

# KLASIFIKASI CITRA KELAINAN KULIT MENGGUNAKAN RESNET50 DAN MOBILENETV2 PADA DATASET HAM10000

*(Skin Lesion Image Classification Using RESNET50 and MOBILENETV2 On The  
HAM10000)*

Ida Ayu Dewi Purnama Anjani<sup>[1]</sup>, Fitri Bimantoro<sup>[1]</sup>, Halil Akhyar<sup>[1]</sup>

<sup>[1]</sup>Dept Informatics Engineering, Mataram University

Jl. Majapahit 62, Mataram, Lombok NTB, INDONESIA

Email: dayudewi37@gmail.com, bimo@unram.ac.id, halil@staff.unram.ac.id

## Abstract

Penyakit kulit merupakan salah satu masalah kesehatan dengan prevalensi yang cukup tinggi di Indonesia. Proses diagnosis kelainan kulit umumnya dilakukan melalui biopsi atau pemeriksaan visual menggunakan dermoskop yang memerlukan waktu serta ketelitian tinggi. Perkembangan deep learning, khususnya Convolutional Neural Network (CNN), memungkinkan proses klasifikasi citra medis dilakukan secara otomatis dengan tingkat akurasi yang lebih baik melalui pendekatan transfer learning. Penelitian ini bertujuan untuk menganalisis dan membandingkan performa model ResNet50 dan MobileNetV2 dalam klasifikasi tujuh jenis kelainan kulit pada dataset HAM10000. Tahapan penelitian meliputi preprocessing data, augmentasi citra, pembagian dataset menjadi data train, validation, dan test, training model menggunakan bobot pretrained ImageNet, fine-tuning, serta evaluasi menggunakan confusion matrix dan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa ResNet50 memperoleh performa terbaik dengan accuracy 85,43%, precision 86,26%, recall 85,43%, dan F1-score 85,67%. Sementara itu, MobileNetV2 memperoleh accuracy 83,70%, precision 83,64%, recall 83,70%, dan F1-score 83,63%. Dari sisi efisiensi komputasi, MobileNetV2 memiliki jumlah parameter yang lebih sedikit dan waktu inferensi yang lebih cepat dibandingkan ResNet50. Hasil penelitian menunjukkan bahwa ResNet50 lebih unggul dalam performa klasifikasi, sedangkan MobileNetV2 lebih efisien untuk implementasi pada perangkat dengan sumber daya komputasi terbatas.

**Keywords:** Klasifikasi Citra, HAM10000, Kelainan Kulit, CNN, ResNet50, MobileNetV2

## 1. PENDAHULUAN

Penyakit kulit adalah salah satu isu kesehatan yang cukup umum di Indonesia dengan prevalensi mencapai 4,60% hingga 12,95% pada tahun 2020 dan termasuk dalam sepuluh penyakit teratas [1]. Penyebab kelainan kulit dapat disebabkan oleh infeksi bakteri atau virus, lingkungan, pola hidup tidak sehat, serta kondisi kesehatan tertentu [2]. Diagnosis kondisi kulit biasanya dilakukan melalui pemeriksaan visual oleh dokter ahli kulit atau melalui prosedur biopsi [3]. Akan tetapi, biopsi memerlukan waktu dan biaya yang cukup tinggi, sementara diagnosis visual tanpa alat bantu seperti dermoskop hanya memiliki akurasi terbatas yaitu

sekitar 60% hingga 84% [4], [5]].

Kemajuan dalam *Artificial Intelligence* (AI) dan *deep learning* membuka peluang baru di sektor kesehatan, terutama dalam pengklasifikasian gambar kelainan kulit melalui *Convolutional Neural Network* (CNN) [6], [7]. Beragam studi sebelumnya menunjukkan bahwa arsitektur CNN dapat memberikan kinerja klasifikasi yang baik pada gambar medis. Studi oleh Supirman dan rekan-rekan memanfaatkan arsitektur VGG16 dengan model ImageNet yang sudah dilatih sebelumnya, mencapai akurasi 82% dalam klasifikasi penyakit kulit [8]. Studi lain oleh Saputra dan Al-Rivan menerapkan ResNet-152

untuk klasifikasi kanker kulit dan mendapatkan akurasi sebesar 87,85% [9]. Hasil tersebut mengindikasikan bahwa pilihan arsitektur CNN berpengaruh besar terhadap kinerja klasifikasi.

Dalam studi ini diterapkan dua arsitektur CNN, yaitu ResNet50 dan MobileNetV2. ResNet50 dipilih karena memiliki mekanisme pembelajaran residual melalui koneksi shortcut yang dapat mendukung *training* jaringan yang lebih dalam tanpa mengalami masalah *vanishing gradient* [10]. Studi sebelumnya mengindikasikan bahwa ResNet50 dapat menghasilkan kinerja klasifikasi yang baik pada gambar penyakit kulit [11]. Sementara itu, MobileNetV2 dipilih karena dikembangkan sebagai arsitektur ringkas dengan struktur *inverted residual* dan *linear bottleneck*, sehingga memiliki jumlah parameter yang lebih sedikit dan efisiensi komputasi yang lebih tinggi [12]. Penelitian yang dilakukan oleh Fransisca dan Matondang menunjukkan bahwa MobileNetV2 dapat mencapai tingkat akurasi 94% dalam mengklasifikasikan citra cacar monyet dengan menggunakan transfer learning [13].

*Dataset* yang digunakan dalam penelitian ini adalah HAM10000 yang mencakup 10.015 gambar dermoskopi dengan tujuh jenis kelainan kulit yaitu melanocytic nevi (NV), melanoma (MEL), benign keratosis (BKL), basal cell carcinoma (BCC), actinic keratosis (AKIEC), vascular lesions (VASC), serta dermatofibroma (DF). *Dataset* HAM10000 adalah salah satu *dataset* referensi yang sering digunakan dalam penelitian klasifikasi penyakit kulit karena memiliki validasi medis yang baik dan variasi gambar yang cukup lengkap [14].

Walaupun ResNet50 dan MobileNetV2 telah banyak digunakan pada klasifikasi citra medis, berdasarkan penelitian yang merujuk pada jurnal klasifikasi, belum ditampilkan perbandingan langsung kedua arsitektur pada *dataset* HAM100000 dengan evaluasi yang mencakup performa klasifikasi dan efisiensi komputasi secara bersamaan. Oleh karena itu, penelitian ini membandingkan ResNet50 dan MobileNetV2 pada *dataset* HAM100000 berdasarkan *accuracy*, *precision*, *recall*, *F1-score*, jumlah parameter, dan waktu inferensi.

## 2. TINJAUAN PUSTAKA

Beberapa penelitian telah dilakukan dalam pengklasifikasian kelainan kulit dengan menggunakan metode *deep learning* yang berbasis *Convolutional Neural Network* (CNN). Supirman dan rekan-rekannya memanfaatkan arsitektur VGG16 yang berbasis *transfer learning* untuk klasifikasi penyakit kulit dan mencapai *accuracy* 82% [8]. Penelitian lain yang dilakukan oleh Tommy Saputra dan Muhammad Ezar Al Rivan memanfaatkan ResNet-152 dan AlexNet untuk mengklasifikasikan kanker kulit, dengan ResNet-152 mencapai *accuracy* 87,85% [9]. Penelitian oleh I Putu Agus dan rekan-rekan menunjukkan bahwa penerapan *transfer learning* pada CNN dapat meningkatkan kinerja klasifikasi serta mempercepat proses *training* hingga 60% dibandingkan dengan *training* dari nol [15].

Penelitian yang dilakukan oleh Talha Mahboob Alam dan rekan-rekannya berfokus pada penanganan ketidakseimbangan kelas dalam *dataset* kanker kulit melalui *deep learning* dan berhasil meraih *accuracy* sekitar 91% dengan memanfaatkan RegNetY-320 [16]. Akan tetapi, model tersebut memiliki tingkat kompleksitas yang tinggi dan memerlukan sumber daya komputasi yang besar. Penelitian lain yang menerapkan ResNet50 untuk klasifikasi gambar penyakit cacar monyet mencapai *accuracy training* sebesar 94,82% dan *testing* sebesar 76,10%, yang mengindikasikan adanya kemungkinan *overfitting* [11]. Pengaplikasian *hybrid deep learning* yang menggabungkan Mask R-CNN dan ResNet50 pada *dataset* ISIC 2020 menghasilkan *accuracy* klasifikasi mencapai 96,75%, namun meningkatkan kompleksitas sistem serta kebutuhan untuk anotasi manual [17].

