

Explainable-AI (XAI) dengan GradCAM terhadap Arsitektur ResNet50V2 pada Klasifikasi Luka Kaki Diabetes

Khairul Hafidh ^{1*}, Eka Ardiani Putri ², Andrean Taufik ³

^{1,3} Program Studi Informatika, Fakultas Teknik, Universitas Tanjungpura, Kota Pontianak, Provinsi Kalimantan Barat, Indonesia.

² Program Studi Pendidikan Dokter, Fakultas Kedokteran, Universitas Tanjungpura, Kota Pontianak, Provinsi Kalimantan Barat, Indonesia.

ABSTRACT

This study developed an image classification system for Diabetic Foot Ulcers (DFU) to distinguish between Abnormal (Ulcer) and Normal (Healthy Skin) classes using deep learning and Explainable Artificial Intelligence (XAI). An initial dataset of 2,855 images underwent resizing to 224×224 pixels, class balancing using SMOTE, partitioning with a 70:15:15 ratio, and data augmentation. SMOTE generated a balanced dataset of 4,624 images, while augmentation expanded the training set to 16,180 images. The comparative evaluation examined ResNet50 and ResNet50V2 based on accuracy, precision, recall, and F1-score. ResNet50 achieved an accuracy of 91.21%, whereas ResNet50V2 attained 98.99% with only seven misclassifications. ResNet50V2 was identified as the superior model, achieving performance metrics of approximately 99% across all evaluation criteria. Grad-CAM was implemented to provide visual interpretability through heatmaps highlighting the regions influencing the model's predictions. For ulcer classifications, Grad-CAM accurately highlighted the wound margins with output probabilities reaching 1.0. These results demonstrate that integrating ResNet50V2 with Grad-CAM produces both high diagnostic accuracy and transparent model reasoning, offering strong potential for computer-aided DFU screening, although validation on external clinical cohorts remains necessary.

ABSTRAK

Penelitian ini mengembangkan sistem klasifikasi citra luka kaki diabetik (diabetic foot ulcer/DFU) untuk membedakan kelas Abnormal (Ulcer) dan Normal (Healthy Skin) menggunakan deep learning dan Explainable Artificial Intelligence (XAI). Dataset awal sebanyak 2.855 citra diproses melalui pengubahan ukuran menjadi 224×224 piksel, penyeimbangan kelas menggunakan Synthetic Minority Over-sampling Technique (SMOTE), pembagian data dengan rasio 70:15:15, dan augmentasi data. Penerapan SMOTE menghasilkan 4.624 citra seimbang, sedangkan augmentasi meningkatkan data latih menjadi 16.180 citra. Penelitian ini membandingkan kinerja arsitektur ResNet50 dan ResNet50V2 berdasarkan metrik akurasi, presisi, recall, dan F1-score. ResNet50 memperoleh akurasi 91,21%, sedangkan ResNet50V2 mencapai akurasi 98,99% dengan hanya tujuh kesalahan klasifikasi. ResNet50V2 terpilih sebagai model terbaik karena seluruh metrik performanya mencapai kisaran 99%. Grad-CAM diterapkan untuk menyediakan interpretabilitas visual melalui heatmap yang memetakan area fokus model dalam membuat keputusan. Pada citra ulkus, Grad-CAM memperlihatkan fokus visual yang tepat pada area lesi luka dengan probabilitas prediksi hingga 1,0. Hasil penelitian menunjukkan bahwa integrasi ResNet50V2 dan Grad-CAM menghasilkan klasifikasi yang akurat sekaligus transparan, sehingga berpotensi mendukung analisis klinis DFU berbasis citra, meskipun validasi eksternal pada populasi yang lebih luas masih diperlukan.

ARTICLE HISTORY

Received: 5 September 2026.

Revised: 11 September 2026.

Accepted: 10 October 2026.

Available Online: 20 October 2026.

Published: 1 December 2026.

KEYWORDS

Diabetic Foot Ulcer (DFU); Deep Learning; ResNet50V2; Explainable Artificial Intelligence (XAI); Grad-CAM.

KATA KUNCI

Diabetic Foot Ulcer; Deep Learning; ResNet50V2; Explainable Artificial Intelligence; Grad-CAM.

1. Pendahuluan

Diabetes melitus merupakan penyakit metabolik kronis yang dapat menimbulkan berbagai komplikasi serius, salah satunya luka kaki diabetik (*diabetic foot ulcer/DFU*) yang berisiko berkembang menjadi infeksi berat, nekrosis jaringan, hingga amputasi apabila tidak diidentifikasi dan ditangani secara dini. Secara visual, karakteristik DFU menunjukkan tingkat variabilitas yang tinggi, mencakup ragam warna luka, morfologi bentuk, dimensi ukuran, tekstur jaringan, derajat keparahan, serta fluktuasi pencahayaan pada citra klinis. Karakteristik tersebut menyebabkan proses diagnosis dan pemantauan lesi secara visual menjadi kompleks. Kemajuan teknologi *computer vision* dan *deep learning* telah membuka peluang klasifikasi lesi secara otomatis dan objektif melalui analisis citra digital (Hafidh & Asrin, 2023). Kendati demikian, kompleksitas arsitektur model *deep learning* sering kali memunculkan sifat kotak hitam (*black box*). Hal ini menimbulkan kendala kepercayaan di ranah medis karena klinisi membutuhkan transparansi dan interpretabilitas atas area lesi yang menjadi dasar penentuan keputusan prediksi model (Muralidhara *et al.*, 2022; Xu *et al.*, 2022).

