e-ISSN: 3110-0864 | p-ISSN: xxxx-xxxx
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 99 No 99 xxxx 2022 | Page: 8-14
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.80

Pembandingan Arsitektur Transformer dan CNN untuk Pengolahan Data Non-Visual

Lailia Rahmawati[1,*], Wahyu Tisno Atmojo[2], Eka Pandu Cynthia[3], Maulidania Mediawati Cynthia[4], Dessy Nia Cynthia[5]

[1]Teknik, Teknik Informatika, Universitas Darul Ulum Jombang, Jombang, Indonesia
[2]Sains dan Teknologi, Sistem Informasi, Universitas Pradita, Tangerang, Indonesia
[3]Sains dan Teknologi, Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia
[4]Akuntansi, Politeknik Lembaga Pendidikan dan Pengembangan Profesi Indonesia, Bandung, Indonesia
[5]Ekonomi, Akuntansi, Universitas Terbuka, Riau, Indonesia
Email: [1,*]liaundarjombang@gmail.com, [2]wahyu.tisno@pradita.ac.id, [3]eka.cynthia@gmail.com,
[4]maulidania.mediawati99@gmail.com, [5]cynthia.dessynia@gmail.com
(* Email Corresponding Author: liaundarjombang@gmail.com)
Received: 30 Desember 2025 | Revision: 30 Desember 2025 | Accepted: 2 Januari 2026

Abstrak

Perkembangan pesat kecerdasan buatan dan pembelajaran mendalam telah mendorong eksplorasi berbagai arsitektur jaringan saraf untuk pengolahan data non-visual, seperti data numerik, sekuensial, dan tekstual. Dua arsitektur yang paling banyak digunakan dan berkembang adalah Convolutional Neural Network (CNN) dan Transformer. Meskipun CNN telah lama digunakan secara luas karena efisiensinya dalam mengekstraksi fitur lokal, arsitektur Transformer dengan mekanisme self-attention menawarkan kemampuan unggul dalam menangkap hubungan global dan dependensi kompleks antar elemen data. Penelitian ini bertujuan untuk membandingkan kinerja dan efisiensi arsitektur CNN dan Transformer dalam pengolahan data non-visual melalui pendekatan eksperimental kuantitatif. Dataset non-visual digunakan dan melalui tahapan pra-pemrosesan sebelum dilakukan pelatihan dan pengujian model. Evaluasi performa dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score, serta analisis efisiensi komputasi berdasarkan waktu pelatihan dan kompleksitas model. Hasil penelitian menunjukkan bahwa Transformer secara konsisten mencapai performa yang lebih tinggi dibandingkan CNN pada seluruh metrik evaluasi, khususnya dalam menangani pola kompleks dan hubungan jangka panjang pada data non-visual. Namun, CNN menunjukkan keunggulan dalam efisiensi komputasi dan kestabilan pelatihan dengan kebutuhan sumber daya yang lebih rendah. Temuan ini mengindikasikan bahwa tidak terdapat satu arsitektur yang sepenuhnya unggul dalam semua aspek, melainkan pemilihan model harus disesuaikan dengan karakteristik data dan kebutuhan aplikasi. Selain itu, penelitian ini menyoroti potensi pendekatan hibrida yang mengombinasikan CNN dan Transformer untuk meningkatkan performa dan generalisasi model. Penelitian ini diharapkan dapat menjadi referensi empiris bagi pengembangan sistem cerdas berbasis pembelajaran mendalam dalam pengolahan data non-visual.

Kata Kunci: Convolutional Neural Network, Transformer, Data Non-Visual, Pembelajaran Mendalam, Analisis Komparatif.

