ISSN (Online) 2747-0563
Volume 6 Tahun 2026

Studi Literatur Mengenai Arsitektur Transformer dalam Klasifikasi Citra Medis

Muh. Yassar Nurfajri Dharmawan[1], Faisal Muttaqin[2*], Budi Mukhamad Mulyo[3]

[1,2,3] Ilmu Komputer, UPN "Veteran" Jawa Timur
[1]22081010088@student.upnjatim.ac.id
[2]faisalmuttaqin.if@upnjatim.ac.id
[3]budi.m.mulyo.fasilkom@upnjatim.ac.id
*Corresponding author email: 22081010088@student.upnjatim.ac.id

***Abstrak*— Perkembangan deep learning telah memberikan kontribusi signifikan dalam analisis citra medis, khususnya dalam meningkatkan akurasi sistem computer-aided diagnosis. Meskipun *Convolutional Neural Network* (CNN) telah banyak digunakan, keterbatasannya dalam menangkap dependensi global antar fitur citra masih menjadi tantangan utama. Dalam beberapa tahun terakhir, arsitektur berbasis *Transformer* muncul sebagai alternatif yang mampu mengatasi keterbatasan tersebut melalui mekanisme *self-attention*. Penelitian ini bertujuan untuk melakukan studi literatur sistematis terhadap 20 publikasi ilmiah dalam lima tahun terakhir guna menganalisis hubungan antara arsitektur, strategi pelatihan, dan karakteristik dataset terhadap performa model. Hasil penelitian menunjukkan bahwa *Vision Transformer* (ViT) dan *Swin Transformer* merupakan pendekatan dominan dengan performa yang tinggi dan konsisten. Berdasarkan hasil analisis literatur, model Swin Transformer mampu mencapai akurasi hingga 99.57%–99.85% pada beberapa studi, sementara pendekatan hybrid CNN–*Transformer* menunjukkan akurasi sekitar 97.0% dengan precision 97.5%, recall 97.2%, dan F1-score 97.4%. Selain itu, *Vision Transformer* juga menunjukkan performa yang baik dengan precision mencapai 93.3%, recall 92.8%, dan F1-score 93.0%. Secara umum, sebagian besar penelitian menunjukkan performa model dengan akurasi di atas 90% bahkan mendekati 99%, yang menegaskan efektivitas arsitektur *Transformer* dalam klasifikasi citra medis. Kebaruan penelitian ini terletak pada penyajian analisis terintegrasi yang menegaskan bahwa performa model tidak hanya ditentukan oleh arsitektur, tetapi juga oleh strategi pelatihan dan kompleksitas dataset yang digunakan.**

Kata Kunci— **Transformer, Swin Transformer, Vision Transformer, klasifikasi citra medis, deep learning, self-attention, computer-aided diagnosis**

I. PENDAHULUAN

Analisis citra medis telah mengalami perkembangan signifikan seiring dengan kemajuan kecerdasan buatan, terutama dalam bidang *deep learning* [1]. Citra medis seperti MRI, CT scan, dan X-ray memiliki fitur kompleks dan dimensi data yang tinggi, sehingga memerlukan metode analisis yang dapat secara efektif mengekstraksi fitur tersebut [2]. Dalam konteks ini, salah satu tugas penting dalam sistem *computer-aided diagnosis* (CAD) adalah klasifikasi gambar medis; ini membantu tenaga medis menemukan dan mendiagnosis penyakit dengan lebih cepat dan tepat [3].

Selama beberapa tahun terakhir, pendekatan berbasis *Convolutional Neural Network* (CNN) telah mendominasi penelitian di bidang ini. CNN dikenal efektif dalam menangkap pola lokal melalui mekanisme *convolution* dan *pooling* [4]. Namun demikian, keterbatasan utama *Convolutional Neural Network* (CNN) terletak pada kemampuannya yang terbatas dalam menangkap hubungan global antar fitur, karena bergantung pada *local receptive field* [5]. Hal ini menjadi tantangan dalam analisis citra medis yang sering kali membutuhkan pemahaman konteks global, terutama untuk mengidentifikasi pola anatomi yang kompleks dan tersebar.

