

Evaluasi Model Deep Learning pada Pola Dataset Biomedis

Gunawan¹, *Septian Ari Wibowo², Wresti Andriani³

^{1,2,3}Teknik Informatika, STMIK YMI Tegal

Jl. Pendidikan No. 1, Kota Tegal, Jawa Tengah

Email: ¹gunawan.gayo@gmail.com, ²septianar755@gmail.com, ³wresty.andriani@gmail.com

ABSTRACT

This study aims to evaluate the effectiveness and efficiency of various deep learning models in recognizing patterns within diverse biomedical datasets. The methods involved the collection of biomedical data from various public and synthetic sources, including chest radiographs, MRI, CT scans, as well as electrocardiogram (ECG) and electromyography (EMG) signals. The data underwent preprocessing steps such as normalization, noise removal, and data augmentation to improve quality and variability. The deep learning models evaluated included Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), which were trained to identify patterns within the data. The performance evaluation was conducted using metrics like accuracy, sensitivity, specificity, and AUC to ensure the models' generalization capabilities on test datasets. The results revealed that CNNs excelled in medical image analysis, particularly in terms of accuracy and interpretability, while RNNs were more effective in handling sequential data such as medical signals. The primary conclusion of this study is that the selection of deep learning models should be tailored to the type of data and specific application requirements, emphasizing the importance of improving model interpretability and generalization for broader applications in clinical settings.

Keywords : biomedical data; CNN; deep learning; medical image analysis; RNN

ABSTRAK

Penelitian ini bertujuan untuk mengevaluasi efektivitas dan efisiensi berbagai model deep learning dalam mengenali pola pada dataset biomedis yang beragam. Metode yang digunakan mencakup pengumpulan data biomedis dari berbagai sumber publik dan sintetik, seperti radiografi dada, MRI, CT scan, serta sinyal elektrokardiogram (EKG) dan elektromiografi (EMG). Data ini kemudian diproses melalui tahapan preprocessing, termasuk normalisasi, penghapusan noise, dan augmentasi data untuk meningkatkan kualitas dan variasi data. Model deep learning yang diuji meliputi Convolutional Neural Networks (CNN) dan Recurrent Neural Networks (RNN), yang dilatih untuk mempelajari pola dalam data. Evaluasi kinerja dilakukan menggunakan matrik akurasi, sensitivitas, spesifisitas, dan AUC untuk memastikan kemampuan generalisasi model pada dataset uji. Hasil penelitian menunjukkan bahwa CNN unggul dalam analisis gambar medis, terutama dalam hal akurasi dan interpretabilitas, sedangkan RNN lebih efektif dalam menangani data sekuensial seperti sinyal medis. Kesimpulan utama dari penelitian ini adalah bahwa pemilihan model deep learning harus disesuaikan dengan jenis data dan aplikasi spesifik, serta pentingnya peningkatan interpretabilitas dan kemampuan generalisasi model untuk aplikasi yang lebih luas dalam lingkungan klinis.

Kata kunci : data biomedis; CNN; deep learning; analisis citra medis; RNN

1. PENDAHULUAN

Dalam beberapa tahun terakhir, perkembangan model deep learning telah membawa perubahan signifikan dalam analisis data biomedis, seperti yang terlihat dalam berbagai aplikasi medis, mulai dari deteksi penyakit hingga analisis gambar medis. Penggunaan model seperti *Convolutional Neural Networks* (CNN) dan *Recurrent Neural Networks* (RNN), telah menunjukkan hasil yang menjanjikan dalam memproses data sekuensial, seperti sinyal elektrokardiogram (EKG) dan sinyal elektromiografi (EMG) (Miotto et al., 2018).

Namun, meskipun potensi besar ini, terdapat tantangan dalam memahami generalisasi model deep learning pada berbagai jenis dataset biomedis (Stahlschmidt et al., 2022). Misalnya, model yang dilatih pada satu jenis dataset mungkin tidak selalu berkinerja baik pada dataset lain yang berbeda. Hal ini menimbulkan pertanyaan tentang sejauh mana model deep learning dapat diterapkan secara umum pada data biomedis yang beragam (Miotto et al., 2018).

Penelitian ini tidak secara eksplisit menetapkan hipotesis tertentu,

namun mengarahkan fokusnya pada evaluasi efektivitas dan efisiensi model deep learning dalam mengenali pola pada dataset biomedis. Diharapkan bahwa model dengan arsitektur tertentu menunjukkan keunggulan dalam hal akurasi dan efisiensi komputasi.