Eksplorasi penggunaan *deep learning* pada klasifikasi DFU telah menunjukkan perkembangan yang signifikan. Xu *et al.* (2022) memanfaatkan *class knowledge banks* untuk membedakan kondisi infeksi dan iskemia dengan mengekstraksi representasi patologis visual. Pendekatan lain dilakukan oleh Muralidhara *et al.* (2022) yang memanfaatkan citra termal plantar untuk mengklasifikasikan tingkat keparahan ulkus. Evaluasi komparatif berbagai arsitektur *Convolutional Neural Network* (CNN), termasuk ResNet50, juga dilaporkan mampu menghasilkan kinerja yang tinggi pada deteksi komplikasi luka (Ahsan *et al.*, 2023). Upaya pengembangan berlanjut pada optimasi inferensi waktu nyata (*real-time*) (Fadhel *et al.*, 2024) serta pemanfaatan *transfer learning* pada skenario keterbatasan data (*few-shot learning*) yang mencatatkan akurasi sebesar 98,67% (C. Wang *et al.*, 2024). Landasan teoretis arsitektur *residual network* generasi kedua (ResNetV2) didasarkan pada perancangan *pre-activation residual block*, yaitu penempatan *Batch Normalization* dan ReLU sebelum lapisan konvolusi dengan tetap mempertahankan *identity mapping* pada jalur *shortcut* (He *et al.*, 2016) [cek referensi: pastikan apakah merujuk ke He *et al.*, 2016 atau Gevers & Smeulders, 2016]. Karakteristik struktural ini memfasilitasi propagasi sinyal informasi dan gradien secara langsung tanpa hambatan pada jaringan yang sangat dalam, sehingga berpotensi mempercepat optimasi konvergensi fitur visual citra luka kaki diabetik yang memiliki kompleksitas tekstur tinggi.

Seiring tuntutan keandalan klinis, penelitian DFU terkini mulai menaruh perhatian besar pada aspek interpretabilitas model (*explainability*). Sathya Preiya dan Kumar (2023) menunjukkan bahwa pola fitur visual yang diekstraksi lapisan dalam CNN berkorelasi langsung dengan karakteristik patogenesis ulkus. Pengelompokan kondisi klinis luka menjadi status normal, terinfeksi, berulkus, dan nekrosis juga telah berhasil dipetakan menggunakan ResNet50 (Z. Wang *et al.*, 2024). Guna membuka mekanisme internal prediksi, Rathore *et al.* (2025) mengembangkan kerangka kerja DFU_XAI yang mengintegrasikan teknik SHAP, LIME, dan *Gradient-weighted Class Activation Mapping* (Grad-CAM), di mana Grad-CAM terbukti efektif memvisualisasikan lokalisasi ulkus. Pada lingkup nasional, integrasi *deep transfer learning* dan metode atribusi visual serupa juga diimplementasikan pada citra dermoskopi untuk klasifikasi lesi kulit (Ramadhan & Zeniarja, 2025).

Meskipun kajian klasifikasi DFU berbasis *deep learning* telah berkembang pesat, sebagian besar penelitian terdahulu masih memusatkan evaluasi pada pencapaian metrik akurasi melalui modifikasi arsitektur yang kompleks, augmentasi data, maupun mekanisme *few-shot*. Kajian komprehensif yang mengkorelasikan akurasi tinggi dengan interpretabilitas visual yang efisien masih terbatas. Model mutakhir seperti ConMatFormer telah mengintegrasikan modul *attention*, arsitektur Transformer, dan teknik XAI majemuk (Grad-CAM, Grad-CAM++, dan LIME) guna memperjelas keputusan diagnosis DFU (Rifat *et al.*, 2025). Akan tetapi, tingginya kompleksitas parameter pada model berbasis *multi-attention* tersebut meningkatkan beban komputasi secara signifikan serta mempersulit verifikasi hubungan langsung antara pola gradien fitur dengan batas morfologi luka.

Untuk mengatasi kesenjangan tersebut, penelitian ini mengusulkan penerapan *Explainable Artificial Intelligence* (XAI) berbasis Grad-CAM yang diintegrasikan secara langsung pada arsitektur ResNet50V2 untuk klasifikasi biner citra DFU antara kelas abnormal (ulkus) dan normal (kulit sehat), dengan membandingkannya terhadap ResNet50 standar sebagai pembanding *baseline*. ResNet50V2 dipilih karena mekanisme *pre-activation residual learning* mampu mengekstraksi representasi fitur hierarkis secara stabil, sementara Grad-CAM dimanfaatkan untuk memproyeksikan *heatmap* atensi visual guna memastikan bahwa

keputusan klasifikasi didasarkan pada karakteristik patologis luka, bukan pada artefak latar belakang citra. Integrasi ini bertujuan menghasilkan model klasifikasi yang tidak hanya memiliki akurasi diagnosis yang unggul, tetapi juga transparan, akuntabel, dan dapat diinterpretasi secara visual oleh tenaga medis, sehingga meningkatkan penerimaan klinis terhadap penerapan kecerdasan artifisial dalam penanganan luka diabetik (Hüsers *et al.*, 2022; Zhang *et al.*, 2022).

2. Metodologi

Penelitian ini menerapkan kerangka kerja komparatif berbasis pembelajaran mendalam (*deep learning*) dan teknik atribusi visual untuk mengklasifikasikan citra luka kaki diabetik secara akurat dan terinterpretasi. Alur kerja penelitian dirancang secara sistematis mulai dari penyiapan data klinis, penyeimbangan distribusi kelas, pelatihan arsitektur *residual network*, hingga pengujian transparansi keputusan model melalui pemetaan gradien spasial. Keseluruhan tahapan pelaksanaan penelitian tersebut disajikan secara skematis pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

Secara keseluruhan, metodologi penelitian terdiri atas empat tahapan utama, yaitu: pengumpulan dataset, praproses data, pembangunan model, dan evaluasi sistem. Uraian rinci dari masing-masing tahapan dijelaskan pada subbagian berikut.

