

Klasifikasi Ulkus Kaki Diabetik Berbasis Transfer Learning: Studi Komparatif Arsitektur CNN

Nur Nafiiyah^{1,*}, Rifqi Firdaus Al Jauhari¹, Agus Harjoko², Achmad Nizar Hidayanto³

¹Teknik Informatika, Fakultas Teknik, Universitas Islam Lamongan, Lamongan, Indonesia

²Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Gadjah Mada, Yogyakarta, Indonesia

³Sistem Informasi, Fakultas Ilmu Komputer, Universitas Indonesia, Jakarta, Indonesia

Email: ^{1,*}mynaff@unisla.ac.id, ²rifqifirdausaljauhari@gmail.com, ³aharjoko@ugm.ac.id, ⁴nizar@cs.ui.ac.id

Email Penulis Korespondensi: mynaff@unisla.ac.id*

Submitted: 07/02/2026; Accepted: 23/02/2026; Published: 31/03/2026

Abstrak– Proses klasifikasi citra medis secara otomatis sangat penting untuk membantu proses diagnosis yang cepat dan objektif. Deep learning, terutama *Convolutional Neural Network* (CNN), telah menunjukkan kemampuan yang lebih baik untuk mengidentifikasi dan mengklasifikasikan gambar. Namun, model yang tepat masih sulit dipilih karena perbedaan dalam desain arsitektur CNN dapat berdampak besar pada kinerja. Tujuan dari penelitian ini adalah untuk mengevaluasi dan membandingkan kinerja lima arsitektur CNN (VGG16, VGG19, ResNet101, InceptionV3, dan DenseNet121) ketika menggunakan dataset Ulkus Kaki Diabetes (Diabetic Foot Ulcer/DFU) untuk mengklasifikasikan gambar medis kelas normal dan abnormal. Metode berbasis dataset digunakan untuk menjalankan semua eksperimen. Akurasi, ketepatan, *recall*, dan *F1-score* digunakan untuk menilai model. Hasil eksperimen menunjukkan bahwa ResNet101 adalah yang terbaik dengan akurasi dan *F1-score* sebesar 99,87%, diikuti oleh VGG19 dan VGG16, masing-masing dengan *F1-score* di atas 99%. DenseNet121 dan InceptionV3 juga menunjukkan kinerja yang kompetitif, meskipun sedikit di bawah model berbasis residual dan VGG. Hasil ini menunjukkan betapa pentingnya melakukan analisis komparatif saat memilih arsitektur CNN untuk klasifikasi citra medis berbasis deep learning.

Kata Kunci: Diabetes; DFU; *Convolutional Neural Network*; Transfer Learning; Ulkus Kaki

Abstract– The automatic classification of medical images is crucial for rapid and objective diagnosis. Deep learning, particularly *Convolutional Neural Networks* (CNNs), has demonstrated improved image identification and classification capabilities. However, selecting the right model remains challenging as differences in CNN architecture design can significantly impact performance. The purpose of this study is to evaluate and compare the performance of five CNN architectures (VGG16, VGG19, ResNet101, InceptionV3, and DenseNet121) using the Diabetic Foot Ulcer (DFU) dataset to classify normal and abnormal medical images. A dataset-based approach was used to conduct all experiments. Accuracy, precision, recall, and *F1-score* were used to evaluate the models. The experimental results showed that ResNet101 performed best with an accuracy and *F1-score* of 99.87%, followed by VGG19 and VGG16, each with an *F1-score* above 99%. DenseNet121 and InceptionV3 also demonstrated competitive performance, although slightly below residual-based and VGG-based models. These results demonstrate the importance of conducting a comparative analysis when selecting a CNN architecture for deep learning-based medical image classification.

Keywords: Diabetes; DFU; *Convolutional Neural Network*; Transfer Learning; Foot Ulcer

1. PENDAHULUAN

Perkembangan *Artificial Intelligence* (AI) dan *Convolutional Neural Network* (ML) dalam beberapa tahun terakhir telah memberikan dampak yang sangat signifikan terhadap berbagai bidang, termasuk kesehatan, industri, pendidikan, dan sistem informasi. Dalam bidang medis, pemanfaatan AI semakin luas karena mampu membantu proses analisis data yang kompleks secara lebih cepat, konsisten, dan objektif. Salah satu bentuk penerapan AI yang berkembang pesat adalah pada analisis citra medis, di mana komputer dilatih untuk mengenali pola visual tertentu yang berhubungan dengan kondisi kesehatan pasien. Teknologi ini menjadi sangat penting karena citra medis sering kali mengandung informasi visual yang kompleks dan membutuhkan interpretasi yang teliti oleh tenaga ahli [1], [2], [3].

