

Deep learning-based object detection application for medical images using the vision transformer (ViT) method

Ade Rinanda Wahyuni Hasibuan¹, Fatma Sari Hutagalung²

^{1,2}Universitas Muhammadiyah Sumatera Utara, Indonesia

Email: wahyunihasibuanaderinanda@gmail.com; fatmasari@umsu.ac.id

ABSTRAK

Stroke iskemik merupakan salah satu kondisi medis yang memerlukan proses diagnosis secara cepat dan akurat karena dapat menyebabkan kerusakan otak permanen. Citra medis, seperti CT-Scan, dapat digunakan sebagai salah satu sumber informasi untuk membantu tenaga medis dalam mengidentifikasi kondisi otak pasien. Namun, analisis citra secara manual membutuhkan waktu dan dapat dipengaruhi oleh keterbatasan manusia dalam melakukan interpretasi. Penelitian ini bertujuan mengembangkan aplikasi klasifikasi citra CT-Scan untuk membantu mengidentifikasi kondisi stroke iskemik menggunakan metode *Vision Transformer* (ViT). Tahapan penelitian meliputi *data preprocessing*, *patch embedding*, *positional encoding*, *transformer encoder*, dan *classification head* untuk mengklasifikasikan citra ke dalam dua kategori, yaitu *stroke* dan *normal*. Dataset yang digunakan berupa citra CT-Scan pasien yang diperoleh dari Rumah Sakit Hendra, Binjai. Implementasi sistem dilakukan menggunakan bahasa pemrograman Python dan *framework* FastAPI. Hasil pengujian menunjukkan bahwa model ViT memperoleh akurasi sebesar 82,93% pada pengujian akhir. Hasil tersebut menunjukkan bahwa model ViT dapat digunakan untuk melakukan klasifikasi citra CT-Scan ke dalam kategori *stroke* dan *normal* pada dataset yang digunakan. Penelitian ini dapat menjadi salah satu pendekatan dalam pengembangan sistem pendukung keputusan berbasis *deep learning* untuk membantu proses analisis citra medis.

Kata Kunci: deep learning, vision transformer (vit), citra medis, stroke iskemik

ABSTRACT

Ischemic stroke is a medical condition that requires rapid and accurate diagnosis because it can cause permanent brain damage. Medical images, such as CT scans, can serve as a source of information to assist healthcare professionals in identifying patients' brain conditions. However, manual image analysis is time-consuming and may be affected by human limitations in image interpretation. This study aims to develop a CT scan image classification application to assist in identifying ischemic stroke using the Vision Transformer (ViT) method. The research stages include data preprocessing, patch embedding, positional encoding, transformer encoder, and classification head to classify images into two categories: stroke and normal. The dataset consists of patient CT scan images obtained from Hendra Hospital, Binjai. The system was implemented using the Python programming language and the FastAPI framework. The evaluation results show that the ViT model achieved an accuracy of 82.93% in the final testing. These results indicate that the ViT model can be used to classify CT scan images into stroke and normal categories on the dataset used in this study. This research may serve as an approach for developing deep learning-based decision support systems to assist in the analysis of medical images.

Keywords: deep learning; vision transformer (vit); medical images; ischemic stroke

Corresponding Author:

Ade Rinanda Wahyuni Hasibuan,
Universitas Muhammadiyah Sumatera Utara,
Jl. Kapten Muchtar Basri No.3, Glugur Darat II, Kec. Medan Tim., Kota
Medan, Sumatera Utara 20238, Indonesia
Email: wahyunihasibuanaderinanda@gmail.com



1. INTRODUCTION

Stroke iskemik merupakan kondisi medis yang terjadi akibat terhambatnya aliran darah menuju jaringan otak sehingga pasokan oksigen dan nutrisi yang dibutuhkan oleh sel-sel otak menjadi berkurang. Gangguan tersebut dapat menyebabkan perubahan atau kerusakan pada jaringan otak sehingga diperlukan proses diagnosis dan penanganan yang tepat. Pemeriksaan pencitraan medis merupakan salah satu bagian penting dalam evaluasi pasien stroke karena dapat memberikan informasi mengenai kondisi jaringan otak. Salah satu modalitas pencitraan yang banyak digunakan dalam pemeriksaan stroke adalah *Computed Tomography* (CT-Scan) kepala karena dapat memberikan gambaran struktur otak dan membantu proses identifikasi kondisi patologis pada pasien stroke (Juniada et al., 2022; Yusastra et al., 2021).

