



Department of Digital Business

**Journal of Artificial Intelligence and Digital Business (RIGGS)**

Homepage: <https://journal.ilmudata.co.id/index.php/RIGGS>

Vol. 4 No. 4 (2026) pp: 14959-14968

P-ISSN: 2963-9298, e-ISSN: 2963-914X

---

## Implementasi CNN Mobilenetv2 untuk Klasifikasi Kanker Kulit Dermatoskopi Digital Medis

Desi Permata Sari<sup>1</sup>, Bambang Irawan<sup>2</sup>

Program Studi Teknik Informatika Fakultas Teknik Universitas Muhadi Setiabudi

[desii.permatas06@gmail.com](mailto:desii.permatas06@gmail.com), [bambangumus@gmail.com](mailto:bambangumus@gmail.com)

### Abstract

*Skin cancer is one of the most dangerous diseases that can become fatal if not detected accurately at an early stage. Conventional visual diagnosis performed by dermatologists still faces limitations due to subjectivity and dependence on clinical experience. Therefore, an objective, fast, and efficient diagnostic support system is required. Recent advances in artificial intelligence, particularly deep learning, enable the utilization of dermatoscopic images to assist in automated skin cancer detection. This study aims to implement and evaluate the performance of the Convolutional Neural Network (CNN) MobileNetV2 architecture for classifying skin cancer images into two classes, namely benign and malignant. The research adopts an experimental deep learning approach using the HAM10000 dataset, which consists of 10,015 dermatoscopic images. The data undergo preprocessing stages including image resizing, normalization, and data augmentation, followed by stratified splitting into training, validation, and testing sets. The MobileNetV2 model is trained using a transfer learning strategy through two stages, namely head training and fine-tuning. Experimental results show that the proposed model achieves an accuracy of 82.42%, with a malignant class precision of 64.04% and a recall of 19.93%. Although the model demonstrates strong performance in identifying benign lesions, its ability to detect malignant cases remains limited. These findings indicate that MobileNetV2 has potential as a clinical decision support system for skin cancer classification, but further improvements are required to enhance sensitivity toward malignant lesions.*

**Keywords:** Skin Cancer, Dermatoscopic Images, Convolutional Neural Network, Mobilenetv2, Deep Learning

### 1. Latar Belakang

Kulit merupakan lapisan terluar tubuh yang dihuni oleh berbagai mikroba yang berperan secara fisik, kimia, dan mikrobiologis, serta melalui jalur imunitas baik bawaan maupun yang diperoleh, dalam menjalankan fungsinya sebagai pelindung antara tubuh dan lingkungan luar. Kulit terdiri atas tiga lapisan, yaitu lapisan paling luar (epidermis), lapisan di bawah epidermis (dermis), dan lapisan paling dalam (hipodermis) [1]. Karena posisinya yang berada di bagian terluar, kulit secara langsung menerima berbagai rangsangan seperti sentuhan, rasa sakit, serta pengaruh lain dari lingkungan. Oleh sebab itu, kulit menjadi bagian tubuh yang rentan terhadap berbagai penyakit [2]. Salah satu penyakit yang sering menyerang kulit adalah kanker kulit.

Kanker kulit merupakan kondisi abnormal pada struktur penyusun kulit yang terjadi akibat hilangnya kemampuan kulit dalam melakukan regenerasi secara berkala [3]. Kondisi ini dapat menyebabkan terbentuknya tumor yang bersifat jinak maupun ganas apabila tidak segera ditangani. Lesi kulit yang bersifat ganas (malignant) dapat menyebar ke jaringan lain dan berpotensi fatal jika tidak terdiagnosis secara dini dan tepat. Diagnosis kondisi kulit tidak boleh dilakukan secara asal, sebab kesalahan dalam perawatan dan penanganan bisa berakibat serius. Oleh karena itu, deteksi dini menjadi salah satu upaya penting dalam mencegah dan mengatasi kanker kulit secara efektif [4]. Oleh karena itu, dibutuhkan metode diagnosis yang objektif, cepat, dan efisien untuk mendukung proses klinis tersebut.

Perkembangan teknologi kecerdasan buatan (artificial intelligence) mendorong lahirnya sistem computer-aided diagnosis (CAD) berbasis citra medis yang banyak dimanfaatkan untuk membantu diagnosis penyakit kulit. Salah satu pendekatan yang paling banyak digunakan dalam pengolahan citra medis adalah Convolutional Neural Network (CNN). CNN merupakan bagian dari metode deep learning yang dirancang khusus untuk memproses data berbentuk citra melalui operasi konvolusi sehingga mampu mengekstraksi fitur visual secara otomatis. Arsitektur CNN tersusun atas neuron-neuron dengan bobot dan bias, di mana setiap neuron memproses input melalui operasi matematis sebelum diteruskan ke lapisan berikutnya [5]. Keunggulan utama CNN terletak pada kemampuannya dalam mempelajari pola visual yang kompleks tanpa memerlukan proses ekstraksi fitur manual.

Berbagai penelitian menunjukkan bahwa CNN mampu memberikan performa yang kompetitif dalam klasifikasi citra dermatoskopi kanker kulit, khususnya dalam membedakan lesi benign dan malignant. Penerapan CNN tidak hanya meningkatkan akurasi klasifikasi, tetapi juga membantu mempercepat proses diagnosis dan mengurangi subjektivitas dalam penilaian visual, sehingga berpotensi besar sebagai sistem pendukung diagnosis klinis berbasis citra medis [6].