Salah satu cabang utama dari ML yang banyak digunakan dalam analisis citra adalah Deep Learning (DL). Berbeda dengan metode pembelajaran mesin konvensional yang umumnya memerlukan ekstraksi fitur secara manual, DL mampu melakukan ekstraksi fitur secara otomatis langsung dari data mentah. Kemampuan ini membuat DL sangat efektif dalam menangani data berdimensi tinggi, seperti gambar, video, dan sinyal medis. Dalam berbagai penelitian, DL telah menunjukkan performa yang sangat baik pada tugas-tugas seperti klasifikasi, segmentasi, deteksi objek, serta pengenalan pola dalam citra medis [4], [5]. Dengan kemampuan tersebut, DL menjadi salah satu pendekatan yang menjanjikan untuk mendukung sistem diagnosis berbantuan komputer (*computer-aided diagnosis*), sehingga dapat meningkatkan efisiensi dan akurasi dalam pengambilan keputusan klinis [6].

Dalam domain analisis citra, arsitektur DL yang paling banyak digunakan adalah *Convolutional Neural Network* (CNN). CNN dirancang khusus untuk memproses data berbentuk *grid*, seperti citra, dengan memanfaatkan operasi konvolusi untuk mengekstraksi fitur visual secara bertingkat. Lapisan-lapisan awal CNN

biasanya mempelajari fitur dasar seperti tepi, tekstur, dan bentuk sederhana, sedangkan lapisan yang lebih dalam mampu menangkap pola yang lebih kompleks dan spesifik terhadap objek yang dianalisis [7]. Oleh karena itu, CNN telah menjadi metode utama dalam berbagai aplikasi klasifikasi citra medis, termasuk deteksi kanker kulit, analisis citra retina, identifikasi pneumonia dari citra X-ray, dan klasifikasi luka kronis. Meskipun CNN memiliki kemampuan yang sangat baik dalam pembelajaran fitur, penerapan model ini secara langsung pada data medis sering menghadapi kendala utama, yaitu keterbatasan jumlah dataset berlabel. Dalam konteks medis, proses anotasi data umumnya memerlukan keterlibatan tenaga ahli seperti dokter atau spesialis, sehingga memakan waktu, biaya, dan sumber daya yang tidak sedikit. Selain itu, variasi kondisi pencahayaan, sudut pengambilan gambar, kualitas kamera, dan karakteristik luka dapat menambah kompleksitas dalam proses pelatihan model. Untuk mengatasi keterbatasan tersebut, pendekatan Transfer Learning (TL) menjadi solusi yang banyak digunakan dalam penelitian citra medis [8], [9]. Transfer learning merupakan teknik yang memanfaatkan model CNN yang telah dilatih sebelumnya pada dataset berskala besar, seperti ImageNet, kemudian menyesuaikannya kembali untuk tugas klasifikasi yang lebih spesifik. Dengan pendekatan ini, model tidak perlu mempelajari fitur dari awal, tetapi dapat memanfaatkan representasi visual umum yang sudah dipelajari sebelumnya. Hal ini sangat bermanfaat ketika jumlah data latih terbatas, karena model tetap dapat mencapai performa yang baik dengan waktu pelatihan yang lebih efisien [10], [11]. Oleh sebab itu, transfer learning menjadi salah satu strategi utama dalam pengembangan sistem klasifikasi citra medis modern. Salah satu permasalahan medis yang sangat relevan untuk dikaji dengan pendekatan ini adalah ulkus kaki diabetik.

Ulkus Kaki Diabetes (Diabetic Foot Ulcer, DFU) [12], [13] adalah salah satu komplikasi serius yang diderita oleh penderita diabetes dan memerlukan pengobatan segera. CNN dapat membantu klasifikasi otomatis citra DFU [14], [15]. Meskipun berbagai arsitektur CNN telah digunakan dalam penelitian klasifikasi citra medis, pemilihan arsitektur yang paling sesuai untuk suatu dataset tertentu masih menjadi tantangan. Setiap arsitektur CNN memiliki karakteristik desain yang berbeda, seperti kedalaman jaringan, jumlah parameter, mekanisme koneksi antarlapisan, serta efisiensi ekstraksi fitur. Perbedaan ini menyebabkan performa model dapat bervariasi meskipun diuji pada dataset yang sama. Meskipun berbagai jenis arsitektur CNN telah digunakan untuk tugas klasifikasi citra medis, masih sulit untuk memilih yang paling cocok [16], [17], [18]. Ini karena berbagai desain jaringan dapat menyebabkan dataset yang sama bekerja dengan cara yang berbeda.

Berdasarkan konteks ini, tujuan penelitian ini adalah untuk melakukan analisis perbandingan kinerja beberapa arsitektur CNN berbasis transfer learning (CNN), termasuk VGG16, VGG19, ResNet101, InceptionV3, dan DenseNet121, ketika diklasifikasikan gambar DFU menjadi kelas normal dan abnormal. Metode berbasis dataset digunakan untuk melakukan seluruh eksperimen, dan metrik akurasi, ketepatan, *recall*, dan *F1-score* digunakan untuk mengevaluasi kinerja. Fokus utama penelitian ini adalah memberikan evaluasi empiris yang dapat diukur mengenai perbedaan dalam kinerja antar arsitektur CNN. Diharapkan bahwa evaluasi ini akan membantu dalam menentukan model yang tepat untuk membangun sistem klasifikasi citra medis berbasis *deep learning*.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan utama yang saling berurutan, mulai dari persiapan data hingga evaluasi performa model klasifikasi citra Ulkus Kaki Diabetes (Diabetic Foot Ulcer/DFU). Tahapan-tahapan penelitian tersebut digambarkan dalam Gambar 1. Tahapan-tahapan tersebut meliputi pengambilan dan persiapan sampel data, pengembangan sampel data, desain arsitektur model, proses pelatihan, dan evaluasi performa model.