Untuk mengatasi keterbatasan tersebut, arsitektur berbasis *Transformer* mulai diperkenalkan dalam domain *computer vision*, termasuk pada analisis citra medis [6]. *Transformer*, yang awalnya dikembangkan dalam bidang *natural language processing*, memiliki kemampuan unggul dalam memodelkan dependensi jangka panjang melalui mekanisme *self-attention* [7]. Mekanisme ini memungkinkan model untuk menangkap hubungan global antar bagian citra secara lebih efektif dibandingkan pendekatan konvensional.

Seiring perkembangannya, berbagai varian Transformer telah dikembangkan untuk meningkatkan kinerja pada tugas visi komputer, seperti *Vision Transformer* (ViT), *Data-efficient Image Transformer* (DeiT), serta *Swin Transformer* yang mengadopsi pendekatan *hirarkis* dan efisien [8]. Berbagai penelitian menunjukkan bahwa model berbasis Transformer memiliki potensi dalam meningkatkan performa klasifikasi citra medis, terutama dalam menangani kompleksitas fitur, variasi skala, serta hubungan spasial yang luas [9]. Meskipun demikian, penerapan *Transformer* dalam domain ini masih menghadapi beberapa tantangan, seperti kebutuhan data yang besar, kompleksitas komputasi, serta keterbatasan interpretabilitas model.

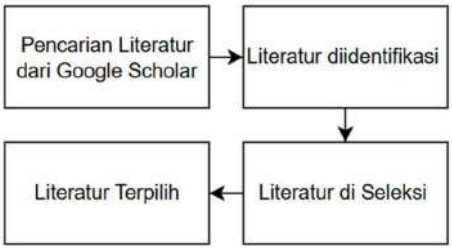
Berdasarkan hal tersebut, penelitian ini bertujuan untuk melakukan studi literatur mengenai penerapan arsitektur Transformer dalam klasifikasi citra medis. Kajian ini mencakup analisis terhadap berbagai jenis arsitektur *Transformer*, keunggulan dan keterbatasannya, serta perkembangan penelitian terkini di bidang tersebut. Dengan demikian, penelitian ini diharapkan dapat memberikan pemahaman komprehensif mengenai peran Transformer dalam meningkatkan kinerja sistem analisis citra medis berbasis kecerdasan buatan.

II. METODOLOGI

Penelitian ini merupakan studi literatur yang mengkaji dan merangkum berbagai sumber relevan terkait klasifikasi citra medis menggunakan arsitektur *Transformer*. Metodologi yang digunakan meliputi dua tahap utama, yaitu identifikasi dan seleksi literatur.

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

Tahap pertama dilakukan dengan mengidentifikasi literatur melalui Google Scholar, sehingga diperoleh sebanyak 20 referensi. Selanjutnya, pada tahap seleksi, literatur yang telah dikumpulkan disaring berdasarkan kesesuaian judul dengan topik penelitian. Selain itu, sumber yang digunakan dibatasi pada publikasi dalam kurun waktu lima tahun terakhir.