Pada kasus stroke iskemik, citra CT-Scan dapat digunakan untuk membantu memperoleh informasi diagnostik mengenai kondisi otak pasien. Kualitas informasi yang diperoleh dari pemeriksaan CT-Scan dapat dipengaruhi oleh berbagai faktor teknis, seperti *slice thickness* dan *interval reconstruction*, yang dapat memengaruhi tampilan struktur pada citra (Yan et al., 2020). Selain itu, pengaturan parameter citra seperti *window width* dan *window level* juga dapat memengaruhi visualisasi struktur dan karakteristik jaringan pada hasil CT-Scan (Puspita et al., 2025). Penelitian mengenai temuan CT-Scan pada pasien stroke juga menunjukkan bahwa pemeriksaan CT-Scan kepala dapat memberikan informasi mengenai karakteristik pencitraan yang ditemukan pada pasien dengan kondisi stroke (Susanto & Sari, 2025).

Interpretasi citra CT-Scan memerlukan kemampuan untuk mengenali karakteristik dan perubahan pada struktur otak berdasarkan informasi visual yang tersedia. Dalam praktiknya, analisis citra medis membutuhkan pengetahuan dan ketelitian sehingga pengembangan teknologi berbasis komputer dapat digunakan sebagai pendekatan untuk membantu proses analisis citra. Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence*) dan *Deep Learning* memberikan peluang untuk mengembangkan sistem yang mampu melakukan analisis dan klasifikasi citra secara otomatis. Penerapan *Deep Learning* pada citra medis telah dilakukan pada berbagai objek, termasuk citra *chest X-ray* dan citra mammografi, untuk membantu proses klasifikasi kondisi medis (Maulana et al., 2025; Ardiyansyah et al., 2025).

Salah satu arsitektur *Deep Learning* yang berkembang dalam bidang klasifikasi citra adalah *Vision Transformer* (ViT). ViT mengadaptasi mekanisme *Transformer* yang sebelumnya banyak digunakan dalam pemrosesan bahasa alami (*Natural Language Processing*) ke dalam pengolahan citra. Berbeda dengan pendekatan CNN yang menggunakan operasi konvolusi untuk mengekstraksi fitur lokal, ViT merepresentasikan citra sebagai rangkaian *patch* yang kemudian diubah menjadi representasi token untuk diproses menggunakan mekanisme *Transformer Encoder* (Dosovitskiy et al., 2021). Mekanisme tersebut memungkinkan model mempelajari hubungan antara bagian-bagian citra melalui mekanisme *self-attention* sehingga informasi dari berbagai wilayah citra dapat dipertimbangkan dalam proses klasifikasi (Putriany & Ariatmanto, 2025).

Penerapan ViT untuk klasifikasi citra telah dilakukan pada berbagai bidang. Pangestu et al. (2024), misalnya, menerapkan ViT untuk melakukan klasifikasi tingkat kematangan buah dan menunjukkan penggunaan arsitektur *Transformer* sebagai pendekatan dalam klasifikasi citra. Penelitian lain juga menerapkan ViT untuk klasifikasi citra medis, seperti klasifikasi pneumonia berdasarkan citra *chest X-ray* (Maulana et al., 2025) dan klasifikasi kanker payudara menggunakan citra mammografi (Ardiyansyah et al., 2025). Pengembangan model berbasis ViT juga dilakukan pada berbagai permasalahan klasifikasi citra lainnya, yang menunjukkan bahwa arsitektur tersebut dapat digunakan sebagai salah satu pendekatan dalam pengembangan sistem klasifikasi berbasis citra (Firdaus et al., 2025). Hal tersebut menunjukkan adanya peluang penerapan ViT pada permasalahan klasifikasi citra medis, termasuk citra CT-Scan otak.