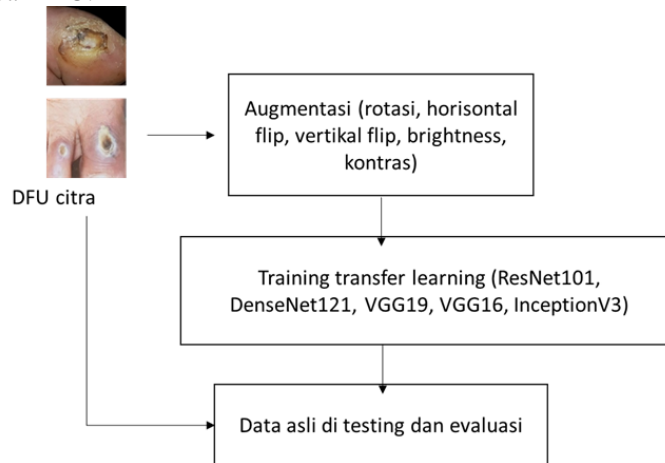
Penelitian ini menggunakan gambar ulkus kaki diabetes (DFU) dari dua kelas: normal dan abnormal. Semua gambar disesuaikan untuk tujuan penelitian dan diambil dari dataset publik. Dataset kemudian dibagi sesuai proporsi ke dalam data pelatihan dan data pengujian.

Proses rotasi, flip horisontal, flip vertikal, kecerahan, dan kontras dilakukan pada seluruh gambar selama tahap augmentasi. Lima arsitektur CNN berbasis transfer learning (CNN) yang digunakan dalam penelitian ini adalah DenseNet121, InceptionV3, ResNet101, VGG16, dan VGG19. Setiap model menggunakan bobot awal hasil pelatihan pada dataset ImageNet. Lapisan Global Average Pooling (GAP) dihapus dari masing-masing arsitektur dan digantikan oleh lapisan klasifikasi bawaan. Lapisan dense dengan 128 neuron berfungsi sebagai lapisan ekstraksi fitur tingkat tinggi. Pada langkah akhir, satu output layer yang memiliki fungsi aktivasi sigmoid untuk klasifikasi biner ditambahkan.

Setiap model dilatih dengan konfigurasi yang sama untuk memastikan perbandingan kinerja yang adil. Pelatihan dilakukan dengan ukuran batch dua dan jumlah epoch dua puluh. Dengan learning rate 0,0001, Stochastic Gradient Descent (SGD) adalah optimizer yang digunakan. Di sisi lain, loss function adalah binary cross-entropy. Ini sesuai dengan skenario klasifikasi dua kelas.

Data pelatihan digunakan untuk melatih setiap model hingga mencapai jumlah epoch yang telah ditetapkan. Untuk mengurangi kesalahan prediksi, bobot model diperbarui selama pelatihan berdasarkan nilai kerugian yang dihasilkan oleh loss function. Setelah pelatihan selesai, masing-masing model diuji dengan data DFU asli. Ini dilakukan dengan menggunakan metrik akurasi (Rumus 1), ketepatan (Rumus 2), *recall* (Rumus 3), dan *F1-score*

(Rumus 4). Rumus-rumus ini dirancang untuk memberikan gambaran mendalam tentang kemampuan model untuk mengklasifikasikan gambar DFU.



Gambar 1. Tahapan Penelitian

2.2 Classification Evaluation Metrics

Pada penelitian ini, metrik evaluasi klasifikasi biner digunakan untuk mengevaluasi kinerja model klasifikasi. Rumus 1 menggunakan metrik akurasi, Rumus 2 menggunakan *precision*, Rumus 3 menggunakan *recall*, dan Rumus 3 menggunakan *F1-score*. Keempat metrik ini dihitung berdasarkan nilai Positif Benar (TP), Negatif Benar (TN), Positif Benar (FP), dan Negatif Benar (FN) yang diperoleh dari matriks *confusion*. Akurasi ditentukan dengan menghitung proporsi prediksi yang benar terhadap seluruh data pengujian. *Precision* (Positive Predictive Value): kemampuan model untuk menemukan prediksi positif dengan tepat diukur oleh *precision*. *Recall* (Sensitivity atau True Positive Rate): kemampuan model untuk menemukan seluruh data positif yang sebenarnya diukur oleh *recall*. Untuk menyeimbangkan *precision* dan *recall*, *F1-score* adalah rata-rata harmonik.

Untuk mengklasifikasikan gambar Ulkus Kaki Diabetes (DFU) ke dalam dua kategori, yaitu normal dan abnormal, keempat metrik ini digunakan. Dalam klasifikasi medis, ketepatan dan *recall* sangat penting untuk mengurangi kesalahan diagnosis, tetapi akurasi memberikan gambaran umum tentang kinerja model. *F1-score* menunjukkan bahwa kemampuan deteksi dan ketepatan prediksi seimbang.