2.1 Pengumpulan Dataset

Penelitian ini memanfaatkan dataset citra luka kaki diabetik (DFU) yang dihimpun dari repositori publik Shafi (2025) dan Siddiqui (2025). Data dari repositori Shafi mencakup 543 citra kelas normal dan 512 citra kelas abnormal (ulkus). Sementara itu, data tambahan dari repositori Siddiqui yang sebelumnya diintegrasikan pada repositori Roboflow dengan akurasi rujukan sistem awal sebesar 69,5% terdiri atas citra dengan derajat keparahan luka tingkat 1 hingga 4. Seluruh citra grade 1 hingga grade 4 tersebut kemudian dilebur menjadi satu kategori kelas abnormal (ulkus), yang dikumpulkan dari direktori uji, direktori validasi, serta direktori pelatihan tanpa perlakuan augmentasi, dengan jumlah masing-masing tingkatan keparahan sebanyak 450 citra. Melalui penggabungan kedua sumber tersebut, diperoleh dataset awal sebanyak 2.855 citra. Volume data ini lebih besar dibandingkan dataset yang digunakan pada penelitian Amjad *et al.* (2024) serta Biswas *et al.* (2023) yang mencapai akurasi 98,75% menggunakan arsitektur ResNet50.

2.2 Praproses Data

Tahap praproses diawali dengan penyeragaman resolusi spasial seluruh citra menjadi 224×224 piksel guna menyesuaikan dimensi masukan baku keluarga arsitektur ResNet serta menjaga efisiensi komputasi konvolusi. Untuk menjaga validitas evaluasi dan mencegah kebocoran data (*data leakage*), dataset awal dipartisi terlebih dahulu menjadi tiga subset independen dengan rasio 70% untuk data pelatihan, 15% untuk data validasi, dan 15% untuk data pengujian. Penanganan ketimpangan distribusi kelas (*class imbalance*) dilakukan secara eksklusif pada subset data pelatihan menggunakan Synthetic Minority Over-sampling Technique (SMOTE) merujuk pada prinsip Phankokkruad (2021). Pada tahap ini, setiap citra RGB berukuran $224 \times 224 \times 3$ diratakan (*flattened*) menjadi vektor fitur satu dimensi sepanjang 150.528 nilai intensitas piksel sebagai representasi numerik masukan algoritma. SMOTE kemudian membangkitkan sampel sintesis baru melalui interpolasi linier antarsampel pada kelas minoritas hingga tercapai keseimbangan jumlah kelas, sebelum direkonstruksi kembali ke matriks citra $224 \times 224 \times 3$. Setelah distribusi seimbang tercapai, variasi data pelatihan diperkaya lebih lanjut menggunakan augmentasi citra. Transformasi dilakukan dengan memilih secara acak lima kombinasi dari delapan teknik yang tersedia, meliputi rotasi (*rotation*), pergeseran lebar (*width shift*), pergeseran tinggi (*height shift*), pergeseran sudut (*shear*), pembesaran (*zoom*), pembalikan horizontal (*horizontal flip*), penyesuaian pencahayaan (*brightness*), serta pengisian area kosong (*fill mode*) (Permana Ellison *et al.*, 2025).

2.3 Pembangunan Model

Pengembangan model klasifikasi diimplementasikan menggunakan kerangka kerja Keras dengan membandingkan dua arsitektur Convolutional Neural Network (CNN), yaitu ResNet50 dan ResNet50V2. Kedua model memanfaatkan bobot pra-latih ImageNet (*transfer learning*) sebagai pengekstraksi fitur dasar (*feature extractor*). Pembangunan model dilakukan tanpa penyesuaian bobot menyeluruh (*fine-tuning*), di mana seluruh lapisan konvolusi dasar dibekukan (*frozen*) dan pelatihan hanya difokuskan pada pembaruan bobot lapisan klasifikasi akhir (*fully connected layer*) yang ditambahkan. Proses komputasi menggunakan algoritma optimasi Adam (*Adaptive Moment Estimation*) dengan tingkat pembelajaran (*learning rate*) konstan sebesar 0,0001. Pemilihan optimizer Adam didasarkan pada kemampuan adaptasi laju gradien per parameter, kebutuhan memori yang efisien, dan stabilitas konvergensinya pada citra medis (Ahsan *et al.*, 2023). Guna memastikan keterulangan (*reproducibility*) hasil eksperimen, inisialisasi bobot dan pengacakan data dikunci menggunakan nilai konstan (*random seed* = 42).

2.4 Evaluasi Sistem

Tahap evaluasi sistem dirancang untuk mengukur performa diagnostik dan tingkat transparansi keputusan model. Evaluasi ini terbagi menjadi dua aspek utama: evaluasi kinerja kuantitatif dan evaluasi interpretasi visual secara kualitatif. Evaluasi kuantitatif dilakukan dengan mengukur metrik akurasi (*accuracy*), presisi (*precision*), sensitivitas (*recall*), dan nilai F1 (*F1-score*) pada kedua arsitektur model menggunakan subset data pengujian yang belum pernah dilihat sebelumnya. Sementara itu, evaluasi kualitatif dilakukan melalui penerapan metode Gradient-weighted Class Activation Mapping (Grad-CAM) untuk menghasilkan visualisasi heatmap yang memetakan proyeksi gradien pada lapisan konvolusi terakhir, sehingga area lesi klinis yang memicu keputusan klasifikasi dapat divalidasi secara visual.

