(2025), 3 (7): 233-257 **Scientica** 3021-8209

Jurnal Ilmiah Sain dan Teknologi

ANALISIS *EXPLAINABILITY* MODEL *CNN* UNTUK KLASIFIKASI LUKA: STUDI KOMPARASI RESNET50V2, DENSENET201, DAN MOBILENETV2 DENGAN GRAD-CAM

Angga Permana Ellion [1], Rudy Dwi Nyoto [2], Khairul Hafidh [3]
Universitas Tanjungpura
Fakultas Teknik, Jurusan Informatika
Bidang Informatika
anggapermanaellion814@gmail.com

Abstract
Skin wounds are a common health issue that, if not properly managed, can lead to serious infections. Artificial intelligence (AI) technology, particularly Convolutional Neural Networks (CNNs), has been widely utilized for medical image analysis, including wound classification. This research aims to analyze and compare the performance of pre-trained CNN models—ResNet50V2, DenseNet201, and MobileNetV2—in classifying six types of wounds (abrasion, burn, laceration, bruise, cut, stab) and to evaluate the explainability quality of each model using Grad-CAM. Wound image data was obtained from Kaggle, comprising a total of 400 original images. These were then resized to 224×224 pixels and oversampled using SMOTE until each class contained 122 images, resulting in a total of 732 images. The dataset was split into 70% training data (510 images), 15% validation data (108 images), and 15% testing data (114 images), with the training data subsequently augmented to 2550 images. The quantitative evaluation results indicate that ResNet50V2 excels in metric performance with an accuracy of 0.93, sensitivity (recall) of 0.80, specificity of 0.96, PPV (precision) of 0.82, NPV of 0.96, and an F1-score of 0.79. In terms of training duration, MobileNetV2 was the most efficient (35 minutes 21 seconds) compared to ResNet50V2 (3 hours 4 minutes 7 seconds) and DenseNet201 (4 hours 49 minutes 7 seconds). The explainability analysis using Grad-CAM revealed that ResNet50V2 was also the most effective in focusing attention on relevant wound areas (33 out of 48 test images focused on the wound area), outperforming DenseNet201 and MobileNetV2, which tended to be more frequently distracted. The "Cut" class demonstrated the best performance in ResNet50V2, while the "Bruise" and "Abrasion" classes exhibited the highest rates of misprediction. Specifically, "Bruise" was frequently confused with "Abrasion," and "Burn" was often mistaken for "Cut," "Abrasion," or "Laceration," representing potentially fatal misclassifications. This research concludes that ResNet50V2 is the best model as it is not only accurate but also possesses superior explainability, allowing it to focus attention on relevant visual features of the wound. For future research, it is recommended to perform fine-tuning on the models, increase data diversity, and conduct Grad-CAM heatmap validation with medical experts.

Article History
Received: 5 Desember 2025
Reviewed: 8 Desember 2025
Published: 9 Desember 2025

Key Words
Wound Classification, Convolutional Neural Network (CNN), ResNet50V2, DenseNet201, MobileNetV2, Explainable AI (XAI), Grad-CAM, Transfer Learning.

Abstrak
Luka kulit merupakan masalah kesehatan umum yang jika tidak ditangani dengan benar dapat menyebabkan infeksi serius. Teknologi kecerdasan buatan (AI), *khususnya Convolutional Neural Network* (*CNN*), telah banyak dimanfaatkan untuk analisis citra medis, termasuk klasifikasi luka. Penelitian ini bertujuan untuk menganalisis dan membandingkan performa model *CNN* pre-trained ResNet50V2, DenseNet201, dan MobileNetV2 dalam klasifikasi enam jenis luka (abrasi, bakar, laserasi, lebam, sayat, tusuk) serta mengevaluasi kualitas *explainability* masing-masing model menggunakan Grad-CAM. Data citra luka diperoleh dari Kaggle, dengan total 400 gambar asli yang kemudian diproses melalui *resize* menjadi 224x224 piksel dan di-*oversampling* menggunakan *SMOTE* hingga setiap

Sejarah Artikel
Received: 5 Desember 2025
Reviewed: 8 Desember 2025
Published: 9 Desember 2025

Kata Kunci
Klasifikasi Luka, Convolutional Neural Network (*CNN*), ResNet50V2, DenseNet201, MobileNetV2, Explainable AI (*XAI*), Grad-CAM, Transfer Learning.

kelas memiliki 122 gambar, sehingga total menjadi 732 gambar. Dataset dibagi menjadi 70% data latih (510 gambar), 15% data validasi (108 gambar), dan 15% data uji (114 gambar), di mana data latih kemudian diaugmentasi menjadi 2550 gambar. Hasil evaluasi kuantitatif menunjukkan bahwa ResNet50V2 unggul dalam performa metrik dengan akurasi 0,93, sensitivitas (*recall*) 0,80, spesifisitas 0,96, PPV (presisi) 0,82, NPV 0,96, dan *F1-score* 0,79. Durasi pelatihan MobileNetV2 adalah yang paling efisien (35 menit 21 detik) dibandingkan ResNet50V2 (3 jam 4 menit 7 detik) dan DenseNet201 (4 jam 49 menit 7 detik). Analisis *explainability* menggunakan Grad-CAM menunjukkan bahwa ResNet50V2 juga paling efektif dalam memfokuskan perhatian pada area luka yang relevan (33 dari 48 gambar uji fokus pada area luka), mengungguli DenseNet201 dan MobileNetV2 yang cenderung lebih sering terdistraksi. Kelas "Sayat" menunjukkan performa terbaik di ResNet50V2, sementara kelas "Lebam" dan "Abrasi" memiliki tingkat kesalahan prediksi tertinggi, dengan "Lebam" sering tertukar dengan "Abrasi", dan "Bakar" sering tertukar dengan "Sayat", "Abrasi", atau "Laserasi" yang berisiko fatal. Penelitian ini menyimpulkan bahwa ResNet50V2 merupakan model terbaik karena tidak hanya akurat tetapi juga memiliki explainability yang superior, memungkinkannya untuk memusatkan perhatian pada fitur-fitur visual luka yang relevan. Disarankan untuk melakukan fine-tuning model, menambah diversitas data, serta melakukan validasi *heatmap* Grad-CAM dengan ahli medis untuk penelitian selanjutnya.

PENDAHULUAN

Luka pada kulit adalah masalah kesehatan yang sangat umum. Hampir setiap orang pernah mengalaminya, terutama anak-anak yang aktif bermain. Penyebabnya bisa macam-macam, mulai dari jatuh, tergores benda tajam, hingga luka bakar. Meski sering dianggap ringan, luka bisa jadi serius jika tidak dirawat dengan benar. Kebersihan yang buruk bisa membuat luka terinfeksi oleh bakteri atau kuman lain, yang akibatnya bisa fatal, bahkan sampai menyebabkan kematian (Guo, S., & DiPietro, L. A., 2010). Di Indonesia sendiri, menurut data Badan Pusat Statistik, jumlah kasus luka ringan terus naik setiap tahunnya, mencapai 160.449 kasus pada tahun 2022 (Badan Pusat Statistik Indonesia, 2024). Ini menunjukkan bahwa penanganan luka yang tepat masih menjadi isu penting.

Untuk membantu proses diagnosis dan perawatan luka, teknologi kecerdasan buatan (*AI*) kini mulai banyak dimanfaatkan, khususnya di bidang analisis gambar medis. Salah satu teknologi *AI* yang paling populer untuk tugas ini adalah *Convolutional Neural Network* atau *CNN*. Kehebatan *CNN* adalah kemampuannya untuk belajar mengenali pola dari gambar secara otomatis, sama seperti cara manusia belajar membedakan objek (Litjens et al., 2017). Dalam penelitian ini, akan digunakan tiga arsitektur *CNN* yang dikenal andal, yaitu ResNet50V2, DenseNet201, dan MobileNetV2. Ketiganya terbukti punya performa tinggi untuk tugas klasifikasi gambar (He et al., 2016; Huang et al., 2017; Sandler et al., 2018).

Penelitian klasifikasi citra medis dengan model *pretrained* telah banyak dilakukan. Contohnya, Tarek et al (2024) mengembangkan BCa*XAI* menggunakan 3D Inception-ResNetV2 dan Grad-CAM untuk klasifikasi kanker payudara, mencapai akurasi 98.53%. Manop Phankokkruad (2021) mengusung *ensemble transfer learning* dengan kombinasi VGG16, ResNet50V2, dan DenseNet201 untuk deteksi kanker paru-paru, mencapai akurasi validasi 91%. Sementara itu, penelitian tentang deteksi luka luar dengan *transfer learning* menggunakan VGG-16, Inception-v3, dan ResNet-50 menunjukkan VGG-16 sebagai model terbaik dengan akurasi uji 97%.

Model *deep learning* yang canggih, meskipun akurat, seringkali berfungsi sebagai "kotak hitam" (*black box*). Artinya, meskipun kita mengetahui hasil prediksinya, mekanisme di balik keputusan tersebut tidak transparan. Dalam konteks medis, ini menjadi masalah serius, karena

dokter atau tenaga medis tidak dapat sepenuhnya memercayai keputusan mesin tanpa memahami dasar pertimbangannya (Holzinger et al., 2019). Hal ini menimbulkan risiko bagi pasien jika model salah mengidentifikasi luka karena fokus pada fitur yang tidak relevan dalam gambar.

Untuk mengatasi permasalahan "kotak hitam" ini, muncul bidang ilmu *Explainable AI* (*XAI*). Tujuan *XAI* adalah meningkatkan transparansi dan pemahaman terhadap cara kerja AI. Salah satu teknik *XAI* yang relevan dan akan digunakan dalam penelitian ini *adalah Gradient-weighted Class Activation Mapping* (Grad-CAM). Grad-CAM mampu menghasilkan "peta panas" (*heatmap*) yang memvisualisasikan bagian gambar mana yang paling penting bagi model CNN saat membuat keputusan (Selvaraju et al., 2017), sehingga memungkinkan kita untuk "mengintip" proses penalaran model.

Banyak penelitian sebelumnya cenderung hanya berfokus pada perbandingan akurasi model. Namun, pertanyaan krusial yang sering terabaikan adalah, di antara model-model yang akurat tersebut, manakah yang memiliki cara kerja paling logis dan dapat dipercaya? Apakah model benar-benar memfokuskan perhatian pada ciri-ciri luka yang relevan secara medis?