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1 - score = \frac{2*Precision*Recall}{Precision+Recall} \tag{4}$$

Hasil evaluasi dari seluruh arsitektur CNN dianalisis secara komparatif untuk mengidentifikasi model dengan performa terbaik. Analisis ini digunakan untuk menilai pengaruh pemilihan arsitektur CNN terhadap performa klasifikasi citra Ulkus Kaki Diabetes berbasis deep learning.

2.3 Dataset

Dataset penelitian ini diambil dari [12], dataset umum dari Kaggle. Kelas penyakit DFU adalah Abnormal (Ulkus) dan Normal (Kulit Sehat), dengan masing-masing kelas Ulkus 512 gambar dan Kulit Sehat 543 gambar (Tabel 1).

Tabel 1. Dataset Asli

No	Kelas	Total
1	Abnormal (Ulcer)	512
2	Normal (Healthy Skin)	543
Total		1055

Model transfer learning CNN dapat melakukan klasifikasi dengan akurasi tinggi jika dataset yang dilatih berukuran besar. Penelitian ini menambahkan dataset dengan augmentasi, metode augmentasinya adalah rotasi 900, pembalikan horizontal, pembalikan vertikal, kontras, dan kecerahan seperti pada penelitian [3]. Total data yang digunakan untuk pelatihan adalah 5275 gambar dan pengujian adalah 1055 gambar (Tabel 2).

Tabel 2. Dataset Penelitian

No	Kelas	Testing	Training	Total
1	Abnormal (Ulcer)	512	2560	3072
2	Normal (Healthy Skin)	543	2715	3258
Total		1055	5275	6330

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Hasil evaluasi performa lima arsitektur *Convolutional Neural Network* (CNN) dalam klasifikasi citra Ulkus Kaki Diabetes (Diabetic Foot Ulcer/DFU) disajikan dalam Tabel 3. Arsitektur-arsitektur ini adalah DenseNet121, InceptionV3, ResNet101, VGG16, dan VGG19. Untuk memberikan gambaran menyeluruh tentang kinerja model pada skenario klasifikasi dua kelas, evaluasi dilakukan menggunakan empat metrik: akurasi, *precision*, *recall*, dan *F1-score*. Hasil eksperimen menunjukkan bahwa ResNet101 menunjukkan performa terbaik dari semua model yang diuji, dengan nilai akurasi 99,87%, ketepatan 100%, dan *recall* 99,75%. Nilai-nilai ini menunjukkan bahwa ResNet101 memiliki kemampuan untuk melakukan klasifikasi dengan tingkat kesalahan yang tinggi.

Selain itu, model berbasis VGG menunjukkan kinerja yang sangat kompetitif. Dengan akurasi 99,52% dan *F1-score* 99,34%, VGG19 dan VGG16 masing-masing memiliki skor akurasi 99,15% dan *F1-score* 99,17%, masing-masing. Hasil yang konsisten pada metrik akurasi dan *recall* menunjukkan bahwa kedua model tersebut mampu mengenali secara seimbang kelas normal dan abnormal.

Namun, DenseNet121 dan InceptionV3 menunjukkan kinerja yang lebih buruk daripada model lain. DenseNet121 mendapatkan *F1-score* sebesar 97,35%, sedangkan InceptionV3 mendapatkan *F1-score* sebesar 95,76%. Namun, kedua model tetap kinerja yang kompetitif dan stabil, terutama dalam hal metrik ketepatan dan *recall*.

Tabel 3. Hasil Evaluasi

No	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
1	DenseNet121	97.23	96.08	98.65	97.35
2	InceptionV3	96.02	98.96	92.76	95.76
3	ResNet101	99.87	100	99.75	99.87
4	VGG16	99.15	98.98	99.37	99.17
5	VGG19	99.52	98.92	99.76	99.34