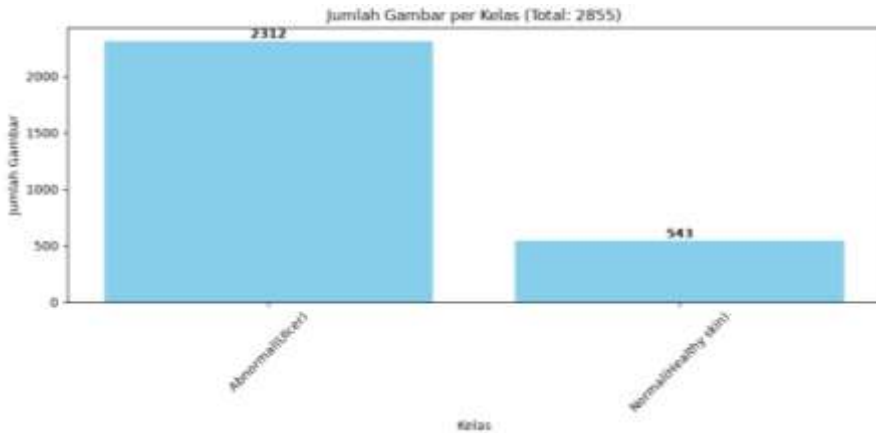
3. Hasil

Tahap pengumpulan dataset menghasilkan 2.855 citra klinis yang terbagi ke dalam dua kategori, yaitu kelas Abnormal (Ulcer) dan Normal (Healthy Skin) sebagaimana ditampilkan pada Gambar 2.



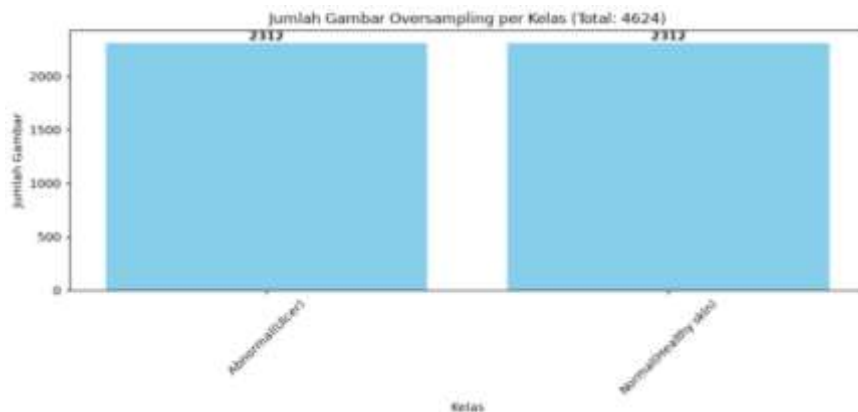
Gambar 2. Sampel Dua Kelas Citra DFU: (a) Abnormal (Ulcer); (b) Normal (Healthy Skin)

Distribusi awal dataset menunjukkan ketidakseimbangan kelas yang signifikan, dengan jumlah sampel normal yang lebih dominan dibandingkan kelas ulkus, sebagaimana dipetakan pada Gambar 3.



Gambar 3. Sebaran Distribusi Dataset Awal

Penerapan penyeimbangan data menggunakan SMOTE menghasilkan dataset sintesis yang terdistribusi secara seimbang antar-kelas dengan total akumulasi sebanyak 4.624 citra, seperti ditunjukkan pada Gambar 4.



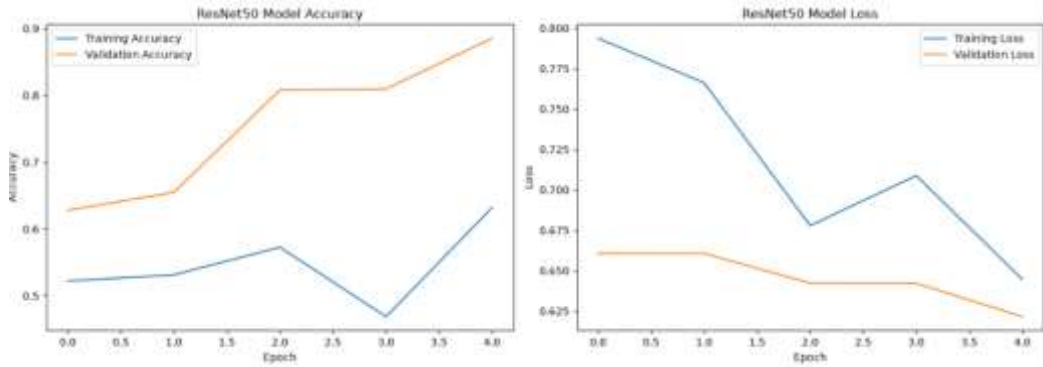
Gambar 4. Distribusi Dataset Setelah Penyeimbangan dengan SMOTE

Dataset seimbang tersebut selanjutnya dipartisi dengan rasio 70% untuk data pelatihan (3.236 citra), 15% untuk data validasi (694 citra), dan 15% untuk data pengujian (694 citra). Subset data latih kemudian diperluas melalui prosedur augmentasi acak, menghasilkan lima citra variasi baru dari setiap satu citra asli. Perluasan ini meningkatkan volume subset pelatihan menjadi 16.180 citra. Komposisi data latih sebelum dan setelah augmentasi dirinci pada Tabel 1.