Gbr 1. Metodologi Penelitian

III. HASIL DAN PEMBAHASAN

Tabel 1. Daftar Literatur

Author	Variabel Independent	Variabel Dependent	Metode	Hasil
Ze Liu et al. (2021) [8]	Menggunakan mekanisme shifted window (SW-MSA), desain hierarchical feature, dan self-attention berbasis window lokal.	Akurasi klasifikasi (Top-1 ImageNet)	Vision Transformer berbasis hierarchical (Swin Transformer)	Model mencapai akurasi klasifikasi hingga 87.3% pada ImageNet-1K
Huang et al. (2022) [10]	Mekanisme shifted window, penggunaan residual Swin Transformer blocks, multi-channel loss dengan sensitivity maps, serta teknik undersampling pada k-space MRI.	Kualitas rekonstruksi citra MRI	SwinMR (*Swin Transformer-based MRI reconstruction*)	Rekonstruksi lebih akurat, robust terhadap noise dan undersampling, serta mendekati ground truth dengan kualitas tinggi
Yustiana et al. (2026)[11]	ViT-B16, fundus dataset, preprocessing, transfer learning, fine-tuning, Adam, learning rate 0.001/0.0001, batch size 16, epoch 15.	Accuracy, PrecisionRecal, F1-score, Confusion matrix	Vision Transformer	Performa model mencapai precision 93.3%, recall 92.8%, dan F1-score 93.0%, dengan kinerja terbaik pada diabetic retinopathy (100%) dan terendah pada glaukoma (87%).
Prastyo et al. (2025) [12]	Mekanisme shifted window (W-MSA/SW-MSA), dan hyperparameter	Metrik accuracy, precision, recall, dan F1-score.	*Swin Transformer*	Akurasi sekitar 99.57% – 99.85%, serta mengungguli CNN dan ViT
Yin et al. (2023)[13]	Deep learning CNN-Transformer dengan preprocessing dan optimasi pelatihan.	Metrik accuracy, recall, precision, dan AUC.	CNN dan Transformer (ViT, Swin)	akurasi 0.994 (validasi) dan 0.980 (test)
Hidayat et al. (2026) [14]	DenseNet-201, Swin Transformer, transfer learning, preprocessing, dan augmentasi citra	Metrik accuracy, precision, sensitivity (recall), F1-score, dan AUC.	DenseNet-201 dan Swin Transformer	Accuracy (88.05 %), sensitivity, F1-score, dan AUC (94.73 %)
Saddique et al. (2025) [15]	Hybrid CNN-Transformer dengan preprocessing, parameter	Metrik accuracy, precision, recall, F1-score,	Hybrid deep learning CNN dan Transformer	Accuracy 97.0%, precision 97.5%,

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

Author	Variabel Independent	Variabel Dependent	Metode	Hasil
	pelatihan, dan transfer learning	specificity, dan loss		recall 97.2%, F1-score 97.4%, specificity 98.0%, serta loss terendah (0.12).
Grace et al. (2025)[16]	Model CNN, ResNet50, dan ViT dengan preprocessing, Adam optimizer, serta dataset CBIS-DDSM	Accuracy, precision, recall, dan F1-score	CNN, ResNet50, dan ViT	ResNet 50Accuracy ≈ 97% (tertinggi) CNN : Accuracy ≈ 92–93% Vision Transformer (ViT) : Accuracy ≈ 71–75%
Wang et al. (2022)[17]	Mekanisme Swin Transformer, SSL, SRCL, dan dataset tanpa label.	Accuracy, F1-score, AUC, top-k accuracy, Dice score, dan F1-score deteksi mitosis.	Hybrid CNN + Swin Transformer	CTrans Path dengan SRCL mencapai performa unggul, melampaui SSL lain dan pretraining ImageNet, serta generalisasi lebih baik
Yan et al. (2022)[18]	MMTrans berbasis Swin GAN dengan komponen utama, self-attention, loss kompleks, dan dataset MRI beragam	terhadap kualitas, kontras, dan anatomi	(MMTrans)	, CycleGAN, RegGAN dengan PSNR, SSIM, MAE lebih baik mendekati ground truth
Bataineh et al. (2024) [19]	Transfer learning, preprocessing augmentasi, parameter pelatihan, dataset MRI	Accuracy, Recall, Precisio,F1-score	Swin Transformer dan ResNet50 V2 (SwT+ResNet50V2)	Akurasi tinggi hingga 99.9%, melampaui model lain dengan peningkatan signifikan kinerja klasifikasi
Ozer et al. (2022)[20]	Swin Transformer, preprocessing augmentasi, pretrained, parameter pelatihan, dataset medis	Accuracy, AUC (Area Under Curve), Kemampuan klasifikasi kualitas citra medis (baik/buruk, ada objek asing, dll)	Swin Transformer	Pada dataset Object-CXR: Accuracy mencapai 87.1%, AUC mencapai 0.922 Pada dataset LVOT MRI: Accuracy hingga 95.59%, Performa sebanding dengan CNN