Di Indonesia, sejumlah penelitian telah mengkaji penerapan CNN dalam klasifikasi kanker kulit. Misalnya, penelitian yang dilakukan oleh [7] menunjukkan bahwa model CNN mampu mencapai akurasi hingga 99,01% dalam membedakan citra kanker kulit benign dan malignant pada dataset yang digunakan. Hasil tersebut mengindikasikan bahwa CNN memiliki potensi tinggi ketika diterapkan secara optimal pada data lokal. Penelitian lain menerapkan pendekatan pra-pemrosesan hybrid dengan mengombinasikan teknik Contrast Limited Adaptive Histogram Equalization (CLAHE) dan median filter untuk meningkatkan kualitas citra dermatoskopi, yang menghasilkan akurasi sebesar 78,19%. Hasil ini menunjukkan bahwa kualitas pra-pemrosesan data berpengaruh signifikan terhadap performa klasifikasi CNN, meskipun akurasinya masih perlu ditingkatkan [8].

Selain itu, penggunaan arsitektur CNN yang lebih dalam seperti ResNet-50 juga menunjukkan hasil yang kuat dalam klasifikasi citra kanker kulit. Penelitian oleh [9] melaporkan bahwa model ResNet-50 mampu mencapai akurasi hingga 92,00% dalam membedakan lesi benign dan malignant, yang menunjukkan bahwa arsitektur CNN yang lebih kompleks mampu mempelajari fitur visual secara lebih mendalam dari citra dermatoskopi..

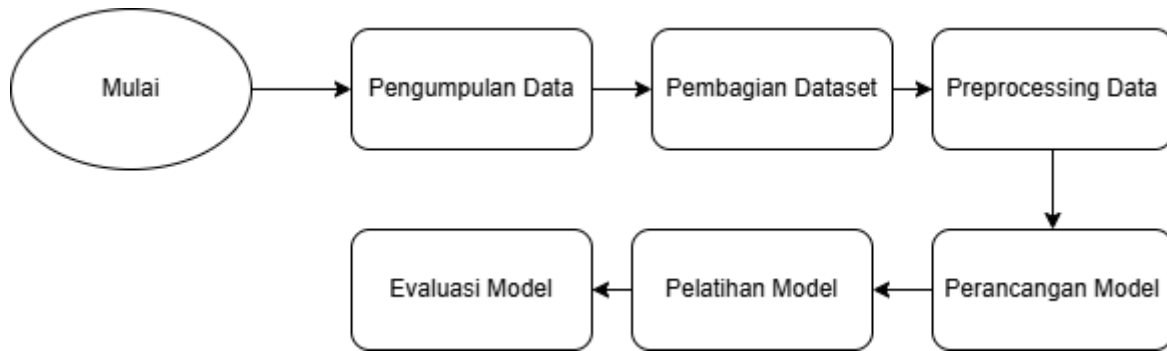
Secara keseluruhan, tinjauan literatur sistematis mengenai penggunaan CNN dalam diagnosis otomatis kanker kulit menyimpulkan bahwa metode deep learning memiliki kemampuan tinggi dalam mengekstraksi fitur relevan dari citra dermatoskopi dan mampu mencapai performa klasifikasi yang mendekati atau bahkan melampaui metode konvensional, dengan beberapa studi melaporkan tingkat akurasi di atas 90% [6].

Meskipun demikian, penerapan CNN konvensional masih menghadapi sejumlah tantangan, seperti kebutuhan komputasi yang relatif tinggi, ketidakseimbangan distribusi data antara kelas benign dan malignant, serta ketergantungan terhadap kualitas dan ukuran dataset. Tantangan ini membuka peluang bagi pengembangan arsitektur CNN yang lebih efisien. Salah satu arsitektur yang dirancang untuk mengatasi keterbatasan tersebut adalah MobileNetV2, yang mengadopsi konsep depthwise separable convolution untuk menekan kompleksitas komputasi tanpa mengorbankan kemampuan ekstraksi fitur.

Berdasarkan uraian tersebut, penelitian ini berfokus pada implementasi dan evaluasi arsitektur MobileNetV2 dalam klasifikasi kanker kulit menggunakan citra dermatoskopi digital medis. Penelitian ini diharapkan dapat menghasilkan model yang efisien secara komputasi dan memiliki performa klasifikasi yang baik, sehingga berpotensi mendukung pengembangan sistem diagnosis kanker kulit berbasis citra medis pada berbagai skenario layanan kesehatan digital. Selain itu, penelitian ini juga memberikan gambaran mengenai tantangan dalam mendeteksi kelas malignant, sehingga dapat menjadi dasar untuk pengembangan metode yang lebih sensitif terhadap kasus kanker kulit.

## 2. Metode Penelitian

Penelitian ini mengadopsi pendekatan eksperimental berbasis pembelajaran mendalam (deep learning) untuk mengembangkan dan mengevaluasi model klasifikasi kanker kulit menggunakan arsitektur Convolutional Neural Network (CNN) MobileNetV2 pada citra dermatoskopi digital. Proses penelitian dirancang secara sistematis melalui serangkaian tahapan yang saling terkait, meliputi akuisisi data, transformasi dan augmentasi citra, stratifikasi dataset, perancangan arsitektur model, pelatihan bertahap, serta penilaian performa yang holistik. Seluruh implementasi dilaksanakan dalam lingkungan pemrograman Python 3.10+ dengan memanfaatkan ekosistem TensorFlow 2.x dan Keras, dijalankan pada platform Google Colab yang terintegrasi dengan penyimpanan Google Drive. Reprodusibilitas dijamin melalui penetapan seed acak global (SEED=42) pada TensorFlow dan NumPy.