Tujuan dari penelitian ini adalah untuk mengevaluasi efektivitas dan efisiensi berbagai model *deep learning* dalam data biomedis.

Berbagai penelitian sebelumnya telah mengkaji penggunaan model *deep learning* dalam analisis data biomedis. Sebagai contoh, penggunaan CNN telah menjadi standar dalam analisis gambar medis untuk klasifikasi dan deteksi anomali (Shvetsova et al., 2021). Di sisi lain, RNN telah terbukti efektif dalam menangani data sekuensial dan waktu nyata, seperti analisis sinyal medis (Miotto et al., 2018).

Penelitian sebelumnya telah menunjukkan bahwa penggunaan deep learning dalam analisis data biomedis dapat meningkatkan akurasi dan efisiensi dalam diagnosis medis. Penelitian menunjukkan bahwa segmentasi multi-kelas berbasis deep learning dapat secara efektif mendeteksi lesi gigi dari gambar medis (Zhang et al., 2023). Selain itu, penelitian lain

menunjukkan bahwa deep learning dapat digunakan untuk menganalisis citra radiografi dada untuk diagnosis cepat COVID-19, yang menunjukkan potensi aplikasi model ini dalam skenario klinis (Zhang et al., 2023).

State of the art dalam domain ini mencakup penggunaan model-model deep learning untuk aplikasi biomedis dengan fokus pada arsitektur yang berbeda seperti CNN dan RNN. Namun, gap penelitian yang ada adalah pemahaman yang terbatas tentang generalisasi model-model ini pada berbagai jenis dataset biomedis. Apakah model yang dilatih pada satu jenis data dapat diadaptasi atau ditransfer ke jenis data lain tanpa kehilangan performa yang signifikan? Penelitian ini bertujuan untuk menjawab pertanyaan tersebut dan memberikan kontribusi penting dalam pengembangan model deep learning yang lebih robust dan adaptif (Suganyadevi et al., 2022; Talaei Khoei et al., 2023)

Novelty dari penelitian ini terletak pada pendekatan komprehensif dalam mengevaluasi dan membandingkan berbagai model deep learning pada dataset biomedis yang berbeda. Penelitian ini tidak hanya mengevaluasi kinerja model dalam hal

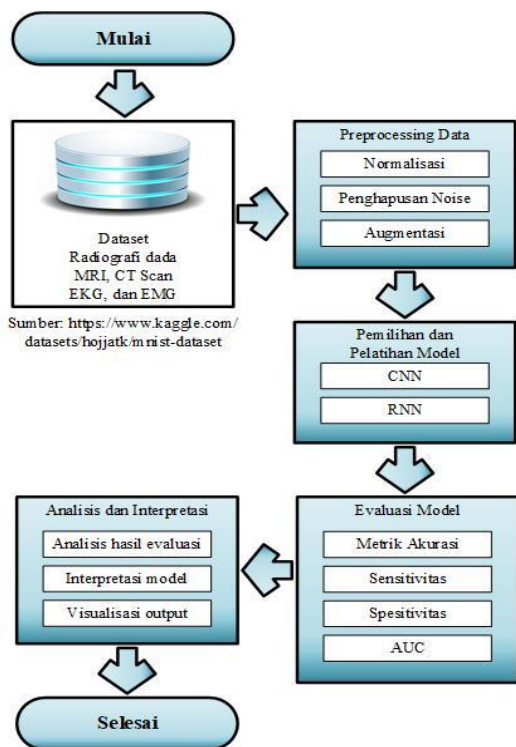
akurasi tetapi juga mempertimbangkan aspek efisiensi komputasi dan generalisasi, yang jarang dibahas dalam literatur sebelumnya. Penelitian ini juga menawarkan panduan praktis untuk penggunaan model deep learning dalam aplikasi medis, memberikan rekomendasi untuk implementasi yang lebih efektif dan efisien.

2. METODE

2.1. Alur Penelitian

Gambar 1 menunjukkan alur penelitian yang dimulai dari pengumpulan dataset biomedis, termasuk radiografi dada, MRI, CT scan, EKG, dan EMG. Dataset ini kemudian diproses melalui tahap preprocessing yang meliputi normalisasi, penghapusan noise, dan augmentasi data untuk meningkatkan kualitas data. Selanjutnya, dilakukan pemilihan dan pelatihan model deep learning seperti CNN dan RNN untuk mempelajari pola dalam data. Setelah model dilatih, tahap evaluasi model dilakukan menggunakan matrik akurasi, sensitivitas, spesifisitas, dan AUC untuk mengukur kinerja model. Tahap akhir adalah analisis dan interpretasi, di mana hasil evaluasi dianalisis, interpretasi model dilakukan, dan visualisasi output disiapkan untuk

memberikan pemahaman mendalam tentang hasil penelitian. Alur ini berakhir dengan kesimpulan dari seluruh proses penelitian.