Dalam konteks stroke, penerapan arsitektur berbasis *Transformer* pada citra CT-Scan telah mulai dikembangkan. Penelitian mengenai klasifikasi dan lokalisasi stroke berbasis CT-Scan menggunakan arsitektur *Vision Transformer* menunjukkan bahwa pendekatan berbasis *Transformer* dapat digunakan untuk menganalisis citra CT-Scan otak pada kasus stroke. Pendekatan tersebut menunjukkan potensi penggunaan mekanisme *attention* untuk memanfaatkan informasi dari berbagai bagian citra dalam proses klasifikasi. Selain itu, penelitian terbaru mengenai klasifikasi stroke berbasis CT-Scan dengan model berbasis *Transformer* juga menunjukkan perkembangan penggunaan arsitektur *Transformer* dalam mendukung analisis citra stroke (Qari & Thafar, 2025). Perkembangan tersebut menjadi dasar bahwa penerapan ViT pada citra CT-Scan otak dapat diteliti lebih lanjut untuk mengembangkan sistem klasifikasi yang dapat membantu proses analisis citra.

Meskipun demikian, penerapan ViT secara khusus untuk klasifikasi citra CT-Scan otak ke dalam kategori *normal* dan *stroke* masih memiliki ruang untuk dikembangkan. Perbedaan karakteristik dataset, jumlah data, metode prapemrosesan, konfigurasi model, serta strategi pelatihan dapat memengaruhi kinerja model yang dihasilkan. Oleh karena itu, diperlukan penelitian yang mengimplementasikan ViT pada dataset CT-Scan otak dan melakukan evaluasi terhadap kemampuan model dalam membedakan citra normal dan citra stroke. Evaluasi kinerja model dapat dilakukan menggunakan metrik klasifikasi seperti *accuracy*, *precision*, *recall*,

dan *F1-score* serta *confusion matrix* untuk mengetahui kemampuan model dalam mengidentifikasi setiap kategori.

Berdasarkan permasalahan tersebut, penelitian ini berfokus pada penerapan *Vision Transformer* (ViT) untuk melakukan klasifikasi citra CT-Scan otak ke dalam dua kategori, yaitu *normal* dan *stroke*. Sistem yang dikembangkan berbentuk aplikasi berbasis web sehingga hasil klasifikasi model dapat ditampilkan melalui antarmuka yang dapat diakses oleh pengguna. Pengujian dilakukan menggunakan dataset citra CT-Scan untuk mengevaluasi kemampuan model dalam mengklasifikasikan citra berdasarkan kategori yang telah ditentukan. Kinerja model selanjutnya dievaluasi berdasarkan metrik klasifikasi yang digunakan dalam penelitian.

Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem pendukung analisis citra medis berbasis *Deep Learning*, khususnya dalam penerapan *Vision Transformer* untuk klasifikasi citra CT-Scan pada kasus *stroke*. Sistem yang dikembangkan tidak dimaksudkan untuk menggantikan proses diagnosis yang dilakukan oleh tenaga medis, tetapi sebagai pendekatan berbasis komputasi yang dapat digunakan untuk mendukung proses analisis dan klasifikasi citra medis.

2. RESEARCH METHOD

Penelitian ini menggunakan pendekatan kuantitatif dan kualitatif. Pendekatan kuantitatif digunakan untuk proses pelatihan, pengujian, dan evaluasi kinerja model dalam melakukan klasifikasi citra medis, sedangkan pendekatan kualitatif digunakan untuk memperoleh informasi melalui observasi dan wawancara di Rumah Sakit Hendra, Binjai. Tahapan penelitian meliputi pengumpulan data, pra-pemrosesan data (*data preprocessing*), perancangan model *Vision Transformer* (ViT), pelatihan model (*training*), serta pengujian dan evaluasi hasil klasifikasi.