3.2 Pembahasan

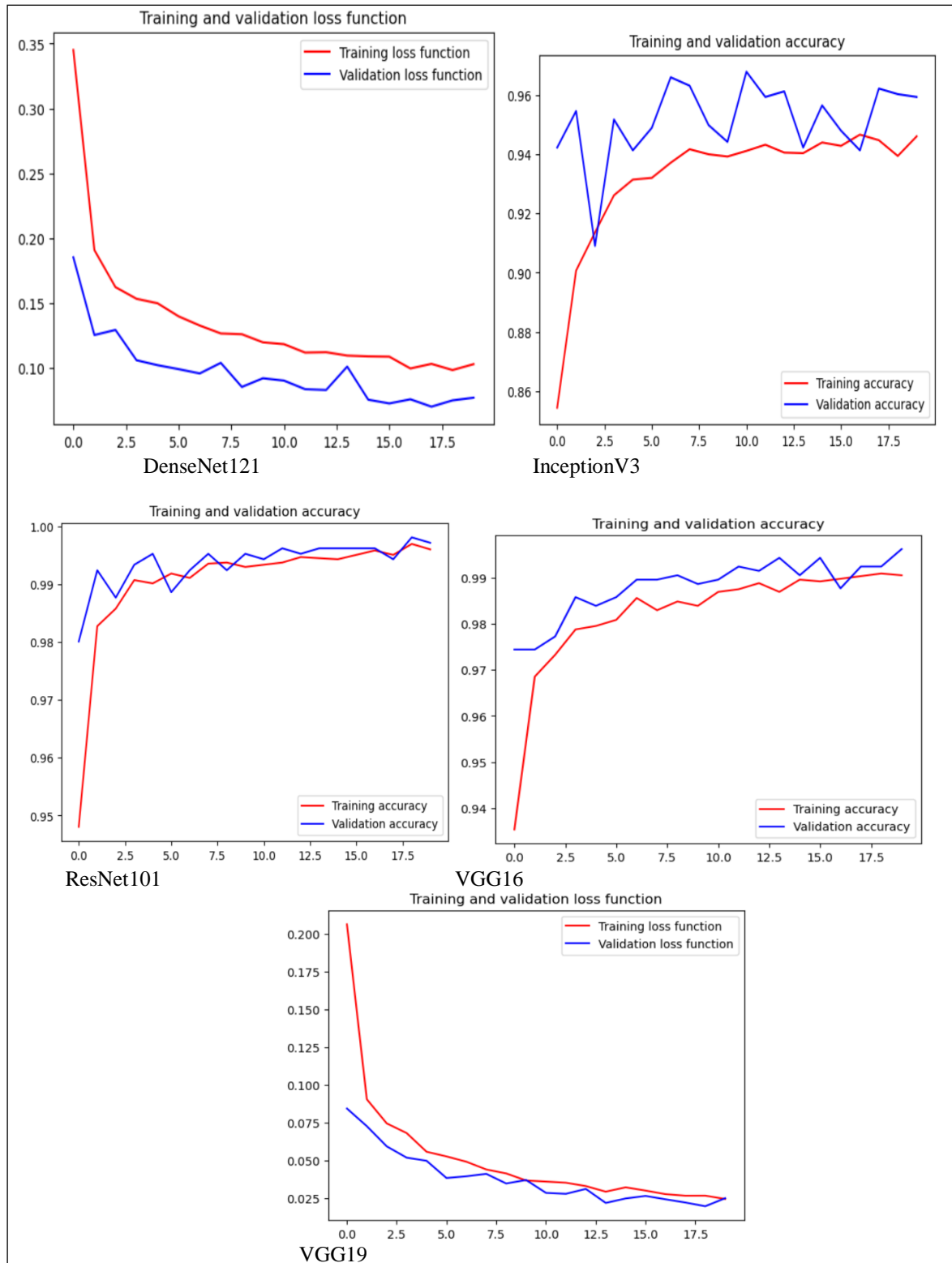
Hasil eksperimen menunjukkan bahwa arsitektur CNN sangat memengaruhi kinerja klasifikasi gambar DFU. Model berbasis pembelajaran residual, terutama ResNet101, memberikan hasil terbaik secara konsisten untuk setiap metrik evaluasi. Penelitian sebelumnya menunjukkan bahwa koneksi residual dapat membantu pembelajaran fitur yang lebih dalam dan stabil, terutama pada jaringan dengan kedalaman tinggi. Penemuan ini sejalan dengan temuan ini.

Meskipun memiliki arsitektur yang lebih sederhana dibandingkan ResNet101, model VGG16 dan VGG19 juga menunjukkan kinerja yang tinggi. Hal ini menunjukkan bahwa, ketika dikombinasikan dengan pendekatan transfer learning yang tepat dan dataset yang tepat, arsitektur CNN konvensional masih dapat berfungsi dengan baik untuk representasi fitur.

Sebaliknya, DenseNet121 dan InceptionV3 menunjukkan hasil yang lebih buruk pada metrik *recall* dan *F1-score*. Perbedaan ini menunjukkan bahwa kemampuan model berbeda dalam menyeimbangkan *precision* dan *recall*. Ini adalah komponen penting dalam klasifikasi gambar medis untuk mengurangi kesalahan prediksi pada kelas abnormal.

Secara keseluruhan, analisis komparatif ini memberikan pemahaman empiris tentang bagaimana arsitektur CNN berbeda dalam klasifikasi gambar DFU berbasis deep learning. Penemuan ini dapat digunakan sebagai referensi saat memilih model yang tepat untuk membangun sistem pendukung diagnosis berbasis gambar medis.

Ada perbedaan yang cukup jelas dalam kinerja antar arsitektur CNN yang diuji, seperti yang ditunjukkan pada Gambar 2 (Grafik Akurasi Model). Semua tiga model menunjukkan nilai akurasi tertinggi dan stabil, ResNet101, VGG19, dan VGG16, masing-masing dengan nilai di atas 99%. Ini menunjukkan bahwa ketiga model tersebut memiliki kemampuan untuk melakukan klasifikasi citra DFU dengan tingkat kesalahan yang sangat rendah. Sebaliknya, DenseNet121 dan InceptionV3 menunjukkan nilai akurasi yang lebih rendah dari kedua model tersebut. Meskipun demikian, kinerja kedua model tersebut masih dianggap baik dan konsisten, menunjukkan bahwa arsitektur tersebut masih mampu mempelajari pola visual citra DFU. Namun, pada dataset yang digunakan, arsitektur berbasis residual dan VGG tidak seoptimal. Grafik ini menunjukkan hasil kuantitatif dari tabel evaluasi bahwa arsitektur CNN memengaruhi akurasi klasifikasi.