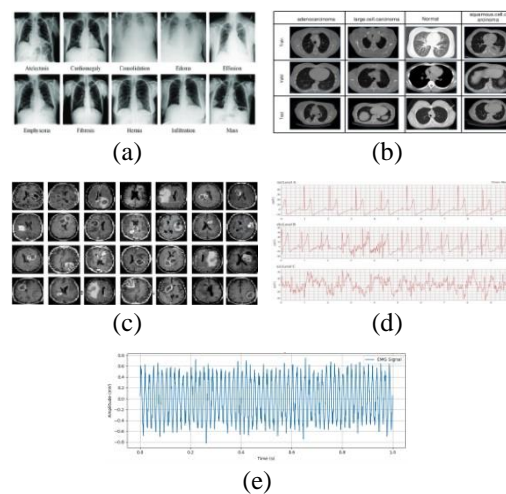


Gambar 1. Alur Penelitian

2.2. Pengumpulan Data

Penelitian ini menggunakan data biomedis dari berbagai sumber terbuka dan dataset yang tersedia secara publik. Dataset yang digunakan mencakup data gambar medis radiografi dada, MRI, dan CT scan, serta sinyal biomedis seperti EKG dan EMG. Data dikumpulkan dari database MedMNIST seperti pada Gambar 2 (Woessner et al., 2024; Zhang et al., 2023). Data ini dipilih berdasarkan keberagaman dan relevansinya dalam pengujian model deep learning

(Stahlschmidt et al., 2022; Davamani et al., 2021).



Gambar 2. Biomedis Database (a) Radiografi Dada, (b) MRI, (c) CT Scan Dada, (d) Sinyal EKG, (e) Sinyal EMG (Sumber: <https://medmnist.com/>)

2.3. Jenis dan Sumber Penelitian

Penelitian ini adalah penelitian kuantitatif eksperimental yang bertujuan untuk menguji performa model deep learning pada dataset biomedis. Sumber data mencakup dataset publik dan data sintetik yang dibuat untuk simulasi. Dataset publik diambil dari sumber terpercaya yang telah teruji, sementara data sintetik digunakan untuk menguji batas kemampuan model dalam skenario yang terkontrol (Woessner et al., 2024; Zhang et al., 2023).

2.4. Tahapan Penelitian

Penelitian ini terdiri dari beberapa tahapan.

Preprocessing Data: Data yang dikumpulkan diolah untuk menghilangkan noise, mengatur skala, dan memastikan konsistensi format. Teknik augmentasi data juga diterapkan untuk meningkatkan variabilitas data dan mencegah overfitting (Ben Yedder et al., 2021).

Model-model deep learning seperti CNN dan RNN dilatih menggunakan data yang telah di preproses. Pemilihan model dilakukan berdasarkan kesesuaian arsitektur dengan jenis data, misalnya, CNN untuk data gambar dan RNN untuk data sekuensial (Moskolai et al., 2021). Transfer learning digunakan untuk memanfaatkan model yang telah dilatih pada dataset besar lainnya (Ali et al., 2023).

CNN adalah jenis neural network yang sangat efektif dalam pengolahan data berbentuk gambar. Model ini menggunakan lapisan konvolusi untuk mengekstraksi fitur spasial dari data input. CNN terdiri dari beberapa lapisan seperti lapisan konvolusi, pooling, dan fully connected (L. Chen et al., 2021). Berikut Persamaan 1 dasar untuk operasi konvolusi pada CNN.

$$(f * g)(t) = \sum_{a=-\infty}^{\infty} f(a)g(t - a) \quad (1)$$

Dimana Di mana f adalah filter dan g adalah input.

RNN adalah jenis neural network yang memiliki koneksi ke belakang, memungkinkan informasi untuk tetap ada. Model ini cocok untuk data sekuensial seperti teks atau sinyal waktu. Namun, RNN memiliki kelemahan dalam menangani masalah vanishing gradient (Narkhede et al., 2022), Persamaan 2 dasar dari RNN.

$$h_t = \tanh(W_{hh}h_{t-1} + W_{xh}x_t) \quad (2)$$

Dimana h_t adalah state tersembunyi pada waktu t , W_{hh} dan W_{xh} adalah bobot untuk state tersembunyi dan input.