Gambar 1. Tahapan Penelitian

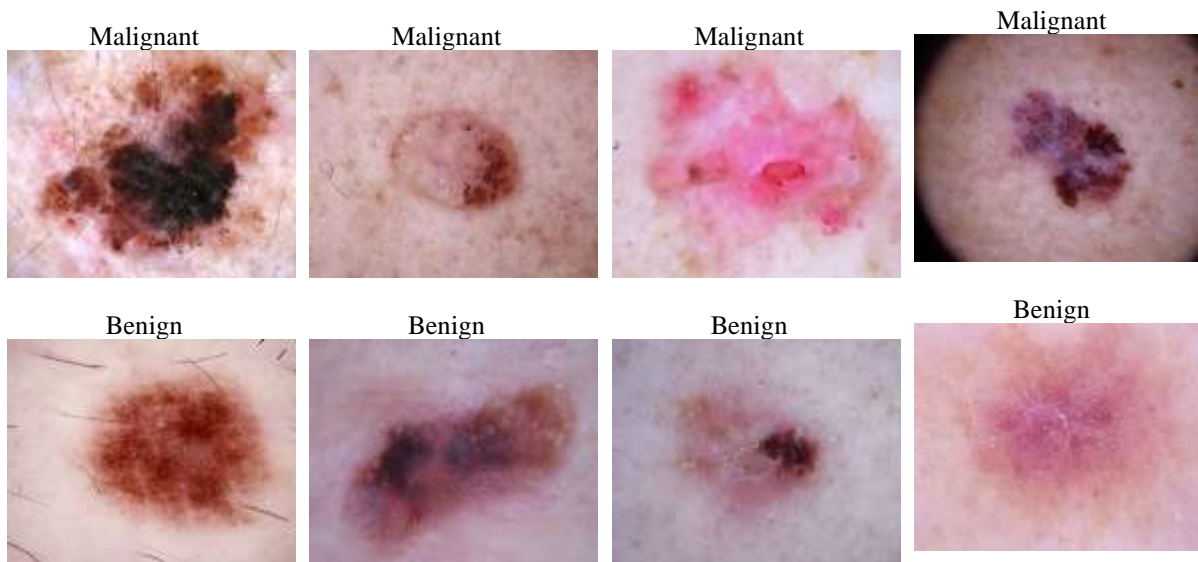
### Pengumpulan Data

Penelitian ini dimulai dengan pengumpulan dataset citra dermatoskopi penyakit kulit yang diperoleh dari sumber terbuka terpercaya, yaitu dataset HAM10000 (Human Against Machine with 10000 training images) yang tersedia melalui platform Kaggle. Dataset ini terdiri atas 10.015 citra dermatoskopi beresolusi tinggi ( $600 \times 450$  piksel) yang dikumpulkan dari berbagai sumber klinis dan telah diverifikasi secara medis oleh pakar dermatologi. Setiap citra disertai metadata klinis dalam format CSV (HAM10000\_metadata.csv) yang memuat informasi diagnosis, image\_id, serta lesion\_id.

Dalam penelitian ini, label diagnosis dikelompokkan menjadi dua kelas utama (klasifikasi biner) untuk mendukung deteksi dini lesi ganas, yaitu:

- Malignant (kelas positif,  $y = 1$ ): mel (melanoma), bcc (basal cell carcinoma), dan akiec (actinic keratoses dan karsinoma intraepitelial)  $\rightarrow$  sebanyak 1.954 citra
- Benign (kelas negatif,  $y = 0$ ): nv (melanocytic nevi), bkl (benign keratosis-like lesions), df (dermatofibroma), dan vasc (vascular lesions)  $\rightarrow$  sebanyak 8.061 citra

Pemetaan label biner ini dipilih untuk menekankan kemampuan model dalam mengidentifikasi lesi berpotensi ganas secara lebih fokus, sesuai dengan tujuan penelitian. Pemetaan tersebut juga disimpan dalam berkas teks untuk menjaga konsistensi interpretasi label di seluruh tahapan eksperimen.



Gambar 2 Sampel Citra yang digunakan

Untuk memberikan gambaran dataset yang digunakan, Gambar 1 menampilkan sampel citra dari kedua kelas (benign dan malignant) yang diambil secara acak dari subset pengujian, sehingga dapat merepresentasikan variasi lesi pada masing-masing kategori.

Selanjutnya, pembagian dataset dilakukan secara stratified berdasarkan lesion\_id guna mencegah kebocoran data (data leakage) serta memastikan citra dari lesi yang sama tidak muncul pada subset yang berbeda. Proporsi

pembagian yang digunakan adalah sekitar 70% data pelatihan, 15% data validasi, dan 15% data pengujian, dengan distribusi sebagai berikut:

Tabel 1. Deskripsi Dataset

	Pelatihan	Validasi	Pengujian
Malignant	1361	307	286
Benign	5.700	1162	1199
Total	7061	1469	1485

Pembagian dataset ini dilakukan agar model tidak hanya mampu mempelajari pola pada data pelatihan, tetapi juga dapat diuji secara objektif pada data yang belum pernah dilihat sebelumnya, sehingga kemampuan generalisasi model dapat dinilai secara lebih valid. [10]

### Pre-Processing

Tahap preprocessing merupakan langkah penting dalam machine learning untuk mempersiapkan citra agar dapat digunakan secara efektif, sehingga model mendapatkan hasil klasifikasi yang optimal[11]. Pada penelitian ini, preprocessing dilakukan untuk memastikan seluruh citra dermatoskopi memiliki format input yang seragam dan sesuai dengan kebutuhan arsitektur MobileNetV2. Proses ini mencakup beberapa tahapan utama, yaitu penataan data, pengubahan ukuran citra (image resizing), normalisasi nilai piksel, serta augmentasi data (data augmentation) guna meningkatkan variasi citra pada data latih [12].