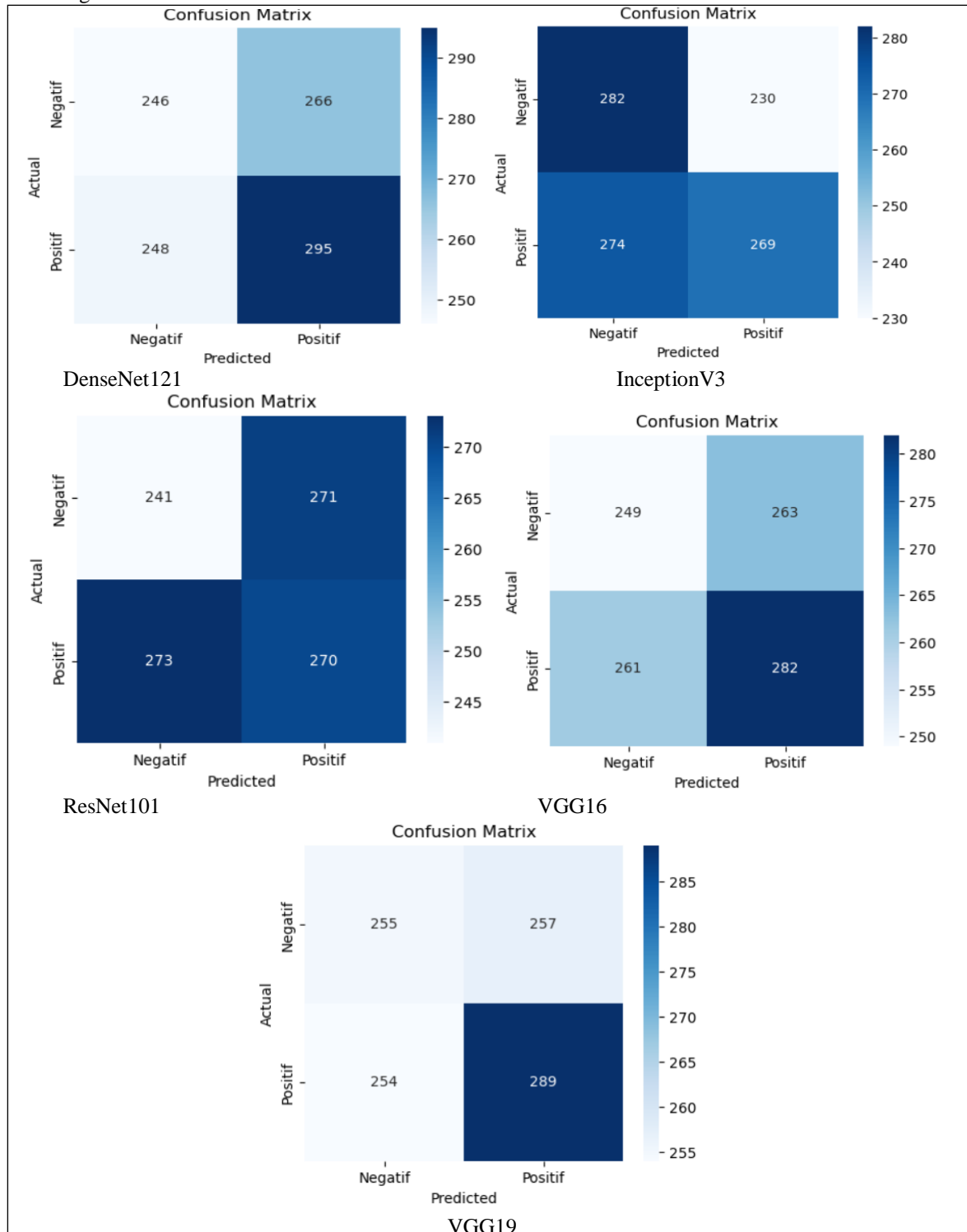


Gambar 2. Grafik Akurasi Model

Kemampuan masing-masing model untuk membedakan kelas normal dan abnormal ditunjukkan dalam Gambar 3 Model Matriks *Confusion*. Mayoritas prediksi pada model yang sangat baik, seperti ResNet101, VGG16, dan VGG19, ditemukan pada diagonal utama confusion matrix. Ini menunjukkan bahwa ada banyak prediksi yang benar untuk kedua kelas. Ini didukung oleh nilai *precision* dan *recall* yang tinggi, serta *F1-score* di atas 99%.

Salah satu kelas mengalami kesalahan klasifikasi yang lebih besar pada DenseNet121, meskipun prediksi benar masih dominan. Ini ditunjukkan oleh nilai *recall* yang lebih rendah dibandingkan model terbaik. Namun, InceptionV3 menunjukkan ketidakseimbangan yang lebih jelas antara ketepatan dan *recall*, menunjukkan bahwa model ini mungkin mengenali satu kelas lebih baik daripada kelas lainnya.