Dalam penelitian lainnya, Nurul Khasanah dan Monikka Nur Winnarto memanfaatkan ResNet50 untuk mendeteksi melanoma dan berhasil mencapai *accuracy* 87% [18]. Sementara itu, Fransisca serta Matondang menerapkan MobileNetV2 dengan *transfer learning* untuk mendeteksi cacar monyet dan mendapatkan *accuracy* sebesar 94% [13]. Penelitian oleh Ram Chandna dan rekan-rekan memanfaatkan kombinasi EfficientNet dan MobileNetV2 dengan metode *two-path* dan mencapai *accuracy* 93% [19]. Di samping itu, Nirupama dan Virupakshappa berinovasi pada MobileNetV2 dengan menambahkan perhatian dan

fitur *multi-scale*, yang menghasilkan *accuracy* mencapai 98,6%, meskipun jumlah parameter model mengalami peningkatan yang signifikan [20].

Menurut penelitian sebelumnya, CNN menunjukkan kinerja yang mengesankan dalam pengklasifikasian citra medis. Akan tetapi, sebagian besar penelitian lebih memusatkan perhatian pada peningkatan *accuracy* klasifikasi tanpa membahas perbandingan efisiensi komputasi antara arsitektur CNN. Sehingga, penelitian ini bertujuan untuk membandingkan kinerja klasifikasi dan efisiensi komputasi antara ResNet50 dan MobileNetV2 pada *dataset* HAM1000.

Penelitian terdahulu menggunakan arsitektur dan *dataset* yang beragam. Pada daftar penelitian berikut belum ditampilkan perbandingan langsung ResNet50 dan MobileNetV2 pada HAM10000 yang sekaligus mengevaluasi performa klasifikasi dan efisiensi komputasi.

Tabel 1 merupakan tabel yang berisi penelitian sebelumnya yang berkaitan dengan penelitian yang akan dilakukan dengan menampilkan nama penulis, model yang digunakan, dan hasil yang diperoleh.

TABEL 1. STATE OF THE ART PENELITIAN KLASIFIKASI CITRA KELAINAN KULIT

| Penelitian                    | Model & Dataset/Tugas                      | Hasil                     |
|-------------------------------|--------------------------------------------|---------------------------|
| Lubis dkk. [8]                | VGG16 + TL; penyakit kulit                 | 82%                       |
| Saputra & Al Rivan [9]        | ResNet-152 + AlexNet; kanker kulit         | 87,85%                    |
| Alam dkk. [16]                | RegNetY-320; kanker kulit imbalanced       | ≈91%                      |
| Saputra dkk. [11]             | ResNet50; cacar monyet                     | Train 94,82%; test 76,10% |
| Akram dkk. [17]               | Mask R-CNN + ResNet50; ISIC 2020           | 96,75%                    |
| Khasanah & Winnarto [18]      | ResNet50; melanoma                         | 87%                       |
| Fransisca & Matondang [13]    | MobileNetV2 + TL; cacar monyet             | 94%                       |
| Chandna dkk. [19]             | EfficientNet + MobileNetV2; penyakit kulit | 93%                       |
| Nirupama & Virupakshappa [20] | MobileNetV2 + attention; penyakit kulit    | 98,6%                     |

## 2.1 Teori Penunjang

Dalam penelitian ini terdapat beberapa teori fundamental yang digunakan sebagai dasar penelitian, yaitu:

### a. Citra Digital Dermatoskopi

Citra dermoskopi adalah gambar digital yang dihasilkan dari pemeriksaan kulit menggunakan alat dermoskop. Gambar ini menyimpan data visual yang mencakup warna, tekstur, dan pola struktur kelainan kulit dalam format RGB [17]. Dalam proses pengklasifikasian kelainan kulit, gambar dermoskopi menghadapi tantangan seperti *noise*, rambut yang menutupi kelainan, variasi pencahayaan, serta perbedaan ukuran dan bentuk kelainan kulit [5]. Sehingga, diperlukan pendekatan pengolahan gambar dan *machine learning* yang dapat mengidentifikasi pola visual secara otomatis.

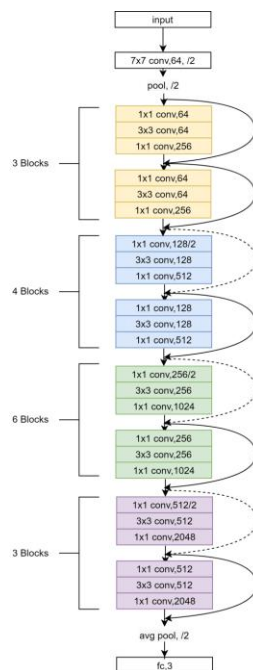
### b. Convolutional Neural Network (CNN)

*Convolutional Neural Network* (CNN) adalah salah satu teknik *deep learning* yang sering diterapkan dalam pengklasifikasian gambar. CNN terdiri dari lapisan *convolution*, *pooling*, *activation*, dan *fully connected layer* [5]. *Convolution layer* dimanfaatkan untuk mengambil fitur lokal seperti tepi dan tekstur, sementara *pooling layer* berfungsi untuk memperkecil ukuran peta fitur agar model lebih efisien. CNN dapat mengenali pola rumit pada gambar medis sehingga sering digunakan dalam pengklasifikasian citra dermoskopi [5], [7].

### c. Transfer Learning

*Transfer learning* adalah metode yang memanfaatkan model CNN yang sudah dilatih sebelumnya dengan *dataset* besar seperti ImageNet [4], [10]. Dalam *transfer learning*, model yang telah dilatih sebelumnya digunakan sebagai pengambil fitur, lalu lapisan terakhir model diadaptasi sesuai dengan jumlah kelas dalam *dataset* penelitian. Metode ini mendukung percepatan proses *training*, menekan *overfitting*, dan meningkatkan kinerja klasifikasi gambar medis [8], [14].

#### d. ResNet50



Gambar 1. Diagram arsitektur ResNet50

ResNet50 adalah arsitektur CNN yang menerapkan *residual learning* dengan menggunakan *shortcut connection* untuk mengatasi masalah *vanishing gradient* pada jaringan yang dalam. Dengan 50 layer, ResNet50 dapat mengekstraksi fitur-fitur kompleks dari citra dermoskopi dan menghasilkan kinerja klasifikasi yang tinggi. Akan tetapi, model ini memiliki banyak parameter sehingga memerlukan sumber daya komputasi yang lebih tinggi [17], [18].

$$H(x) = F(x) + x$$

Persamaan *residual learning*: keluaran blok merupakan transformasi  $F(x)$  yang ditambahkan dengan *shortcut*  $x$ .

MobileNetV2 adalah arsitektur CNN yang dikembangkan untuk efisiensi komputasi di perangkat dengan sumber daya terbatas [12]. MobileNetV2 memanfaatkan residual terbalik dan convolution terpisah kedalaman untuk menurunkan jumlah parameter dan operasi komputasi. Meskipun ringan, MobileNetV2 tetap dapat memberikan kinerja klasifikasi yang baik dalam berbagai penelitian klasifikasi gambar medis [13], [21].

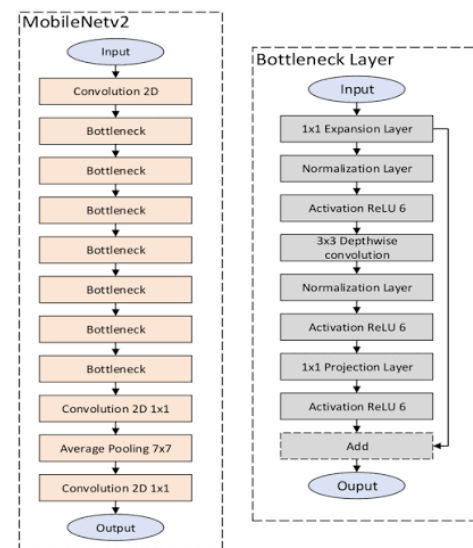
$$y = P_l(DW(P_e(x)))$$

$P_e$ : *pointwise expansion*

$DW$ : *depthwise convolution*

$P_l$ : *linear projection* pada *bottleneck* MobileNetV2.

#### e. MobileNetV2



Gambar 2. Diagram arsitektur MobileNetV2

#### Dataset HAM10000

*Dataset* HAM10000 berisi 10.015 gambar dermoskopi yang mencakup tujuh tipe kelainan kulit yaitu melanocytic nevi (NV), melanoma (MEL), benign keratosis (BKL), basal cell carcinoma (BCC), actinic keratoses (AKIEC), vascular lesions (VASC), dan dermatofibroma (DF) [14]. *Dataset* ini sering digunakan dalam penelitian klasifikasi penyakit kulit karena memiliki validasi medis yang kuat. Akan tetapi, pembagian jumlah data di antara kelas tidak seimbang yang dapat memengaruhi kinerja model klasifikasi [16].

