

Analisis Perbandingan *MobileNetV2* dan *ResNet50* Berbasis *Transfer Learning* untuk Klasifikasi Citra dan Implementasi *Text-to-Speech*

Mega Gloria ^{1,*}, Mujiono Sadikin ¹, Khairunnisa Fadhilla Ramdhanian ¹

¹ Informatika; Universitas Bhayangkara Jakarta Raya; Jl. Raya Perjuangan No.81 Margamulya, Kota Bekasi, Jawa Barat, Telp. (021) 7231948; e-mail: 202210715173@mhs.ubharajaya.ac.id, mujiono@dsn.ubharajaya.ac.id, khairunnisa.fadhilla@dsn.ubharajaya.ac.id

* Korespondensi: e-mail: 202210715173@mhs.ubharajaya.ac.id

Diterima: 8 Juli 2026; Review: 13 Agst 2026; Disetujui: 14 Agst 2026; Diterbitkan: 14 Agst 2026

Abstract

Visual impairment limits the ability of people with blindness to recognize surrounding objects, creating the need for artificial intelligence-based assistive technology. This study compared the performance of MobileNetV2 and ResNet50 based on Transfer Learning for image classification and implemented the best web-based model integrated with Text-to-Speech. The research employed the Cross-Industry Standard Process for Data Mining methodology using the CIFAR-100 dataset. The results showed that ResNet50 achieved the highest accuracy of 79.95%, while MobileNetV2 demonstrated better computational efficiency. ResNet50 was successfully implemented in a web application capable of recognizing objects from external images and delivering the classification results through speech output.

Keywords: *Transfer Learning, MobileNetV2, ResNet50, Image Classification, Text-to-Speech*

Abstrak

Gangguan penglihatan menyebabkan penyandang disabilitas netra mengalami kesulitan mengenali objek sehingga diperlukan teknologi asistif berbasis kecerdasan buatan. Penelitian ini bertujuan membandingkan performa *MobileNetV2* dan *ResNet50* berbasis *Transfer Learning* untuk klasifikasi citra serta mengimplementasikan model terbaik berbasis web yang terintegrasi dengan *Text-to-Speech*. Penelitian menggunakan metodologi *Cross-Industry Standard Process for Data Mining* dengan dataset CIFAR-100. Hasil penelitian menunjukkan bahwa *ResNet50* memperoleh akurasi tertinggi sebesar 79,95%, sedangkan *MobileNetV2* lebih unggul dalam efisiensi komputasi. *ResNet50* kemudian berhasil diimplementasikan pada aplikasi web yang mampu mengenali objek dari citra eksternal dan menyampaikan hasil klasifikasi dalam bentuk suara.

Kata kunci: *Transfer Learning, MobileNetV2, ResNet50, Klasifikasi Citra, Text-to-Speech*

1. Pendahuluan

Gangguan penglihatan masih menjadi salah satu permasalahan kesehatan global yang berdampak terhadap kemampuan seseorang dalam mengenali objek di lingkungan sekitarnya. Penyandang disabilitas netra umumnya bergantung pada indera pendengaran dan peraba untuk

memperoleh informasi visual sehingga diperlukan teknologi asistif yang mampu mengubah informasi visual menjadi bentuk yang lebih mudah diakses. Menurut World Health Organization, gangguan penglihatan memengaruhi jutaan orang di seluruh dunia dan menjadi tantangan penting dalam upaya meningkatkan kualitas hidup serta kemandirian penyandang disabilitas (WHO, 2026).

Perkembangan *Computer Vision* dan *Deep Learning* membuka peluang untuk membangun sistem pengenalan objek yang dapat diintegrasikan dengan teknologi *Google Text-to-Speech (gTTS)* sehingga hasil klasifikasi citra dapat disampaikan dalam bentuk suara secara otomatis. Teknologi *TTS* modern telah mengalami perkembangan signifikan dan mampu menghasilkan keluaran suara yang lebih alami sehingga banyak digunakan pada sistem aksesibilitas dan interaksi manusia-komputer (Tan et al., 2021).

Dalam bidang klasifikasi citra, pendekatan *Transfer Learning* menjadi salah satu metode yang banyak digunakan karena memungkinkan pemanfaatan bobot model yang telah dilatih sebelumnya sehingga proses pelatihan menjadi lebih efisien dan tetap menghasilkan performa yang baik pada dataset dengan jumlah data terbatas (Alzubaidi et al., 2021; Wijaya et al., 2021).