Mengambil inspirasi dari riset-riset sebelumnya dan menyadari pentingnya explainable AI (*XAI*), penelitian ini mengusulkan pendekatan yang menyeluruh. Dalam penelitian ini aku menguji dan membandingkan performa kuantitatif beberapa model pretrained berdasarkan metrik Akurasi, Sensitivitas (*Recall*), Spesifisitas, PPV (Presisi), NPV, dan *F1-score* serta mengintegrasikan uji interpretasi menggunakan Grad-CAM. Pendekatan ini bertujuan untuk tidak hanya mengidentifikasi model dengan kinerja tertinggi, tetapi juga menemukan model yang paling logis dan dapat dijelaskan dalam konteks klasifikasi citra medis.

METODE

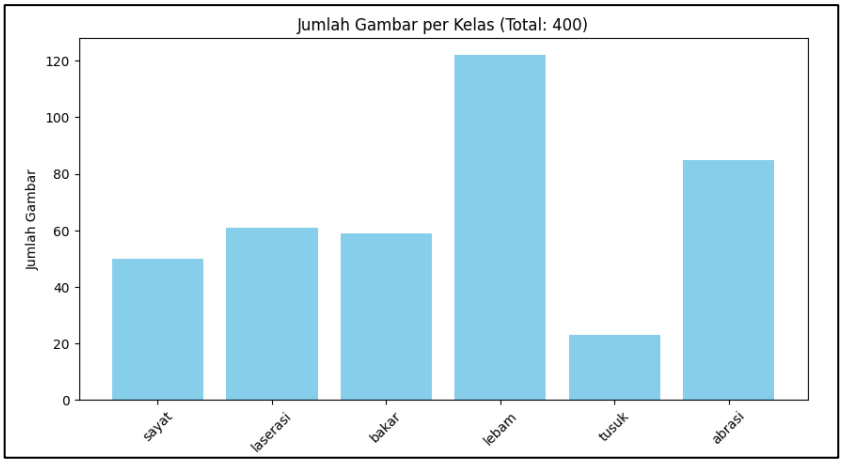
Penelitian ini menggunakan algoritma optimasi Adam (Adaptive Moment Estimation) sebagai *optimizer* dengan learning rate 0,0001. Adam adalah salah satu algoritma optimasi yang populer dalam pengembangan model *deep learning*. Berbagai studi sebelumnya menunjukkan bahwa penggunaan Adam dapat menghasilkan akurasi yang lebih baik dibandingkan *optimizer* lainnya (Bera & Shrivastava, 2020; Irfan et al., 2022; Wikarta et al., 2020). Adam, yang pertama kali diperkenalkan oleh Kingma dan Ba pada tahun 2015, dikenal memiliki keunggulan seperti efisiensi komputasi, kebutuhan memori yang rendah, serta berbagai manfaat lainnya (Kingma & Ba, 2015).

Penelitian ini menerapkan categorical *cross entropy* sebagai fungsi *loss* untuk klasifikasi multi-kelas. Fungsi ini secara khusus digunakan dalam permasalahan klasifikasi dengan lebih dari dua kelas. *Categorical cross entropy* bekerja dengan mengukur selisih antara distribusi probabilitas prediksi model dan distribusi label aktual, lalu berusaha meminimalkan nilai tersebut di mana nilai loss idealnya mendekati nol (Géron, 2017). Pelatihan dilakukan menggunakan 50 epoch dengan batch size 32.

HASIL DAN PEMBAHASAN

Hasil Pengumpulan Data

Berdasarkan skenario yang dijelaskan pada sub-bab 3.2.1, gambar-gambar diperoleh dari website Kaggle.com. Hasil yang telah didapatkan dapat dilihat pada Gambar 3.1.



Gambar 3. 1 Jumlah Dataset Asli

Pada Gambar 3.1 diperlihatkan bahwa dataset ini, hasil dari pengumpulan data, memiliki total 400 gambar yang terdistribusi ke dalam enam (6) kelas luka yang berbeda,yaitu abrasi (85 gambar), bakar (59 gambar), laserasi (60 gambar), lebam (122 gambar), sayat (50 gambar), dan tusuk (23 gambar).

Gambar 3. 2 Contoh Gambar dari Hasil Pengumpulan Data

Gambar 3.2 merupakan contoh gambar yang diperoleh dari *Kaggle.com*, merepresentasikan dataset yang digunakan dalam penelitian ini.

Hasil Pra-proses Data

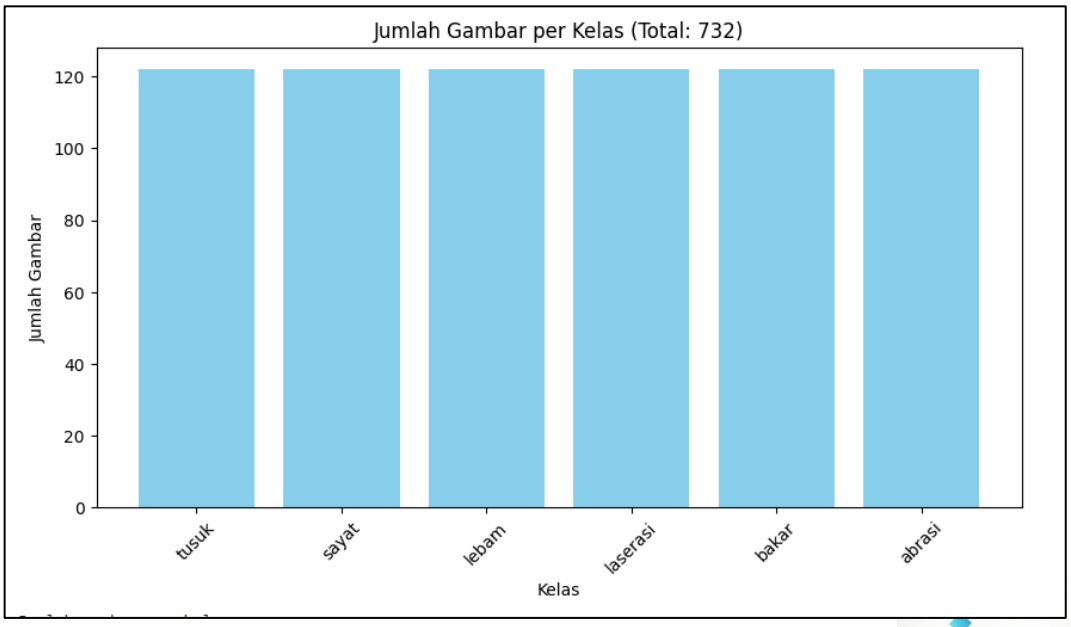
1. Hasil *Resize*

Gambar asli yang telah didapat sebelumnya dari ukuran piksel asli yaitu 640 × 640 piksel menjadi ukuran piksel yang sesuai dengan *input_shape* ketiga model, yaitu 224 x 224 seperti pada Gambar 3.3 berikut.

Gambar 3. 3 Hasil *Resize* Gambar : A. Gambar Asli, B. Setelah Resize

2. Hasil *Oversampling*

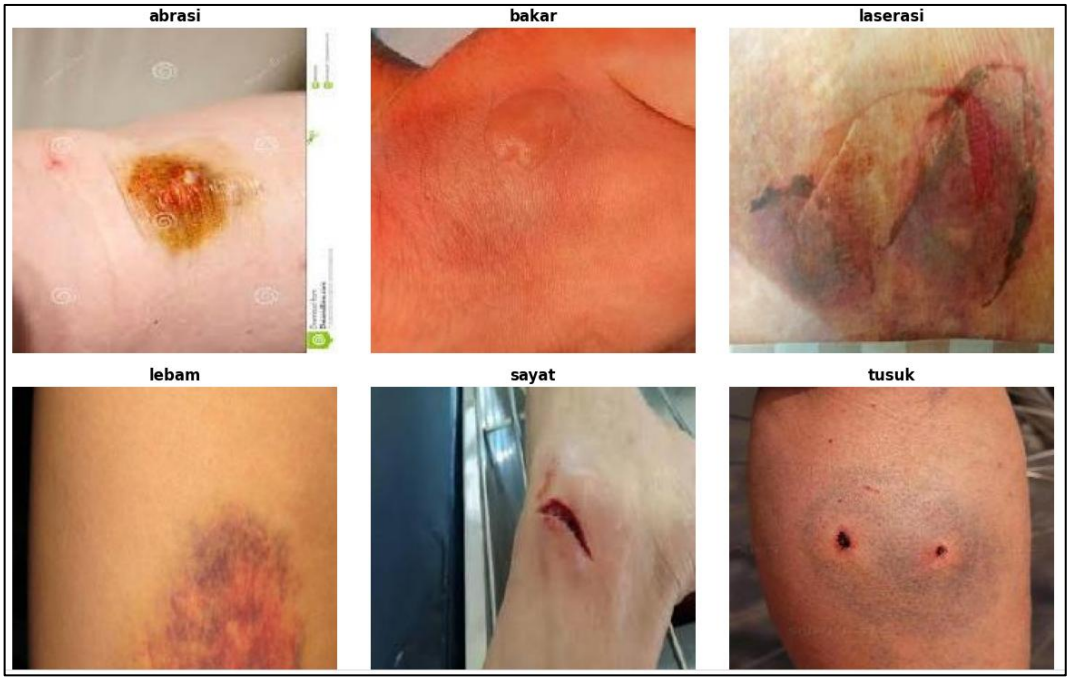
Hasil dari proses *oversampling* menunjukkan bahwa kelas dengan jumlah data tertinggi, yaitu 'lebam' dengan 122 gambar, dijadikan sebagai acuan. Selanjutnya, teknik *oversampling* diaplikasikan untuk meratakan jumlah sampel pada semua kelas hingga mencapai 122 gambar. Hal ini memastikan setiap kelas memiliki representasi data yang setara sebelum proses pelatihan lebih lanjut. Berikut distribusi data hasil penerapan *oversampling* pada Gambar 3.4.



Gambar 3. 4 Hasil *Oversampling*

Gambar 3.4 menunjukkan bahwa seluruh data telah terdistribusi secara merata di setiap kelas, dengan total 732 gambar.

Pada Gambar 3.5, dapat dilihat contoh hasil gambar yang telah di-*oversampling*. Gambar-gambar ini merupakan representasi visual dari data tambahan yang dihasilkan untuk menyeimbangkan dataset.