#### f. Data Augmentation dan Class Imbalance

Augmentasi data digunakan untuk meningkatkan variasi data *training* dan membantu model dalam melakukan generalisasi dengan lebih efektif [13]. Proses augmentasi dilaksanakan melalui transformasi gambar seperti rotasi, *flipping*, pergeseran, *shear*, dan *zoom*. Dalam penelitian ini, augmentasi hanya dilakukan pada data *train*.

*Dataset* HAM10000 juga menunjukkan distribusi kelas yang tidak seimbang, yang membuat model cenderung memprediksi kelas yang lebih dominan. Untuk mengatasi permasalahan itu diterapkan *class weight* agar model lebih fokus pada kelas minoritas selama proses *training*.

#### g. Confusion Matrix dan Metrik Evaluasi

##### 1. Confusion matrix

*Confusion matrix* digunakan untuk menilai kinerja model klasifikasi dengan membandingkan hasil prediksi

model dengan label yang benar [22]. Dalam penelitian ini, *confusion matrix* diterapkan untuk menilai klasifikasi tujuh kategori kelainan kulit pada *dataset* HAM10000 yang menghasilkan matriks berukuran 7×7. Baris dalam matriks menunjukkan kelas yang sebenarnya, sementara kolom menggambarkan kelas yang diprediksi oleh model [23].

|              |         | Kelas Prediksi |         |         |         |         |
|--------------|---------|----------------|---------|---------|---------|---------|
|              |         | Kelas 1        | Kelas 2 | Kelas 3 | Kelas 4 | Kelas 5 |
| Kelas Aktual | Kelas 1 | TP11           | FP12    | FN13    | FP14    | FP15    |
|              | Kelas 2 | FP21           | TP22    | FP23    | FN24    | FP25    |
|              | Kelas 3 | FN31           | FP32    | TP33    | FP34    | FP35    |
|              | Kelas 4 | FP42           | FN43    | FP43    | TP44    | FP45    |
|              | Kelas 5 | FP51           | FP52    | FN53    | FP54    | TP55    |

Gambar 3 *Confusion Matrix Multiclass*

Angka pada diagonal utama *confusion matrix* mencerminkan total prediksi yang akurat atau *True Positive* (TP) untuk setiap kelas. Sementara itu, nilai di luar diagonal menggambarkan kesalahan klasifikasi antara kelas. Berdasarkan matriks kebingungan, beberapa parameter evaluasi dapat dihasilkan, yaitu *True Positive* (TP), *False Positive* (FP), *False Negative* (FN), dan *True Negative* (TN) [22].

## II. Metrik evaluasi

Berdasarkan *confusion matrix*, kinerja model dievaluasi melalui beberapa metrik seperti akurasi, presisi, recall, dan F1-score [22], [23], [24].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

*Accuracy* digunakan untuk menilai kemampuan model dalam mengklasifikasikan semua data dengan tepat.

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

*Precision* menggambarkan seberapa akurat model dalam memprediksi kelas positif.

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

*Recall* digunakan untuk menilai kemampuan model dalam mengidentifikasi data positif yang sesungguhnya.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

F1-score adalah rata-rata harmonis dari *precision* dan *recall* yang digunakan untuk memberikan evaluasi yang lebih seimbang, terutama pada *dataset* dengan distribusi kelas yang tidak merata.

Keterangan: TP (True Positive) menunjukkan data positif yang terprediksi benar, TN (True Negative) adalah data negatif yang terprediksi benar, FP (False Positive) adalah data negatif yang keliru diprediksi

sebagai positif, dan FN (False Negative) adalah data positif yang keliru diprediksi sebagai negatif.

## 3. METODE PENELITIAN

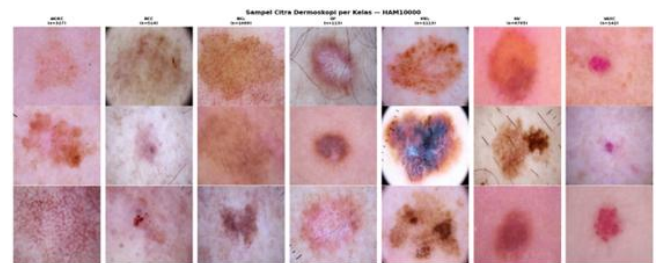
### 3.1 Alat dan Bahan

Alat dan bahan digunakan dalam penelitian ini untuk mendukung proses *testing* sistem klasifikasi kelainan kulit.

#### a. Alat

Alat yang digunakan dalam penelitian ini meliputi laptop ASUS Vivobook 14 i7, RAM 8GB, *operation system* Windows 11 *home single language*, Kaggle Notebook, Bahasa pemrograman Python, dan penyimpanan Google Drive.

#### b. Bahan



Gambar 4. Citra per Kelas *Dataset* HAM10000

Data yang digunakan berasal dari *dataset* gambar dermoskopi HAM10000 (*Human Against Machine* 10000) yang diperoleh dari platform Kaggle. *Dataset* ini berisi 10.015 gambar berformat JPEG dengan tujuh label subclass kelainan kulit dengan rincian seperti Tabel II.

TABEL II. DISTRIBUSI JUMLAH CITRA PADA *DATASET* HAM10000

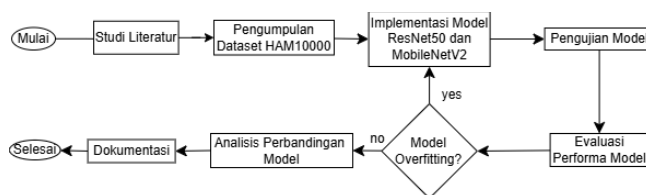
| No    | Jenis Kelainan Kulit | Label | Jumlah |
|-------|----------------------|-------|--------|
| 1.    | Melanocytic Nevi     | NV    | 6.705  |
| 2.    | Melanoma             | MEL   | 1.113  |
| 3.    | Benign Keratosis     | BKL   | 1.099  |
| 4.    | Basal Cell Carcinoma | BCC   | 514    |
| 5.    | Actinic Keratoses    | AKIEC | 327    |
| 6.    | Vascular Lesions     | VASC  | 142    |
| 7.    | Dermatofibroma       | DF    | 115    |
| Total |                      |       | 10.015 |

*Dataset* HAM10000 memiliki ketidakseimbangan kelas yang signifikan, di mana kelas NV mendominasi dibandingkan kelas lainnya. Kondisi ini berpotensi menyebabkan model bias terhadap kelas mayoritas dan menurunkan kemampuan klasifikasi pada kelas minoritas. Oleh karena itu, diterapkan teknik augmentasi data untuk meningkatkan variasi citra pada kelas minoritas dan membantu meningkatkan kemampuan generalisasi

model. *Dataset* yang digunakan berupa citra dermoskopi berwarna dari berbagai jenis kelainan kulit yang digunakan untuk proses *training* dan evaluasi model klasifikasi.

### 3.2 Alur Penelitian

Penelitian ini dilakukan melalui beberapa langkah yang disusun secara sistematis, mulai dari studi literatur hingga melakukan evaluasi model. Tujuan alur penelitian adalah untuk membangun sistem klasifikasi kelainan kulit dengan menggunakan arsitektur CNN ResNet50 dan MobileNetV2 pada *dataset* HAM10000. Tahapan penelitian digambarkan pada Gambar alur penelitian.



Gambar 5. Alur Penelitian

#### a. Studi Literatur

Proses studi literatur dilaksanakan dengan cara melihat dan memahami penelitian yang telah ada sebelumnya mengenai klasifikasi penyakit kulit menggunakan CNN, khususnya yang melibatkan *dataset* HAM10000 serta struktur arsitektur ResNet50 dan MobileNetV2. Studi ini menjadi dasar untuk merancang metode dan strategi penelitian.

#### b. Pengumpulan *Dataset* HAM10000

Dalam tahap pengumpulan *dataset* HAM10000, *dataset* HAM1000 diambil dari Kaggle yang berisi gambar dermoskopi dengan tujuh kelas jenis kelainan kulit. Selanjutnya, data tersebut dipersiapkan dengan membaginya menjadi data *train*, data *val*, dan data *test*.

#### c. Tahap Implementasi Model

Pada tahap implementasi model, dilakukan penerapan model klasifikasi dengan menggunakan arsitektur ResNet50 dan MobileNetV2.

#### d. *Testing* Model

Tahap *testing* model berlangsung dengan melakukan uji coba pada model menggunakan data *testing* untuk menilai kemampuan generalisasi model.