Setiap citra terlebih dahulu disesuaikan ukurannya (resize) mengikuti spesifikasi input model, kemudian nilai piksel dinormalisasi agar berada pada skala yang lebih konsisten. Normalisasi ini dilakukan untuk membantu mempercepat proses konvergensi selama pelatihan sekaligus menjaga kestabilan pembaruan bobot, sehingga pembelajaran tidak mudah terganggu oleh perbedaan intensitas citra yang terlalu ekstrem. Selanjutnya, dataset disusun dalam struktur direktori berbasis kelas (benign dan malignant) agar proses pemanggilan data dapat berlangsung lebih teratur dan otomatis melalui mekanisme data generator.

Untuk meningkatkan kemampuan generalisasi model terhadap variasi kondisi citra dermatoskopi, penelitian ini menerapkan augmentasi menggunakan ImageDataGenerator. Teknik augmentasi yang digunakan meliputi rotasi, pergeseran (shift), pembesaran (zoom), serta pembalikan citra (flipping). Augmentasi hanya diterapkan pada data pelatihan, sedangkan data validasi dan pengujian dibiarkan dalam kondisi asli agar evaluasi performa tetap mencerminkan kemampuan model pada data nyata secara objektif.

Pelatihan model dilakukan melalui strategi dua tahap, yaitu head training dan fine-tuning, untuk memaksimalkan pemanfaatan transfer learning. Pada tahap awal (head training), lapisan klasifikasi tambahan dilatih menggunakan batch size = 32, learning rate =  $3 \times 10^{-4}$ , dan maksimum 10 epoch. Tahap ini bertujuan membangun kemampuan dasar model dalam membedakan pola visual benign dan malignant secara cepat. Selanjutnya, pada tahap fine-tuning, sebagian lapisan MobileNetV2 dilatih ulang dengan learning rate =  $1 \times 10^{-5}$  dan maksimum 20 epoch, sehingga penyesuaian bobot berlangsung lebih halus tanpa merusak fitur umum yang telah dipelajari sebelumnya dari ImageNet. Untuk mengurangi risiko overfitting, diterapkan early stopping yang menghentikan pelatihan secara otomatis ketika performa validasi tidak menunjukkan peningkatan dalam beberapa epoch berturut-turut.

### Perancangan Model



Gambar 3. Perancangan Model MobileNetV2

Tahap perancangan model dilakukan untuk membangun sistem klasifikasi kanker kulit berbasis citra dermatoskopi dengan memanfaatkan arsitektur MobileNetV2 sebagai backbone utama. MobileNet adalah arsitektur CNN ringan

yang dirancang untuk perangkat mobile dan embedded. Dengan menggunakan depthwise separable convolution, MobileNet mengurangi beban komputasi tanpa mengorbankan akurasi secara signifikan. Versi lanjutannya, seperti MobileNetV3, mendukung penggunaan di aplikasi real-time[13]. Pada penelitian ini, model dilatih menggunakan dataset yang telah dibagi menjadi tiga subset, yaitu data pelatihan (training), data validasi (validation), dan data pengujian (testing). Pada penelitian ini, model dilatih menggunakan dataset yang telah dibagi menjadi tiga subset, yaitu data pelatihan (training), data validasi (validation), dan data pengujian (testing). Pembagian ini dilakukan agar model tidak hanya mampu mempelajari pola pada data latih, tetapi juga dapat dievaluasi secara objektif menggunakan data yang belum pernah dilihat sebelumnya, sehingga kemampuan generalisasi model dapat diukur secara lebih valid.

MobileNetV2 diimplementasikan menggunakan pendekatan transfer learning, yaitu memanfaatkan bobot awal (pre-trained weights) dari dataset ImageNet agar model memiliki kemampuan dasar dalam mengenali pola visual umum. Strategi ini dipilih untuk mempercepat proses pembelajaran sekaligus meningkatkan performa, terutama ketika dataset penelitian memiliki keterbatasan ukuran serta distribusi kelas yang tidak sepenuhnya seimbang.

Arsitektur model dibangun dengan menjadikan MobileNetV2 sebagai ekstraktor fitur (feature extractor), kemudian ditambahkan beberapa lapisan klasifikasi (classification head) pada bagian akhir untuk menyesuaikan model dengan tugas klasifikasi biner benign dan malignant. Lapisan tambahan tersebut meliputi Global Average Pooling untuk merangkum fitur spasial menjadi representasi vektor yang lebih ringkas, kemudian diikuti oleh lapisan Dense sebagai pengambil keputusan akhir. Untuk mengurangi risiko overfitting, diterapkan Dropout sebagai mekanisme regularisasi agar model tidak terlalu bergantung pada pola tertentu pada data pelatihan.

### Pelatihan Model

Tahap pelatihan model dilakukan untuk membentuk kemampuan MobileNetV2 dalam mengklasifikasikan citra lesi kulit ke dalam dua kelas, yaitu benign dan malignant. Pelatihan menggunakan pendekatan transfer learning, dengan memanfaatkan bobot awal (pre-trained weights) dari dataset ImageNet. Strategi ini dipilih karena dapat mempercepat proses pembelajaran, meningkatkan stabilitas pelatihan, serta membantu model memperoleh representasi fitur visual yang lebih baik.

Proses pelatihan dilaksanakan dalam dua tahap, yaitu head training dan fine-tuning. Pada tahap head training, seluruh lapisan dasar MobileNetV2 dibekukan (frozen), sehingga pelatihan hanya difokuskan pada lapisan klasifikasi tambahan (classification head). Tahap ini menggunakan batch size 32, learning rate  $3 \times 10^{-4}$ , dan maksimum 10 epoch.