Gambar 3. 5 Contoh Hasil *Oversampling*

3. Hasil Pembagian Data

Hasil pembagian data dapat dilihat pada Tabel 3.1, di mana dataset dibagi menjadi 70% untuk data pelatihan, 15% untuk data validasi, dan 15% untuk data uji. Masing-masing bagian data ini kemudian disimpan dalam direktori yang berbeda sesuai peruntukannya.

Tabel 3. 1 Hasil Pembagian Data

Dataset	Jumlah
Latih	510
Validasi	108
Uji	114

Secara rinci, data pelatihan terdiri dari 510 gambar, data validasi sebanyak 108 gambar, dan data uji mencakup 114 gambar, kemudian data uji untuk Grad-CAM dipinjam dari data uji sebanyak 48 gambar.

4. Hasil Augmentasi pada *Dataset* Latih

Hasil dari augmentasi data latih menggunakan delapan teknik yang diacak oleh *ImageDataGenerator* adalah terbentuknya lima gambar baru dari setiap satu gambar asli, dengan setiap gambar hasil augmentasi tidak sama persis. Berikut adalah contoh hasil augmentasi pada tiap kelas yang di lakukan :

Gambar 3. 6 Hasil Augmentasi

Setelah proses augmentasi, jumlah data pada *dataset* latih menjadi 2550 gambar, seperti yang bisa dilihat pada Tabel 3.2.

Tabel 3. 2 Tabel Hasil Augmentasi *Dataset* Latih

Label	Sebelum Augmentasi	Sesudah Augmentasi
Abrasi	85	425
Sayat	85	425
Laserasi	85	425
Lebam	85	425
Tusuk	85	425
Bakar	85	425
Total	510	**2550**

Hasil Pembangunan Model

Berikut merupakan hasi implementasi kode program berbasis model arsitektur ResNet50V2, DenseNet201, MobileNetV2 dengan menggunakan pendekatan *transfer learning* yang bisa dilihat pada Gambar 3.7.

```
base_model = ResNet50V2(
    weights='imagenet',
    include_top=False,
    input_shape=(224, 224, 3)
)
base_model.trainable = False  # Freeze base layers

x = GlobalAveragePooling2D()(base_model.output)
x = Dropout(0.5)(x)  # Dropout layer untuk reduce overfitting
predictions = Dense(num_classes, activation='softmax')(x)

model = Model(inputs=base_model.input, outputs=predictions)
```

ResNet50V2

```
base_model = MobileNetV2(
    weights='imagenet',
    include_top=False,
    input_shape=(224, 224, 3)
)
base_model.trainable = False  # Freeze base layers

x = GlobalAveragePooling2D()(base_model.output)
x = Dropout(0.5)(x)  # Dropout layer untuk reduce overfitting
predictions = Dense(num_classes, activation='softmax')(x)

model = Model(inputs=base_model.input, outputs=predictions)
```

Densenet201

```
base_model = DenseNet201(
    weights='imagenet',
    include_top=False,
    input_shape=(224, 224, 3)
)
base_model.trainable = False  # Freeze base layers

x = GlobalAveragePooling2D()(base_model.output)
x = Dropout(0.5)(x)  # Dropout layer untuk reduce overfitting
predictions = Dense(num_classes, activation='softmax')(x)

model = Model(inputs=base_model.input, outputs=predictions)
```

MobileNetV2

Gambar 3. 7 Kode Program Pembangunan Model

Gambar 3.7 menggunakan arsitektur model yang dibangun di atas *feature extractor* ResNet50V2, DenseNet201, MobileNetV2 yang telah dilatih pada dataset ImageNet.

```
Total params: 23,577,094 (89.94 MB)
Trainable params: 12,294 (48.02 KB)
Non-trainable params: 23,564,800 (89.89 MB)
```

ResNet50V2

```
Total params: 18,333,510 (69.94 MB)
Trainable params: 11,526 (45.02 KB)
Non-trainable params: 18,321,984 (69.89 MB)
```

Densenet201

```
Total params: 2,265,670 (8.64 MB)
Trainable params: 7,686 (30.02 KB)
Non-trainable params: 2,257,984 (8.61 MB)
```

MobileNetV2

Gambar 3. 8 Jumlah Parameter dari Hasil Latih Model

Pada Gambar 3.8 bisa dilihat ResNet50V2 memiliki total 23.577.094 parameter, diikuti oleh Densenet201 dengan 18.333.510 parameter, dan MobileNetV2 yang jauh lebih efisien dengan 2.265.670 parameter. Total lama pelatihan dari masing-masing model dapat dilihat pada Tabel 3.3.

Tabel 3. 3 Tabel Lama Pelatihan Model

Pre-trained Model	**Lama Pelatihan**	**Terbilang**
ResNet50V2	`Duration: 3:04:07.159984`	3 Jam 4 Menit 7 Detik
DenseNet201	`Duration: 4:49:07.077378`	4 Jam 49 Menit 7 Detik
MobileNetV2	`Duration: 0:35:21.323556`	35 Menit 21 Detik

Data menunjukkan MobileNetV2 paling efisien, hanya butuh 35 Menit 21 Detik pelatihan. Bandingkan dengan ResNet50V2 (3 Jam 4 Menit 7 Detik) dan DenseNet201 (4 Jam 49 Menit 7 Detik).

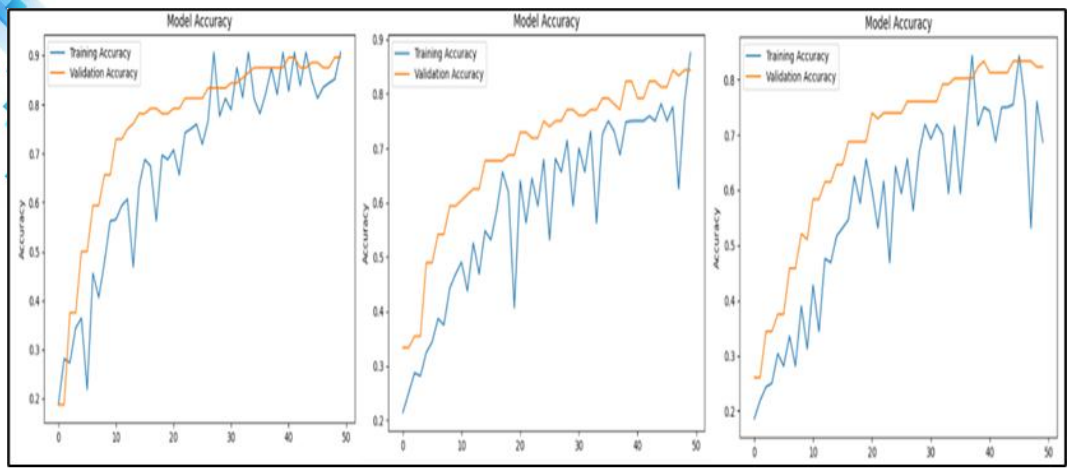
Evaluasi

Evaluasi Performa Model

1. **Akurasi dan *Loss* Pelatihan**

a. Akurasi

Berdasarkan analisis grafik pada Gambar 3.9, ketiga arsitektur model berhasil menunjukkan proses pembelajaran yang efektif, namun dengan tingkat kinerja yang sedikit berbeda. ResNet50V2 (kiri) tampil sebagai model dengan performa terbaik, yang akurasi validasinya mencapai puncak mendekati 90%.



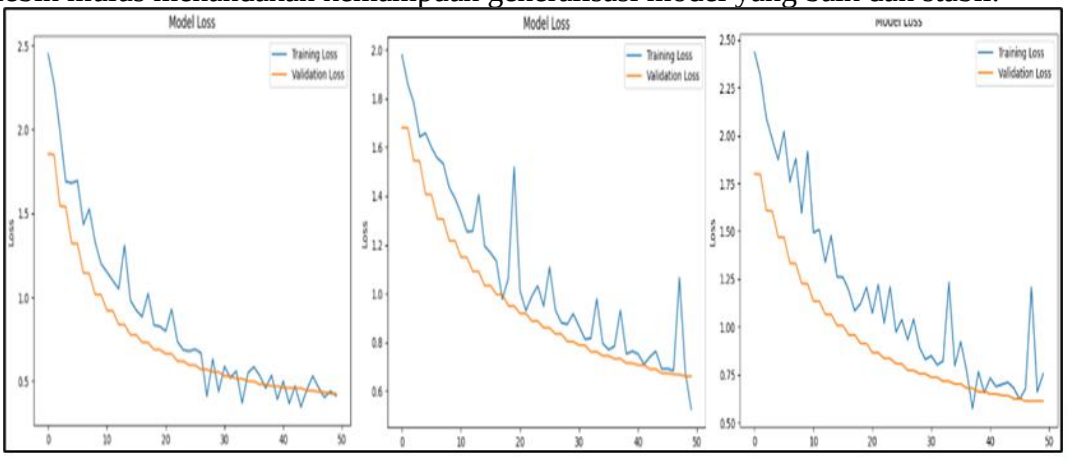
a b c

Gambar 3. 9 Perbandingan Visualisasi Hasil Pelatihan (akurasi) : a. ResNet50V2, b. DenseNet201, c. MobileNetV2

Sementara itu, DenseNet201 (tengah) dan MobileNetV2 (kanan) juga menunjukkan kinerja yang sangat baik dan sebanding, dengan akurasi validasi yang sama-sama stabil di sekitar 85%. Pola yang paling menonjol pada ketiga model adalah kurva akurasi validasi yang secara konsisten lebih tinggi daripada akurasi pelatihan. Hal ini mengindikasikan bahwa teknik regularisasi yang diterapkan sangat efektif mencegah overfitting dan membantu model melakukan generalisasi dengan baik, meskipun proses pelatihan itu sendiri menunjukkan fluktuasi yang cukup tinggi.