Penelitian dilaksanakan di Rumah Sakit Hendra, Binjai, Sumatera Utara, selama tiga bulan, yaitu dari Desember 2024 hingga Februari 2025. Data yang digunakan merupakan data sekunder yang diperoleh dari rekam medis pasien *stroke* iskemik berupa citra CT-Scan otak. Dataset diklasifikasikan ke dalam dua kategori, yaitu *Stroke* dan *Normal*. Selanjutnya, dataset dibagi menjadi dua subset, yaitu data latih (*train*) dan data validasi (*validation*) dengan perbandingan 80:20. Data latih digunakan untuk proses pembelajaran model, sedangkan data validasi digunakan untuk mengevaluasi kinerja model terhadap data yang tidak digunakan secara langsung dalam proses pembaruan parameter.

Metode *Vision Transformer* (ViT) bekerja dengan memanfaatkan mekanisme *self-attention* dan *patch embedding*. Citra berukuran 224×224 piksel dibagi menjadi *patch* berukuran 16×16 piksel sehingga menghasilkan 196 *patch*. Setiap *patch* kemudian direpresentasikan dalam bentuk vektor berdimensi 768 dan dikombinasikan dengan *positional encoding* untuk memberikan informasi mengenai posisi masing-masing *patch* dalam citra. Selanjutnya, seluruh representasi *patch* diproses melalui *Transformer Encoder* yang terdiri atas *Multi-Head Self-Attention* (MHSA) dan *Feed Forward Network* (FFN) untuk menghasilkan representasi fitur yang digunakan dalam proses klasifikasi.

Sistem dikembangkan menggunakan bahasa pemrograman Python dengan *framework* FastAPI serta beberapa *library* pendukung, yaitu PyTorch, TorchVision, dan Timm. Dataset diorganisasikan ke dalam direktori *train* dan *validation* berdasarkan kategori klasifikasi. Model dilatih selama 5 *epochs* dengan *batch size* sebesar 8, kemudian hasil pelatihan disimpan pada direktori *checkpoints*. Antarmuka aplikasi dikembangkan dalam bentuk aplikasi berbasis web yang memungkinkan pengguna mengunggah citra CT-Scan dan memperoleh hasil klasifikasi berdasarkan model yang telah dilatih.

3. RESULTS AND DISCUSSION

A. Implementasi Sistem

Aplikasi klasifikasi citra medis dikembangkan dalam bentuk aplikasi berbasis web menggunakan *framework* FastAPI. Aplikasi ini memungkinkan pengguna, khususnya tenaga medis, untuk mengunggah citra CT-Scan otak dan memperoleh hasil klasifikasi secara otomatis. Implementasi sistem dilakukan melalui beberapa tahapan, meliputi instalasi *library* Python, penyiapan dan pemrosesan dataset, pelatihan model (*training*), serta pengujian aplikasi (*testing*).

Spesifikasi perangkat keras yang digunakan dalam penelitian ini meliputi RAM sebesar 4 GB, kapasitas penyimpanan sebesar 80 GB, dan prosesor Intel. Perangkat lunak yang digunakan terdiri atas Python versi 3.10, Notepad++, serta beberapa *library* pendukung *Deep Learning*, yaitu *torch*, *torchvision*, *timm*, *numpy*, *Pillow*, dan *scikit-learn*. Dataset terdiri atas dua kelas, yaitu *Stroke* dan *Normal*, yang masing-masing diorganisasikan ke dalam *train set* dan *validation set*. *Train set* digunakan untuk proses pelatihan model, sedangkan *validation set* digunakan untuk mengevaluasi kinerja model terhadap data yang tidak digunakan secara langsung dalam proses pembaruan parameter model.