Selanjutnya, pada tahap fine-tuning, sebagian lapisan MobileNetV2 dibuka kembali (unfreeze) mulai dari layer ke-160, agar model dapat menyesuaikan fitur dengan karakteristik citra dermatoskopi yang digunakan. Pada tahap ini, nilai learning rate diturunkan menjadi  $1 \times 10^{-5}$  dengan maksimum 20 epoch, sehingga pembaruan bobot berlangsung lebih halus dan tidak mengganggu fitur dasar yang telah dipelajari sebelumnya.

Selama proses pelatihan, performa model dipantau menggunakan data validasi. Untuk mengurangi risiko overfitting, diterapkan mekanisme early stopping yang akan menghentikan pelatihan apabila metrik validasi tidak menunjukkan peningkatan dalam beberapa epoch berturut-turut.

Parameter	Nilai	Keterangan
Optimizer	Adam	Algoritma optimasi yang digunakan untuk memperbarui bobot jaringan secara adaptif
Learning Rate (Tahap 1 – Head Training)	0.0003 ( $3 \times 10^{-4}$ )	Kecepatan pembelajaran pada tahap pelatihan lapisan klasifikasi awal
Learning Rate (Tahap 2 – Fine-tuning)	0.00001 ( $1 \times 10^{-5}$ )	Kecepatan pembelajaran yang sangat kecil pada tahap penyesuaian lapisan dasar MobileNetV2
Batch Size	32	Jumlah sampel data yang diproses dalam satu iterasi forward-backward
Epoch (Tahap 1)	Maksimum 10	Jumlah iterasi maksimum pada tahap head training (dengan early stopping)
Epoch (Tahap 2)	Maksimum 20	Jumlah iterasi maksimum pada tahap fine-tuning (dengan early stopping)
Fine-tune dari Layer	Mulai dari layer ke-160	Lapisan awal MobileNetV2 (0–159) tetap dibekukan, sisanya di-unfreeze
Loss Function	Binary Crossentropy	Fungsi kerugian yang sesuai untuk klasifikasi biner (malignant vs benign)

Metrik Utama	Binary Accuracy, AUC	Digunakan untuk memantau performa selama pelatihan dan validasi
Class Weight	Benign (0): ~0.619 Malignant (1): ~2.594	Pembobotan untuk menangani ketidakseimbangan kelas (balanced weighting)
Dropout Rate	0.3	Diterapkan sebelum lapisan output untuk mencegah overfitting
Early Stopping	Patience = 6 (monitor: val_auc)	Penghentian otomatis jika tidak ada peningkatan AUC validasi selama 6 epoch

### Evaluasi Model

Tahap evaluasi dilakukan untuk menilai sejauh mana model MobileNetV2 mampu mengklasifikasikan citra lesi kulit ke dalam dua kelas, yaitu benign dan malignant, secara akurat dan konsisten. Evaluasi dilakukan setelah proses pelatihan selesai, dengan menggunakan data pengujian (testing) yang tidak pernah dilibatkan selama proses training maupun validasi. Pendekatan ini penting untuk memastikan bahwa hasil evaluasi benar-benar mencerminkan kemampuan model dalam menghadapi data baru (generalization).

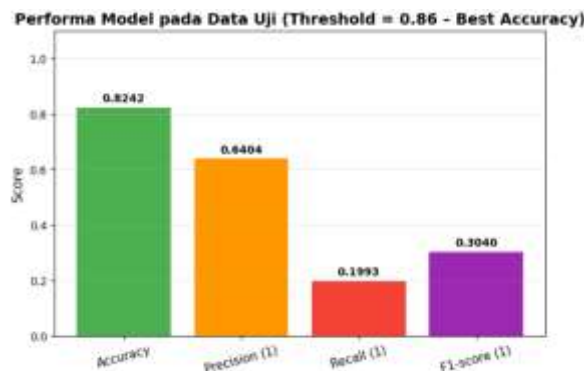
Pada penelitian ini, performa model dianalisis menggunakan beberapa metrik utama, yaitu accuracy, precision, recall, F1-score, serta AUC (Area Under Curve). Pemilihan metrik tersebut bertujuan untuk memberikan gambaran performa secara lebih menyeluruh, terutama karena dataset memiliki kecenderungan ketidakseimbangan jumlah data antara kelas benign dan malignant. Dalam konteks deteksi kanker kulit, penggunaan metrik seperti recall dan AUC menjadi penting karena model tidak hanya dituntut akurat secara umum, tetapi juga mampu mengenali kasus malignant dengan baik.

Untuk memperoleh evaluasi yang lebih mendalam, penelitian ini menerapkan beberapa pendekatan pengukuran. Pertama, dilakukan evaluasi berdasarkan masing-masing kelas (benign dan malignant) untuk mengetahui performa model secara spesifik pada setiap kategori. Kedua, digunakan confusion matrix untuk menggambarkan distribusi prediksi benar dan salah, sekaligus melihat pola kesalahan model pada kedua kelas. Ketiga, digunakan classification report yang memuat nilai precision, recall, dan F1-score per kelas, sehingga memberikan gambaran yang lebih komprehensif mengenai performa model pada seluruh kategori klasifikasi [14].