#### e. Evaluasi Performa Model

Untuk tahap evaluasi performa model, digunakan metrik *accuracy*, *precision*, *recall*, dan *F1-Score*, serta dilakukan analisis *confusion matrix* untuk melihat penyebaran prediksi pada setiap kelas. Jika evaluasi menunjukkan bahwa model mengalami *overfitting*, maka akan dilakukan perbaikan dengan kembali ke langkah implementasi model. Jika hasilnya baik, maka akan dilanjutkan ke tahap berikutnya.

#### f. Analisis Perbandingan Model

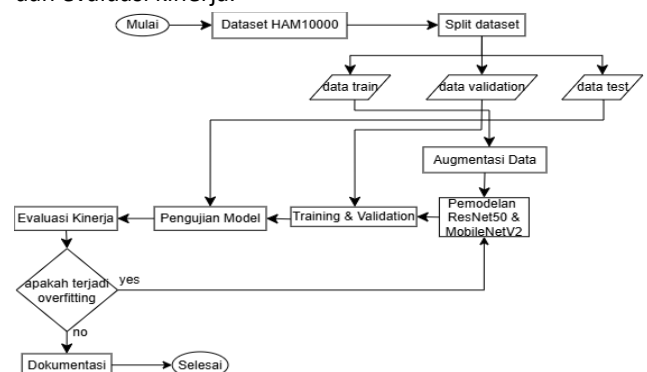
Pada tahap analisis perbandingan model, dilakukan analisis perbandingan performa antara arsitektur ResNet50 dan MobileNetV2 berdasarkan hasil evaluasi yang telah didapatkan.

#### g. Dokumentasi

Tahap dokumentasi mencakup penyimpanan hasil analisis yang kemudian disajikan dalam bentuk laporan sebagai rekaman dari penelitian yang dilakukan.

### 3.3 Metode Klasifikasi

Dalam penelitian ini, metode klasifikasi memanfaatkan arsitektur CNN yang berbasis *transfer learning*, yaitu ResNet50 dan MobileNetV2 untuk mengklasifikasikan tujuh tipe kelainan kulit dalam *dataset* HAM10000. Langkah-langkah dalam klasifikasi meliputi *preprocessing*, augmentasi data, *training* model, *testing*, dan evaluasi kinerja.



Gambar 6. Metode Klasiifikasi

#### a. *Preprocessing*

Tahap *preprocessing* bertujuan untuk menyiapkan citra agar sesuai sebagai *input* untuk model CNN. Semua gambar diubah ukurannya agar seragam. Setelah itu, dilakukan *preprocessing* dengan menggunakan fungsi *preprocess\_input* untuk menyesuaikan distribusi nilai piksel dengan model yang telah dilatih sebelumnya pada ImageNet.

b. **Pembagian Dataset**

*Dataset* dibagi menjadi tiga bagian, yaitu 70% untuk data *training*, 15% untuk data validasi, dan 15% untuk data *testing*. Data *training* digunakan untuk melatih model, data validasi digunakan untuk memantau proses *training* dan mendeteksi apakah terjadi *overfitting*, sedangkan data *testing* digunakan untuk melakukan evaluasi akhir model.

c. **Augmentasi Data**

Augmentasi data diterapkan hanya pada data *training* untuk meningkatkan variasi dan membantu mengatasi ketidakseimbangan kelas. Teknik augmentasi yang diterapkan meliputi rotasi, *flip horizontal*, *flip vertical*, *width shift*, *height shift*, *shear*, dan *zoom*.

d. **Transfer Learning ResNet50 dan MobileNetV2**

Penelitian ini memanfaatkan *transfer learning* dengan bobot yang diambil dari ImageNet. ResNet50 dipilih karena kemampuannya dalam *residual learning* yang efektif dalam mengekstraksi fitur kompleks, sementara MobileNetV2 dipilih karena lebih ringan dalam jumlah parameter dan lebih efisien dalam perhitungan.

e. **Training Model**

Proses *training* dilaksanakan menggunakan optimasi Adam dengan berbagai pengaturan *hyperparameter* seperti *learning rate*, *batch size*, dan *dropout*. Data validasi dimanfaatkan untuk memantau perubahan akurasi dan kerugian selama proses *training* model.

f. **Testing Model**

Model yang telah menjalani proses *training* selanjutnya diuji dengan data *testing* untuk menilai kemampuan model dalam mengklasifikasikan data yang belum pernah dipakai sebelumnya. Prediksi yang dihasilkan dari data *testing* selanjutnya dijadikan landasan untuk menganalisis kinerja model ResNet50 dan MobileNetV2.

g. **Evaluasi Kinerja Model**

Penilaian kinerja model dilakukan dengan memanfaatkan *confusion matrix* dan metrik *accuracy*, *precision*, *recall*, serta *F1-score* [22]. Metrik tersebut digunakan untuk menilai kinerja klasifikasi model secara menyeluruh serta untuk tiap kelas kelainan kulit dalam *dataset* HAM10000.

h. **Pemeriksaan Overfitting**

*Testing overfitting* dilakukan dengan membandingkan nilai *accuracy* dan *loss* pada data *training* dan validasi. Apabila terdapat perbedaan mencolok antara performa *training* dan *validation*, maka perlu dilakukan penyesuaian pada proses *training* seperti augmentasi data, *dropout*, *fine-tuning*, atau pengaturan *hyperparameter* untuk mengurangi *overfitting* dan memperbaiki kemampuan generalisasi model.

i. **Dokumentasi**

Semua proses *preprocessing*, *training*, *testing*, dan evaluasi model didokumentasikan secara sistematis untuk membantu analisis hasil penelitian dan penyusunan laporan penelitian.

### 3.4 Testing

*Testing* dilakukan untuk membandingkan kinerja ResNet50 dan MobileNetV2 dalam klasifikasi citra kelainan kulit. Untuk mempercepat proses *training* dan *inference* model, *testing* dilakukan pada *platform* Kaggle Notebook menggunakan GPU NVIDIA Tesla T4. Selain menilai performa klasifikasi, penelitian ini juga menganalisis efisiensi komputasi. Hal ini dilakukan dengan melihat jumlah parameter model dan waktu inferensi per gambar. Waktu inferensi dihitung dengan data *test* untuk mengetahui kecepatan prediksi masing-masing model terhadap data yang belum pernah dilihat sebelumnya.

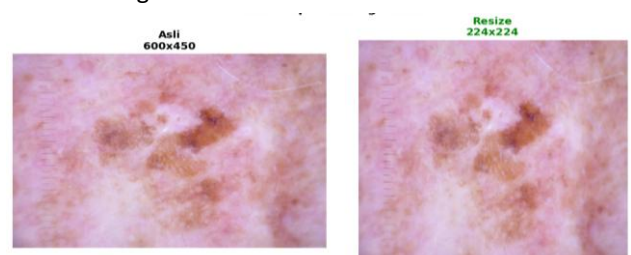
## 4. HASIL DAN PEMBAHASAN

### 4.1 Implementasi Sistem

#### 4.1.1 Preprocessing Data

Proses persiapan data dilakukan sebelum gambar digunakan untuk melatih model ResNet50 dan MobileNetV2. Langkah-langkah persiapan ini meliputi *resize* gambar dan normalisasi menggunakan fungsi yang disediakan oleh masing-masing arsitektur CNN.

a. **Resize gambar**



Gambar 7. Visualisasi Resize Gambar

Seluruh gambar disesuaikan ukurannya menjadi 224×224 piksel agar cocok dengan dimensi *input* standar ResNet50 dan MobileNetV2. Pengubahan ukuran ini dilakukan dengan fungsi `load_img()` dari TensorFlow. Penyesuaian ukuran gambar bertujuan agar semua data dapat diproses dengan seragam selama tahap *training*, validasi, dan *testing*.

#### b. Normalisasi

Normalisasi dilakukan menggunakan fungsi `preprocess_input` yang ada pada setiap arsitektur CNN. Fungsi ini secara otomatis menyelaraskan distribusi nilai piksel sesuai dengan format *input pretrained* ImageNet sehingga proses *transfer learning* dapat dilakukan dengan lebih efektif. Oleh karena itu, pada penelitian ini tidak dilakukan pembagian nilai piksel secara manual dengan 255.