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

Author	Variabel Independent	Variabel Dependent	Metode	Hasil
Üzen et al. (2024)[21]	ArsitekturSC-MP-Mixer: ConvMixer, Swin Transformer, multipath, parameter pelatihan, dataset BCCD, PBC, Raabin	Accuracy, Precisio, Recall, F1-score	SC-MP-Mixer : Multipath ConvMixer, Multipath Swin Transformer, Classification layer	BCCD dataset : Accuracy: 95.66% PBC dataset: Accuracy: 99.65% Raabin dataset: Accuracy: 98.68%
Halim et al. (2024)[22]	Hyperparameter tuning (GridSearch), augmentasi, preprocessing, dataset 12.500 citra sel darah putih, parameter learning rate, weight decay, heads, dan layer transformer.	Output berupa accuracy (training dan validation) serta performa klasifikasi model dalam membedakan jenis sel darah putih.	Vision Transformer	Training accuracy mencapai sekitar 98% dan validation accuracy 83.44%, menunjukkan performa baik dengan indikasi overfitting ringan.
Chen et al. (2023)[23]	Strategi training fine-tuning penuh, layer akhir, dan from scratch, dengan dataset BCCD_Dataset.	Accuracy, Precisio, Recall, F1-Score	Deep Learning (Vision Transformer – SW-ViT) + Transfer Learning	SW-ViT mencapai akurasi 98.03% dan F1-score 98.04%, lebih baik dibanding metode lain serta lebih efisien dalam waktu training
Mahajaya et al. (2024)[24]	Optimizer Adam, AdamW, SGD, LAMB dengan epoch 50, batch size 32, learning rate 0.0001, dataset citra X-ray paru-paru.	Accuracy, PrecisionRecall, F1-score, epoch	Vision Transformer	Optimizer Adam terbaik (akurasi 93.9%, F1-score 93.9%, waktu tercepat), diikuti AdamW, LAMB, dan SGD terendah.

A. Variabel

Variabel dalam klasifikasi citra medis berbasis deep learning terdiri dari variabel independen dan dependen. Variabel independen yang paling umum digunakan meliputi arsitektur model seperti *Convolutional Neural Network* (CNN) + *Transformer*, *Vision Transformer*, dan *Swin Transformer*, serta strategi pelatihan seperti preprocessing, augmentasi, transfer learning, dan pengaturan hyperparameter. Selain itu, jenis dataset seperti MRI, X-ray, fundus, dan citra sel darah juga menjadi faktor penting dalam pembentukan pola fitur.

Variabel dependen berupa performa model yang diukur menggunakan *accuracy*, *precision*, *recall*, *F1-score*, dan *AUC*. Semakin baik pengolahan variabel independen, maka semakin tinggi performa model dalam melakukan klasifikasi citra medis.

B. Metode

Tabel 2. Daftar Metode

No	Metode yang Digunakan	Jumlah
1	Swin Transformer	3
2	Vision Transformer (ViT)	3
3	CNN + Transformer (ViT/Swin)	2
4	CNN + Transformer (Hybrid)	1
5	CNN + ResNet50 + Vision Transformer	1
6	CNN + Swin Transformer (CTransPath)	1
7	DenseNet-201 + Swin Transformer	1
8	Swin Transformer + ResNet50V2	1
9	ConvMixer + Swin Transformer	1
10	GAN berbasis Swin Transformer (MMTrans)	1
11	SW-ViT (Swin Vision Transformer)	1
12	SwinMR (Swin Transformer MRI)	1

ISSN (Online) 2747-0563
Volume 6 Tahun 2026

Berdasarkan Tabel 2, *Swin Transformer* dan *Vision Transformer* (ViT) merupakan metode yang paling sering digunakan dalam studi literatur yang sudah di dapatkan. Hal ini disebabkan karena kedua metode tersebut mampu menangkap hubungan global antar bagian citra melalui mekanisme *self-attention*, sehingga lebih efektif dalam memahami struktur kompleks pada citra medis dibandingkan *Convolutional Neural Network* (CNN) yang cenderung fokus pada fitur lokal.

Berdasarkan keseluruhan literatur yang telah dianalisis, terlihat bahwa performa model klasifikasi citra medis sangat dipengaruhi oleh kombinasi antara arsitektur model, strategi pelatihan, serta karakteristik dataset yang digunakan. Secara umum, metode berbasis *Transformer*, khususnya *Swin Transformer* dan *Vision Transformer* (ViT), menunjukkan performa yang konsisten lebih tinggi dibandingkan metode konvensional seperti *Convolutional Neural Network* (CNN). Hal ini ditunjukkan oleh sebagian besar penelitian yang menghasilkan nilai akurasi, precision, recall, dan F1-score yang berada pada kategori tinggi hingga sangat tinggi, bahkan pada beberapa kasus mencapai di atas 99%.