### 3. Hasil

Penelitian ini menghasilkan analisis performa model CNN MobileNetV2 dalam melakukan klasifikasi citra dermatoskopi menjadi dua kelas utama, yaitu benign dan malignant. Implementasi penelitian dilakukan secara bertahap, dimulai dari pengumpulan dataset HAM10000, dilanjutkan dengan pemetaan label menjadi klasifikasi biner, proses pembagian data secara stratified berdasarkan lesion\_id, serta tahap preprocessing yang mencakup resize, normalisasi, dan augmentasi citra. Setelah itu, model dibangun menggunakan pendekatan transfer learning dengan bobot awal dari ImageNet, kemudian dilatih melalui dua tahap, yaitu head training dan fine-tuning. Tahap akhir adalah evaluasi model menggunakan data pengujian untuk menilai kemampuan generalisasi model melalui metrik seperti accuracy, precision, recall, F1-score, confusion matrix, dan AUC.

#### Hasil Pengujian

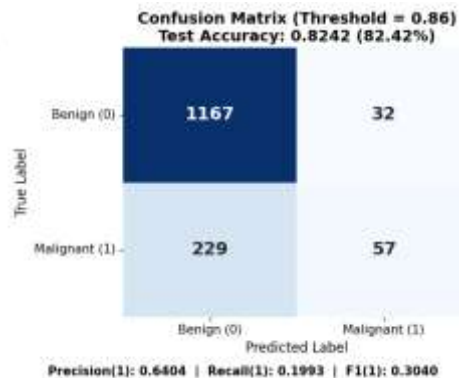


Gambar 4. Hasil pengujian

Pada tahap pengujian, model CNN berbasis MobileNetV2 dievaluasi menggunakan data uji untuk mengukur kemampuan generalisasi dalam mengklasifikasikan citra dermatoskopi kanker kulit ke dalam dua kelas, yaitu benign (0) dan malignant (1). Pengujian dilakukan dengan menerapkan threshold 0,86 yang dipilih sebagai threshold terbaik karena menghasilkan akurasi tertinggi. Hasil evaluasi menunjukkan bahwa model memperoleh nilai accuracy sebesar 0,8242, yang mengindikasikan bahwa sebagian besar citra pada data uji berhasil diklasifikasikan dengan benar. Selain itu, model menghasilkan nilai precision pada kelas malignant sebesar 0,6404, yang menandakan bahwa prediksi malignant yang dihasilkan model cukup tepat. Namun, nilai recall kelas malignant sebesar 0,1993 menunjukkan bahwa kemampuan model dalam mendeteksi seluruh kasus malignant masih terbatas. Kondisi tersebut turut tercermin pada nilai F1-score sebesar 0,3040, yang memperlihatkan bahwa keseimbangan antara precision dan recall pada kelas malignant belum optimal, meskipun akurasi keseluruhan model tergolong baik.

### Analisis Confusion Matrix

Evaluasi juga dilakukan menggunakan confusion matrix untuk melihat detail prediksi benar dan salah pada masing-masing kelas. Confusion matrix yang diperoleh adalah sebagai berikut:

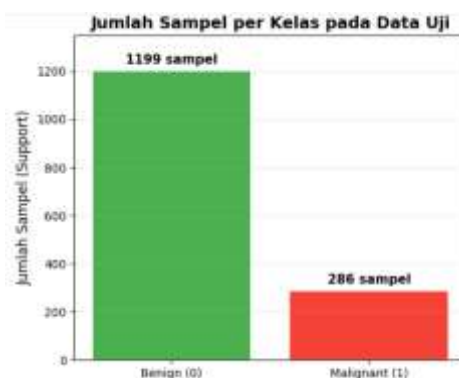


Gambar 5. Confusion matrix

Berdasarkan confusion matrix yang diperoleh, model mampu mengklasifikasikan 1167 citra benign secara tepat sebagai benign (True Negative). Selain itu, terdapat 32 citra benign yang keliru diprediksi sebagai malignant (False Positive). Pada kelas malignant, model menghasilkan 57 prediksi benar (True Positive), namun masih terdapat 229 citra malignant yang salah diklasifikasikan sebagai benign (False Negative).

Hasil ini menunjukkan bahwa model memiliki performa yang sangat baik dalam mengenali kelas benign, tetapi masih belum optimal dalam mendeteksi kelas malignant. Hal tersebut terlihat dari tingginya jumlah false negative, yang mengindikasikan bahwa sebagian besar kasus malignant belum berhasil teridentifikasi dengan baik. Dalam konteks medis, kondisi ini menjadi perhatian penting karena false negative dapat menyebabkan kasus kanker kulit tidak terdeteksi dan berpotensi menghambat penanganan lebih lanjut.

### Pembahasan Classification Report



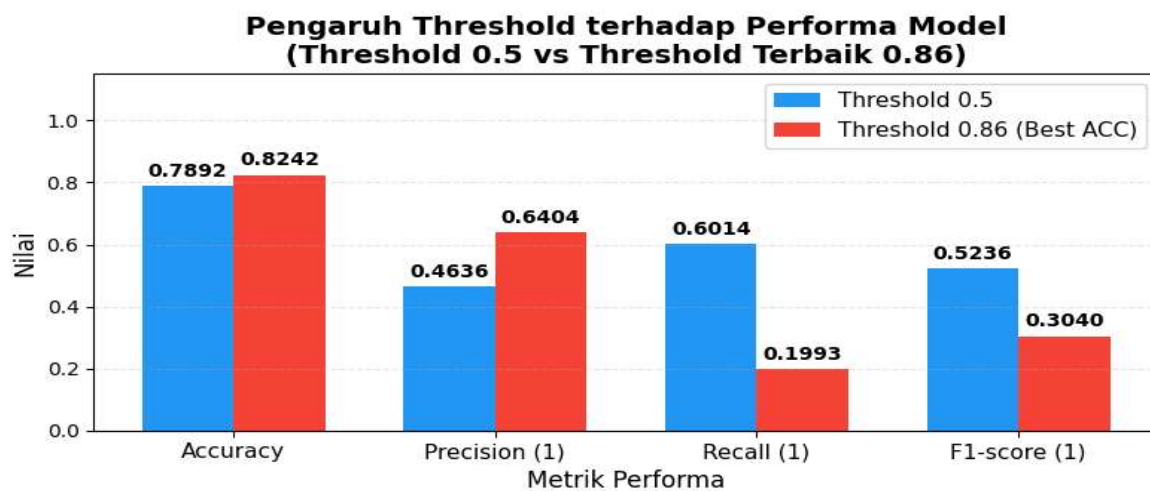
Gambar 6. Classification report

Berdasarkan classification report, performa model pada masing-masing kelas menunjukkan perbedaan yang cukup jelas. Pada kelas benign (0), model memperoleh nilai precision sebesar 0,8360, recall sebesar 0,9733, dan F1-score