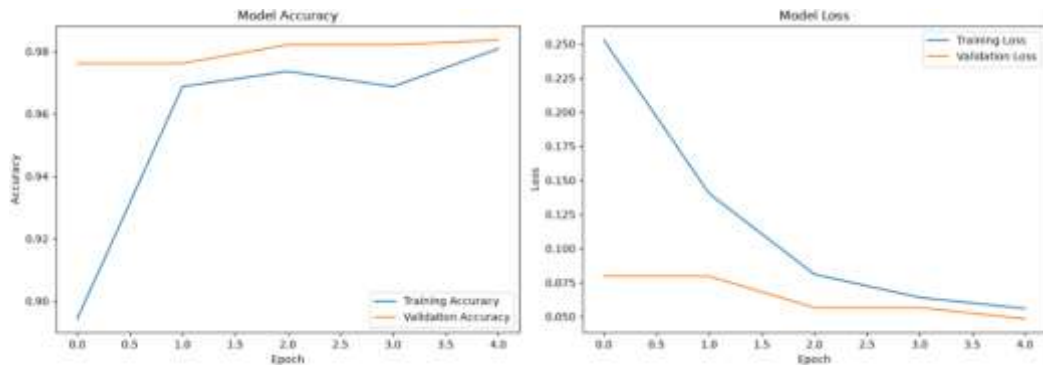
Tabel 1. Distribusi Jumlah Citra Data Latih Sebelum dan Setelah Augmentasi

Kelas	Sebelum Augmentasi	Setelah Augmentasi
Abnormal (<i>Ulcer</i>)	1.618	8.090
Normal (<i>Healthy Skin</i>)	1.618	8.090
Total	3.236	16.180

Proses komputasi pelatihan model diimplementasikan pada perangkat komputasi dengan spesifikasi prosesor Intel Core i5 dan memori 12 GB RAM. Pelatihan model ResNet50 memerlukan waktu komputasi 1 jam 53 menit 8 detik untuk menyelesaikan lima epoch, sedangkan ResNet50V2 membutuhkan durasi 12 menit 38 detik lebih lama. Pola konvergensi nilai akurasi dan fungsi *loss* kedua arsitektur selama proses pelatihan disajikan pada Gambar 5 dan Gambar 6.

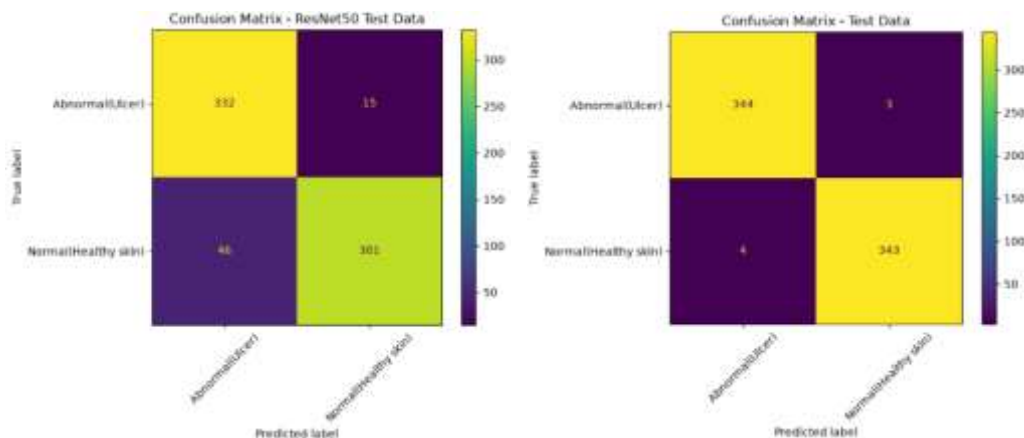


Gambar 5. Grafik Akurasi dan Loss Pelatihan Model ResNet50



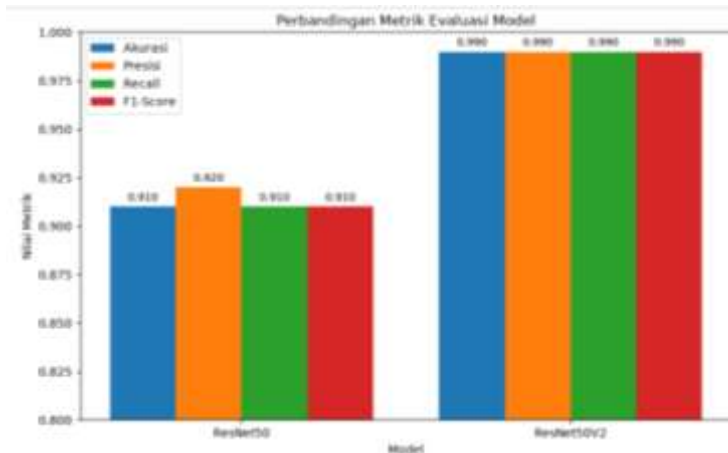
Gambar 6. Grafik Akurasi dan Loss Pelatihan Model ResNet50V2

Berdasarkan kurva pembelajaran pada Gambar 5, arsitektur ResNet50 mencatatkan peningkatan akurasi validasi dari 62,80% pada epoch 1 menjadi 88,54% pada epoch 5, disertai penurunan *validation loss* dari 0,6607 menjadi 0,6215. Sementara itu, Gambar 6 memperlihatkan ResNet50V2 mencapai konvergensi yang lebih cepat dan stabil; akurasi pelatihan meningkat dari 89,44% menjadi 98,09% dan akurasi validasi mencapai 98,36%. Nilai *loss* pelatihan ResNet50V2 tereduksi dari 0,2528 menjadi 0,0560, dengan *validation loss* akhir sebesar 0,0485. Pengujian generalisasi model dilakukan terhadap 694 citra uji independen. Distribusi prediksi benar dan galat klasifikasi dari kedua model direpresentasikan melalui matriks konfusi pada Gambar 7.