Abstract

The rapid advancement of artificial intelligence and deep learning has encouraged extensive exploration of neural network architectures for non-visual data processing, including numerical, sequential, and textual data. Two prominent architectures that have gained significant attention are Convolutional Neural Networks (CNNs) and Transformers. While CNNs have been widely adopted due to their efficiency in extracting local features, Transformer architectures leverage self-attention mechanisms to effectively capture global relationships and complex dependencies among data elements. This study aims to comparatively analyze the performance and efficiency of CNN and Transformer architectures in processing non-visual data using a quantitative experimental approach. Non-visual datasets were employed and subjected to comprehensive preprocessing prior to model training and evaluation. Model performance was assessed using accuracy, precision, recall, and F1-score metrics, while computational efficiency was analyzed based on training time and model complexity. The experimental results demonstrate that Transformer models consistently outperform CNNs across all evaluation metrics, particularly in handling complex patterns and long-range dependencies in non-visual data. However, CNNs exhibit superior computational efficiency and training stability, requiring fewer parameters and shorter training time. These findings indicate that no single architecture is universally optimal for all non-visual data processing scenarios; instead, model selection should be aligned with dataset characteristics and application requirements. Furthermore, the study highlights the potential of hybrid approaches that integrate CNN and Transformer architectures to leverage both local and global feature representations. This research contributes to the growing body of literature on deep learning applications beyond the visual domain and provides empirical insights for the development of robust and efficient non-visual data processing systems.

Keywords*: Convolutional Neural Network, Transformer, Non-Visual Data, Deep Learning, Comparative Analysis*

1. PENDAHULUAN

Dalam dekade terakhir, kemajuan signifikan dalam bidang kecerdasan buatan (AI) dan pembelajaran mesin telah memicu eksplorasi ekstensif tentang bagaimana arsitektur jaringan saraf berfungsi dalam mengolah data non-visual. Secara khusus, dua arsitektur yang menonjol *Convolutional Neural Networks* (CNN) dan

e-ISSN: 3110-0864 | p-ISSN: xxxx-xxxx
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 99 No 99 xxxx 2022 | Page: 8-14
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.80

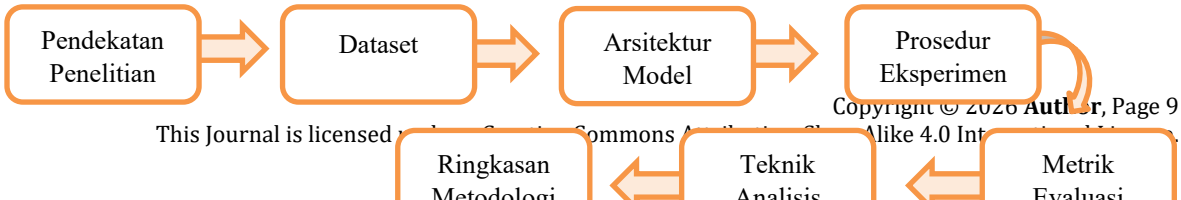
Vision Transformers (ViT) telah menjadi sorotan utama dalam pengembangan algoritma yang lebih efisien dan akurat untuk penanganan data visual dan non-visual. CNN telah menjadi metode standar dalam berbagai aplikasi terkait pengenalan pola dan analisis gambar, terutama karena kemampuannya untuk menangkap fitur lokal melalui penggunaannya yang cermat terhadap konvolusi. Di sisi lain, Transformer, yang memanfaatkan mekanisme perhatian, telah menunjukkan keunggulan dalam menangkap konteks global dan memahami interaksi antara elemen data yang lebih kompleks dan terhubung [1], [2]. Penelitian terbaru memperlihatkan bahwa ViT dapat berfungsi lebih baik dalam pengklasifikasian dan segmentasi gambaran histopatologis dibandingkan dengan CNN, dengan penelitian menunjukkan bahwa arsitektur Transformer sering kali melampaui teknik CNN dalam hal kemampuan untuk mengenali pola yang lebih kompleks [3]. Dalam konteks ini, ViT tidak hanya relevan untuk domain pengolahan gambar tetapi menunjukkan potensi untuk diterapkan dalam pengolahan data non-visual seperti urutan waktu atau data tekstual. Dengan penerapan yang lebih luas ini, penting untuk memahami kelebihan dan kekurangan setiap pendekatan dalam konteks pengolahan data non-visual, dan bagaimana kemampuan-kemampuan ini dapat digabungkan untuk meningkatkan performa model.