B. Metode Pengumpulan Data

Model *Vision Transformer* (ViT) dibangun dengan ukuran citra input 224×224 piksel. Pada arsitektur ViT, citra dapat direpresentasikan sebagai rangkaian *patch* berukuran tertentu yang kemudian diubah menjadi

representasi vektor untuk diproses oleh *Transformer Encoder*. Konsep tersebut merupakan karakteristik utama ViT dalam melakukan klasifikasi citra (Dosovitskiy et al., 2021). Pada penelitian ini, citra berukuran 224×224 piksel dibagi menggunakan *patch size* 16×16 piksel sehingga menghasilkan 196 *patch*. Setiap *patch* direpresentasikan dalam bentuk vektor berdimensi 768 dan selanjutnya diproses dengan *positional encoding* untuk memberikan informasi mengenai posisi masing-masing *patch* dalam citra. Penggunaan pendekatan berbasis *patch* dan *positional encoding* memungkinkan informasi spasial dari citra tetap dipertimbangkan dalam proses pemodelan menggunakan Transformer (Dosovitskiy et al., 2021).

Dalam penelitian ini, proses pelatihan model menggunakan *optimizer* Adam dengan *learning rate* sebesar 0,0001 dan *loss function cross-entropy*. Proses pelatihan dilakukan selama 5 *epochs* dengan *batch size* sebesar 8, menggunakan 80% data sebagai data latih dan 20% sebagai data validasi. Konfigurasi tersebut merupakan parameter eksperimen yang ditetapkan dalam penelitian untuk memperoleh model yang dapat melakukan klasifikasi terhadap dataset CT-Scan yang digunakan. Penggunaan ViT sebagai model klasifikasi citra telah diterapkan pada berbagai permasalahan klasifikasi dengan konfigurasi model dan dataset yang berbeda, sehingga parameter pelatihan perlu disesuaikan dengan karakteristik dataset dan tujuan penelitian (Pangestu et al., 2024). Selama proses pelatihan, perubahan nilai akurasi dan *loss* diamati untuk mengevaluasi perkembangan kinerja model pada setiap *epoch*.

Penerapan ViT pada penelitian berbasis citra medis juga telah dilakukan pada beberapa jenis data. Maulana et al. (2025) menerapkan ViT untuk klasifikasi penyakit pneumonia menggunakan citra *chest X-ray*, sedangkan Ardiyansyah et al. (2025) menggunakan ViT untuk klasifikasi kanker payudara berdasarkan citra mammografi. Kedua penelitian tersebut menunjukkan bahwa ViT dapat digunakan sebagai pendekatan *Deep Learning* dalam klasifikasi citra medis, meskipun karakteristik citra dan permasalahan klasifikasinya berbeda dengan penelitian ini (Maulana et al., 2025; Ardiyansyah et al., 2025). Selain itu, penerapan ViT pada klasifikasi citra dengan objek yang berbeda juga telah dilakukan oleh Pangestu et al. (2024), yang menggunakan beberapa varian ViT untuk klasifikasi tingkat kematangan buah pisang. Hal tersebut menunjukkan bahwa konfigurasi model, dataset, dan skenario pengujian merupakan faktor yang perlu diperhatikan ketika menerapkan ViT pada suatu permasalahan klasifikasi citra (Pangestu et al., 2024).

Berdasarkan hasil pengujian, model ViT memperoleh akurasi sebesar 82,93% pada pengujian akhir dalam mengklasifikasikan citra CT-Scan ke dalam dua kategori, yaitu *Stroke* dan *Normal*. Hasil tersebut menunjukkan bahwa model mampu melakukan klasifikasi pada dataset yang digunakan dalam penelitian. Nilai akurasi tersebut merupakan hasil eksperimen pada dataset dan konfigurasi yang digunakan dalam penelitian ini sehingga tidak dapat secara langsung digeneralisasikan pada dataset CT-Scan lainnya. Selain itu, nilai akurasi tersebut belum dapat digunakan untuk menyimpulkan bahwa ViT lebih unggul dibandingkan metode konvensional karena penelitian ini tidak melakukan pengujian komparatif menggunakan model pembanding pada dataset dan kondisi eksperimen yang sama.