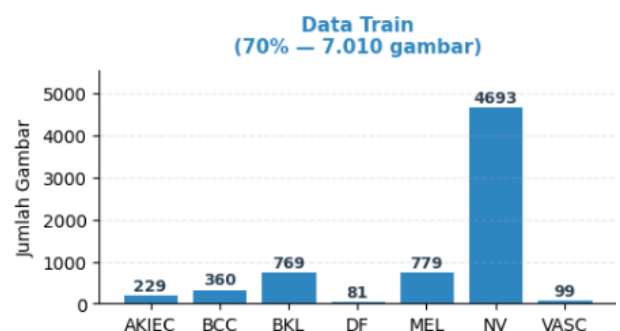
#### 4.1.2 Pembagian Dataset

*Dataset* HAM100000 dibagi menjadi data *training*, validasi, dan *testing* menggunakan *stratified split* dengan *random state* 42 dan proporsi 70:15:15. Pembagian dilakukan pada citra asli sebelum proses augmentasi sehingga citra hasil augmentasi hanya berasal dari data *training* dan tidak masuk ke data validasi maupun *testing*. Dengan demikian, kebocoran data (*data leakage*) akibat citra hasil augmentasi masuk ke *subset* validasi atau *testing* dapat dihindari. Namun, pembagian pada eksperimen utama dilakukan pada tingkat citra dan belum menggunakan pengelompokan berdasarkan *lesion\_id*. Pemeriksaan terhadap pembagian data menunjukkan bahwa terdapat 485 *lesion\_id* yang muncul pada *training* dan validasi, 512 *lesion\_id* yang muncul pada *training* dan *testing*, serta 115 *lesion\_id* yang muncul pada validasi dan *testing*. Selain itu, dari 7.470 *lesion\_id* unik, sebanyak 1.956 *lesion\_id* (26,2%) memiliki lebih dari satu citra. Temuan tersebut menunjukkan adanya potensi kemiripan citra antar-*subset* yang dapat memengaruhi estimasi kemampuan generalisasi model. Sebagai verifikasi, dilakukan pengujian tambahan menggunakan *GroupShuffleSplit* berdasarkan *lesion\_id* tanpa melakukan *training* ulang. Hasil verifikasi menunjukkan *overlap lesion\_id* sebesar 0 pada pasangan *training*-validasi, *training*-*testing*, dan validasi-*testing*. Namun, hasil utama penelitian tetap menggunakan model yang dilatih dengan *split* awal sehingga hasil pengujian tidak diklaim bebas sepenuhnya dari potensi *leakage* berbasis *lesion\_id*. Penggunaan *group-based split* berdasarkan *lesion\_id* direkomendasikan untuk eksperimen lanjutan.

TABEL III. PEMBAGIAN DATASET

| Subset     | Jumlah Data | Persentase | Target Akhir Data |
|------------|-------------|------------|-------------------|
| Training   | 7010        | 70%        | 16693             |
| Validation | 1502        | 15%        | 1502              |
| Testing    | 1503        | 15%        | 1503              |

Data *training* digunakan untuk melatih model, data *validation* digunakan untuk memantau performa model selama *training*, sedangkan data *testing* digunakan untuk evaluasi akhir terhadap data yang belum pernah dilihat sebelumnya. Pembagian *dataset* dilakukan sebelum augmentasi untuk mencegah kebocoran data. Augmentasi hanya diterapkan pada data *training*, sedangkan data *validation* dan *testing* tetap menggunakan citra asli.



Gambar 8. Grafik Distribusi Data Train

Distribusi *dataset* menunjukkan adanya ketidakseimbangan kelas yang cukup tinggi. Kelas NV mendominasi dengan jumlah data terbesar, sedangkan kelas DF memiliki jumlah data paling sedikit. Ketidakseimbangan ini menjadi dasar penerapan augmentasi data dan *class weight* selama *training*.

#### 4.1.3 Augmentasi Data



Gambar 9. Contoh Visualisasi Augmentasi Data

Augmentasi *offline* dilakukan menggunakan *ImageDataGenerator* pada data *training* dan hasil citra augmentasi disimpan sebagai *file* JPEG pada penyimpanan sementara sebelum digunakan dalam *training*. Transformasi yang digunakan meliputi rotasi  $\pm 30^\circ$ , *horizontal flip*, *vertical flip*, *width shift* 15%, *height shift* 15%, *shear* 0,1, *zoom* 15%, serta *brightness range* 0,8–1,2. Augmentasi *offline* diterapkan pada kelas yang jumlah datanya kurang dari 2.000 citra hingga mencapai target tersebut.

Selain augmentasi *offline*, augmentasi *online* diterapkan secara dinamis pada *pipeline* tf.data selama *training* dan tidak disimpan sebagai *file*. Transformasi *online* meliputi *random horizontal flip*, *random vertical flip*, *random brightness* dengan *max\_delta* 30, *random contrast* pada rentang 0,8–1,2, *random saturation* pada rentang 0,8–1,2, dan *random hue* dengan rentang  $\pm 0,05$ . Augmentasi *online* dilakukan pada *batch* data *training* sehingga variasi citra dapat berubah pada setiap proses *training*.

Augmentasi *offline* hanya diterapkan pada kelas minoritas. Berdasarkan data *training* sebelum augmentasi, AKIEC berjumlah 229, BCC 360, BKL 769, DF 81, MEL 779, VASC 99, sedangkan NV berjumlah 4.693 citra sehingga tidak diaugmentasi. Setelah augmentasi *offline*, setiap kelas minoritas mencapai 2.000 citra dan total data *training* menjadi 16.693 citra. Selain augmentasi, digunakan *class weight* untuk memberi bobot lebih besar pada kelas dengan jumlah data yang lebih sedikit.

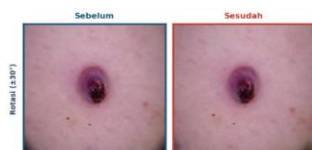
TABEL IV. CLASS WEIGHT MASING-MASING KELAS

| No | Kelas | Class Weight |
|----|-------|--------------|
| 1  | AKIEC | 1.1924       |
| 2  | BCC   | 1.1924       |
| 3  | BKL   | 1.1924       |
| 4  | DF    | 1.1924       |
| 5  | MEL   | 1.1924       |
| 6  | NV    | 0.5081       |
| 7  | VASC  | 1.1924       |

*Class Weight* masing-masing kelas yaitu diperoleh 1.1924 kecuali kelas NV dengan bobot 0.5081 karena merupakan kelas yang dominan. *Class weight* ditentukan dengan rumus berikut:

$$w_i = N / (K \times n_i)$$

dengan N adalah jumlah seluruh data *training* setelah augmentasi, K adalah jumlah kelas, dan  $n_i$  adalah jumlah data pada kelas ke-i.



Gambar 10. Augmentasi Rotasi

Gambar 10 memperlihatkan hasil rotasi augmentasi pada citra dermoskop dilakukan secara acak dalam *range*

–30° hingga 30° untuk mendukung model dalam mengenali kelainan kulit dari berbagai sudut pandang pengambilan gambar.



Gambar 11. Hasil Augmentasi Horizontal Flip

Gambar 11 memperlihatkan hasil dari augmentasi *flip* horizontal. Transformasi ini membalik gambar secara horizontal sehingga model tidak terpengaruh oleh orientasi kanan atau kiri dari kelainan kulit.



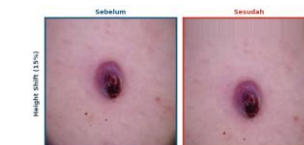
Gambar 12. Augmentasi Vertical Flip

Gambar 12 memperlihatkan hasil dari augmentasi *vertical flip*. Metode ini membalik gambar secara vertikal untuk meningkatkan kemampuan model dalam mendeteksi berbagai variasi arah pengambilan gambar.



Gambar 13 Hasil Augmentasi Width Shift

Gambar 13 memperlihatkan hasil dari pergeseran lebar augmentasi. Transformasi ini memindahkan posisi citra ke kiri atau kanan hingga 15% dari lebar gambar guna meniru posisi kelainan yang tidak selalu terletak di tengah citra.



Gambar 14. Hasil Augmentasi Height Shift

Gambar 14 memperlihatkan hasil dari pergeseran tinggi yang diolah. Metode ini menggeser gambar ke atas atau ke bawah hingga 15% dari tinggi gambar agar model lebih responsif terhadap variasi posisi vertikal objek.



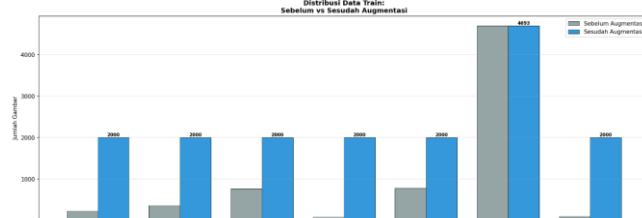
Gambar 15. Hasil Augmentasi Vertical Flip

Gambar 15 memperlihatkan hasil dari augmentasi *shear*. Perubahan ini menimbulkan efek miring pada gambar untuk meniru pergeseran perspektif saat melakukan pengambilan gambar dermoskopi.