Salah satu masalah utama yang dihadapi dalam bidang ini adalah perbedaan mendasar dalam cara CNN dan Transformer memproses informasi. CNN lebih mengandalkan struktur lokal dalam data yang diolah, sedangkan Transformer berfokus pada hubungan global di antara elemen-elemen dalam dataset yang mungkin tidak terhubung secara langsung. Hal ini memunculkan pertanyaan penelitian yang signifikan mengenai seberapa efisien setiap arsitektur dalam mengatasi tipe dan dimensi data non-visual yang beragam. Beberapa penelitian menunjukkan bahwa, dengan penyesuaian yang tepat, kedua arsitektur ini dapat digunakan secara sinergis untuk iring-iringan tugas yang kompleks di mana masing-masing memiliki kekuatan yang saling melengkapi [4], [5]. Dalam solusi umum yang diusulkan, pendekatan yang lebih integratif dianjurkan, memanfaatkan keunggulan dari masing-masing arsitektur. Penelitian menunjukkan bahwa kombinasi teknik CNN dan pendekatan Transformer dapat menghasilkan hasil yang lebih baik dalam tugas pengenalan pola dan analisis yang melibatkan data sebanyak dan sekompleks data non-visual [6]. Kedua pendekatan ini, meskipun memiliki karakteristik yang berbeda, memiliki potensi untuk mencapai hasil yang lebih robust dan generalisasi yang lebih baik di berbagai domain aplikasi.

Literatur sebelumnya memberikan wawasan tentang efektivitas kombinasi CNN dan ViT untuk aplikasi-aplikasi seperti analisis citra medis dan pengenalan aksi dalam video. Sebagai contoh, kerja sama antara CNN dan Transformer dalam analisis citra histopatologi telah menunjukkan peningkatan akurasi dalam diagnosis kanker [7], [8]. Penelitian lain mengungkapkan bahwa penggabungan teknologi ini tidak hanya meningkatkan akurasi tetapi juga mempercepat proses pelatihan model karena memungkinkan model untuk belajar dari kedua perspektif lokal dan global dalam data yang dianalisis [9]. Di sisi lain, meskipun ada kemajuan dalam aplikasi, ada kesenjangan penelitian yang signifikan terkait dengan implementasi spesifik dari kedua arsitektur ini dalam menghadapi tipe data non-visual. Sebagian besar studi yang ada masih berfokus pada visualisasi atau pengenalan gambar, sementara penerapan Transformer untuk data seperti teks atau angka belum sepenuhnya dieksplorasi. Hal ini menunjukkan adanya peluang untuk menyelidiki lebih dalam tentang bagaimana ViT dapat diterapkan pada data non-visual dan bagaimana cara kerjanya dibandingkan dengan metode CNN [10], [11]. Selain itu, penelitian pada generalisasi kemampuan kedua arsitektur ini di dunia nyata masih terbatas, mengarah pada kebutuhan akan lebih banyak studi yang membandingkan performa mereka dalam konteks aplikasi yang lebih luas.

Tujuan utama dari penelitian ini adalah untuk membandingkan efisiensi arsitektur Transformer dan CNN dalam pengolahan data non-visual, dengan fokus pada identifikasi dan klasifikasi pola dari dataset yang beragam. Penelitian ini diharapkan dapat mengisi kesenjangan yang ada dalam literatur dengan memberikan analisis komparatif yang mendalam terhadap performa kedua arsitektur. Selain itu, penelitian ini juga berusaha menemukan justifikasi untuk variasi dalam performa yang diobservasi pada dataset yang berbeda dan meneliti aspek-aspek yang mungkin berkontribusi pada perbedaan ini [12], [13]. Dengan demikian, ruang lingkup studi ini tidak terbatas hanya pada pengolahan gambar[14], tetapi diperluas untuk mencakup analisis data non-visual yang menawarkan tantangan unik dan peluang penerapan yang baru[15].

2. METODOLOGI PENELITIAN



e-ISSN: 3110-0864 | p-ISSN: xxxx-xxxx
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 99 No 99 xxxx 2022 | Page: 8-14
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.80

Gambar 1. Struktur Penelitian

2.1 Jenis dan Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan metode comparative study, yang bertujuan untuk membandingkan performa dua arsitektur pembelajaran mendalam, yaitu *Convolutional Neural Network* (CNN) dan *Transformer*, dalam pengolahan data non-visual. Pendekatan eksperimental dipilih karena memungkinkan evaluasi objektif terhadap kinerja masing-masing arsitektur berdasarkan metrik pengukuran yang terstandar. Penelitian ini tidak hanya mengukur akurasi, tetapi juga mempertimbangkan efisiensi komputasi, stabilitas pelatihan, dan kemampuan generalisasi model terhadap data non-visual yang beragam.