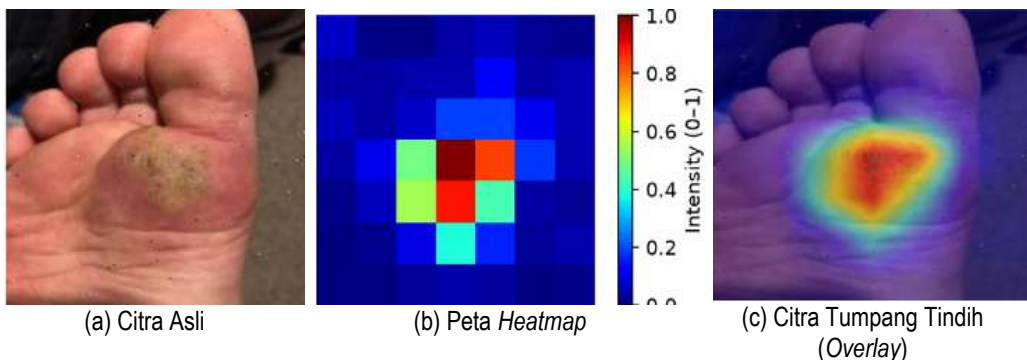
(a) ResNet50 (b) ResNet50V2
Gambar 7. Matriks Konfusi pada Data Uji: (a) ResNet50; (b) ResNet50V2

Berdasarkan matriks konfusi Gambar 7(a), ResNet50 mengklasifikasikan 332 citra ulkus dan 301 citra kulit normal secara benar, namun menghasilkan 15 *false negative* dan 46 *false positive*, sehingga mencatatkan akurasi pengujian sebesar 91,21%. Sebaliknya, matriks konfusi ResNet50V2 pada Gambar 7(b) menunjukkan peningkatan ketepatan prediksi dengan 344 citra ulkus dan 343 citra normal terprediksi secara akurat, serta hanya menghasilkan 3 *false negative* dan 4 *false positive*, menghasilkan akurasi sebesar 98,99%. Perbandingan ringkasan performa kedua model pada data uji disajikan pada Gambar 8.

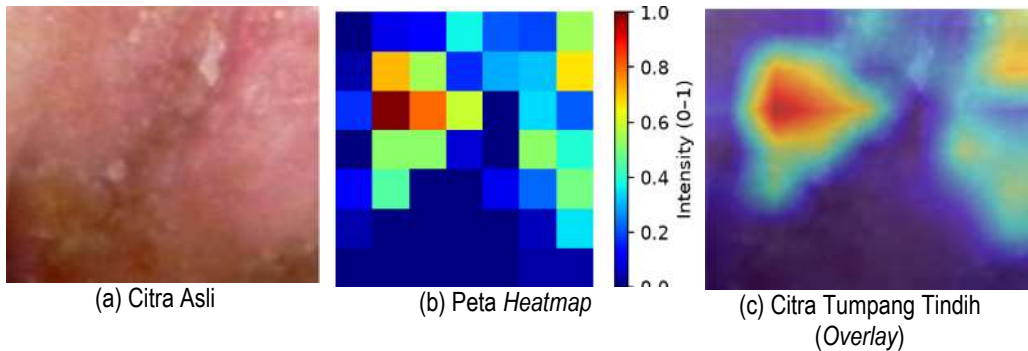


Gambar 8. Perbandingan Metrik Kinerja ResNet50 dan ResNet50V2 terhadap Data Uji

Secara komparatif, ResNet50 memperoleh skor presisi 0,920, sensitivitas (*recall*) 0,910, dan skor F1 sebesar 0,910. Pada sisi lain, ResNet50V2 secara konsisten mencapai skor 0,990 pada seluruh parameter evaluasi (akurasi, presisi, *recall*, dan skor F1), yang mencerminkan penurunan total kesalahan prediksi dari 61 citra pada ResNet50 menjadi hanya 7 citra pada ResNet50V2. Evaluasi interpretabilitas visual model terbaik (ResNet50V2) dilakukan menggunakan teknik Grad-CAM untuk memetakan distribusi bobot atensi visual. Representasi visual pada kasus prediksi benar (*true positive*) dan kesalahan prediksi (*false negative*) disajikan pada Gambar 9 dan Gambar 10.



Gambar 9. Implementasi Grad-CAM pada Prediksi Benar (*True Positive*): (a) Citra Asli; (b) Peta *Heatmap*; (c) Citra Tumpang Tindih (*Overlay*)



Gambar 10. Implementasi Grad-CAM pada Kasus Kesalahan Prediksi (*False Negative*): (a) Citra Asli; (b) Peta *Heatmap*; (c) Citra Tumpang Tindih (*Overlay*)

Pada Gambar 9, visualisasi Grad-CAM pada sampel prediksi benar memperlihatkan konsentrasi intensitas warna merah pekat yang terfokus tepat pada batas lesi ulkus dengan probabilitas keluaran model sebesar 1,0. Sebaliknya, pada kasus kegagalan klasifikasi Gambar 10, peta aktivasi model tidak terfokus pada lesi luka melainkan tersebar ke jaringan perifer di sekitarnya dengan nilai probabilitas prediksi yang menurun menjadi 0,68.

4. Pembahasan

Hasil eksperimen menunjukkan bahwa ResNet50V2 menghasilkan performa diagnostik yang lebih unggul dibandingkan arsitektur pendahulunya, ResNet50, dengan margin peningkatan akurasi sebesar 7,78 poin persentase pada data uji yang sama. Keunggulan ini berkaitan langsung dengan perbedaan struktural blok residual pada kedua arsitektur. ResNet50V2 menerapkan modifikasi pra-aktivasi (*pre-activation*), di mana fungsi normalisasi *Batch Normalization* dan aktivasi non-linier ReLU ditempatkan sebelum lapisan konvolusi. Pengaturan urutan ini memungkinkan koneksi pintas (*shortcut connection*) beroperasi sebagai pemetaan identitas (*identity mapping*) murni tanpa distorsi sinyal, sehingga memfasilitasi propagasi gradien yang lebih stabil dan langsung melintasi lapisan-lapisan dalam jaringan tanpa degradasi informasi. Sebaliknya, pada ResNet50 standar, kehadiran operasi aktivasi pasca-penjumlahan (*post-activation*) cenderung meredam besaran gradien selama proses propagasi balik, yang membatasi kemampuan model dalam menangkap fitur morfologi lesi yang subtil. Stabilitas arsitektur ResNet50V2 juga tecermin pada kurva pelatihan yang menunjukkan konvergensi mulus tanpa anomali osilasi maupun gejala *overfitting* yang ekstrem. Tingkat akurasi validasi yang sedikit melampaui akurasi data latih pada fase konvergensi akhir menandakan bahwa penerapan transformasi augmentasi data geometris dan fotometris berhasil berfungsi sebagai regularisator spasial yang efektif. Augmentasi acak tersebut mempersulit tugas optimasi pada data latih secara proporsional, sehingga model terdorong untuk mempelajari representasi pola patologis yang invarian terhadap perubahan sudut, skala, dan variasi pencahayaan klinis.