Gambar 16. Hasil Augmentasi Zoom

Gambar 16 memperlihatkan hasil dari augmentasi *zoom*. Metode ini menambah atau mengurangi ukuran citra hingga 20% untuk mensimulasikan variasi jarak pengambilan foto terhadap permukaan kulit.



Gambar 17. Bar Chart Distribusi Data Train

Perbandingan distribusi data *training* sebelum dan setelah augmentasi ditampilkan pada Gambar 14. Setelah proses augmentasi selesai, jumlah data pada kelas minoritas meningkat menjadi 2000 citra per kelas, sementara kelas NV tetap menjadi kelas dengan jumlah data terbanyak tanpa penambahan augmentasi.

#### 4.1.4 Implementasi Model ResNet50 dan MobileNetV2

Penelitian ini menggunakan dua arsitektur *Convolutional Neural Network* (CNN) berbasis transfer learning yaitu ResNet50 dan MobileNetV2. Kedua model memanfaatkan *pretrained weights* dari ImageNet dengan parameter `include_top=False` sehingga lapisan klasifikasi bawaan dihapus dan digantikan dengan *classification head* baru.

TABEL V. PERBANDINGAN PARAMETER KEDUA MODEL

| Komponen        | ResNet50            | MobileNetV2            |
|-----------------|---------------------|------------------------|
| <i>Backbone</i> | ResNet50 + ImageNet | MobileNetV2 + ImageNet |
| Input           | 224×224×3           | 224×224×3              |
| Pooling         | GAP 2D              | GAP 2D                 |
| Dense 1         | 512 ReLU            | 256 ReLU               |
| Dropout 1       | 0,45                | 0,50                   |
| Dense 2         | 256 ReLU            | 128 ReLU               |
| Dropout 2       | 0,225               | 0,25                   |

| Output             | 7 softmax             | 7 softmax             |
|--------------------|-----------------------|-----------------------|
| <i>Fine-tuning</i> | 50/175 layer terakhir | 40/154 layer terakhir |
| Parameter          | 24.772.999            | 2.621.255             |
| Epoch maksimum     | 25 + 50               | 25 + 30               |

Berdasarkan Tabel V, kedua model menggunakan *classification head* yang terdiri dari *Global Average Pooling* (GAP), lapisan *dense*, normalisasi *batch*, *dropout*, dan lapisan *output* dengan aktivasi softmax. Perbedaan signifikan antara kedua arsitektur terletak pada jumlah neuron serta nilai *dropout* yang diterapkan.

ResNet50 menggunakan *dense layer* 512 dan 256 neuron dengan regularisasi L2 sebesar 0,0001 serta *dropout* masing-masing 0,45 dan 0,225. MobileNetV2 menggunakan *dense layer* 256 dan 128 neuron dengan regularisasi L2 sebesar 0,0001 serta *dropout* masing-masing 0,50 dan 0,25. Kedua model menggunakan *batch normalization* dan *output* 7 neuron dengan aktivasi softmax.

Penggunaan *batch normalization* bertujuan untuk mempertahankan stabilitas distribusi data selama proses pelatihan, sedangkan *dropout* digunakan untuk mengurangi kemungkinan *overfitting* dengan menonaktifkan beberapa neuron secara acak selama proses pelatihan. Pada kedua model, lapisan *output* memanfaatkan fungsi aktivasi softmax dengan 7 neuron yang sesuai dengan jumlah kelas dalam *dataset* HAM100000.

#### 4.1.5 Training Model

*Training* dilakukan dalam dua tahap, yaitu pelatihan *classification head* dan *fine-tuning*. Pada tahap pertama seluruh *backbone* dibekukan sehingga hanya *classification head* yang dilatih. Pada tahap kedua, ResNet50 membuka 50 *layer* terakhir dari 175 *layer backbone*, sedangkan MobileNetV2 membuka 40 *layer* terakhir dari 154 *layer backbone*. Pemilihan jumlah *layer* yang dibuka ditetapkan sebagai konfigurasi eksperimen untuk memberikan ruang adaptasi fitur pretrained tanpa membuka seluruh *backbone*.

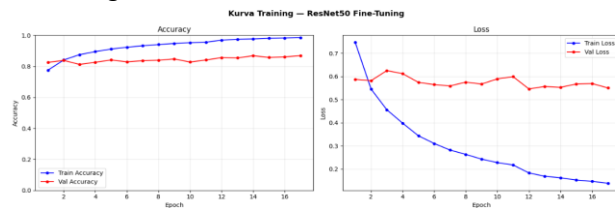
Pada proses *training* digunakan optimizer Adam, fungsi *loss categorical crossentropy* dengan label *smoothing* 0,1, serta *callback* EarlyStopping, ReduceLROnPlateau, dan ModelCheckpoint.

##### a. Training ResNet50



Gambar 18. Grafik Training Classification Head ResNet50

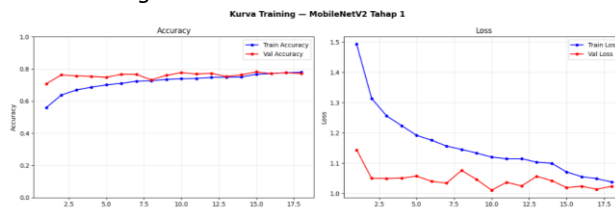
Pada tahap classification head, model ResNet50 memperoleh *training accuracy* sebesar 84,26% dan *validation accuracy* sebesar 81,76%. Selisih *accuracy* sebesar 2,51% menunjukkan belum terdapat indikasi *overfitting*.



Gambar 19. Grafik Training Fine Tuning ResNet50

Pada tahap *fine-tuning*, *training accuracy* meningkat menjadi 98,56% dan *validation accuracy* mencapai 86,88%. Selisih *accuracy* sebesar 11,67% menunjukkan munculnya indikasi *overfitting* ringan, namun model masih memiliki kemampuan generalisasi yang baik.

#### b. Training MobileNetV2



Gambar 20. Grafik Training Classification Head MobileNetV2

Pada tahap *classification head*, MobileNetV2 memperoleh *training accuracy* sebesar 77,79% dan *validation accuracy* sebesar 77,10%. Selisih *accuracy* sebesar 0,08% menunjukkan model memiliki stabilitas *training* yang baik.



Gambar 21. Grafik Training Fine Tuning MobileNetV2

Pada tahap *fine-tuning*, *training accuracy* meningkat menjadi 95,64% dan *validation accuracy* mencapai 85,09%. Walaupun terjadi peningkatan performa, *validation loss* cenderung lebih fluktuatif dibandingkan ResNet50. Hal ini menunjukkan bahwa kapasitas ekstraksi fitur MobileNetV2 masih lebih terbatas dibandingkan ResNet50.

Secara keseluruhan, kedua model menunjukkan *validation accuracy* yang cukup tinggi namun *validation loss* tidak turun secara signifikan. Hal ini dipengaruhi oleh ketidakseimbangan kelas *dataset* HAM10000 dan tingginya kemiripan visual antar kelas seperti MEL dan NV.

## 4.2 Testing dan Pembahasan

### 4.2.1 Parameter Training Model

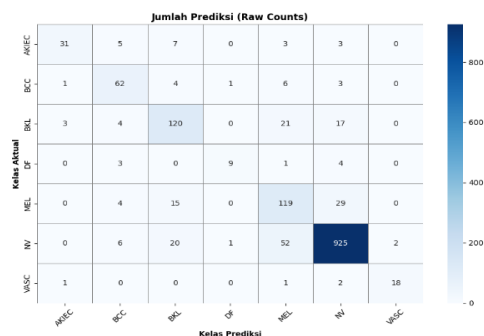
Parameter *training* yang digunakan pada kedua model ditunjukkan pada Tabel 4.4.