sebesar 0,8994 dengan jumlah data (support) sebanyak 1199. Nilai recall yang sangat tinggi mengindikasikan bahwa model mampu mengenali hampir seluruh citra benign dengan baik. Hal ini diperkuat oleh nilai F1-score yang tinggi, sehingga dapat disimpulkan bahwa performa model pada kelas benign tergolong sangat optimal.

Sementara itu, pada kelas malignant (1) model menghasilkan precision sebesar 0,6404, recall sebesar 0,1993, dan F1-score sebesar 0,3040 dengan jumlah data (support) sebanyak 286. Nilai precision yang relatif baik menunjukkan bahwa ketika model memprediksi suatu citra sebagai malignant, prediksi tersebut cukup sering benar. Namun, nilai recall yang rendah menandakan bahwa model hanya mampu mendeteksi sekitar 19,93% dari keseluruhan citra malignant, sehingga sebagian besar kasus malignant masih keliru diklasifikasikan sebagai benign. Dalam konteks medis, kondisi ini perlu menjadi perhatian karena rendahnya recall pada kelas malignant berkaitan langsung dengan tingginya risiko false negative, yang dapat menyebabkan kasus kanker kulit tidak teridentifikasi secara tepat.

### Pengaruh Threshold Terhadap Performa Model



Gambar 7. Pengaruh threshold terhadap performa model

Penggunaan threshold terbaik sebesar 0,86 terbukti mampu meningkatkan nilai akurasi model menjadi 82,42%. Namun, penerapan threshold yang relatif tinggi membuat model menjadi lebih selektif atau “ketat” dalam memberikan prediksi malignant, sehingga jumlah prediksi malignant yang dihasilkan cenderung lebih sedikit. Kondisi ini berdampak pada penurunan kesalahan false positive, yang terlihat dari nilai FP hanya 32. Akan tetapi, konsekuensi lainnya adalah meningkatnya jumlah false negative, yaitu mencapai 229, yang menunjukkan bahwa masih banyak kasus malignant yang justru diprediksi sebagai benign [15].

Temuan ini mengindikasikan bahwa penentuan threshold berdasarkan akurasi semata belum tentu menjadi pilihan paling optimal dalam konteks klasifikasi medis. Hal tersebut karena akurasi yang tinggi dapat tercapai dengan lebih mengutamakan prediksi pada kelas mayoritas (benign), sementara kemampuan model dalam mendeteksi kelas malignant dapat menurun secara signifikan. Oleh karena itu, pemilihan threshold sebaiknya juga mempertimbangkan metrik lain seperti recall pada kelas malignant untuk meminimalkan risiko kesalahan deteksi kasus kanker kulit.

### Analisis Performa Rata-Rata (Macro Avg Dan Weighted Avg)



Gambar 8. Performa Rata - Rata

Hasil evaluasi juga memperlihatkan bahwa model memperoleh macro average F1-score sebesar 0,6017, weighted average F1-score sebesar 0,7847, serta accuracy sebesar 0,8242. Perbedaan yang cukup signifikan antara nilai macro dan weighted average menunjukkan adanya indikasi ketidakseimbangan jumlah data (class imbalance) pada dataset pengujian. Hal ini terlihat dari distribusi data yang didominasi oleh kelas benign sebanyak 1199 data, sedangkan kelas malignant hanya 286 data.

Kondisi tersebut menyebabkan model cenderung lebih optimal dalam memprediksi kelas benign sebagai kelas mayoritas, sehingga nilai akurasi keseluruhan tampak tinggi. Namun, performa pada kelas malignant tetap lebih rendah karena kontribusinya terhadap perhitungan metrik weighted average lebih kecil. Dengan demikian, hasil ini menegaskan bahwa evaluasi model pada kasus medis tidak cukup hanya mengandalkan akurasi, melainkan perlu mempertimbangkan metrik lain yang lebih representatif terhadap kelas minoritas, seperti recall dan F1-score pada kelas malignant.