Dari perspektif interpretabilitas, visualisasi atribusi fitur melalui Grad-CAM membuktikan bahwa keunggulan kuantitatif ResNet50V2 ditopang oleh dasar penalaran visual yang valid secara medis. Pada sampel klasifikasi benar (*true positive*), konsentrasi gradien aktivasi model terkunci secara presisi pada jaringan granulasi dan tepi ulkus, bukan pada artefak visual latar belakang seperti seprai, jari kaki yang sehat, atau bayangan pencahayaan. Fenomena ini membuktikan bahwa jaringan secara mandiri mampu mengisolasi biomarker visual lesi patologis DFU. Sebaliknya, analisis kesalahan (*error analysis*) pada sampel *false negative* (Gambar 10) mengungkap keterbatasan model saat berhadapan dengan lesi berukuran kecil, kontras rendah, atau batas lesi yang menyerupai pigmentasi kulit normal. Pada kondisi tersebut, mekanisme atensi visual model mengalami disorientasi spasial dan gagal memusatkan bobot gradien pada area lesi yang sebenarnya, menghasilkan penurunan probabilitas kepercayaan menjadi 0,68. Capaian akurasi ResNet50V2 sebesar 98,99% dalam penelitian ini sejalan dan kompetitif dengan temuan literatur terkini. Kinerja ini melampaui hasil penelitian Wang *et al.* (2024) yang membukukan akurasi 98,67% melalui skema *few-shot learning*, serta melampaui akurasi model ResNet50 pada studi Biswas *et al.* (2023)

yang mencapai 98,75%. Meskipun demikian, integrasi ResNet50V2 dan Grad-CAM menawarkan keunggulan komparatif dari segi kompleksitas komputasi yang lebih efisien bila dibandingkan dengan model hibrida multi-atensi transformer seperti ConMatFormer (Rifat *et al.*, 2025). Dengan mempertahankan arsitektur residual konvolusional tunggal yang diimbui eksplanasi visual transparan, model ini menyediakan keseimbangan optimal antara kecepatan komputasi inferensi dan akuntabilitas diagnostik. Temuan ini menegaskan potensi ResNet50V2 terintegrasi Grad-CAM sebagai instrumen penapisan dini DFU yang andal, dengan catatan bahwa validasi prospektif pada data kohort eksternal dengan variasi kondisi klinis yang lebih heterogen tetap menjadi prasyarat mutlak sebelum implementasi klinis nyata.

5. Kesimpulan

Penelitian ini berhasil membuktikan bahwa integrasi arsitektur ResNet50V2 dan metode Explainable Artificial Intelligence (XAI) berbasis Grad-CAM efektif meningkatkan akurasi serta transparansi pada klasifikasi citra luka kaki diabetik (DFU). Berdasarkan hasil evaluasi komparatif terhadap 694 citra uji, ResNet50V2 menunjukkan keunggulan yang konsisten dibandingkan model pembanding ResNet50. ResNet50V2 mencatatkan tingkat akurasi pengujian sebesar 98,99%, dengan nilai presisi, sensitivitas (*recall*), dan skor F1 yang seluruhnya mencapai 99,0%, serta berhasil menekan jumlah galat prediksi menjadi hanya 7 citra (3 *false negative* dan 4 *false positive*). Sebaliknya, model ResNet50 standar hanya memperoleh akurasi 91,21% dengan 61 kesalahan klasifikasi. Keunggulan tersebut dipengaruhi oleh desain *pre-activation residual block* pada ResNet50V2 yang memfasilitasi propagasi sinyal gradien secara lebih langsung, sehingga model lebih efektif dalam mempelajari karakteristik morfologis lesi yang kompleks. Selain keunggulan pada metrik kuantitatif, penerapan Grad-CAM mampu mengatasi sifat kotak hitam (*black box*) model dengan menghasilkan visualisasi *heatmap* atensi visual yang terpusat tepat pada batas lesi ulkus pada kasus prediksi benar. Kemampuan atribusi visual ini memberikan bukti empiris bahwa keputusan klasifikasi didasarkan pada karakteristik patologi luka, bukan pada artefak latar belakang citra. Meskipun demikian, evaluasi terhadap kasus kegagalan klasifikasi menunjukkan bahwa model masih rentan mengalami disorientasi spasial pada lesi berukuran kecil atau lesi dengan batas pigmentasi yang menyerupai jaringan kulit normal. Untuk menunjang kesiapan implementasi klinis yang nyata, penelitian lanjutan perlu difokuskan pada pengujian validasi eksternal menggunakan data prospektif dari berbagai fasilitas layanan kesehatan dengan populasi pasien dan kondisi pencahayaan yang lebih heterogen, serta pengembangan deteksi bertingkat untuk mengklasifikasikan derajat keparahan luka secara lebih komprehensif.

Ucapan Terima Kasih

Ucapan terima kasih kepada Lembaga Penelitian dan Pengabdian Kepada Masyarakat (LPPM) UNTAN atas dukungan pendanaan Penelitian Dosen Muda Universitas Tanjungpura Tahun 2026.