b. *Loss*

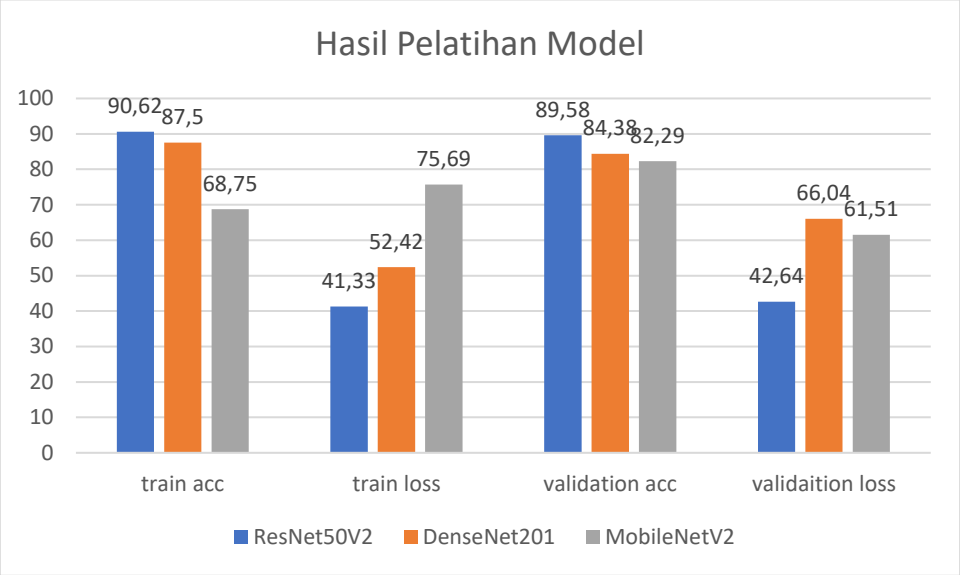
Analisis pada grafik loss ini selaras dengan temuan dari grafik akurasi sebelumnya. Ketiga model, yaitu ResNet50V2, DenseNet201, dan MobileNetV2 menunjukkan proses pembelajaran yang sukses, ditandai dengan tren penurunan kurva loss yang konsisten baik pada data pelatihan maupun validasi. ResNet50V2 kembali menunjukkan kinerja paling unggul dengan mencapai nilai *validation loss* (tingkat kesalahan) terendah di antara ketiganya, yaitu di bawah 0.5. Fenomena yang paling menonjol dan terjadi pada ketiga model adalah nilai *validation loss* yang secara konsisten lebih rendah daripada training loss (biru). Hal ini sekali lagi mengonfirmasi bahwa tidak terjadi *overfitting* dan teknik regularisasi yang digunakan sangat efektif. Meskipun kurva training loss tampak tidak stabil, kurva *validation loss* yang jauh lebih mulus menandakan kemampuan generalisasi model yang baik dan stabil.



a b c

Gambar 3. 10 Perbandingan Visualisasi Hasil Pelatihan (*loss*) : a. ResNet50V2, b. DenseNet201, c. MobileNetV2

Berdasarakan hasil pelatihan ketiga model yang dibangun , berikut adalah hasil dari pelatihan ke tiga model yang bisa dilihat pada Gambar 3.11.



Gambar 3. 11 Hasil Pelatihan Model

Dari hasil ini, ResNet50V2 menunjukkan kinerja paling unggul. Model ini memiliki akurasi paling tinggi baik saat latihan (90.62%) maupun saat diuji dengan data baru (89.58%). Angka "*loss*" atau kesalahannya juga paling rendah di kedua fase. Ini berarti ResNet50V2 tidak hanya pintar menghafal data latihan, tetapi juga sangat baik dalam mengenali pola pada data baru. Ini adalah tanda model yang bagus dan tidak mengalami *overfitting* (yaitu, tidak terlalu fokus menghafal data latihan sehingga tidak bisa mengenali data baru).

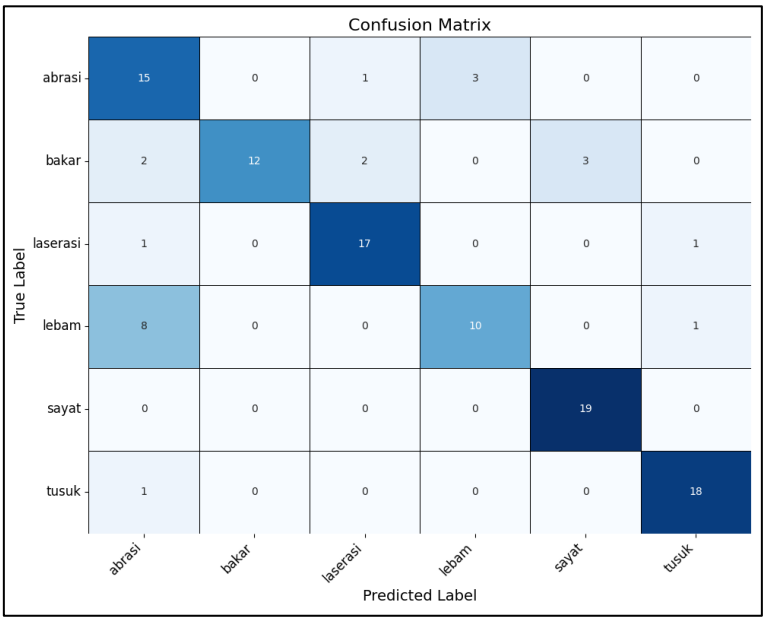
DenseNet201 berada di posisi kedua. Akurasi latihannya cukup baik (87.5%) dan akurasi validasinya juga lumayan (84.38%). Namun, ada sedikit penurunan akurasi dari latihan ke validasi, dan nilai "*loss*" -nya lebih tinggi dibandingkan ResNet50V2. Ini bisa mengindikasikan sedikit *overfitting*, di mana model ini sedikit menghafal data latihan sehingga performanya sedikit menurun saat dihadapkan pada data baru. Meskipun demikian, DenseNet201 masih menunjukkan kinerja yang solid.

Dari hasil pelatihan model MobileNetV2. Model ini memiliki akurasi latihan yang paling rendah (68.75%) dan "*loss*" latihan yang paling tinggi (75.69%). Ini seolah-olah MobileNetV2 tidak begitu pintar saat belajar di awal. Namun, saat diuji dengan data validasi, akurasinya meningkat pesat menjadi 82.29%, dan "loss" validasinya juga jauh lebih rendah (61.51%) dibandingkan saat latihan, bahkan lebih baik dari DenseNet201 dalam hal "loss" validasi. Fenomena ini menunjukkan bahwa MobileNetV2 kemungkinan mengalami *underfitting* saat latihan (tidak cukup belajar dari data latihan secara mendalam).

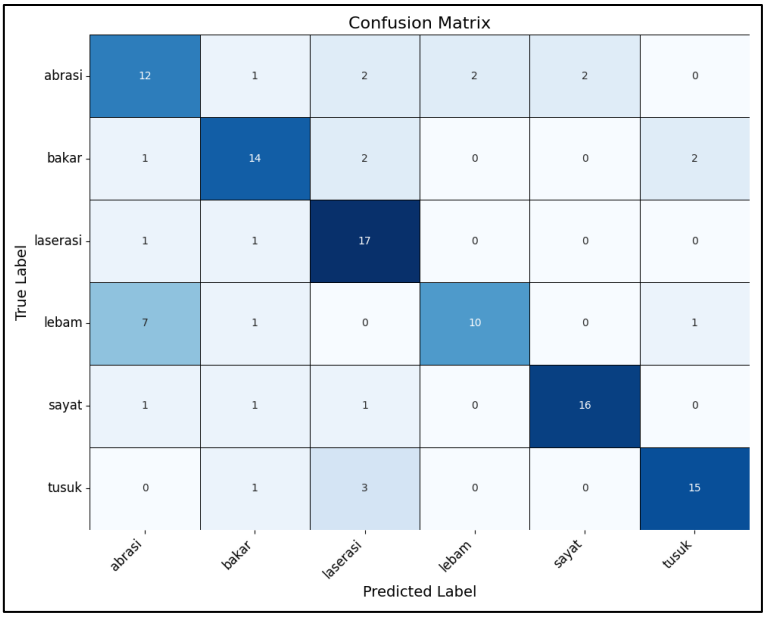
2. *Confussion Matrix* dan *Classifcation Report*

Setelah dilatih, model akan diuji dengan data uji. Hasil pengujian berupa *confusion matrix* ini dapat dilihat pada Gambar 3.12 berikut..

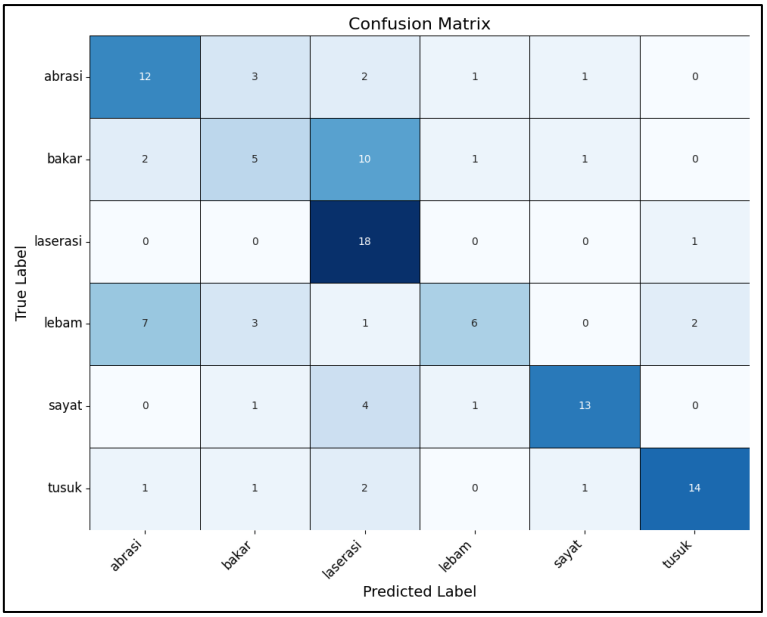
a. Confussion Matrix



A



B



C

Gambar 3. 12 Perbandingan *Heatmap Confussion Matrix* pada DataValidasi: A. ResNet50V2, B. DenseNet201, C. MobileNetv2

Berdasarkan *confussion matrix* pada Gambar 3.12 diatas maka didapatkanlah masing-masing nilai TP, TN, FN, FP dsari model ResNet50V2, DensenNet201, dan MobileNetV2 yang nanti akan menjadi faktor untuk perhitungan performa dari masing-masing model, berikut adalah nilai TP, TN, FN, FP dapat dilihat secara berturut-turut dilihat pada Tabel 3. 4 model ResNet50V2, Tabel 3.5 model DenseNet201, dan 4.6 model MobileNetV2, berikut adalah nilai TP, TN, FN, FP dari setiap model.