#### 4. Kesimpulan

Berdasarkan hasil pengujian, model CNN MobileNetV2 dengan threshold terbaik sebesar 0,86 mampu mencapai akurasi 82,42% pada data uji. Hasil ini menunjukkan bahwa model telah mampu mempelajari pola visual pada citra dermatoskopi dan melakukan klasifikasi secara cukup konsisten, terutama pada kelas benign. Hal tersebut terlihat dari tingginya nilai recall pada kelas benign serta jumlah prediksi benar yang dominan pada confusion matrix, yang menandakan bahwa model memiliki kemampuan yang kuat dalam mengenali karakteristik lesi non-kanker. Namun demikian, performa model pada kelas malignant masih belum optimal. Nilai recall yang rendah pada kelas malignant menunjukkan bahwa model hanya mampu mendeteksi sebagian kecil kasus kanker kulit, sementara sisanya masih salah diklasifikasikan sebagai benign. Kondisi ini juga diperkuat oleh tingginya jumlah false negative, yang berarti banyak kasus malignant tidak teridentifikasi oleh model. Dalam konteks medis, kesalahan false negative menjadi aspek yang sangat krusial karena berpotensi menyebabkan keterlambatan diagnosis atau pengambilan keputusan klinis yang kurang tepat. Selain itu, penggunaan threshold yang tinggi memang dapat meningkatkan akurasi keseluruhan dan menekan false positive, namun di sisi lain membuat model menjadi lebih selektif dalam memberikan label malignant sehingga kemampuan deteksi kasus kanker justru menurun. Dengan demikian, dapat disimpulkan bahwa model ini sudah cukup baik sebagai langkah awal dalam penerapan klasifikasi kanker kulit berbasis citra dermatoskopi, tetapi masih memerlukan peningkatan untuk menghasilkan sistem yang lebih andal dalam mendeteksi kelas malignant. Pengembangan lanjutan dapat dilakukan melalui beberapa pendekatan, seperti penyeimbangan jumlah data antar kelas (misalnya oversampling atau class balancing), penerapan class weight agar model lebih sensitif terhadap kelas minoritas, serta peningkatan variasi data melalui augmentasi citra yang lebih terarah. Selain itu, strategi evaluasi dan penentuan threshold juga sebaiknya tidak hanya berorientasi pada akurasi, tetapi lebih mempertimbangkan metrik yang relevan untuk kasus medis seperti recall kelas malignant atau pendekatan optimasi berbasis F1-score dan AUC, sehingga model dapat meminimalkan risiko kesalahan deteksi pada kasus kanker kulit.

#### Referensi

1. R. Yohannes and Muhammad Ezar Al Rivian, "Klasifikasi Jenis Kanker Kulit Menggunakan CNN-SVM," vol. 2, no. 2, pp. 133–144, 2022.
2. F. Bimantoro, "Sistem Pakar Diagnosa Penyakit Kulit pada Manusia dengan Metode Dempster Shafer," vol. 4, no. 2, pp. 129–138, 2020.
3. F. Zulfa et al., "Profil Kanker Kulit Melanoma dan Non-Melanoma berdasarkan Hasil Histopatologi Profile of Melanoma and Non-Melanoma Skin Cancer based on Histopathology Results," vol. 14, no. September, pp. 1749–1755, 2024.
4. D. P. Mawardi, M. Novita, N. D. Saputro, F. Teknik, and U. Pgri, "Deteksi Awal Klasifikasi Jenis Penyakit Kanker Kulit Dengan Algoritma Convolutional Neural Network ( CNN ) Berbasis Mobile Apps," vol. 3, no. 2, pp. 1–6, 2024.
5. M. Z. Eryad, K. N. Ramadhani, A. Arifianto, and L. Belakang, "Pengenalan Bentuk Tangan dengan Convolutional Neural Network ( CNN ) 2 Kajian Pustaka," vol. 7, no. 2, pp. 8212–8222, 2020.
6. A. Muhajirin, A. A. Hasibuan, A. Antoni, A. A. Nst, and N. R. Wahyuni, "Systematic Literature Review : Advancements in Skin Cancer Diagnosis Using Convolutional Neural Networks and Dermatoscopic Imaging," vol. 3, no. 2, pp. 33–46, 2024, doi: 10.56313/jictas.v1i1.3i2.
7. Findriyani and R. A. Saputra, "Klasifikasi Kanker Kulit Berdasarkan Data Citra Benign Dan Malignant Menggunakan Convolutional Neural Network," vol. 3, no. April, 2024.
8. F. Dartiko, R. J. Pradana, R. E. Sari, W. Syahputra, and W. K. Z. Oktoeberza, "Klasifikasi Kanker Kulit Berbasis CNN dengan Metode Hybrid Preprocessing," 2024.
9. F. Y. Permama, C. Sri, K. Aditya, and D. R. Chandranegara, "Segmentasi dan Klasifikasi Gambar Citra pada Kanker Kulit Menggunakan Metode Convolutional Neural Network ( CNN ) dengan Arsitektur ResNet-50," vol. 6, no. 4, pp. 391–404, 2024.
10. H. Luqman Hakim, Zamah Sari, "Klasifikasi Citra Pigmen Kanker Kulit Menggunakan Convolutional Neural Network," vol. 1, no. 10, pp. 379–385, 2021.

11. J. Z. Muhammad Eky Ramadhan, "Implementasi Deep Transfer Learning dan Explainable AI dalam Klasifikasi Kanker Kulit Implementation of Deep Transfer Learning and Explainable AI in Skin," vol. 14, pp. 2266–2279, 2025.
12. D. G. Shindy Maheswari, "Deteksi dini kanker kulit menggunakan cnn, dnn, dan efficientnet: pendekatan deep learning berbasis web 1) 1,2)," vol. 10, no. 2, pp. 932–944, 2025.
13. E. P. Alief Prima Gani, "Pengembangan Aplikasi Android Berbasis Machine Learning Untuk," vol. 9, no. 6, pp. 9551–9557, 2025.
14. H. Aryanda, L. S. Harahap, and D. A. Aptanta, "Analisis Performa CNN Berbasis MobileNetV2 pada Dataset Citra Dokumen Laporan Praktikum," vol. 2, pp. 115–128, 2025.
15. H. T. Nguyen, D. D. Nguyen, and T. T. Le, "Skin Lesion Classification Using MobileNetV2 and Transfer Learning," *Journal of Imaging*, vol. 7, no. 8, pp. 1–14, 2021, doi: 10.3390/jimaging7080123.