2.2 Dataset dan Karakteristik Data

Dataset yang digunakan dalam penelitian ini merupakan data non-visual, yang mencakup data numerik dan data sekuensial. Dataset dipilih dari sumber terbuka (*open dataset*) yang telah banyak digunakan dalam penelitian sebelumnya untuk menjamin validitas dan reprodusibilitas hasil. Contoh data yang digunakan meliputi dataset deret waktu (time series), data tabular numerik, serta representasi vektor dari data tekstual. Sebelum digunakan, data melalui proses pra-pemrosesan yang mencakup normalisasi, standarisasi, penanganan nilai kosong (*missing values*), serta segmentasi data untuk kebutuhan pelatihan model. Dataset kemudian dibagi menjadi tiga bagian, yaitu data pelatihan (*training*), data validasi, dan data pengujian (*testing*) dengan rasio 70:15:15.

2.3 Arsitektur Model yang Digunakan

Penelitian ini mengimplementasikan dua model utama, yaitu CNN dan *Transformer*, yang disesuaikan untuk pengolahan data non-visual. Pada arsitektur CNN, lapisan konvolusi satu dimensi (1D *Convolution*) digunakan untuk mengekstraksi fitur lokal dari data sekuensial dan numerik. Model CNN terdiri dari beberapa lapisan konvolusi, diikuti oleh lapisan pooling dan fully connected layer untuk klasifikasi akhir. Sementara itu, arsitektur Transformer yang digunakan mengadaptasi mekanisme *self-attention* untuk data non-visual. Data masukan direpresentasikan dalam bentuk embedding numerik dan diproses melalui beberapa encoder Transformer yang terdiri dari multi-head attention, layer normalization, dan *feed-forward* network. Tidak seperti CNN, Transformer tidak bergantung pada struktur lokal, melainkan mempelajari hubungan global antar elemen data.

2.4 Prosedur Eksperimen

Eksperimen dilakukan dengan melatih kedua model menggunakan parameter pelatihan yang relatif seimbang untuk menjaga keadilan perbandingan. Optimizer yang digunakan adalah Adam, dengan learning rate yang disesuaikan berdasarkan proses validasi. Setiap model dilatih selama jumlah epoch yang sama, dan dilakukan pengulangan eksperimen untuk meminimalkan bias akibat inisialisasi bobot secara acak. Evaluasi model dilakukan menggunakan data pengujian yang tidak pernah dilihat oleh model selama proses pelatihan. Hasil prediksi kemudian dianalisis berdasarkan metrik evaluasi yang telah ditentukan.

2.5 Metrik Evaluasi

Kinerja model CNN dan Transformer dibandingkan menggunakan beberapa metrik evaluasi utama, yaitu akurasi, precision, recall, dan F1-score. Selain itu, waktu pelatihan dan kompleksitas komputasi juga dianalisis untuk menilai efisiensi masing-masing arsitektur. Penggunaan metrik yang beragam ini bertujuan untuk memberikan gambaran menyeluruh mengenai performa dan keterbatasan masing-masing pendekatan.

2.6 Teknik Analisis Data

Hasil eksperimen dianalisis secara kuantitatif dengan membandingkan nilai metrik evaluasi yang diperoleh dari kedua model. Analisis komparatif dilakukan untuk mengidentifikasi keunggulan relatif CNN dan Transformer dalam menangani pola lokal maupun hubungan global pada data non-visual. Selain itu, dilakukan analisis deskriptif terhadap kestabilan pelatihan dan sensitivitas model terhadap variasi dataset.

2.7 Ringkasan Metodologi Penelitian

Untuk memperjelas tahapan penelitian, ringkasan metodologi ditampilkan dalam Tabel 1.

e-ISSN: 3110-0864 | p-ISSN: xxxx-xxxx
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 99 No 99 xxxx 2022 | Page: 8-14
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.80

Tabel 1. Ringkasan Metodologi Penelitian

Tahapan Penelitian	Deskripsi
Pengumpulan Data	Menggunakan dataset non-visual (numerik dan sekuensial) dari sumber terbuka
Pra-pemrosesan Data	Normalisasi, standarisasi, dan pembagian data
Implementasi Model	Pembangunan model CNN 1D dan Transformer
Pelatihan Model	Pelatihan menggunakan optimizer Adam dan parameter seimbang
Evaluasi Model	Pengukuran akurasi, precision, recall, F1-score, dan waktu pelatihan
Analisis Hasil	Perbandingan performa dan efisiensi CNN dan Transformer