Tabel 3. 4 Model ResNet50V2

Kelas	TP	FN	FP	TN
abrasi	15	4	12	83
bakar	12	7	0	95
laserasi	17	2	3	92
lebam	10	9	3	92
sayat	19	0	3	92
tusuk	18	1	2	93

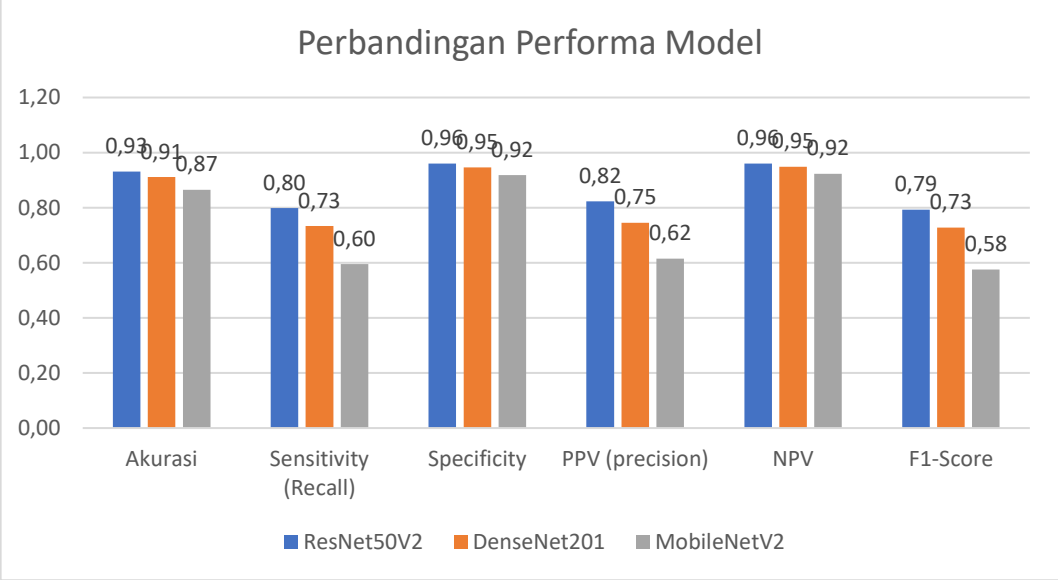
Tabel 3. 5 Model DenseNet201

Kelas	TP	FN	FP	TN
abrasi	12	7	10	85
bakar	14	5	5	90
Kelas	**TP**	**FN**	**FP**	**TN**
laserasi	17	2	8	87
lebam	10	9	2	93
sayat	16	3	2	93
tusuk	15	4	3	92

Tabel 3. 6 Model MobileNetv2

Kelas	TP	FN	FP	TN
abrasi	12	7	10	85
bakar	5	14	8	87
laserasi	18	1	19	76
lebam	6	13	3	92
sayat	13	6	3	92
tusuk	14	5	3	92

Berdasarkan Tabel 3.4, Tabel 3.5 dan 4.6 didapatkan bahwa model ResNet50V2 yang terbaik dari segi akurasi, Sensitivitas (*recall*), Spesifisitas, PPV (presisi), NPV, dan juga *F1-score*. Sehingga performa model dapat dilihat pada Gambar 3.13 dibawaah ini.



Gambar 3. 13 Perbandingan Perfoma Model

Dilihat dari Gambar 3.13, MobileNetV2 menunjukkan nilai terendah pada akurasi, recall, spesifisitas, PPV (presisi), NPV, dan juga F1-score. Berdasarkan hal ini, dapat disimpulkan bahwa urutan model terbaik adalah ResNet50V2 di posisi pertama, diikuti oleh DenseNet201 di posisi kedua, dan MobileNetV2 di posisi terakhir.

Selanjutnya, model ResNet50V2 akan dianalisis lebih mendalam untuk mengevaluasi performa antarkelas yang diprediksi, berdasarkan Tabel 3.7 yang menampilkan hasil pengujian terhadap 114 data uji.

Tabel 3. 7 Hasil Pengujian ResNet50V2

No	Gambar	Fakta	Sistem	Hasil
1	abrasi_tes[1]	abrasi	abrasi	TP
2	abrasi_tes[2]	abrasi	abrasi	TP
3	abrasi_tes[3]	abrasi	lebam	FN
4	abrasi_tes[4]	abrasi	abrasi	TP
5	abrasi_tes[5]	abrasi	abrasi	TP
6	abrasi_tes[6]	abrasi	abrasi	TP
7	abrasi_tes[7]	abrasi	abrasi	TP

8	abrasi_tes[8]	abrasi	lebam	FN
9	abrasi_tes[9]	abrasi	abrasi	TP
10	abrasi_tes[10]	abrasi	abrasi	TP
11	abrasi_tes[11]	abrasi	abrasi	TP
12	abrasi_tes[12]	abrasi	lebam	FN
13	abrasi_tes[13]	abrasi	abrasi	TP
14	abrasi_tes[14]	abrasi	abrasi	TP
15	abrasi_tes[15]	abrasi	abrasi	TP
16	abrasi_tes[16]	abrasi	abrasi	TP
17	abrasi_tes[17]	abrasi	laserasi	FN
18	abrasi_tes[18]	abrasi	abrasi	TP
19	abrasi_tes[19]	abrasi	abrasi	TP
20	bakar_tes[1]	bakar	bakar	TP
21	bakar_tes[2]	bakar	sayat	FN
No	**Gambar**	**Fakta**	**Sistem**	**Hasil**
22	bakar_tes[3]	bakar	bakar	TP
23	bakar_tes[4]	bakar	bakar	TP
24	bakar_tes[5]	bakar	bakar	TP
25	bakar_tes[6]	bakar	bakar	TP
26	bakar_tes[7]	bakar	bakar	TP
27	bakar_tes[8]	bakar	abrasi	FN
28	bakar_tes[9]	bakar	bakar	TP
29	bakar_tes[10]	bakar	bakar	TP
30	bakar_tes[11]	bakar	laserasi	FN
31	bakar_tes[12]	bakar	bakar	TP
32	bakar_tes[13]	bakar	bakar	TP
33	bakar_tes[14]	bakar	bakar	TP
34	bakar_tes[15]	bakar	sayat	FN
35	bakar_tes[16]	bakar	sayat	FN
36	bakar_tes[17]	bakar	laserasi	FN
37	bakar_tes[18]	bakar	abrasi	FN
38	bakar_tes[19]	bakar	bakar	TP
39	laserasi_tes[1]	laserasi	laserasi	TP
40	laserasi_tes[2]	laserasi	laserasi	TP
41	laserasi_tes[3]	laserasi	laserasi	TP
42	laserasi_tes[4]	laserasi	laserasi	TP
43	laserasi_tes[5]	laserasi	laserasi	TP
44	laserasi_tes[6]	laserasi	laserasi	TP
No	**Gambar**	**Fakta**	**Sistem**	**Hasil**
45	laserasi_tes[7]	laserasi	laserasi	TP
46	laserasi_tes[8]	laserasi	laserasi	TP
47	laserasi_tes[9]	laserasi	laserasi	TP
48	laserasi_tes[10]	laserasi	laserasi	TP

49	laserasi_tes[11]	laserasi	laserasi	TP
50	laserasi_tes[12]	laserasi	laserasi	TP
51	laserasi_tes[13]	laserasi	laserasi	TP
52	laserasi_tes[14]	laserasi	abrasi	FN
53	laserasi_tes[15]	laserasi	laserasi	TP
54	laserasi_tes[16]	laserasi	laserasi	TP
55	laserasi_tes[17]	laserasi	laserasi	TP
56	laserasi_tes[18]	laserasi	tusuk	FN
57	laserasi_tes[19]	laserasi	laserasi	TP
58	lebam_tes[1]	lebam	abrasi	FN
59	lebam_tes[2]	lebam	lebam	TP
60	lebam_tes[3]	lebam	lebam	TP
61	lebam_tes[4]	lebam	abrasi	FN
62	lebam_tes[5]	lebam	lebam	TP
63	lebam_tes[6]	lebam	lebam	TP
64	lebam_tes[7]	lebam	abrasi	FN
65	lebam_tes[8]	lebam	lebam	TP
66	lebam_tes[9]	lebam	abrasi	FN
67	lebam_tes[10]	lebam	abrasi	FN
No	**Gambar**	**Fakta**	**Sistem**	**Hasil**
68	lebam_tes[11]	lebam	tusuk	FN
69	lebam_tes[12]	lebam	abrasi	FN
70	lebam_tes[13]	lebam	lebam	TP
71	lebam_tes[14]	lebam	lebam	TP
72	lebam_tes[15]	lebam	lebam	TP
73	lebam_tes[16]	lebam	lebam	TP
74	lebam_tes[17]	lebam	abrasi	FN
75	lebam_tes[18]	lebam	abrasi	FN
76	lebam_tes[19]	lebam	abrasi	FN
77	sayat_tes[1]	sayat	sayat	TP
78	sayat_tes[2]	sayat	sayat	TP
79	sayat_tes[3]	sayat	sayat	TP
80	sayat_tes[4]	sayat	sayat	TP
81	sayat_tes[5]	sayat	sayat	TP
82	sayat_tes[6]	sayat	sayat	TP
83	sayat_tes[7]	sayat	sayat	TP
84	sayat_tes[8]	sayat	sayat	TP
85	sayat_tes[9]	sayat	sayat	TP
86	sayat_tes[10]	sayat	sayat	TP
87	sayat_tes[11]	sayat	sayat	TP
88	sayat_tes[12]	sayat	sayat	TP
89	sayat_tes[13]	sayat	sayat	TP
90	sayat_tes[14]	sayat	sayat	TP