Secara keseluruhan, analisis confusion matrix menunjukkan bahwa model dengan *F1-score* tinggi dan akurasi juga memiliki distribusi prediksi yang lebih seimbang antar kelas; ini adalah elemen penting dalam klasifikasi gambar medis.



Gambar 3. Confusion Matrix Model

Untuk menginterpretasikan temuan penelitian ini, ada beberapa keterbatasan. Pertama, dataset yang digunakan tidak mewakili sepenuhnya berbagai kondisi klinis Diabetic Foot Ulcer (DFU) di lapangan. Berbagai variabel seperti variasi pencahayaan, sudut pengambilan gambar, tingkat keparahan luka, dan perbedaan perangkat akuisisi gambar semuanya termasuk dalam batas jumlah dan variasi gambar yang digunakan. Kedua, penelitian ini tidak menyelidiki klasifikasi multi-level berdasarkan kategori klinis yang lebih rinci atau tingkat keparahan luka, sehingga hanya berkonsentrasi pada skenario klasifikasi dua kelas: normal dan abnormal. Namun, dalam praktik medis, pengelompokan tingkat keparahan sangat penting untuk membantu dalam membuat keputusan tentang diagnosis dan prosedur penanganan. Ketiga, evaluasi model hanya menggunakan metrik kuantitatif seperti ketepatan, ketepatan, ulang, dan *F1-score*. Tidak ada analisis tambahan seperti uji statistik signifikansi

perbandingan model atau evaluasi ketahanan terhadap data eksternal. Keempat, penelitian ini belum melakukan analisis menyeluruh tentang elemen efisiensi komputasi, seperti waktu pelatihan, waktu inferensi, dan jumlah memori yang dibutuhkan. Semua elemen ini sangat penting untuk menerapkan sistem atau perangkat bergerak berbasis klinis. Terakhir, model belum diuji baik pada skenario real-time maupun integrasi langsung dalam sistem pendukung keputusan klinis. Oleh karena itu, evaluasi lebih lanjut diperlukan untuk memastikan kinerjanya dalam lingkungan operasional yang sebenarnya.

4. KESIMPULAN

Dengan menggunakan pendekatan berbasis dataset, penelitian ini mengevaluasi efektivitas penggunaan teknik transfer learning berbasis *Convolutional Neural Network* (CNN) untuk mengklasifikasikan gambar ulkus kaki diabetes (DFU) ke dalam dua kelas, yaitu normal dan abnormal. Untuk memastikan perbandingan performa yang adil dan objektif, lima arsitektur CNN (DenseNet121, InceptionV3, ResNet101, VGG16, dan VGG19) diuji menggunakan konfigurasi pelatihan yang sama. Seluruh model diubah dengan menambah lapisan Global Average Pooling (GAP) dan lapisan Dense dengan 128 neuron. Selain itu, optimizer Stochastic Gradient Descent (SGD) dilatih dengan learning rate 0,0001, ukuran batch 2, dan jumlah epoch 20. Hasil eksperimen menunjukkan bahwa pemilihan arsitektur CNN sangat memengaruhi kinerja klasifikasi citra DFU. ResNet101 menunjukkan kinerja terbaik secara konsisten pada setiap metrik evaluasi yang digunakan, dengan nilai akurasi dan *F1-score* tertinggi. VGG19 dan VGG16 juga menunjukkan kinerja yang sangat kompetitif. Meskipun mereka berada di bawah model berbasis residual dan VGG pada dataset yang digunakan, DenseNet121 dan InceptionV3 masih menunjukkan kinerja yang stabil. Hasil menunjukkan bahwa evaluasi multimetrik, terutama *F1-score*, penting untuk menilai keseimbangan antara *precision* dan *recall* dalam klasifikasi gambar medis. Secara keseluruhan, penelitian ini memberikan analisis komparatif yang terukur terhadap berbagai arsitektur CNN yang populer untuk klasifikasi DFU berbasis pembelajaran mendalam. Hasil penelitian ini diharapkan dapat menjadi referensi empiris bagi peneliti dan praktisi dalam memilih arsitektur CNN yang sesuai untuk mengembangkan sistem pendukung diagnosis berbasis citra medis. Untuk meningkatkan generalisasi dan efisiensi model, penelitian dapat dilanjutkan dengan memperluas variasi dataset, melakukan pengujian pada data lapangan, dan mengeksplorasi arsitektur yang lebih ringan.