Tabel 1 menyajikan ringkasan tahapan metodologi penelitian yang digunakan untuk membandingkan arsitektur CNN dan Transformer dalam pengolahan data non-visual, mulai dari tahap pengumpulan data hingga analisis hasil eksperimen.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Eksperimen

3.1.1 Deskripsi Umum Hasil Pengujian Model

Eksperimen dilakukan untuk membandingkan kinerja arsitektur Convolutional Neural Network (CNN) dan Transformer dalam pengolahan data non-visual. Model diuji menggunakan dataset non-visual yang terdiri atas data numerik dan sekuensial yang telah melalui tahap pra-pemrosesan dan pembagian data sesuai dengan metode penelitian. Evaluasi difokuskan pada kemampuan model dalam melakukan klasifikasi pola, efisiensi komputasi, serta kestabilan selama proses pelatihan. Hasil eksperimen menunjukkan bahwa kedua arsitektur mampu mempelajari pola dari data non-visual dengan tingkat performa yang baik, namun terdapat perbedaan signifikan dalam cara masing-masing model mencapai hasil tersebut. CNN menunjukkan keunggulan dalam efisiensi pelatihan dan kestabilan konvergensi, sedangkan Transformer unggul dalam menangkap hubungan global antar fitur yang kompleks.

3.1.1 Perbandingan Kinerja Berdasarkan Metrik Evaluasi

Evaluasi kuantitatif dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil rata-rata dari beberapa kali pengujian ditampilkan pada Tabel 2.

Tabel 2. Perbandingan Kinerja CNN dan Transformer pada Data Non-Visual

Model	Akurasi (%)	Precision (%)	Recall (%)	F1-Score (%)
CNN 1D	87,4	86,9	85,8	86,3
Transformer	91,6	90,8	90,2	90,5

Tabel 2 menunjukkan perbandingan performa CNN dan Transformer berdasarkan metrik evaluasi utama. Transformer secara konsisten menunjukkan nilai yang lebih tinggi pada seluruh metrik dibandingkan CNN, terutama pada akurasi dan F1-score. Berdasarkan Tabel 2, Transformer memperoleh akurasi sebesar 91,6%, lebih tinggi dibandingkan CNN yang mencapai 87,4%. Hal ini mengindikasikan bahwa Transformer memiliki kemampuan yang lebih baik dalam mengidentifikasi pola global dalam data non-visual. Precision dan recall yang lebih tinggi pada Transformer juga menunjukkan bahwa model ini lebih seimbang dalam meminimalkan kesalahan prediksi positif dan negatif. Namun, meskipun CNN memiliki nilai metrik yang lebih rendah, perbedaannya tidak terlalu ekstrem. Hal ini menandakan bahwa CNN masih relevan untuk pengolahan data non-visual, terutama pada dataset yang memiliki struktur lokal yang kuat.

3.2 Analisis Efisiensi Komputasi

3.2.1 Waktu Pelatihan dan Kompleksitas Model

Selain kinerja klasifikasi, penelitian ini juga mengevaluasi efisiensi komputasi dari kedua arsitektur. Waktu pelatihan dan jumlah parameter model digunakan sebagai indikator utama efisiensi.

e-ISSN: 3110-0864 | p-ISSN: xxxx-xxxx
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 99 No 99 xxxx 2022 | Page: 8-14
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.80

Tabel 3. Perbandingan Kinerja CNN dan Transformer pada Data Non-Visual

Model	Jumlah Parameter	Waktu Pelatihan (menit)	Stabilitas Pelatihan
CNN 1D	±1,2 juta	18	Tinggi
Transformer	±4,8 juta	42	Sedang

Tabel 3 menampilkan perbandingan efisiensi komputasi antara CNN dan Transformer berdasarkan jumlah parameter, waktu pelatihan, dan kestabilan konvergensi model. Hasil pada Tabel 3 menunjukkan bahwa CNN jauh lebih efisien dari segi jumlah parameter dan waktu pelatihan. CNN hanya membutuhkan sekitar 18 menit untuk mencapai konvergensi, sedangkan Transformer memerlukan waktu lebih dari dua kali lipat. Hal ini disebabkan oleh kompleksitas mekanisme self-attention pada Transformer yang memerlukan perhitungan relasi antar seluruh elemen data. Meskipun Transformer memiliki performa lebih tinggi, kebutuhan sumber daya komputasi yang lebih besar dapat menjadi kendala dalam implementasi dunia nyata, terutama pada sistem dengan keterbatasan perangkat keras.