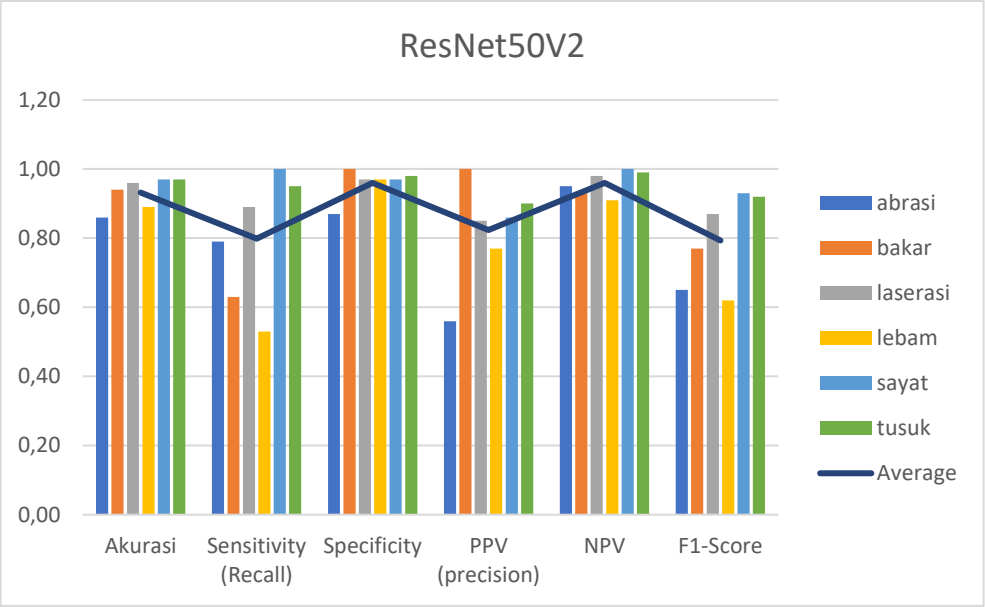
No	Gambar	Fakta	Sistem	Hasil
91	sayat_tes[15]	sayat	sayat	TP
92	sayat_tes[16]	sayat	sayat	TP
93	sayat_tes[17]	sayat	sayat	TP
94	sayat_tes[18]	sayat	sayat	TP
95	sayat_tes[19]	sayat	sayat	TP
96	tusuk_tes[1]	tusuk	tusuk	TP
97	tusuk_tes[2]	tusuk	tusuk	TP
98	tusuk_tes[3]	tusuk	tusuk	TP
99	tusuk_tes[4]	tusuk	tusuk	TP
100	tusuk_tes[5]	tusuk	abrasi	FN
101	tusuk_tes[6]	tusuk	tusuk	TP
102	tusuk_tes[7]	tusuk	tusuk	TP
103	tusuk_tes[8]	tusuk	tusuk	TP
104	tusuk_tes[9]	tusuk	tusuk	TP
105	tusuk_tes[10]	tusuk	tusuk	TP
106	tusuk_tes[11]	tusuk	tusuk	TP
107	tusuk_tes[12]	tusuk	tusuk	TP
108	tusuk_tes[13]	tusuk	tusuk	TP
109	tusuk_tes[14]	tusuk	tusuk	TP
110	tusuk_tes[15]	tusuk	tusuk	TP
111	tusuk_tes[16]	tusuk	tusuk	TP
112	tusuk_tes[17]	tusuk	tusuk	TP
113	tusuk_tes[18]	tusuk	tusuk	TP
No	**Gambar**	**Fakta**	**Sistem**	**Hasil**
114	tusuk_tes[19]	tusuk	tusuk	TP

Tabel 3. 8 Ringkasan *Classification Report* ResNet50V2

ResNet50V2										
Kelas	**TP**	**FN**	**FP**	**TN**	**Akurasi**	**Sensitivity (Recall)**	**Specificity**	**PPV (precision)**	**NPV**	**F1-Score**
abrasi	15	4	12	83	0,86	0,79	0,87	0,56	0,95	0,65
bakar	12	7	0	95	0,94	0,63	1,00	1,00	0,93	0,77
laserasi	17	2	3	92	0,96	0,89	0,97	0,85	0,98	0,87
lebam	10	9	3	92	0,89	0,53	0,97	0,77	0,91	0,62
sayat	19	0	3	92	0,97	1,00	0,97	0,86	1,00	0,93

tusuk	18	1	2	93	0,97	0,95	0,98	0,90	0,99	0,92
Rata-rata					**0,93**	**0,80**	**0,96**	**0,82**	**0,96**	**0,79**

Data dari Tabel 3.8 di atas dapat divisualisasikan dalam bentuk grafik seperti yang ditunjukkan pada Gambar 3.14 dibawah ini.

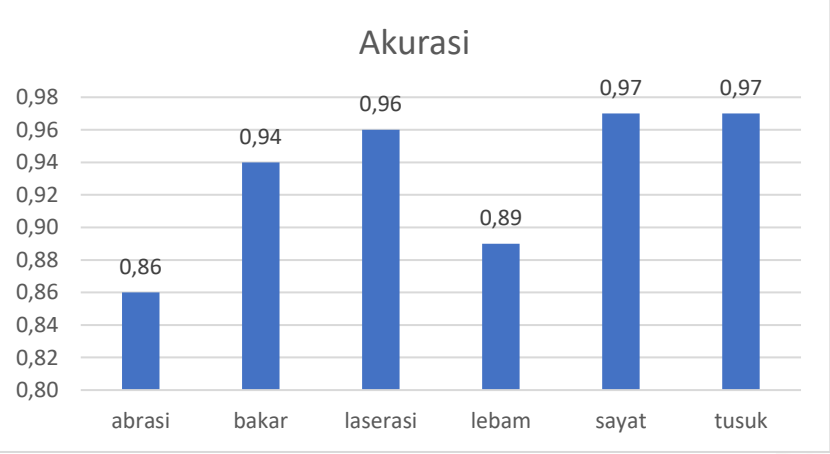


Gambar 3. 14 Grafik Hasil Perhitungan *Classification Report* ResNet50V2

Seperti yang ditunjukkan pada Gambar 3.14, disajikan hasil performa ResNet50V2 untuk masing-masing kelas, meliputi abrasi, bakar, laserasi, lebam, sayat, dan tusuk. Berikut adalah penjelasan lebih lanjut mengenai parameter akurasi, recall, spesifisitas, PPV, NPV, dan F1-score dari hasil tersebut.

a. Akurasi

Akurasi menghitung proporsi prediksi benar model secara keseluruhan, berikut akurasi dari setiap kelas yang bisa dilihat pada Gambar 3.15.

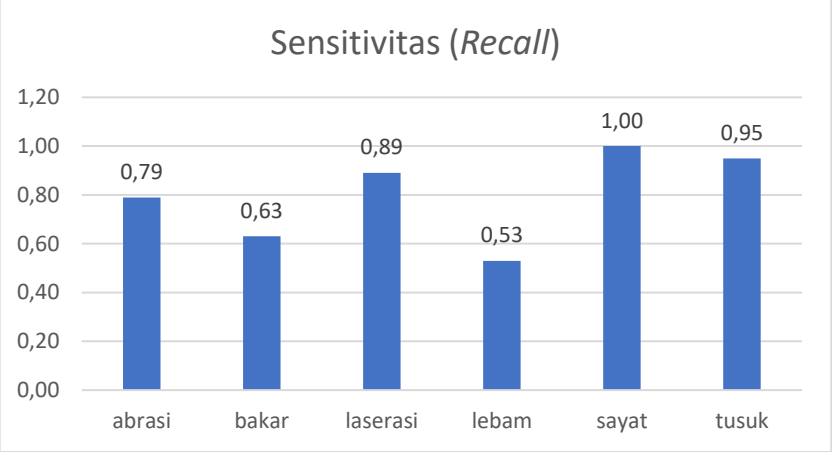


Gambar 3. 15 Akurasi pada Setiap Kelas Model ResNet50V2

Seperti yang ditunjukkan pada Gambar 3.15 bahwa kategori Sayat dan Tusuk menunjukkan akurasi tertinggi, yakni 0,97, mengindikasikan kemampuan model yang sangat baik dalam mengklasifikasikan kedua jenis luka tersebut. Meskipun masih dalam batas yang baik, kategori Abrasi tercatat memiliki akurasi terendah pada angka 0,86.

b. Sensitivitas (*recall*)

Sensitivitas menghitung proporsi kasus positif (true positive) yang berhasil diprediksi benar oleh model dari seluruh kasus positif aktual. Berikut sensitivitas dari setiap kelas yang bisa dilihat pada Gambar 3.16.

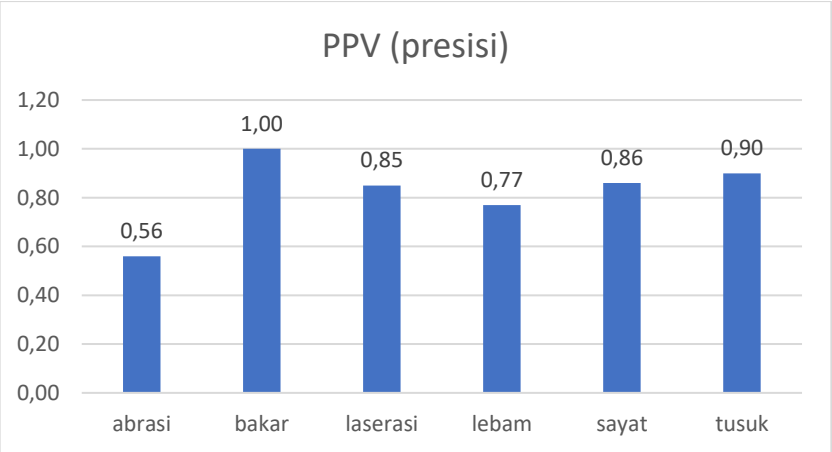


Gambar 3. 16 Sensitivitas pada Setiap Kelas Model ResNet50V2

Data sensitivitas (*recall*) yang disajikan pada Gambar 3.16, terlihat bahwa nilai sensitivitas bervariasi antar kelas. Kategori Sayat menunjukkan nilai sensitivitas tertinggi yaitu 1,00, yang berarti model mampu mengidentifikasi semua kasus "sayat" dengan benar. Sebaliknya, kategori Lebam memiliki sensitivitas terendah dengan nilai 0,53, menunjukkan bahwa model masih sering gagal mengidentifikasi kasus "lebam" yang sebenarnya positif.

c. PPV (Presisi)

PPV (Positive Predictive Value), atau presisi, mengukur proporsi hasil prediksi positif yang benar-benar positif. Ini menunjukkan seberapa dapat diandalkan model ketika memprediksi suatu kasus sebagai positif. Berikut sensitivitas dari setiap kelas yang bisa dilihat pada Gambar 3.17.