TABEL VI. PARAMETER TRAINING MODEL

| Parameter        | ResNet50           | MobileNetV2        |
|------------------|--------------------|--------------------|
| Optimizer        | Adam               | Adam               |
| Batch Size       | 16                 | 16                 |
| LR Tahap 1       | $1 \times 10^{-4}$ | $5 \times 10^{-4}$ |
| LR Fine-tuning   | $3 \times 10^{-5}$ | $3 \times 10^{-5}$ |
| Loss Tahap 1     | CCE + LS 0,1       | CCE + LS 0,1       |
| Loss Fine-tuning | CCE                | CCE + LS 0,1       |
| Dropout          | 0,45; 0,225        | 0,50; 0,25         |
| Epoch maksimum   | 25; 50             | 25; 30             |

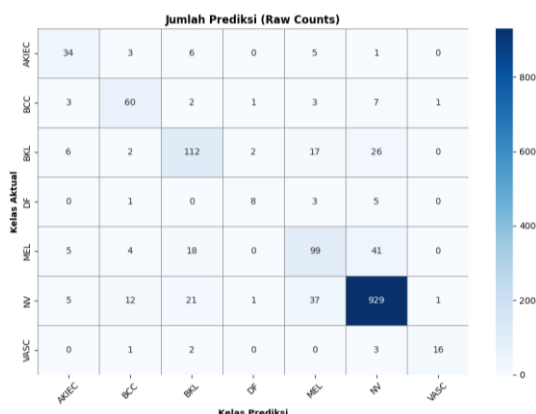
Perbedaan *hyperparameter* ditetapkan berdasarkan karakteristik arsitektur dan kebutuhan regularisasi pada eksperimen. ResNet50 memiliki *backbone* dengan parameter jauh lebih banyak sehingga digunakan *learning rate* tahap awal yang lebih kecil ( $1 \times 10^{-4}$ ) dan *dropout* lebih tinggi (0,45 dan 0,225) untuk membatasi pembaruan bobot yang terlalu agresif dan membantu mengurangi *overfitting*. MobileNetV2 menggunakan *learning rate* tahap awal  $5 \times 10^{-4}$  dengan *dropout* 0,50 dan 0,25 sebagai konfigurasi regularisasi pada *backbone* yang lebih ringan. Pada *fine-tuning*, kedua model menggunakan *learning rate*  $3 \times 10^{-5}$  agar perubahan bobot *pretrained* berlangsung lebih kecil. Konfigurasi tersebut merupakan konfigurasi eksperimen yang digunakan, bukan hasil pencarian *hyperparameter* otomatis.

#### 4.2.2 Confusion Matrix



Gambar 22. Confusion Matrix ResNet50

Berdasarkan *confusion matrix*, ResNet50 menunjukkan performa terbaik pada kelas NV dengan recall sebesar 91,9% dan kelas VASC sebesar 81,8%. Sebaliknya, performa terendah terdapat pada kelas DF dengan recall 52,9% akibat jumlah data yang sangat terbatas.



Gambar 23. Confusion Matrix MobileNetV2

Pada MobileNetV2, performa terbaik juga diperoleh pada kelas NV dengan recall sebesar 92,3%. Namun performa kelas MEL dan DF masih relatif rendah dibandingkan ResNet50. MobileNetV2 cenderung lebih sering memprediksi citra ke kelas mayoritas yaitu NV.

Kesalahan klasifikasi paling banyak terjadi antara kelas MEL dan NV karena kedua kelas memiliki pola pigmen dan distribusi warna yang serupa. Selain itu, beberapa citra BKL juga sering salah diprediksi sebagai MEL atau NV akibat kemiripan tekstur dan bentuk lesi.

#### 4.2.3 Misklasifikasi Model

Pada ResNet50, misklasifikasi terbanyak terjadi pada kelas MEL yang diprediksi sebagai NV sebanyak 29 citra. Selain itu terdapat 17 citra BKL yang diprediksi sebagai NV dan 21 citra BKL yang diprediksi sebagai MEL.

Pada MobileNetV2, kesalahan terbesar juga terjadi pada kelas MEL yang diprediksi sebagai NV sebanyak 41

citra. Hal ini menunjukkan bahwa MobileNetV2 masih mengalami kesulitan membedakan fitur visual halus antara melanoma dan nevus dibandingkan ResNet50.

#### 4.2.4 Performa Model Tiap Kelas

Tabel VII menunjukkan perbandingan precision, recall, dan F1-score masing-masing kelas.

TABEL VII. PARAMETER *TRAINING* MODEL

| Kelas | <i>Precision</i> | <i>Recall</i> | F1-Score | <i>Precision</i> | <i>Recall</i> | F1-Score |
|-------|------------------|---------------|----------|------------------|---------------|----------|
| AKIEC | 0.86             | 0.63          | 0.73     | 0.64             | 0.69          | 0.67     |
| BCC   | 0.74             | 0.81          | 0.77     | 0.72             | 0.78          | 0.75     |
| BKL   | 0.72             | 0.73          | 0.73     | 0.70             | 0.68          | 0.69     |
| DF    | 0.82             | 0.53          | 0.64     | 0.67             | 0.47          | 0.55     |
| MEL   | 0.59             | 0.71          | 0.64     | 0.60             | 0.59          | 0.60     |
| NV    | 0.94             | 0.92          | 0.93     | 0.92             | 0.92          | 0.92     |
| VASC  | 0.90             | 0.82          | 0.86     | 0.89             | 0.73          | 0.80     |

Hasil evaluasi menunjukkan bahwa ResNet50 memberikan performa lebih baik dibandingkan MobileNetV2 pada sebagian besar kelas. Perbedaan paling signifikan terlihat pada kelas MEL dan DF. Hal ini menunjukkan bahwa kemampuan ekstraksi fitur ResNet50 lebih baik dalam membedakan kelas dengan kemiripan visual tinggi.

#### 4.2.5 Hasil Metrik Evaluasi

Perbandingan hasil evaluasi akhir kedua model ditunjukkan pada Tabel 4.6.

TABEL VIII. PERBANDINGAN HASIL EVALUASI MODEL

| Model       | <i>Accuracy</i> | <i>Precision</i> | <i>Recall</i> | F1-Score |
|-------------|-----------------|------------------|---------------|----------|
| ResNet50    | 85.43%          | 86.26%           | 85.43%        | 85.67%   |
| MobileNetV2 | 83.70%          | 83.64%           | 83.70%        | 83.63%   |

Berdasarkan hasil evaluasi, ResNet50 memperoleh performa terbaik dengan *accuracy* 85,43%, *precision* 86,26%, *recall* 85,43%, dan F1-score 85,67%. Sementara itu, MobileNetV2 memperoleh *accuracy* sebesar 83,70% dengan *precision* 83,64%, *recall* 83,70%, dan F1-score 83,63%. Hasil ini menunjukkan bahwa ResNet50 lebih efektif dalam mengekstraksi fitur kompleks pada citra kelainan kulit dibandingkan MobileNetV2.

#### 4.2.6 Perbandingan Efisiensi Komputasi

Perbandingan efisiensi komputasi kedua model ditunjukkan pada Tabel IX.

TABEL IX. PERBANDINGAN EFISIENSI KOMPUTASI MODEL

| Metrik                    | ResNet50   | MobileNetV2 |
|---------------------------|------------|-------------|
| Jumlah Parameter          | 24,772,999 | 2,621,255   |
| Inferensi per Gambar (ms) | 8.78       | 6.94        |

MobileNetV2 memiliki jumlah parameter yang jauh lebih kecil dibandingkan ResNet50 sehingga lebih ringan dan lebih efisien secara komputasi. Waktu inferensi MobileNetV2 juga lebih cepat yaitu 6,94 ms dibandingkan ResNet50 sebesar 8,78 ms.

Walaupun demikian, ResNet50 tetap memberikan performa klasifikasi lebih tinggi. Hal ini menunjukkan adanya trade-off antara *accuracy* dan efisiensi komputasi. ResNet50 lebih unggul dalam kemampuan ekstraksi fitur, sedangkan MobileNetV2 lebih sesuai digunakan pada perangkat dengan sumber daya komputasi terbatas seperti *mobile device* atau *edge device*.

## 5. KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa metode *transfer learning* menggunakan arsitektur CNN mampu mengklasifikasikan citra kelainan kulit pada *dataset* HAM10000 dengan baik. Tahapan *preprocessing* seperti *resize citra*, *preprocess\_input*, augmentasi data, dan penerapan *class weight* membantu meningkatkan kemampuan model dalam menangani ketidakseimbangan data.

Hasil *testing* menunjukkan bahwa model ResNet50 memperoleh performa terbaik dengan *accuracy* sebesar 85,43%, *precision* 86,26%, *recall* 85,43%, dan *F1-score* 85,67%. Sementara itu, MobileNetV2 memperoleh *accuracy* 83,70%, *precision* 83,64%, *recall* 83,70%, dan *F1-score* 83,63%. Hasil tersebut menunjukkan bahwa ResNet50 lebih unggul dalam mengekstraksi fitur dan membedakan pola visual antar kelas kelainan kulit dibandingkan MobileNetV2.