Secara lebih mendalam, hasil tersebut dapat dijelaskan dari sisi kemampuan representasi fitur. Model *Transformer* mampu menangkap hubungan global antar bagian citra melalui mekanisme *self-attention*, sehingga lebih efektif dalam memahami pola kompleks yang tersebar pada citra medis. Hal ini berbeda dengan *Convolutional Neural Network* (CNN) yang cenderung terbatas pada ekstraksi fitur lokal, sehingga pada kasus tertentu kurang mampu menangkap keterkaitan antar area citra yang berjauhan. Oleh karena itu, pada dataset dengan kompleksitas tinggi seperti MRI, X-ray, dan citra sel darah, *Transformer* cenderung menghasilkan performa yang lebih baik.

Jika dibandingkan antar varian *Transformer*, *Swin Transformer* menunjukkan keunggulan yang lebih konsisten dibandingkan *Vision Transformer*. Hal ini terlihat dari beberapa penelitian yang menunjukkan bahwa *Swin Transformer* mampu mengungguli *Convolutional Neural Network* (CNN) dan ViT baik dari segi akurasi maupun stabilitas performa. Keunggulan ini disebabkan oleh penggunaan mekanisme *shifted window* dan *struktur hierarkis*, yang memungkinkan model melakukan ekstraksi fitur secara bertingkat (*multi-scale*) serta menjaga efisiensi komputasi. Dengan demikian, *Swin Transformer* tidak hanya unggul dalam akurasi, tetapi juga dalam kemampuan generalisasi terhadap berbagai jenis dataset medis.

Selain itu, hasil literatur juga menunjukkan bahwa performa model tidak hanya ditentukan oleh arsitektur, tetapi juga oleh strategi pelatihan yang digunakan. Penelitian yang menerapkan *transfer learning*, *fine-tuning*, augmentasi data, serta pengaturan *hyperparameter* yang optimal cenderung menghasilkan performa yang lebih tinggi. Hal ini menunjukkan bahwa keberhasilan model *Transformer* sangat bergantung pada bagaimana model tersebut dilatih dan disesuaikan dengan karakteristik dataset. Pada beberapa kasus, meskipun menggunakan arsitektur yang sama, perbedaan strategi pelatihan dapat menghasilkan performa yang cukup signifikan.

Di sisi lain, pendekatan hybrid yang mengombinasikan CNN dan *Transformer* juga menunjukkan hasil yang sangat kompetitif. Model hybrid mampu menggabungkan keunggulan CNN dalam menangkap fitur lokal dengan keunggulan *Transformer* dalam memahami hubungan global. Hal ini terlihat pada beberapa penelitian yang menghasilkan akurasi di atas 97%, yang menunjukkan bahwa integrasi kedua arsitektur dapat menjadi solusi optimal untuk meningkatkan performa klasifikasi citra medis.

Namun demikian, tidak semua hasil menunjukkan performa yang sempurna. Beberapa penelitian juga mengindikasikan adanya potensi *overfitting*, terutama pada *Vision Transformer* yang dilatih pada dataset terbatas. Hal ini menunjukkan bahwa meskipun *Transformer* memiliki kemampuan yang tinggi, model ini tetap memerlukan jumlah data yang cukup besar agar dapat bekerja secara optimal. Oleh karena itu, penggunaan teknik seperti augmentasi data dan *transfer learning* menjadi sangat penting untuk menjaga keseimbangan antara performa dan generalisasi model.