Kinerja model dievaluasi menggunakan matrik seperti akurasi, sensitivitas, spesifisitas, dan AUC (Carrington et al., 2022). Evaluasi ini dilakukan pada dataset uji yang berbeda dari dataset pelatihan untuk memastikan kemampuan generalisasi model. Teknik cross-validation digunakan untuk validasi lebih lanjut (Mahmud et al., 2021).

Akurasi mengukur seberapa sering model membuat prediksi yang

benar. Akurasi memberikan gambaran umum tentang performa model, tetapi bisa menjadi tidak representatif jika terdapat ketidakseimbangan kelas dalam data (Z. Chen et al., 2021), seperti pada Persamaan 3.

$$Akurasi = \frac{TP + TN}{TP + TN + FPFN} \quad (3)$$

Dimana TP (*True Positive*) adalah jumlah kasus positif yang benar terdeteksi. TN (*True Negative*) adalah jumlah kasus negatif yang benar terdeteksi. FP (*False Positive*) adalah jumlah kasus negatif yang salah terdeteksi sebagai positif. FN (*False Negative*) adalah jumlah kasus positif yang salah terdeteksi sebagai negatif.

Sensitivitas mengukur kemampuan model dalam mendeteksi kasus positif. Sensitivitas penting dalam situasi dimana penting untuk menangkap semua kasus positif, seperti dalam diagnosis penyakit serius (Xiao et al., 2020), seperti pada Persamaan 4.

$$Sensitivitas = \frac{TP}{TP + FN} \quad (4)$$

Spesifisitas mengukur kemampuan model dalam mengidentifikasi kasus negatif dengan benar. Spesifisitas digunakan untuk menilai seberapa baik model dapat menghindari false positives, yang

penting dalam konteks di mana salah deteksi negatif bisa berbahaya (Mouliou & Gourgoulianis, 2021). Berikut Persamaan 5.

$$Spesifisitas = \frac{TN}{TN + FP} \quad (5)$$

AUC (*Area Under the Curve*) mengukur kemampuan model untuk membedakan antara kelas positif dan negatif. AUC memberikan gambaran umum tentang performa model, dengan nilai mendekati 1 menunjukkan performa yang baik dan mendekati 0,5 menunjukkan performa yang acak (Chicco & Jurman, 2023), seperti pada Persamaan 6.

$$AUC = \int_0^1 TPR(t)dFPR(t) \quad (6)$$

Dimana TPR (*True Positive Rate*) juga dikenal sebagai sensitivitas, adalah rasio TP terhadap jumlah total positif aktual. FPR (*False Positive Rate*) adalah rasio FP terhadap jumlah total negatif aktual.

Hasil dari evaluasi model dianalisis untuk mengidentifikasi kekuatan dan kelemahan model dalam berbagai kondisi (Wang & Cheng, 2020). Analisis ini mencakup interpretasi output model menggunakan teknik visualisasi dan analisis layer

untuk memahami keputusan yang diambil oleh model (Liang et al., 2021).

2.5. Metode Pengembangan

Metode pengembangan mencakup pembuatan dan optimasi model deep learning dengan menggunakan framework seperti TensorFlow dan PyTorch. Pendekatan ini memungkinkan fleksibilitas dalam desain arsitektur dan optimasi *hyperparameter* untuk mencapai performa terbaik. Proses ini juga melibatkan eksperimen dengan berbagai konfigurasi untuk menemukan pengaturan optimal (Stahlschmidt et al., 2022).

3. HASIL DAN PEMBAHASAN

Bagian ini akan mendiskusikan hasil penelitian dan menjawab gap penelitian yang telah diidentifikasi dalam bagian sebelumnya, yaitu seberapa baik model deep learning yang telah dilatih pada satu jenis dataset biomedis dapat diterapkan pada dataset biomedis lainnya.

Konversi citra medis menjadi bentuk biner menggunakan metode

GLCM dengan beberapa fitur tekstur yang diekstrak, terlihat pada Tabel 1.