Hasil penelitian ini dapat dibandingkan secara deskriptif dengan beberapa penelitian terdahulu yang menerapkan ViT pada bidang pengolahan citra medis. Maulana et al. (2025) menerapkan ViT pada klasifikasi pneumonia menggunakan citra *chest X-ray* dan melaporkan akurasi sebesar 80% pada konfigurasi *learning rate* tertentu. Sementara itu, Ardiyansyah et al. (2025) menerapkan ViT pada klasifikasi kanker payudara menggunakan citra mammografi. Perbedaan karakteristik dataset dan permasalahan klasifikasi menyebabkan hasil akurasi dari penelitian tersebut tidak dapat digunakan sebagai pembanding langsung terhadap akurasi penelitian ini (Maulana et al., 2025; Ardiyansyah et al., 2025). Penelitian Pangestu et al. (2024) juga menunjukkan bahwa kinerja ViT dapat berbeda berdasarkan varian model yang digunakan, dengan penelitian tersebut memperoleh akurasi tertinggi 91,61% pada salah satu varian ViT untuk klasifikasi kematangan pisang (Pangestu et al., 2024). Oleh karena itu, nilai akurasi sebesar 82,93% dalam penelitian ini menunjukkan kinerja model pada dataset CT-Scan yang digunakan, tetapi tidak secara langsung menunjukkan bahwa model lebih baik atau lebih buruk dibandingkan hasil penelitian sebelumnya.

Model yang dikembangkan juga memiliki beberapa keterbatasan. Ukuran dan keragaman dataset dapat memengaruhi kemampuan model dalam melakukan generalisasi terhadap data di luar dataset penelitian. Penelitian sebelumnya juga menunjukkan bahwa evaluasi ViT dapat dilakukan pada dataset yang berbeda dari dataset pelatihan untuk melihat kemampuan generalisasi model (Pangestu et al., 2024). Selain itu, penerapan ViT perlu mempertimbangkan kebutuhan komputasi karena model Transformer melakukan pemrosesan terhadap representasi *patch* citra. Pengembangan ViT pada berbagai penelitian menunjukkan bahwa pemilihan arsitektur, konfigurasi model, dan karakteristik dataset merupakan aspek yang perlu dipertimbangkan dalam proses pengembangan model (Dosovitskiy et al., 2021; Pangestu et al., 2024).

Penelitian selanjutnya dapat mempertimbangkan penggunaan dataset yang lebih besar dan beragam serta melakukan pengujian dengan data eksternal untuk mengevaluasi kemampuan generalisasi model. Penggunaan model pembanding dan evaluasi dengan metrik tambahan, seperti *precision*, *recall*, *F1-score*, dan *confusion matrix*, juga dapat dilakukan untuk memberikan gambaran kinerja klasifikasi yang lebih

komprehensif. Selain itu, penelitian lebih lanjut dapat mengeksplorasi variasi arsitektur dan konfigurasi ViT untuk mengetahui pengaruh parameter model terhadap kinerja klasifikasi citra CT-Scan.

4. CONCLUSION

Penelitian ini menghasilkan aplikasi berbasis *Deep Learning* dengan metode *Vision Transformer* (ViT) yang digunakan untuk melakukan klasifikasi citra CT-Scan otak ke dalam dua kategori, yaitu *Stroke* dan *Normal*. Model ViT memanfaatkan mekanisme *self-attention* untuk mempelajari hubungan antara bagian-bagian citra dalam proses klasifikasi. Hasil pengujian menunjukkan bahwa model memperoleh akurasi sebesar 82,93% pada dataset yang digunakan dalam penelitian. Hasil tersebut menunjukkan bahwa model ViT dapat digunakan untuk melakukan klasifikasi citra CT-Scan berdasarkan kategori yang telah ditentukan. Aplikasi yang dikembangkan dapat digunakan sebagai salah satu pendekatan dalam sistem pendukung analisis citra medis berbasis *Deep Learning*. Namun, hasil penelitian ini belum dapat digunakan untuk menyatakan bahwa sistem dapat menggantikan proses diagnosis atau secara langsung mempercepat diagnosis dan mengurangi *human error*, karena diperlukan pengujian lebih lanjut dalam lingkungan klinis dengan melibatkan tenaga medis serta dataset yang lebih besar dan beragam.

