancet Oncology, 23*(5), e197–e206.