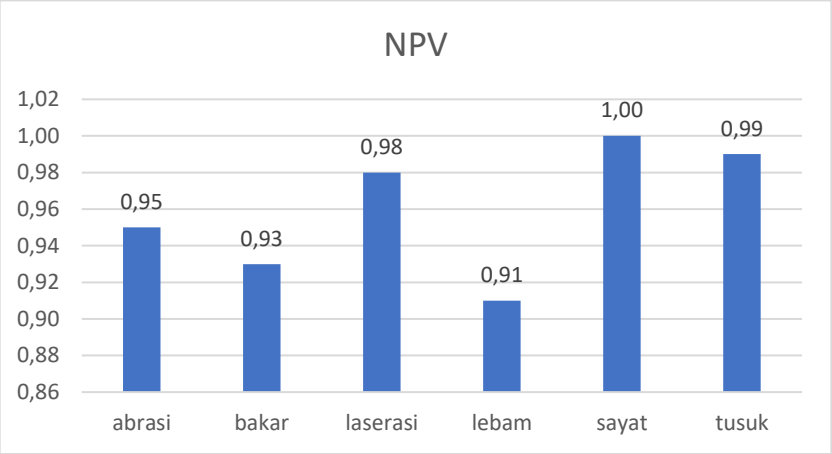


Gambar 3. 17 PPV (presisi) pada Setiap Kelas Model ResNet50V2

Data Positive Predictive Value (PPV) atau presisi pada Gambar 3.17 menunjukkan variasi nilai yang signifikan antar kelas. Kelas Bakar mencatat presisi tertinggi dengan nilai sempurna 1,00, mengindikasikan akurasi sempurna dalam prediksi positif untuk kategori ini. Sebaliknya, kelas Abrasi menunjukkan nilai presisi terendah yaitu 0,56, menandakan proporsi prediksi positif yang tidak tepat cukup tinggi pada kategori tersebut.

d. NPV

NPV (Negative Predictive Value) mengukur proporsi hasil prediksi negatif yang benar-benar negatif. Berikut NPV dari setiap kelas yang bisa dilihat pada Gambar 3.18.

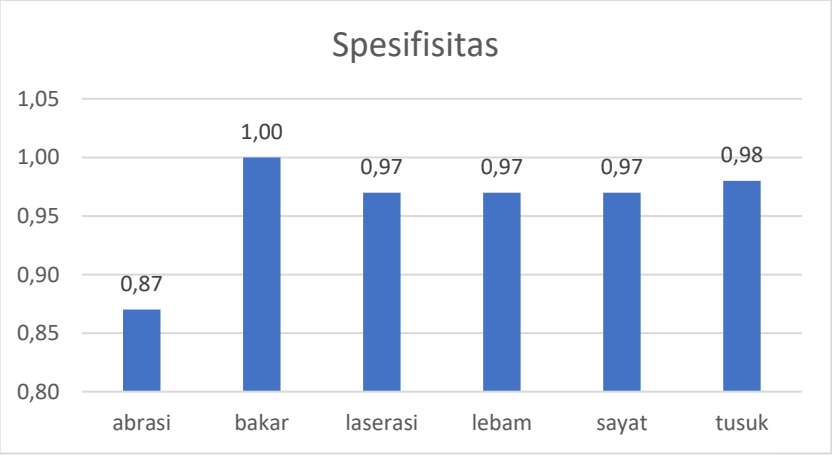


Gambar 3. 18 NPV pada Setiap Kelas Model ResNet50V2

Data Negative Predictive Value (NPV) pada Gambar 3.18 menunjukkan kinerja model dalam mengidentifikasi kasus negatif secara benar. Kelas Sayat mencatat nilai NPV tertinggi sebesar 1,00, mengindikasikan bahwa semua prediksi negatif untuk kelas ini adalah benar. Sementara itu, kelas Lebam memiliki nilai NPV terendah yaitu 0,91, menunjukkan bahwa terdapat sedikit potensi kesalahan dalam prediksi negatif pada kategori tersebut dibandingkan kelas lainnya.

e. Spesifisitas

Spesifisitas mengukur proporsi kasus negatif yang diprediksi dengan benar oleh model. Ini menunjukkan kemampuan model untuk secara tepat mengidentifikasi ketika suatu kasus bukan positif (atau merupakan negatif sejati). Berikut spesifisitas dari setiap kelas yang bisa dilihat pada Gambar 3.19.

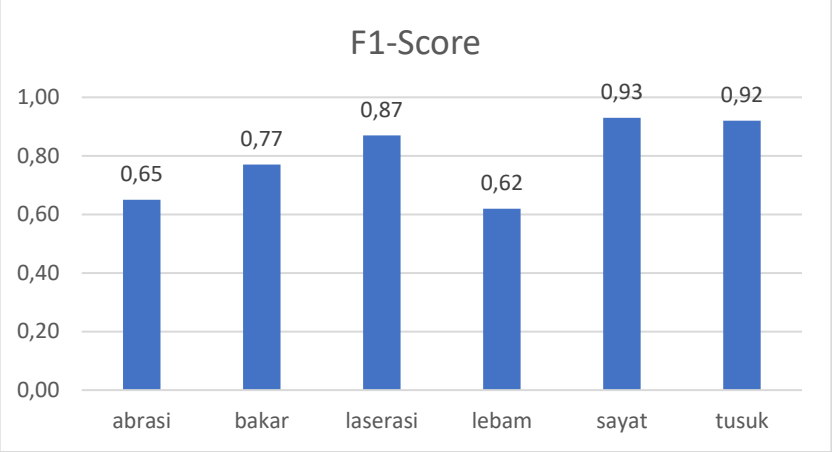


Gambar 3. 19 Spesifisitas pada Setiap Kelas Model ResNet50V2

Data Spesifisitas pada Gambar 3.19 menunjukkan bahwa kelas Bakar mencatat nilai spesifisitas tertinggi yang sempurna, yaitu 1,00, menandakan tidak ada kesalahan dalam memprediksi kasus non-"bakar". Sementara itu, kelas Abrasi memiliki nilai spesifisitas terendah yaitu 0,87, menunjukkan bahwa model memiliki sedikit lebih banyak kesulitan dalam mengidentifikasi kasus non-"abrasi" dibandingkan kategori lainnya.

f. F1-Score

F1-score adalah rata-rata PPV (presisi) dan sensitivitas(recall) . Ini menunjukkan seberapa baik model secara keseluruhan, terutama jika jumlah data per kelas tidak seimbang. Berikut F1-score dari setiap kelas yang bisa dilihat pada Gambar 3.20.



Gambar 3. 20 pada Setiap Kelas Model ResNet50V2

Data F1-Score pada Gambar 3.20 menunjukkan bahwa kelas Sayat mencatat F1-Score tertinggi yaitu 0,93, mengindikasikan kinerja keseluruhan yang paling optimal dalam mengidentifikasi dan memprediksi kelas ini. Sebaliknya, kelas Lebam memiliki F1-Score terendah dengan nilai 0,62, menunjukkan adanya ruang perbaikan yang paling signifikan dalam keseimbangan antara presisi dan recall untuk kategori tersebut.

Dari seluruh hasil analisis metrik Akurasi, Sensitivitas (Recall), Presisi (PPV), NPV, Spesifisitas, dan F1-Score, dapat disimpulkan bahwa kelas Sayat menunjukkan performa keseluruhan yang paling baik. Kelas ini secara konsisten mencatat nilai tinggi atau tertinggi pada berbagai metrik, termasuk Akurasi (0.97), Sensitivitas (1.00), NPV (1.00), dan F1-Score (0.93). Ini mengindikasikan kemampuan model yang sangat kuat dalam mengidentifikasi dan memprediksi kelas "Sayat" secara komprehensif.

Sementara itu, kelas Lebam dan Abrasi menunjukkan performa yang relatif lebih rendah dibanding kelas lainnya. Kelas Lebam memiliki Sensitivitas (0.53) dan F1-Score (0.62) terendah, menunjukkan kesulitan dalam mendeteksi kasus positif "lebam" dan kinerja yang kurang seimbang. Sedangkan kelas Abrasi mencatat Akurasi (0.86), Presisi (0.56), dan Spesifisitas (0.87) terendah, mengindikasikan adanya ruang perbaikan terbesar pada kedua kategori ini. Meski begitu, jika dilihat dari F1-Score sebagai metrik keseimbangan, Lebam sedikit lebih rendah dari Abrasi.

KESIMPULAN & SARAN

5.1 Kesimpulan

Berdasarkan penelitian yang dilakukan, maka dapat disimpulkan bahwa ResNet50V2 adalah model terbaik di antara ResNet50V2, DenseNet201, dan MobileNetV2. Model ini menunjukkan keunggulan signifikan dalam performa metrik, dengan akurasi 0,93, sensitivitas 0,80, spesifisitas 0,96, PPV (presisi) 0,82, NPV 0,96, dan F1-score 0,79. Tidak hanya unggul dalam performa metrik, ResNet50V2 juga memperlihatkan keunggulan pada *analisis explainability*. Melalui *heatmap* Grad-CAM, model ini terbukti fokus pada area luka di 33 dari

48 gambar uji. Hal ini menunjukkan bahwa ResNet50V2 tidak hanya mampu memberikan prediksi yang akurat, tetapi juga memiliki pemahaman visual yang lebih baik dan lebih terfokus pada fitur-fitur relevan (area luka) saat membuat keputusannya.

ResNet50V2 unggul dalam klasifikasi luka luar berkat arsitektur dalamnya dan penggunaan *skip connections*, seperti dijelaskan pada sub-bab 2.2.10 dan 2.2.13. Fitur ini membantu model mengatasi masalah penurunan kinerja pada jaringan yang sangat dalam. Jadi, model bisa mempelajari detail dan pola kompleks dari gambar luka dengan lebih baik. Ini penting banget untuk membedakan luka-luka yang mirip satu sama lain. Perlu diketahui juga, ResNet50V2, DenseNet201, dan MobileNetV2 sama-sama dilatih di *ImageNet*.

Sementara itu, DenseNet201 menempati posisi kedua karena arsitektur *dense connectivity*-nya yang mendorong penggunaan kembali fitur secara efisien dan memastikan aliran informasi yang sangat baik di seluruh jaringan yang diceritakan pada sub-bab 2.2.12 dan sub-bab 2.2.13. DenseNet201 memang bagus dalam memanfaatkan informasi dari data luka. Tapi, dalam penelitian ini, model ini masih sedikit di bawah ResNet50V2 dalam menangkap detail rumit luka. Di sisi lain, MobileNetV2 itu paling efisien dan cocok untuk perangkat kecil karena desain awal nya ditujukan untuk komputasi ringan yang dijelaskan pada sub-bab 2.2.11 dan sub-bab 2.2.13 , tapi karena desainnya yang ringan, model ini mendapat posisi terakhir. MobileNetv2 kurang bisa menangkap fitur-fitur penting dan detail luka yang rumit dibanding dua model yang lebih "berat" seperti ResNet50V2 dan DenseNet201.