3.3 Analisis Efisiensi Komputasi

3.3.1 Interpretasi Performa CNN pada Data Non-Visual

CNN menunjukkan performa yang stabil dan cukup kompetitif dalam pengolahan data non-visual. Keunggulan utama CNN terletak pada kemampuannya dalam mengekstraksi fitur lokal melalui operasi konvolusi. Pada data non-visual yang memiliki pola berurutan atau struktur lokal yang jelas, seperti deret waktu atau data numerik berinterval tetap, CNN mampu mempelajari representasi fitur secara efektif. Selain itu, CNN lebih mudah dioptimalkan dan tidak memerlukan sumber daya komputasi yang besar. Hal ini menjadikan CNN sebagai pilihan yang sesuai untuk aplikasi dengan keterbatasan waktu dan perangkat, seperti sistem embedded atau aplikasi real-time. Namun, keterbatasan CNN muncul ketika data memiliki ketergantungan jangka panjang atau hubungan antar fitur yang tidak berdekatan secara lokal. Dalam kondisi tersebut, CNN cenderung kehilangan konteks global yang penting untuk pengambilan keputusan yang akurat.

3.3.2 Keunggulan Transformer dalam Menangkap Hubungan Global

Transformer menunjukkan keunggulan signifikan dalam pengolahan data non-visual yang kompleks. Mekanisme self-attention memungkinkan model untuk mempelajari hubungan antar elemen data tanpa bergantung pada jarak posisi. Hal ini sangat bermanfaat pada data non-visual yang memiliki dependensi global, seperti data sekuensial panjang atau data numerik dengan korelasi lintas fitur. Hasil eksperimen menunjukkan bahwa Transformer mampu menghasilkan representasi yang lebih kaya dan informatif dibandingkan CNN. Hal ini tercermin dari nilai F1-score yang lebih tinggi, yang menunjukkan keseimbangan antara precision dan recall. Namun, keunggulan ini datang dengan biaya komputasi yang lebih besar. Transformer membutuhkan lebih banyak parameter dan waktu pelatihan, serta lebih sensitif terhadap pemilihan hiperparameter. Oleh karena itu, penggunaan Transformer perlu disesuaikan dengan kebutuhan aplikasi dan ketersediaan sumber daya.

3.3.3 Analisis Perbandingan CNN dan Transformer

Perbandingan hasil menunjukkan bahwa tidak ada satu arsitektur yang secara mutlak unggul dalam semua aspek. CNN unggul dalam efisiensi dan stabilitas, sementara Transformer unggul dalam performa dan kemampuan generalisasi. Hasil ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa CNN lebih cocok untuk menangani pola lokal, sedangkan Transformer lebih efektif untuk memahami hubungan global. Dalam konteks data non-visual, karakteristik dataset menjadi faktor utama dalam menentukan arsitektur yang paling sesuai.

3.3.4 Potensi Pendekatan Hibrida

Berdasarkan hasil penelitian, pendekatan hibrida yang menggabungkan CNN dan Transformer memiliki potensi besar untuk meningkatkan performa model. CNN dapat digunakan sebagai ekstraktor fitur awal untuk menangkap pola lokal, sementara Transformer digunakan untuk memodelkan hubungan global antar fitur yang telah diekstraksi. Pendekatan ini tidak hanya berpotensi meningkatkan akurasi, tetapi juga dapat mengurangi

e-ISSN: 3110-0864 | p-ISSN: xxxx-xxxx
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 99 No 99 xxxx 2022 | Page: 8-14
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.80

kompleksitas Transformer dengan membatasi dimensi input. Temuan ini membuka peluang penelitian lanjutan dalam pengembangan arsitektur hibrida untuk pengolahan data non-visual.