IV. KESIMPULAN

Berdasarkan hasil studi literatur yang telah dilakukan , dapat disimpulkan bahwa arsitektur berbasis *Transformer*, khususnya *Vision Transformer* (ViT) dan *Swin Transformer*, menunjukkan performa yang unggul dalam klasifikasi citra medis dibandingkan pendekatan konvensional seperti *Convolutional Neural Network* (CNN). Hal ini dibuktikan melalui berbagai penelitian yang menunjukkan nilai akurasi tinggi, bahkan mencapai rentang 99.57%–99.85% pada Swin Transformer, serta performa F1-score hingga 97.4% pada model hybrid CNN–*Transformer*. *Swin Transformer* secara konsisten menunjukkan kinerja yang lebih stabil dan efisien dibandingkan ViT, berkat struktur hierarkis dan mekanisme *shifted window* yang memungkinkan ekstraksi fitur multi-skala secara optimal. Selain itu, *Vision Transformer* juga menunjukkan performa yang baik dengan nilai precision hingga 93.3%, recall 92.8%, dan F1-score 93.0%, meskipun pada beberapa kasus memerlukan dataset yang lebih besar untuk menghindari overfitting. Secara keseluruhan, hasil literatur menunjukkan bahwa sebagian besar model berbasis Transformer mampu mencapai akurasi di atas 90% dan bahkan mendekati 99%, yang menegaskan keunggulan metode ini dalam menangani kompleksitas citra medis. Namun demikian, performa model tidak hanya dipengaruhi oleh arsitektur, tetapi juga sangat bergantung pada strategi pelatihan seperti preprocessing, augmentasi data, transfer learning, serta pengaturan hyperparameter.Oleh karena itu, penelitian selanjutnya disarankan untuk mengoptimalkan kombinasi arsitektur dan strategi pelatihan guna meningkatkan generalisasi model serta efisiensi komputasi dalam implementasi sistem diagnosis berbasis citra medis.