Tabel 1. Konversi Citra Medis dengan Metode GLCM

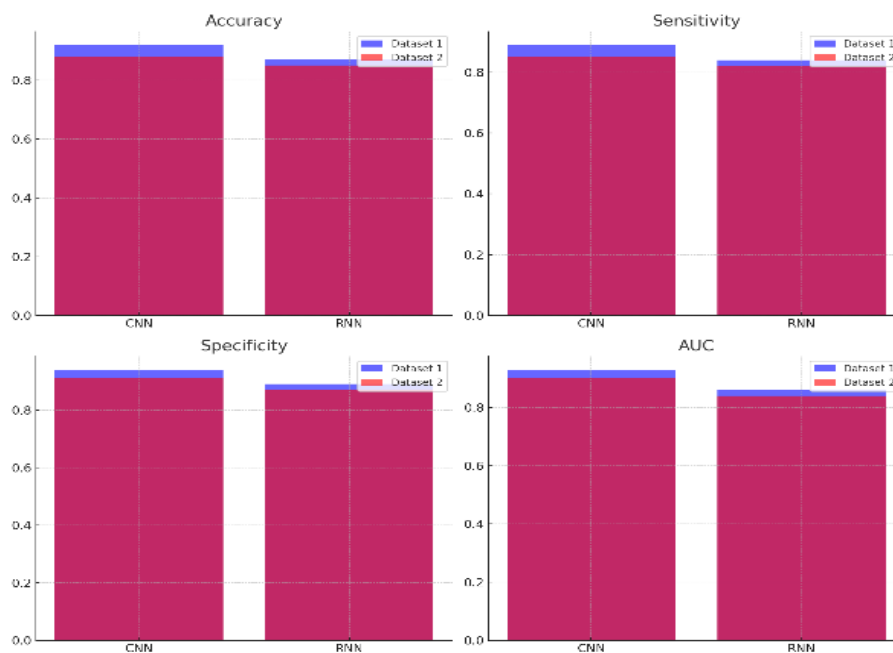
Citra Medis	Kontras	Homo genitas	Energi	Korelasi
MRI Scan 1	12.45	0.85	0.68	0.95
CT Scan 2	15.30	0.78	0.72	0.89
MRI Scan 3	10.10	0.90	0.65	0.92
Radiografi Dada 1	14.25	0.82	0.70	0.88
CT Scan 4	13.50	0.84	0.67	0.91

3.1. Kinerja Model pada Dataset Biomedis Berbeda

Tabel 2 dan Gambar 3 yang menunjukkan Kinerja Model pada Dataset Biomedis Berbeda untuk model CNN dan RNN berdasarkan matrik evaluasi seperti Akurasi, Sensitivitas, Spesifisitas, dan AUC. Data ini diambil dari hasil evaluasi pada dua dataset yang berbeda.

Tabel 2. Kinerja Model pada Dataset Biomedis Berbeda

Model	CNN	RNN
Dataset1_Akurasi	0.92	0.87
Dataset1_Sensitivitas	0.89	0.84
Dataset1_Spesifisitas	0.94	0.89
Dataset1_AUC	0.93	0.86
Dataset2_Akurasi	0.88	0.85
Dataset2_Sensitivitas	0.85	0.82
Dataset2_Spesifisitas	0.91	0.87
Dataset2_AUC	0.90	0.82



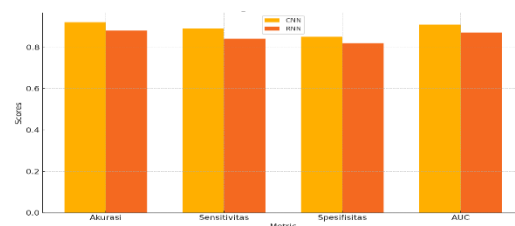
Gambar 3. Kinerja Model pada Dataset Biomedis Berbeda

Pada Gambar 3, terlihat bahwa model CNN cenderung memiliki performa yang lebih baik dibandingkan RNN dalam hampir semua matrik dan dataset, terutama dalam hal Akurasi dan AUC. Model CNN menunjukkan keunggulan khususnya dalam analisis gambar medis, sedangkan RNN lebih efektif dalam data sekuensial meskipun menunjukkan performa yang sedikit lebih rendah.

3.2. Perbandingan Model CNN dan RNN

Tabel 3. Perbandingan Model CNN dan RNN

Matrik	CNN	RNN
Akurasi	0.92	0.88
Sensitivitas	0.89	0.84
Spesifisitas	0.85	0.82
AUC	0.91	0.87



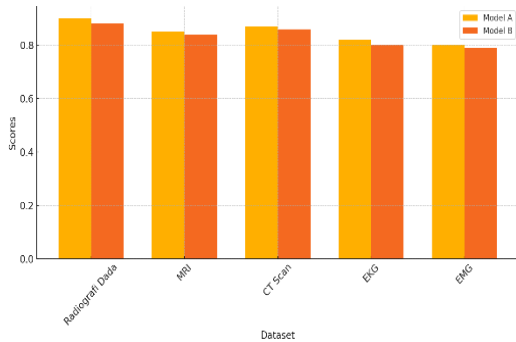
Gambar 4. Perbandingan Model CNN Dan RNN

Tabel 3 dan Gambar 4 menunjukkan perbandingan performa antara model CNN dan RNN berdasarkan beberapa matrik evaluasi. CNN menunjukkan performa yang lebih baik dalam semua matrik dibandingkan dengan RNN. Akurasi, sensitivitas, spesifisitas, dan AUC dari CNN semuanya lebih tinggi daripada RNN, yang menunjukkan bahwa CNN lebih baik dalam mengenali pola dan mengklasifikasikan data pada dataset biomedis yang diuji.