REFERENSI

- Ardiyansyah, M. S., Umbara, F. R., & Melina. (2025). Klasifikasi kanker payudara berbasis deep learning menggunakan Vision Transformer dengan teknik augmentasi data citra. *JURIKOM (Jurnal Riset Komputer)*, 12(3), 241–250. <https://doi.org/10.30865/jurikom.v12i3.8619>
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16 × 16 words: Transformers for image recognition at scale. *International Conference on Learning Representations*. <https://doi.org/10.48550/arXiv.2010.11929>
- Firdaus, D., Sumardi, I., Chazar, C., & Dafy, M. Z. (2025). Image-based malware multiclass classification using Vision Transformer architecture. *Cyber Security dan Forensik Digital*, 8(1), 72–79. <https://doi.org/10.14421/csecurity.2025.8.1.5107>
- Juniada, P. H., Asih, M. W., Margiani, N. N., Eka Putra, I. W. G. A., Martadiani, E. D., & Mahendra Wijaya, I. G. A. G. (2022). Korelasi ASPECT score dengan NIHSS, GCS, dan lama dirawat pada pasien stroke non hemoragik akut di RSUP Sanglah. *E-Jurnal Medika Udayana*, 11(10), 7–15. <https://doi.org/10.24843/MU.2022.V11.i10.P02>
- Maulana, R. T. N., Via, Y. V., & Mandyartha, E. P. (2025). Implementasi Vision Transformer untuk klasifikasi penyakit pneumonia melalui citra chest X-ray. *Journal CERITA: Creative Education of Research in Information Technology and Artificial Informatics*, 11(2), 249–254. <https://doi.org/10.33050/cerita.v11i2.3599>
- Pangestu, A., Purnama, B., & Risnandar. (2024). Vision Transformer untuk klasifikasi kematangan pisang. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 11(1), 75–84. <https://doi.org/10.25126/jtiik.20241117389>
- Puspita, M. I., Aini, V. N., Isnoviasih, S. T., & Darmawan, E. (2025). Optimizing window width and window level variations in head CT scans in hemorrhagic stroke cases. *Journal of Applied Health Management and Technology*, 7(2), 116–119.
- Putriany, E., & Ariatmanto, D. (2025). Comparative analysis of green snake identification using head structure and body patterns with Vision Transformer. *Telematika*, 18(1), 61–73. <https://doi.org/10.35671/telematika.v18i1.2992>
- Susanto, A., & Sari, A. Z. C. (2025). Non contrast head CT scan findings in stroke patients at the Radiology Department of Tiarasella Hospital Bengkulu 2024. *Journal of Applied Health Management and Technology*, 7(1), 27–32. <https://doi.org/10.31983/jahmt.v7i1.12832>
- Yan, M. D., Ardiyanto, J., & Sulaksono, N. (2020). Analisa perbedaan informasi diagnostik CT scan kepala pada kasus stroke iskemik dengan pilihan kombinasi slice thickness dan interval reconstruction. *Jurnal Radiografer Indonesia*, 3(1), 51–55.
- Yusastra, P., Indriyani, & Utama, B. (2021). Overview of the head CT-scan in stroke patients who was treated at Muhammadiyah Hospital Palembang. *Muhammadiyah Medical Journal*, 2(1), 24–34. <https://doi.org/10.24853/mmj.2.1.24-34>