REFERENCES

- [1] L. Alzubaidi, M. A. Fadhel, S. R. Oleiwi, O. Al-Shamma, and J. Zhang, "DFU_QUTNet: diabetic foot ulcer classification using novel deep *Convolutional Neural Network*," *Multimed. Tools Appl.*, vol. 79, no. 21–22, 2020, doi: 10.1007/s11042-019-07820-w.
- [2] M. S. A. Toofanee *et al.*, "DFU-SIAM a Novel Diabetic Foot Ulcer Classification With Deep Learning," *IEEE Access*, vol. 11, 2023, doi: 10.1109/ACCESS.2023.3312531.
- [3] S. Biswas, R. Mostafiz, B. K. Paul, K. M. M. Uddin, M. A. Hadi, and F. Khanom, "DFU_XAI: A Deep Learning-Based Approach to Diabetic Foot Ulcer Detection Using Feature Explainability," *Biomedical Materials and Devices*, vol. 2, no. 2, 2024, doi: 10.1007/s44174-024-00165-5.
- [4] H. A. Ismael, N. H. Al-A'araji, and B. K. Shukur, "An Enhanced Diabetic Foot Ulcer Classification Approach Using GLCM and Deep Convolution Neural Network," *Karbala International Journal of Modern Science*, vol. 8, no. 4, 2022, doi: 10.33640/2405-609X.3268.
- [5] N. Al-Garaawi, R. Ebsim, A. F. H. Alharan, and M. H. Yap, "Diabetic foot ulcer classification using mapped binary patterns and *Convolutional Neural Networks*," *Comput. Biol. Med.*, vol. 140, 2022, doi: 10.1016/j.compbiomed.2021.105055.
- [6] R. Nanda, A. Nath, S. Patel, and E. Mohapatra, "*Convolutional Neural Network* algorithm to evaluate risk factors of diabetic foot ulcers and its severity," *Med. Biol. Eng. Comput.*, vol. 60, no. 8, 2022, doi: 10.1007/s11517-022-02617-w.
- [7] M. H. Yap, B. Cassidy, J. M. Pappachan, C. O'Shea, D. Gillespie, and N. D. Reeves, "Analysis towards classification of infection and ischaemia of diabetic foot ulcers," in *BHI 2021 - 2021 IEEE EMBS International Conference on Biomedical and Health Informatics, Proceedings*, 2021. doi: 10.1109/BHI50953.2021.9508563.
- [8] M. Ahsan, S. Naz, R. Ahmad, H. Ehsan, and A. Sikandar, "A Deep Learning Approach for Diabetic Foot Ulcer Classification and Recognition," *Information (Switzerland)*, vol. 14, no. 1, 2023, doi: 10.3390/info14010036.
- [9] S. Nagaraju *et al.*, "Automated Diabetic Foot Ulcer Detection and Classification Using Deep Learning," *IEEE Access*, vol. 11, 2023, doi: 10.1109/ACCESS.2023.3332292.
- [10] P. Xie *et al.*, "An explainable *Convolutional Neural Network* model for predicting in-hospital amputation rate of patients with diabetic foot ulcer," *Int. Wound J.*, vol. 19, no. 4, 2022, doi: 10.1111/iwj.13691.
- [11] A. Khandakar *et al.*, "A Novel *Convolutional Neural Network* Approach for Severity Classification of Diabetic Foot Complications Using Thermogram Images," *Sensors*, vol. 22, no. 11, 2022, doi: 10.3390/s22114249.
- [12] M. A. Fadhel, L. Alzubaidi, Y. Gu, J. Santamaría, and Y. Duan, "Real-time diabetic foot ulcer classification based on deep learning & parallel hardware computational tools," *Multimed. Tools Appl.*, vol. 83, no. 27, 2024, doi: 10.1007/s11042-024-18304-x.



- [13] Z. Schäfer, A. Mathisen, K. Svendsen, S. Engberg, T. Rolighed Thomsen, and K. Kirketerp-Møller, “Toward Machine-Learning-Based Decision Support in Diabetes Care: A Risk Stratification Study on Diabetic Foot Ulcer and Amputation,” *Front. Med. (Lausanne)*, vol. 7, 2021, doi: 10.3389/fmed.2020.601602.
- [14] M. H. Yap *et al.*, “Deep learning in diabetic foot ulcers detection: A comprehensive evaluation,” *Comput. Biol. Med.*, vol. 135, 2021, doi: 10.1016/j.combiomed.2021.104596.
- [15] S. K. Das, S. Namasudra, A. Kumar, and N. R. Moparthy, “AESPNet: Attention Enhanced Stacked Parallel Network to improve automatic Diabetic Foot Ulcer identification,” *Image Vis. Comput.*, vol. 138, 2023, doi: 10.1016/j.imavis.2023.104809.
- [16] L. Alzubaidi, A. A. Abbood, M. A. Fadhel, O. Al-Shamma, and J. Zhang, “Comparison of hybrid *Convolutional Neural Networks* models for diabetic foot ulcer classification,” *Journal of Engineering Science and Technology*, vol. 16, no. 3, 2021.
- [17] F. Arnia, K. Saddami, R. Roslidar, R. Muharar, and K. Munadi, “Towards accurate Diabetic Foot Ulcer image classification: Leveraging CNN pre-trained features and extreme learning machine,” *Smart Health*, vol. 33, Sep. 2024, doi: 10.1016/j.smhl.2024.100502.
- [18] E. Santos, F. Santos, J. Dallyson, K. Aires, J. M. R. S. Tavares, and R. Veras, “Diabetic Foot Ulcers Classification using a fine-tuned CNNs Ensemble,” in *Proceedings - IEEE Symposium on Computer-Based Medical Systems*, 2022. doi: 10.1109/CBMS55023.2022.00056.