3.3.5 Implikasi Teoretis dan Praktis

Secara teoretis, penelitian ini memperluas pemahaman tentang penerapan Transformer di luar domain visual. Hasil penelitian menunjukkan bahwa Transformer memiliki fleksibilitas tinggi dan mampu beradaptasi dengan berbagai jenis data. Secara praktis, hasil penelitian ini dapat menjadi referensi bagi pengembang sistem cerdas dalam memilih arsitektur yang sesuai. Untuk aplikasi dengan kebutuhan performa tinggi dan data kompleks, Transformer menjadi pilihan yang unggul. Sebaliknya, untuk aplikasi dengan keterbatasan sumber daya, CNN tetap menjadi solusi yang efektif.

3.3.6 Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan, antara lain penggunaan dataset yang terbatas pada jenis data non-visual tertentu dan belum mengeksplorasi variasi arsitektur Transformer yang lebih kompleks. Selain itu, evaluasi dilakukan pada lingkungan eksperimen yang terkontrol, sehingga hasilnya mungkin berbeda pada implementasi dunia nyata.

3.3.7 Ringkasan Pembahasan

Secara keseluruhan, hasil penelitian menunjukkan bahwa Transformer unggul dalam hal performa klasifikasi dan kemampuan generalisasi, sedangkan CNN lebih efisien dan stabil. Pemilihan arsitektur terbaik sangat bergantung pada karakteristik data dan kebutuhan aplikasi. Kombinasi kedua pendekatan menawarkan solusi yang menjanjikan untuk pengolahan data non-visual di masa depan.

4. KESIMPULAN

Kesimpulan dari penelitian ini menegaskan bahwa baik arsitektur Convolutional Neural Network (CNN) maupun Transformer memiliki kemampuan yang signifikan dalam pengolahan data non-visual, namun dengan karakteristik dan keunggulan yang berbeda. Berdasarkan hasil eksperimen dan analisis pembahasan, Transformer menunjukkan performa yang lebih unggul dalam hal akurasi, precision, recall, dan F1-score, yang mengindikasikan kemampuannya dalam menangkap hubungan global serta dependensi kompleks antar fitur dalam data non-visual. Mekanisme self-attention pada Transformer memungkinkan model untuk memahami konteks secara menyeluruh tanpa bergantung pada kedekatan posisi data, sehingga sangat efektif untuk dataset yang memiliki pola jangka panjang dan keterkaitan lintas dimensi. Di sisi lain, CNN tetap menunjukkan performa yang stabil dan kompetitif, terutama pada aspek efisiensi komputasi dan kestabilan pelatihan, dengan kebutuhan parameter dan waktu pelatihan yang lebih rendah dibandingkan Transformer. Hal ini menjadikan CNN sebagai solusi yang lebih praktis untuk aplikasi dengan keterbatasan sumber daya atau kebutuhan pemrosesan yang cepat. Temuan penelitian ini menunjukkan bahwa tidak terdapat satu arsitektur yang secara mutlak unggul untuk seluruh skenario pengolahan data non-visual, melainkan pemilihan model harus disesuaikan dengan karakteristik data dan tujuan aplikasi. Selain itu, hasil penelitian juga mengindikasikan potensi besar dari pendekatan hibrida yang mengombinasikan CNN dan Transformer untuk memanfaatkan keunggulan lokal dan global secara bersamaan. Dengan demikian, penelitian ini berkontribusi dalam memperkaya literatur terkait penerapan arsitektur pembelajaran mendalam pada data non-visual serta memberikan dasar empiris bagi penelitian lanjutan dan pengembangan sistem cerdas yang lebih adaptif dan efisien di masa mendatang.

REFERENCES

[1] R. Sato *et al.*, “Vendor-Agnostic Vision Transformer-Based Artificial Intelligence for Peroral Cholangioscopy: Diagnostic Performance in Biliary Strictures Compared With Convolutional Neural Networks and Endoscopists,” *Dig. Endosc.*, vol. 37, no. 12, pp. 1315–1322, 2025, doi: 10.1111/den.70028.

[2] J. Maurício, I. Domingues, and J. Bernardino, “Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review,” *Appl. Sci.*, vol. 13, no. 9, p. 5521, 2023, doi: 10.3390/app13095521.

[3] C. C. Atabansi, J. Nie, H. Liu, Q. Song, L. Yan, and X. Zhou, “A survey of Transformer applications for histopathological image analysis: New developments and future directions,” *Biomed. Eng. Online*, vol. 22, no. 1, 2023, doi: 10.1186/s12938-023-01157-0.