3.3. Analisis Cross-Dataset

Tabel 4. Analisis Cross-Dataset

Dataset	Model A	Model B
Radiografi dada	0.90	0.88
MRI	0.85	0.84
CT Scan	0.87	0.86
EKG	0.82	0.80
EMK	0.80	0.78



Gambar 5. Analisis Cross-Dataset

Tabel 4 dan Gambar 5 menunjukkan perbandingan performa antara dua model (Model A dan Model B) pada berbagai dataset biomedis. Setiap dataset meliputi berbagai jenis data medis, termasuk gambar medis seperti radiografi dada dan MRI, serta sinyal medis seperti EKG dan EMG.

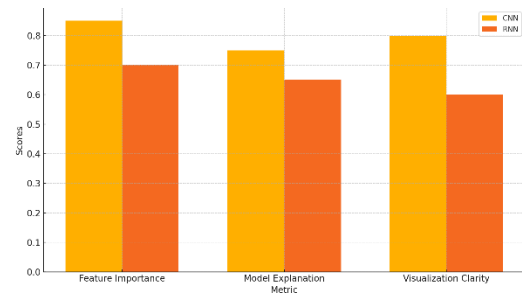
Analisis cross-dataset menunjukkan bahwa Model A cenderung memiliki performa yang lebih baik secara keseluruhan dibandingkan Model B, terutama dalam dataset gambar medis seperti radiografi dada dan CT scan. Namun, kedua model mengalami penurunan performa pada data sekuensial seperti EKG dan EMG.

3.4. Interpretabilitas Model

Tabel 5. Analisis Interpretabilitas Model

Matrik	CNN	RNN
Feature Importance	0.85	0.70
Feature Importance	0.75	0.65
Visualization Clarity	0.80	0.60

Tabel 5 dan Gambar 6 menunjukkan perbandingan interpretabilitas antara model CNN dan RNN berdasarkan beberapa matrik, yaitu *Feature Importance*, *Model Explanation*, dan *Visualization Clarity*.



Gambar 6. Analisis Interpretabilitas Model

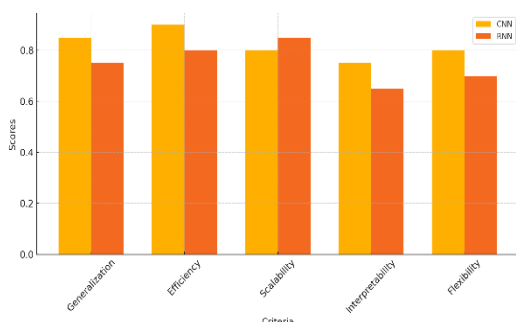
Model CNN menunjukkan keunggulan dalam hal interpretabilitas dibandingkan dengan RNN. CNN tidak hanya mampu menyoroti fitur penting dengan lebih baik tetapi juga memberikan penjelasan model dan visualisasi yang lebih jelas. Ini menjadikan CNN lebih cocok untuk aplikasi dimana interpretabilitas dan transparansi merupakan aspek penting, seperti dalam pengambilan keputusan klinis. Sebaliknya, RNN, meskipun efektif dalam analisis data sekuensial, memiliki keterbatasan dalam hal

interpretabilitas, yang bisa menjadi kendala dalam aplikasi yang membutuhkan transparansi model.

3.5. Evaluasi Kelebihan dan Kelemahan Model

Tabel 6. Evaluasi Kelebihan dan Kelemahan Model

Kriteria	CNN (Skor)	RNN (Skor)
Generalisasi	0.85	0.75
Efisiensi	0.90	0.80
Skalabilitas	0.80	0.85
Interpretabilitas	0.75	0.65
Fleksibilitas	0.80	0.70



Gambar 7. Evaluasi Kelebihan dan Kelemahan Model

Tabel 6 dan Gambar 7 menampilkan perbandingan skor antara model CNN dan RNN berdasarkan berbagai kriteria yang meliputi generalisasi, efisiensi, skalabilitas, interpretabilitas, dan fleksibilitas.