Di antara berbagai arsitektur *Convolutional Neural Network* (CNN), *MobileNetV2* dan *ResNet50* merupakan dua model yang memiliki karakteristik berbeda. *MobileNetV2* dirancang sebagai arsitektur ringan dengan efisiensi komputasi yang tinggi sehingga banyak digunakan pada aplikasi berbasis perangkat bergerak dan *embedded system* (Hermanto et al., 2024; Yong et al., 2023). Sebaliknya, *ResNet50* memiliki kemampuan ekstraksi fitur yang lebih kompleks sehingga sering menghasilkan performa klasifikasi yang lebih tinggi pada berbagai penelitian klasifikasi citra berbasis *transfer learning* (Berliani et al., 2023; Utami et al., 2025).

Beberapa penelitian terdahulu telah membandingkan performa kedua arsitektur tersebut pada berbagai kasus klasifikasi citra. Penelitian Khatri menunjukkan bahwa *ResNet50* menghasilkan tingkat akurasi yang lebih tinggi dibandingkan *MobileNetV2* pada klasifikasi citra medis, sedangkan penelitian Utami melaporkan bahwa *MobileNetV2* memiliki efisiensi komputasi yang lebih baik sehingga lebih sesuai untuk implementasi pada perangkat dengan spesifikasi terbatas. Hasil tersebut menunjukkan bahwa setiap arsitektur memiliki keunggulan masing-masing sehingga pemilihan model tidak dapat hanya didasarkan pada nilai akurasi.

Namun demikian, sebagian besar penelitian sebelumnya masih berfokus pada peningkatan performa klasifikasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Analisis terhadap aspek efisiensi komputasi seperti waktu pelatihan, waktu inferensi, jumlah parameter, dan ukuran model masih relatif terbatas. Selain itu, hasil klasifikasi umumnya berhenti pada tahap evaluasi model tanpa diimplementasikan ke dalam sistem yang dapat dimanfaatkan secara langsung sebagai teknologi asistif. Dengan demikian, masih terdapat research gap berupa belum adanya analisis yang menggabungkan evaluasi performa klasifikasi, efisiensi komputasi, serta implementasi hasil klasifikasi menjadi keluaran suara dalam satu penelitian yang terintegrasi.