Referensi

- Ahsan, M., Naz, S., Ahmad, R., Ehsan, H., & Sikandar, A. (2023). A deep learning approach for diabetic foot ulcer classification and recognition. *Information*, 14(1), Article 36. <https://doi.org/10.3390/info14010036>
- Amjad, K., Asif, S., Waheed, Z., & Guo, Y. (2024). A novel lightweight deep learning framework with knowledge distillation for efficient diabetic foot ulcer detection. *Applied Soft Computing*, 167, Article 112296. <https://doi.org/10.1016/j.asoc.2024.112296>

- Biswas, S., Mostafiz, R., Paul, B. K., Mohi Uddin, K. M., Rahman, M. M., & Shariful, F. N. U. (2023). DFU_MultiNet: A deep neural network approach for detecting diabetic foot ulcers through multi-scale feature fusion using the DFU dataset. *Intelligence-Based Medicine*, 8, Article 100128. <https://doi.org/10.1016/j.ibmed.2023.100128>
- Fadhel, M. A., Alzubaidi, L., Gu, Y., Santamaria, J., & Duan, Y. (2024). Real-time diabetic foot ulcer classification based on deep learning & parallel hardware computational tools. *Multimedia Tools and Applications*, 83(27), 70369–70394. <https://doi.org/10.1007/s11042-024-18304-x>
- Hafidh, K., & Asrin, F. (2023). Pemodelan sistem untuk identifikasi stadium *Plasmodium falciparum* pada citra mikroskopis malaria dengan teknik *object counting*. *Jurnal Jaringan Sistem Informasi Robotik (JSR)*, 7(2), 153–157.
- Hüsers, J., Hafer, G., Heggemann, J., Wiemeyer, S., Przysucha, M., Dissemond, J., & Hübner, U. (2022). Automatic classification of diabetic foot ulcer images - A transfer-learning approach to detect wound maceration. *Studies in Health Technology and Informatics*, 289, 301–304. <https://doi.org/10.3233/SHTI210919>
- Muralidhara, S., Lucieri, A., Dengel, A., & Ahmed, S. (2022). Holistic multi-class classification & grading of diabetic foot ulcerations from plantar thermal images using deep learning. *Health Information Science and Systems*, 10(1), Article 20. <https://doi.org/10.1007/s13755-022-00194-8>
- Permana Ellion, A., Dwi Nyoto, R., & Hafidh, K. (2025). Analisis explainability model CNN untuk klasifikasi luka: Studi komparasi ResNet50V2, DenseNet201, dan MobileNetV2 dengan Grad-CAM. *Jurnal Sistem dan Teknologi Informasi*, 13(1), 45–54.
- Phankokkrud, M. (2021). Ensemble transfer learning for lung cancer detection. Dalam *Proceedings of the 2021 5th International Conference on Digital Technology in Education (ICDTE 2021)* (hlm. 438–442). Association for Computing Machinery. <https://doi.org/10.1145/3478905.3478995>
- Ramadhan, M. E., & Zeniarja, J. (2025). Implementasi deep transfer learning dan explainable AI dalam klasifikasi kanker kulit. *Jurnal SISTEMASI: Sistem Informasi*, 14(1), 112–121.
- Rathore, P. S., Kumar, A., Nandal, A., Dhaka, A., & Sharma, A. K. (2025). A feature explainability-based deep learning technique for diabetic foot ulcer identification. *Scientific Reports*, 15(1), Article 4512. <https://doi.org/10.1038/s41598-025-90780-z>
- Rifat, R. A., Bhoyan, F. H., Mehedi, M. H. K., Hossain, M. K., Hossen, M. J., & Mridha, M. F. (2025). ConMatFormer: A multi-attention and transformer integrated ConvNext based deep learning model for enhanced diabetic foot ulcer classification. *Results in Engineering*, 28, Article 108248. <https://doi.org/10.1016/j.rineng.2025.108248>
- Sathya Preiya, V., & Kumar, V. D. A. (2023). Deep learning-based classification and feature extraction for predicting pathogenesis of foot ulcers in patients with diabetes. *Diagnostics*, 13(12), Article 1983. <https://doi.org/10.3390/diagnostics13121983>
- Shafi, M. (2025). *Diabetic foot ulcer predictions*. Kaggle. <https://www.kaggle.com/code/shafi4/diabetic-foot-ulcer-predictions/input>
- Siddiqui, K. (2025). *DFU dataset annotated into 4 classes*. Kaggle. <https://www.kaggle.com/datasets/khalidsiddiqui2003/dfu-dataset-annotated-into-4-classes>

- Wang, C., Yu, Z., Long, Z., Zhao, H., & Wang, Z. (2024). A few-shot diabetes foot ulcer image classification method based on deep ResNet and transfer learning. *Scientific Reports*, 14(1), Article 30691. <https://doi.org/10.1038/s41598-024-80691-w>
- Wang, Z., Tan, X., Xue, Y., Xiao, C., Yue, K., Lin, K., & Zhang, J. (2024). Smart diabetic foot ulcer scoring system. *Scientific Reports*, 14(1), Article 12076. <https://doi.org/10.1038/s41598-024-62076-1>
- Xu, Y., Han, K., Zhou, Y., Wu, J., Xie, X., & Xiang, W. (2022). Classification of diabetic foot ulcers using class knowledge banks. *Frontiers in Bioengineering and Biotechnology*, 9, Article 811028. <https://doi.org/10.3389/fbioe.2021.811028>
- Zhang, J., Qiu, Y., Peng, L., Zhou, Q., Wang, Z., & Qi, M. (2022). A comprehensive review of methods based on deep learning for diabetes-related foot ulcers. *Frontiers in Endocrinology*, 13, Article 945020. <https://doi.org/10.3389/fendo.2022.945020>