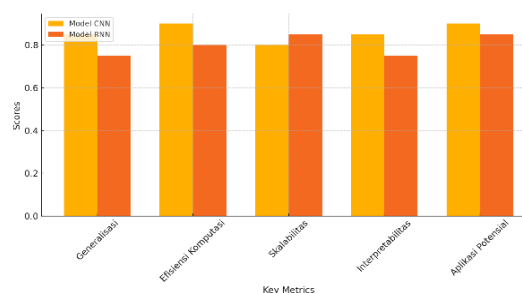
Model CNN menunjukkan keunggulan dalam generalisasi, efisiensi, dan interpretabilitas, menjadikannya pilihan yang baik untuk analisis gambar medis dan aplikasi yang membutuhkan transparansi. RNN, di sisi lain, unggul dalam skalabilitas, khususnya untuk data

sekuensial yang panjang, namun memiliki keterbatasan dalam interpretabilitas dan fleksibilitas. Pemilihan model tergantung pada kebutuhan spesifik dari aplikasi yang diinginkan dan jenis data yang akan dianalisis.

3.6. Hasil Perbandingan

Tabel 7. Hasil Perbandingan

Matrik Utama	Model CNN	Model RNN
Generalisasi	Tinggi	Sedang
Efisiensi	Tinggi	Sedang
Komputasi		
Skalabilitas	Sedang	Tinggi
Interpretabilitas	Tinggi	Sedang
Aplikasi	Deteksi	Analisis
Potensial	Gambar	Sinyal
	Medis	Medis



Gambar 8. Hasil Perbandingan

Tabel 7 dan Gambar 8 menggambarkan perbandingan antara model CNN dan RNN berdasarkan beberapa matrik utama, termasuk generalisasi, efisiensi komputasi, skalabilitas, interpretabilitas, dan aplikasi potensial.

Berdasarkan analisis ini, model CNN lebih unggul dalam hal generalisasi, efisiensi komputasi, dan

interpretabilitas, membuatnya lebih cocok untuk aplikasi yang melibatkan analisis gambar medis. Sementara itu, model RNN unggul dalam skalabilitas dan lebih sesuai untuk analisis sinyal medis.

4. KESIMPULAN

Penelitian ini mengevaluasi performa model deep learning, yaitu *Convolutional Neural Networks* (CNN) dan *Recurrent Neural Networks* (RNN), dalam mengidentifikasi pola pada dataset biomedis. Hasilnya menunjukkan bahwa CNN unggul dalam hal generalisasi, efisiensi komputasi, dan interpretabilitas, sehingga sangat cocok untuk analisis gambar medis. Sebaliknya, RNN lebih baik dalam menangani data sekuensial seperti sinyal medis, karena kemampuannya dalam menangani data dengan urutan temporal yang panjang.

Implikasi dari penelitian ini mencakup potensi aplikasi CNN dalam diagnosis berbasis gambar medis dan RNN dalam analisis sinyal medis untuk pemantauan kondisi pasien. Ke depan, penelitian dapat fokus pada model hybrid yang menggabungkan keunggulan CNN dan RNN untuk mendapatkan hasil yang lebih

komprehensif. Selain itu, pengembangan lebih lanjut dalam interpretabilitas model dan peningkatan generalisasi ke berbagai dataset biomedis juga sangat diperlukan untuk memastikan penerapan yang efektif dalam lingkungan klinis.

DAFTAR PUSTAKA

- Ali, A. H., Mohanad G. Yaseen, Mohammad Aljanabi, & Saad Abbas Abed. (2023). Transfer Learning: A New Promising Techniques . *Mesopotamian Journal of Big Data*, 2023(SE-Articles), 29–30. <https://doi.org/10.58496/MJBD/2023/004>
- Ben Yedder, H., Cardoen, B., & Hamarneh, G. (2021). Deep learning for biomedical image reconstruction: A survey. *Artificial Intelligence Review*, 54(1), 215–251. <https://doi.org/10.1007/s10462-020-09861-2>
- Carrington, A. M., Manuel, D. G., Fieguth, P. W., Ramsay, T., Osmani, V., Wernly, B., Bennett, C., Hawken, S., Magwood, O., & Sheikh, Y. (2022). Deep ROC analysis and AUC as balanced average accuracy, for improved classifier selection, audit and explanation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1), 329–341. <https://doi.org/10.1109/TPAMI.2022.3145392>
- Chen, L., Li, S., Bai, Q., Yang, J., Jiang, S., & Miao, Y. (2021). Review of image classification algorithms based on convolutional neural networks. *Remote Sensing*, 13(22),