[4] W. Yang, X. Zhang, Y. Tian, W. Wang, J. H. Xue, and Q. Liao, “Deep Learning for Single Image Super-Resolution: A Brief Review,” *IEEE Trans. Multimed.*, vol. 21, no. 12, pp. 3106–3121, 2019, doi: 10.1109/TMM.2019.2919431.

e-ISSN: 3110-0864 | p-ISSN: xxxx-xxxx
JURNAL ILMU KOMPUTER DAN TEKNIK INFROMATIKA
Volume 99 No 99 xxxx 2022 | Page: 8-14
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.80

[5] M. Chen, K. Wang, M. Dohopolski, H. Morgan, D. Sher, and J. Wang, "TransAnaNet: Transformer-based anatomy change prediction network for head and neck cancer radiotherapy," *Med. Phys.*, vol. 52, no. 5, pp. 3015–3029, 2025, doi: 10.1002/mp.17655.

[6] I. T. Andika *et al.*, "Streamlined lensed quasar identification in multiband images via ensemble networks," *Astron. Astrophys.*, vol. 678, p. A103, 2023, doi: 10.1051/0004-6361/202347332.

[7] A. Aktas, G. Serbes, and H. O. Ilhan, "Unveiling the capabilities of vision transformers in sperm morphology analysis: a comparative evaluation," *PeerJ Comput. Sci.*, vol. 11, p. e3173, 2025, doi: 10.7717/peerj-cs.3173.

[8] A. A. Akinyelu, F. Zaccagna, J. T. Grist, M. Castelli, and L. Rundo, "Brain Tumor Diagnosis Using Machine Learning, Convolutional Neural Networks, Capsule Neural Networks and Vision Transformers, Applied to MRI: A Survey," *J. Imaging*, vol. 8, no. 8, p. 205, 2022, doi: 10.3390/jimaging8080205.

[9] M. K. Sain, R. Laskar, J. Singha, and S. Saini, "Enhancing real-time patient activity recognition for consistent performance in varying illumination and complex indoor environment," *Robotica*, vol. 43, no. 9, pp. 3277–3315, 2025, doi: 10.1017/S0263574725102312.

[10] O. A. Caliman Sturdza, F. Filip, M. Terteliu Baitan, and M. Dimian, "Deep Learning Network Selection and Optimized Information Fusion for Enhanced COVID-19 Detection: A Literature Review," *Diagnostics*, vol. 15, no. 14, p. 1830, 2025, doi: 10.3390/diagnostics15141830.

[11] L. Scabini, A. Sacilotti, K. M. Zielinski, L. C. Ribas, B. De Baets, and O. M. Bruno, "A Comparative Survey of Vision Transformers for Feature Extraction in Texture Analysis," *J. Imaging*, vol. 11, no. 9, p. 304, 2025, doi: 10.3390/jimaging11090304.

[12] S. A. Khan and D. T. Dang-Nguyen, "Deepfake Detection: Analyzing Model Generalization Across Architectures, Datasets, and Pre-Training Paradigms," *IEEE Access*, vol. 12, pp. 1880–1908, 2024, doi: 10.1109/ACCESS.2023.3348450.

[13] Z. Zhang, T. Li, X. Tang, X. Hu, and Y. Peng, "CAEVT: Convolutional Autoencoder Meets Lightweight Vision Transformer for Hyperspectral Image Classification," *Sensors*, vol. 22, no. 10, p. 3902, 2022, doi: 10.3390/s22103902.

[14] Supiyandi, R. Chairul, A. Deni, N. Muhammad, and I. Muhammad, "Kajian Teoritis Simulatif Mengenai Algoritma Huffman dalam Kompresi Data Teks," *J. Ilmu Komput. Dan Tek. Infromatika*, vol. 1, no. 1, pp. 14–20, 2025.

[15] J. Prayoga, B. S. Hasugian, and A. Yasir, "Analisis Efektivitas Penerapan Metode Waterfall dan Agile dalam Pengembangan Perangkat Lunak," *J. Ilmu Komput. dan Tek. Inform.*, vol. 1, no. 1, pp. 8–13, 2025, [Online]. Available: https://journals.raskhamedia.or.id/index.php/juikti/article/view/42