Pada evaluasi per kelas, kedua model menunjukkan performa terbaik pada kelas NV karena jumlah data yang dominan, sedangkan performa terendah terdapat pada kelas DF akibat keterbatasan jumlah data. Kesalahan klasifikasi juga sering terjadi antara kelas MEL dan NV karena memiliki karakteristik visual yang serupa pada citra dermoskopi. Hal ini menunjukkan bahwa ketidakseimbangan data dan kemiripan visual antar kelas masih menjadi tantangan utama dalam klasifikasi kelainan kulit menggunakan CNN.

Dari sisi efisiensi komputasi, MobileNetV2 memiliki jumlah parameter yang lebih sedikit dan waktu inferensi yang lebih cepat dibandingkan ResNet50, sehingga lebih sesuai untuk perangkat dengan sumber daya terbatas. Namun, ResNet50 tetap memberikan performa klasifikasi yang lebih baik secara keseluruhan.

Penelitian selanjutnya disarankan untuk menggunakan *dataset* tambahan agar distribusi data antar kelas menjadi lebih seimbang, terutama pada kelas minoritas seperti DF dan VASC. Selain itu, dapat diterapkan *preprocessing* tambahan seperti *hair removal*, *segmentasi lesi*, atau peningkatan kualitas citra untuk membantu model memperoleh fitur yang lebih relevan.

Pengembangan penelitian juga dapat dilakukan dengan membandingkan arsitektur lain seperti EfficientNet, DenseNet, atau Vision Transformer untuk meningkatkan performa klasifikasi. Selain itu, implementasi model dalam bentuk aplikasi berbasis web atau mobile dapat dikembangkan sebagai sistem pendukung diagnosis awal kelainan kulit.

## UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah memberikan dukungan dan bantuan selama proses penelitian ini. Penulis berharap penelitian ini dapat memberikan manfaat bagi pengembangan teknologi klasifikasi citra kelainan kulit berbasis deep learning.

## DAFTAR PUSTAKA

- [1] N. Sri Rahayu, A. Dita Puteri, and L. Mufti Azzahri Isnaeni, "HUBUNGAN PERILAKU MASYARAKAT DAN PENGGUNAAN AIR SUNGAI DENGAN GANGGUAN PENYAKIT KULIT DI DESA KAMPUNG PINANG WILAYAH KERJA PUSKESMAS PANTAI RAJA," *Jurnal Imliah Ilmu Kesehatan*, vol. 1, no. 3, p. 2023, 2023.
- [2] D. D. Putri, M. Tanzil Furqon, and R. S. Perdana, "Klasifikasi Penyakit Kulit Pada Manusia Menggunakan Metode Binary Decision Tree Support Vector Machine (BDTSVM) (Studi Kasus: Puskesmas Dinoyo Kota Malang)," 2018. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [3] M. E. M. H. B. dan Y. A. B. F. Mahyudin, "Diagnosis dan Terapi TUMOR MUSKULOSKELETAL (Multidiciplinary Approach)," 2017.
- [4] A. Lawi and A. Muh Amil Siddik, "Impelementasi Arsitektur Dengan Pemilihan Model Transfer Learning CNN," 2023.
- [5] I. Kousis, I. Perikos, I. Hatzilygeroudis, and M. Virvou, "Deep Learning Methods for Accurate Skin Cancer Recognition and Mobile Application," *Electronics (Switzerland)*, vol. 11, no. 9, May 2022, doi: 10.3390/electronics11091294.
- [6] F. N. Darmawan, E. P. Silmina, and T. Hardiani, "Sistem Klasifikasi Peyakit Kulit Menggunakan

- Metode *Convolutional Neural Network* (CNN) Berbasis Website,” 2024.
- [7] D. P. Mawardi, M. Novita, and N. Dwi Saputro, “Deteksi Awal Klasifikasi Jenis Penyakit Kanker Kulit Dengan Algoritma *Convolutional Neural Network* (CNN) Berbasis Mobile Apps,” *Adopsi Teknologi dan Sistem Informasi (ATASI)*, vol. 3, no. 2, pp. 1–6, Dec. 2024, doi: 10.30872/atasi.v3i2.2305.
- [8] C. Lubis, D. Yulianto, U. Tarumanagara Jakarta, R. Sakit Tiara Tangerang, and K. Kunci, “KLASIFIKASI PENYAKIT KULIT MENGGUNAKAN *CONVOLUTIONAL NEURAL NETWORK* (CNN) DENGAN ARSITEKTUR VGG16,” vol. 8, no. 1, 2023.
- [9] T. Saputra, M. Ezar, and A. Rivan, “STRING (Satuan Tulisan Riset dan Inovasi Teknologi) ANALISIS PERFORMA RESNET-152 DAN ALEXNET DALAM KLASIFIKASI JENIS KANKER KULIT,” 2023. [Online]. Available: <https://challenge.isic->
- [10] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” 2016. [Online]. Available: <http://image-net.org/challenges/LSVRC/2015/>
- [11] T. O. Saputra, D. Alamsyah, and K. Kunci, “KLASIFIKASI PENYAKIT CACAR MONYET MENGGUNAKAN METODE *CONVOLUTIONAL NEURAL NETWORK*,” 2023.
- [12] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” 2018.
- [13] P. Sarah Fransisca and N. Matondang, “Deteksi Citra Digital Penyakit Cacar Monyet menggunakan Algoritma *Convolutional Neural Network* dengan Arsitektur MobileNetV2,” 2023, [Online]. Available: <http://journal.ipb.ac.id/index.php/jika>
- [14] Nurkhasanah and Murinto, “Klasifikasi Penyakit Kulit Wajah Menggunakan Metode *Convolutional Neural Network* Classification of Facial Skin Diseases Using the Method of the *Convolutional Neural Network*,” *SAINTEKS*, vol. 18, no. 2, 2021, [Online]. Available: <https://www.kaggle.com/datasets>
- [15] I. P. Agus, K. Hidjah, N. Sulistianingsih, G. Hendro, and S. Syahrir, “Implementasi Arsitektur Deep *Convolutional Neural Network* (CNN) dengan Transfer Learning untuk Klasifikasi Penyakit Kulit,” *JTIM : Jurnal Teknologi Informasi dan Multimedia*, vol. 7, no. 3, pp. 461–477, Jun. 2025, doi: 10.35746/jtim.v7i3.734.
- [16] T. M. Alam *et al.*, “An Efficient Deep Learning-Based Skin Cancer Classifier for an Imbalanced Dataset,” *Diagnostics*, vol. 12, no. 9, Sep. 2022, doi: 10.3390/diagnostics12092115.
- [17] A. Akram, J. Rashid, M. A. Jaffar, M. Faheem, and R. ul Amin, “Segmentation and classification of skin lesions using hybrid deep learning method in the Internet of Medical Things,” *Skin Research and Technology*, vol. 29, no. 11, Nov. 2023, doi: 10.1111/srt.13524.
- [18] N. Khasanah and M. N. Winnarto, “Application of Deep Learning with ResNet50 for Early Detection of Melanoma Skin Cancer,” *Journal Medical Informatics Technology*, pp. 16–20, Mar. 2024, doi: 10.37034/medinftech.v2i1.31.
- [19] R. Chandna *et al.*, “Skin disease classification using two path deep transfer learning models,” *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, vol. 3, no. 4, pp. 169–187, Dec. 2024, doi: 10.60087/jklst.v3.n4.p169.
- [20] Nirupama and Virupakshappa, “MobileNet-V2: An Enhanced Skin Disease Classification by Attention and Multi-Scale Features,” *Journal of Imaging Informatics in Medicine*, vol. 38, no. 3, pp. 1734–1754, Jun. 2025, doi: 10.1007/s10278-024-01271-y.
- [21] S. Arnandito and T. B. Sasongko, “Comparison of EfficientNetB7 and MobileNetV2 in Herbal Plant Species Classification Using *Convolutional Neural Networks*,” 2024. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [22] S. Sathyanarayanan, “Confusion Matrix-Based Performance Evaluation Metrics,” *African Journal of Biomedical Research*, pp. 4023–4031, Nov. 2024, doi: 10.53555/ajbr.v27i4s.4345.
- [23] S. Handayani Pusadan, A. Makahrnun, M. Yazdi, and Z. Sakka, “Pattern Recognition untuk Klasifikasi Penyakit Kanker Kulit menggunakan Artificial Intelligence (AI),” 2025. [Online]. Available: <https://jurnal.unw.ac.id/index.php/IKN>
- [24] D. Chicco and G. Jurman, “The advantages of the Matthews correlation coefficient (MCC) over F1 score and *accuracy* in binary classification evaluation,” *BMC Genomics*, vol. 21, no. 1, Jan. 2020, doi: 10.1186/s12864-019-6413-7.