4712.
<https://doi.org/10.3390/rs13224712>
- Chen, Z., Duan, J., Kang, L., & Qiu, G. (2021). A hybrid data-level ensemble to enable learning from highly imbalanced dataset. *Information Sciences*, 554, 157–176.
<https://doi.org/10.1016/j.ins.2020.12.023>
- Chicco, D., & Jurman, G. (2023). The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification. *BioData Mining*, 16(1), 4.
<https://doi.org/10.1186/s13040-023-00322-4>
- Davamani, K. A., Robin, C. R. R., Amudha, S., & Anbarasi, L. J. (2021). Biomedical image segmentation by deep learning methods. *Computational Analysis and Deep Learning for Medical Care: Principles, Methods, and Applications*, 131–154.
<https://doi.org/10.1002/9781119785750.ch6>
- Liang, Y., Li, S., Yan, C., Li, M., & Jiang, C. (2021). Explaining the black-box model: A survey of local interpretation methods for deep neural networks. *Neurocomputing*, 419, 168–182.
<https://doi.org/10.1016/j.neucom.2020.08.011>
- Mahmud, M., Kaiser, M. S., McGinnity, T. M., & Hussain, A. (2021). Deep learning in mining biological data. *Cognitive Computation*, 13(1), 1–33.
<https://doi.org/10.1007/s12559-020-09773-x>
- Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics*, 19(6), 1236–1246.
- Moskolai, W. R., Abdou, W., Dipanda, A., & Kolyang. (2021). Application of deep learning architectures for satellite image time series prediction: A review. *Remote Sensing*, 13(23), 4822.
<https://doi.org/10.3390/rs13234822>
- Mouliou, D. S., & Gourgoulianis, K. I. (2021). False-positive and false-negative COVID-19 cases: respiratory prevention and management strategies, vaccination, and further perspectives. *Expert Review of Respiratory Medicine*, 15(8), 993–1002.
<https://doi.org/10.1080/17476348.2021.1917389>
- Narkhede, M. V., Bartakke, P. P., & Sutaone, M. S. (2022). A review on weight initialization strategies for neural networks. *Artificial Intelligence Review*, 55(1), 291–322.
<https://doi.org/10.1007/s10462-021-10033-z>
- Shvetsova, N., Bakker, B., Fedulova, I., Schulz, H., & Dylov, D. V. (2021). Anomaly detection in medical imaging with deep perceptual autoencoders. *IEEE Access*, 9, 118571–118583.
<https://doi.org/10.1109/ACCESS.2021.3107163>
- Stahlschmidt, S. R., Ulfenborg, B., & Synnergren, J. (2022). Multimodal deep learning for biomedical data fusion: a review. *Briefings in Bioinformatics*, 23(2), bbab569.

- <https://doi.org/10.1093/bib/bbab569>
- Suganyadevi, S., Seethalakshmi, V., & Balasamy, K. (2022). A review on deep learning in medical image analysis. *International Journal of Multimedia Information Retrieval*, *11*(1), 19–38. <https://doi.org/10.1007/s13735-021-00218-1>
- Talaei Khoei, T., Ould Slimane, H., & Kaabouch, N. (2023). Deep learning: systematic review, models, challenges, and research directions. *Neural Computing and Applications*, *35*(31), 23103–23124. <https://doi.org/10.1007/s00521-023-08957-4>
- Wang, X., & Cheng, Z. (2020). Cross-Sectional Studies: Strengths, Weaknesses, and Recommendations. *Chest*, *158*(1, Supplement), S65–S71. <https://doi.org/https://doi.org/10.1016/j.chest.2020.03.012>
- Woessner, A. E., Anjum, U., Salman, H., Lear, J., Turner, J. T., Campbell, R., Beaudry, L., Zhan, J., Cornett, L. E., Gauch, S., & Quinn, K. P. (2024). Identifying and training deep learning neural networks on biomedical-related datasets. *Briefings in Bioinformatics*, *25*(Supplement_1), bbae232. <https://doi.org/10.1093/bib/bbae232>
- Xiao, S., Wu, Y., & Liu, H. (2020). Evolving status of the 2019 novel coronavirus infection: Proposal of conventional serologic assays for disease diagnosis and infection monitoring. *Journal of Medical Virology*, *92*(5), 464. <https://doi.org/10.1002/jmv.25702>
- Zhang, Y., Dong, Z., Li, S., & Cattani, C. (2023). Deep learning methods for biomedical information analysis. *Journal of Ambient Intelligence and Humanized Computing*, *14*(5), 5293–5296. <https://doi.org/10.1007/s12652-023-04617-6>