REFERENSI

[1] L. Maharani and I. Komputer, "KECERDASAN BUATAN DALAM DIAGNOSTIK MEDIS DAN PERAWATAN KESEHATAN," *Logicloom.id*, vol. 1, no. 7, 2024.
[2] S. H. Gulo, "PENERAPAN TEKNIK DEEP LEARNING DALAM PENGENALAN POLA CITRA MEDIS," *informatika*, vol. 1, no. 2, May 2024.
[3] A. Syakuroh, F. Monado, M. Ariani, Hadi, E. Koriyanti, and Erni, "ANALISIS AKURASI MODEL MOBILENETV2 DALAM KLASIFIKASI CITRA X-RAY UNTUK DETEKSI KONDISI PARU-PARU," *Journal Online of Physics*, vol. 10, no. 3, pp. 67–74, Jul. 2025.
[4] N. Puspita Sari, "Analisis Performa Algoritma CNN dalam Klasifikasi Citra Medis Berbasis Deep Learning Analysis Of CNN Algorithm In Deep Learning-Based Medical Image Classification," *Jurnal Komputer*, vol. 2, no. 2, pp. 87–92, 2024.
[5] J. Homepage, P. Dhiyaul, H. Aq, and B. Irawan, "Penerapan Vision T ransformer Untuk Klasifkasi Sampah R umah T angga," *Journal of Innovative and Creativity*, vol. 6, no. 1, p. 2026, Feb. 2026.
[6] S. Takahashi *et al.*, "Comparison of Vision Transformers and Convolutional Neural Networks in Medical Image Analysis: A Systematic Review," Dec. 01, 2024, *Springer*. doi: 10.1007/s10916-024-02105-8.
[7] S. J. Grace and D. Gunawan, "PERBANDINGAN CNN, RESNET50, DAN VISION TRANSFORMER UNTUK KLASIFIKASI KANKER PAYUDARA BERBASIS WEB," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 945–956, Jul. 2025, doi: 10.36341/rabit.v10i2.6420.
[8] Z. Liu *et al.*, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," Aug. 2021, [Online]. Available: http://arxiv.org/abs/2103.14030
[9] M. A. Alohali, N. El-Rashidy, S. Alaklabi, H. Elmannai, S. Alharbi, and H. Saleh, "Swin-GA-RF: genetic algorithm-based Swin Transformer and random forest for enhancing cervical cancer classification," *Front. Oncol.*, vol. 14, 2024, doi: 10.3389/fonc.2024.1392301.
[10] J. Huang *et al.*, "Swin transformer for fast MRI," *Neurocomputing*, vol. 493, pp. 281–304, Jul. 2022, doi: 10.1016/j.neucom.2022.04.051.
[11] I. Yustiana, I. Yudistiansyah, and ; M Ikhsan Thohir, "The Implementation Of Computer Vision For Eye Disease Classification Using Vision Transformer Architecture On Fundus Images," *Jurnal Media Computer Science*, vol. 5, no. 1, pp. 385–400, Jan. 2026.
[12] A. B. Prasetyo, F. T. Anggraeny, and R. Mumpuni, "Kidney Stone Disease Diagnosis Using Shifted-Windows Transformer (Swin Transformer)," *Jati Emas (Jurnal Aplikasi Teknik dan Pengabdian Masyarakat)*, vol. 9, pp. 419–424, Oct. 2025.
[13] M. Yin *et al.*, "Identification of Asymptomatic COVID-19 Patients on Chest CT Images Using Transformer-Based or Convolutional Neural Network–Based Deep Learning Models," *J. Digit. Imaging*, vol. 36, no. 3, pp. 827–836, Jun. 2023, doi: 10.1007/s10278-022-00754-0.
[14] D. Hidayat, A. Musyafa, and M. Handayani, "A Comparative Study of DenseNet-201 and Swin Transformer for Malignant and Benign Skin Lesion Classification," *Jurnal Teknologi Informatika dan Komputer*, vol. 12, no. 1, pp. 169–184, Jan. 2026, doi: 10.37012/jtik.v12i1.3265.
[15] A. Saddique, A. Manan, M. Ali, S. Siddiqui, and M. Rehan, "HYBRID DEEP LEARNING MODELS FOR MULTI-CLASS CLASSIFICATION OF CHEST X-RAY IMAGES: NORMAL, PNEUMONIA, AND COVID-19," *Spectrum of Engineering Sciences*, vol. 3, no. 7, 2025, doi: 10.5281/zenodo.15790393.
[16] S. J. Grace and D. Gunawan, "PERBANDINGAN CNN, RESNET50, DAN VISION TRANSFORMER UNTUK KLASIFIKASI KANKER PAYUDARA BERBASIS WEB," *Rabit : Jurnal Teknologi dan Sistem Informasi Univrab*, vol. 10, no. 2, pp. 945–956, Jul. 2025, doi: 10.36341/rabit.v10i2.6420.
[17] X. Wang *et al.*, "Transformer-based unsupervised contrastive learning for histopathological image classification," *Med. Image Anal.*, vol. 81, Oct. 2022, doi: 10.1016/j.media.2022.102559.
[18] S. Yan, C. Wang, W. Chen, and J. Lyu, "Swin transformer-based GAN for multi-modal medical image translation," *Front. Oncol.*, vol. 12, Aug. 2022, doi: 10.3389/fonc.2022.942511.
[19] A. F. Al Bataineh *et al.*, "Enhanced Magnetic Resonance Imaging-Based Brain Tumor Classification with a Hybrid Swin Transformer and ResNet50V2 Model," *Applied Sciences (Switzerland)*, vol. 14, no. 22, Nov. 2024, doi: 10.3390/app142210154.
[20] C. Ozer, A. Guler, A. T. Cansever, D. Alis, E. Karaarslan, and I. Oksuz, "Shifted Windows Transformers for Medical Image Quality Assessment," in *Machine Learning in Medical Imaging - 13th International Workshop, MLMI 2022, Held in Conjunction with MICCAI 2022, Proceedings*, Singapore: PublisherSpringer Science and Business Media Deutschland GmbH, Aug. 2022, pp. 425–435. [Online]. Available: http://arxiv.org/abs/2208.06034
[21] H. Üzen and H. Fırat, "A hybrid approach based on multipath Swin transformer and ConvMixer for white blood cells classification," *Health Inf. Sci. Syst.*, vol. 12, no. 1, Dec. 2024, doi: 10.1007/s13755-024-00291-w.
[22] J. Daud Halim and Rizal, "Klasifikasi Sel Darah Putih Menggunakan Vision Transformer (ViT)," *Jurnal Strategi*, vol. 6, no. 2, Nov. 2024.
[23] S. Chen, S. Lu, S. Wang, Y. Ni, and Y. Zhang, "Shifted Window Vision Transformer for Blood Cell Classification," *Electronics (Switzerland)*, vol. 12, no. 11, Jun. 2023, doi: 10.3390/electronics12112442.
[24] N. Sarasuartha Mahajaya, P. Desiana, W. Ayu, and R. R. Huizen, "Pengaruh Optimizer Adam, AdamW, SGD, dan LAMB terhadap Model Vision Transformer pada Klasifikasi Penyakit Paru-paru," in *Prosiding Seminar Hasil Penelitian Informatika dan Komputer (SPINTER) 2024*, Denpasar, Indonesia: Institut Teknologi dan Bisnis STIKOM Bali, Apr. 2024, pp. 818–823. [Online]. Available: https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database,