Analisis *explainability* secara keseluruhan memberikan wawasan kritis mengenai kemampuan model dalam memfokuskan perhatiannya pada area yang relevan. Meskipun ResNet50V2 menunjukkan kemampuan terbaik, model lain seperti MobileNetV2, dan terutama kelas Lebam pada ResNet50V2, memperlihatkan kelemahan dalam aspek *explainability*. Berdasarkan hasil yang ditemukan pada sub-bab 4.3.2 model cenderung terdistraksi oleh fitur di luar area luka, bahkan hingga menyebabkan kesalahan prediksi pada beberapa kasus. Oleh karena itu, penelitian ini menekankan pentingnya analisis *explainability* untuk tidak hanya menilai seberapa baik model memprediksi, tetapi juga bagaimana model sampai pada keputusannya, terutama dalam aplikasi kritis seperti diagnosis medis.

5.2 Saran

1. Melakukan *fine-tuning* pada setiap model sangat disarankan untuk mendapatkan hasil pelatihan yang optimal. Hal ini karena dalam penelitian ini, proses pelatihan tidak melibatkan *fine-tuning*, seperti yang sudah dijelaskan pada sub-bab 3.2.3 Pembangunan Model, di mana semua parameter telah ditentukan sejak awal.
2. Penambahan jumlah serta keberagaman data sangat direkomendasikan untuk kelas Lebam dan Bakar. Pada sub-bab 4.3.1 ditemukan bahwa kelas Lebam adalah yang paling sering salah diprediksi, mencapai 9 kasus kesalahan, dan sering tertukar dengan kelas Abrasi 8 kasus. Sementara itu, tingkat kesalahan prediksi yang tinggi juga ditunjukkan oleh kelas Bakar, yaitu sebesar kasus, dengan pola pertukaran yang beragam ke kelas Sayat, Abrasi, atau Laserasi.
3. Melakukan validasi kualitatif *heatmap* Grad-CAM bersama ahli medis sangat disarankan. Hal ini bertujuan untuk memastikan interpretasi model selaras dengan indikator klinis yang relevan. Perlu dicatat bahwa dalam penelitian ini, proses analisis kualitatif *heatmap* Grad-CAM dilakukan berdasarkan interpretasi peneliti karena tidak didampingi oleh ahli medis, seperti yang sudah dijelaskan pada sub-bab 4.3.2. Sehingga, rekomendasi ini menjadi krusial untuk penelitian di masa mendatang.
4. Berdasarkan hasil yang ditemukan pada sub-bab 4.3.2 terkait kasus *heatmap* Grad-CAM yang tidak fokus pada area luka yang relevan, disarankan untuk melakukan optimalisasi lebih lanjut pada proses pelatihan model dan/atau kualitas data. Ini dapat mencakup penajaman anotasi area luka pada dataset atau eksplorasi arsitektur model dengan

mekanisme *attention* yang lebih kuat, guna memastikan model belajar memfokuskan perhatiannya secara akurat pada fitur visual luka yang relevan.

DAFTAR RUJUKAN

Alpaydin, E. (2021). Machine learning. MIT press.

American Academy of Dermatology Association. (n.d.)Burns: Diagnosis and treatment. (n.d.). Burns: Diagnosis and treatment. .Retrieved from https://www.aad.org/public/diseases/a-z/burns-treatment.

Arrieta, A. B.-R. (2020). Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion. *Explainable Artificial Intelligence (XAI)*, 58.

Badan Pusat Statistik Indonesia. (2024). Statistik kesehatan 2023. *BPS-Statistics Indonesia*.

Bengio, Y. G., & Courville, A. (2017). Deep learning (Vol. 1). Cambridge, MA, USA: MIT press.

Brunicardi, F. C., Andersen, D. K., Billiar, T. R., Dunn, D. L., Kao, L. S., Hunter, J. G., Matthews, J. B., & Pollock, R. E. (Eds.). (2022). Schwartz's principles of surgery (11th ed.). *McGraw-Hill*.

Cholissodin, I., Sutrisno, S., Soebroto, A. A., Hasanah, U., & Febiola, Y. I. (2020). AI, machine learning and deep learning. Fakultas Ilmu Komputer, Universitas Brawijaya, Malang.

Doshi-Velez, F. &. (2017). Towards a rigorous science of interpretable machine learning. *Towards a rigorous science of interpretable machine learning.*

Fawcett, T. (2017). Understanding Confusion Matrices. Towards Data Science. *Retrieved from https://towardsdatascience.com/understanding-confusion-matrices-ceb4ec2b3420*.

Fitch, M. T., Abrahamian, F. M., Moran, G. J., & Talan, D. A. (2008). Emergency department management of puncture wounds and foreign bodies. *Infectious Disease Clinics of North America*, 22(1), 125–143 https://doi.org/10.1016/j.idc.2007.11.002.

Goodfellow, I. B., & Courville, A. (2016). Deep learning. MIT press.

Guidotti, R. M. (2018). A survey of methods for explaining black box models. *ACM Computing Surveys,*, 51(5), 1-42.

Guo, S., & DiPietro, L. A. (2010). Factors affecting wound healing. *Journal of Dental Research*, 89(3), 219–229.

He, K. Z.–E. (2016). *Identity mappings in deep residual networks.*, (pp. 630-645).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 770–778).

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(4), e1312.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700-4708.

Hulley, S. B., & Cumming, S. R. (1998). Designing clinical research: An epidemiologic approach. *Williams & Wilkins*.

Keras documentation. (n.d.). Keras: The Python Deep Learning API. *Retrieved from https://keras.io/*.

Kumar, Hemant & Virmani, Amit & Tripathi, Shivneet & Agrawal, Rashi & Kumar, Sunil. (2021). Transfer Learning and Supervised Machine Learning Approach for Detection of Skin Cancer: Performance Analysis and Comparison. *Drugs and Cell Therapies in Hematology*, 10. 1845-1860.

Lin, M., Chen, Q., & Yan, S. (2013). Network In Network. *arXiv preprint arXiv:1312.4400*.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A., van Ginneken, B., & Sánchez, C. I. (2017). A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, 42, 60–88.

Mayo Clinic. (n.d.). (n.d.). Bruises. Retrieved from https://www.mayoclinic.org/diseases-conditions/bruise/symptoms-causes/syc-20355721.

MedlinePlus. (n.d.-a). (n.d.). Wounds - first aid. Retrieved from https://medlineplus.gov/ency/article/000044.htm.

MedlinePlus. (n.d.-b). (n.d.). Abrasion. Retrieved from https://medlineplus.gov/ency/article/000044.html.

Murphy, K. P. (2012). Machine learning: a probabilistic perspective. MIT press.

National Library of Medicine. (n.d.). (n.d.). Wounds: Types and treatment. Retrieved from https://www.nlm.nih.gov/medlineplus/wounds.html.

R., Kishor & Kumar, R. S. (2023). CoC-ResNet - classification of colorectal cancer on histopathologic images using residual networks. *Multimedia Tools and Applications*, 83. 1-25. 10.1007/s11042-023-17740-5.

RAMADHANTI, Z. (2024). DETEKSI DAN IDENTIFIKASI JENIS LUKA LUAR BERDASARKAN IMAGE FEATURE MENGGUNAKAN CNN DENGAN VARIASI PRE-TRAINED MODEL (Doctoral dissertation, UNIVERSITAS JAMBI).

Ribeiro, M. T. (2016). Why Should I Trust You?" Explaining the Predictions of Any Classifier.

Rohmah, A. A.-1.-6. (2023). Smart city achievement through implementation of digital health services in handling COVID-19 Indonesia. Smart Cities, .

S, I. O. (2021). Teknik Resampling untuk Data Tidak Seimbang, https://ivanongko.medium.com/teknik-resampling-untuk-data-tidak-seimbang-5d566661101c.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. *In Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 4510–4520).

Seidaliyeva, Ulzhalgas & Akhmetov, Daryn & Ilipbayeva, Lyazzat & Matson, Eric. (2020). Real-Time and Accurate Drone Detection in a Video with a Static Background. *Sensors*, 20. 3856. 10.3390/s20143856.

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. ,. *In Proceedings of the IEEE international conference on computer vision*, (pp. 618–626).

Shapiro, L. C.-L.-1.-5. (2022). Efficacy of booster doses in augmenting waning immune responses to COVID-19 vaccine in patients with cancer. *Cancer Cell,*, 40(1), 3-5.

Sharma, P. (2023). https://www.analyticsvidhya.com/blog/2022/03/basic-introduction-to-convolutional-neural-network-in-deep-learning/.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 1 15(1), 1929-1958.

Tarek, M., Ghali, K., & Badawi, A. (2024). Grad-CAM Enabled Breast Cancer Classification with a 3D Inception-ResNet V2: Empowering Radiologists with Explainable Insights. *Cancers*, 16(21), 3668. https://www.mdpi.com/2072-6694/16/21/3668.

TensorFlow documentation. (n.d.). TensorFlow. *Retrieved from https://www.tensorflow.org/*.

Tschandl, P., Rosendahl, C., Akay, B. N., Argenziano, G., Blum, A., Braun, R. P., Cabo, H., Gourhant, J.-Y., Kreusch, J., Lallas, A., Lapins, J., Marghoob, A., Menzies, S., Neuber, N., Paoli, J., Rabinovitz, H., Stolz, W., & Kittler, H. . (2020). Transfer learning for dermatology: Overcoming small datasets with deep neural networks. *Journal of Investigative Dermatology*.

Vishwakarma, N. (diakses 12 Mei 2025). A Guide to Grad-CAM in Deep Learning- https://www.analyticsvidhya.com/blog/2023/12/Grad-CAM-in-deep-learning/.

Wang, L. L. (2020). COVID-Net: A tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Scientific Reports, *10*(1), 19549.*

Weiss, K., Khoshgoftaar, T. M., & Wang, D. . (2020). A survey of transfer learning. *Journal of Big Data*, 3(1), 9.

World Health Organization. (2018). Burns. *WHO*, https://www.who.int/news-room/fact-sheets/detail/burns.