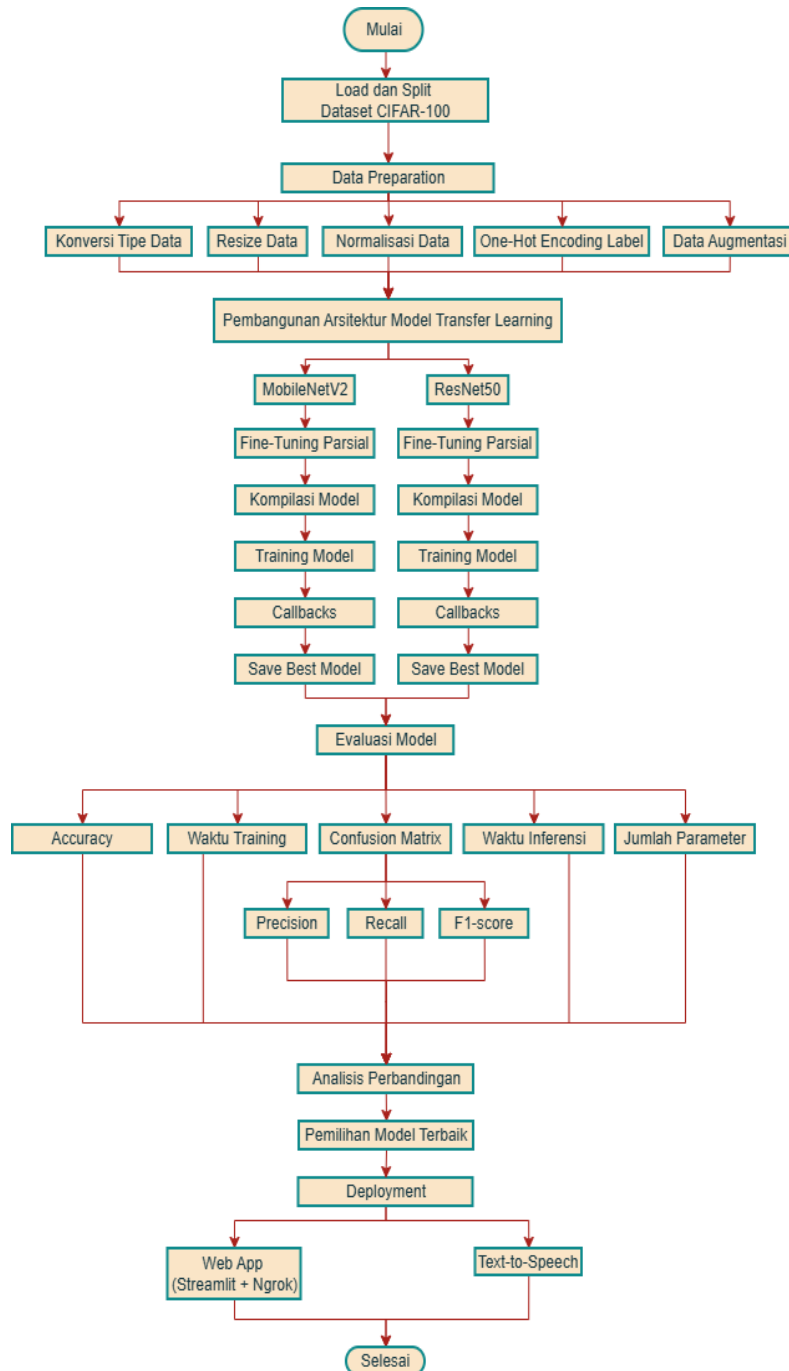
Berdasarkan kesenjangan tersebut, penelitian ini mengusulkan analisis perbandingan *MobileNetV2* dan *ResNet50* berbasis *Transfer Learning* menggunakan dataset CIFAR-100 dengan mengevaluasi performa klasifikasi berdasarkan *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*, serta efisiensi komputasi berdasarkan waktu pelatihan, waktu inferensi, jumlah parameter, dan ukuran model. Penelitian ini menerapkan tahapan pengolahan data secara sistematis mulai dari proses data preparation, pembangunan model *transfer learning*, evaluasi, hingga implementasi model terbaik ke dalam aplikasi berbasis web.

Berdasarkan hasil evaluasi tersebut dilakukan analisis perbandingan antara *MobileNetV2* dan *ResNet50* untuk menentukan model terbaik berdasarkan keseimbangan antara performa klasifikasi dan efisiensi komputasi. Model terbaik selanjutnya diimplementasikan ke dalam aplikasi berbasis web menggunakan Streamlit yang terintegrasi dengan *Google Text-to-Speech (gTTS)* sehingga hasil pengenalan objek dapat disampaikan dalam bentuk suara kepada pengguna.

Kebaruan penelitian ini terletak pada pendekatan evaluasi yang tidak hanya membandingkan performa klasifikasi kedua arsitektur, tetapi juga menganalisis keseimbangan antara akurasi dan efisiensi komputasi sebagai dasar pemilihan model terbaik, kemudian mengimplementasikan model tersebut pada prototipe berbasis web yang dilengkapi fitur *Google Text-to-Speech (gTTS)*. Melalui pendekatan tersebut, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan teknologi asistif berbasis kecerdasan buatan yang lebih efisien, responsif, dan mudah diimplementasikan.

2. Metode Penelitian

Penelitian ini menggunakan metodologi *Cross Industry Standard Process for Data Mining (CRISP-DM)* sebagai kerangka kerja penelitian karena menyediakan tahapan yang sistematis mulai dari pemahaman masalah hingga implementasi model, dapat dilihat pada Gambar 1. Tahapan CRISP-DM meliputi *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Alur penelitian diawali dengan proses *load* dan *split* dataset CIFAR-100, dilanjutkan dengan *data preparation*, pembangunan model *Transfer Learning*, evaluasi performa, analisis perbandingan, pemilihan model terbaik, hingga implementasi model ke dalam aplikasi berbasis web yang terintegrasi dengan *Google Text-to-Speech (gTTS)*.



Sumber: Hasil Penelitian (2026)

Gambar 1. Tahapan Algoritma

Pada gambar 1 menunjukan Dataset yang digunakan adalah CIFAR-100 yang dikembangkan oleh Canadian Institute for Advanced Research (Krizhevsky, 2009). Dataset ini terdiri atas 60.000 citra berwarna berukuran 32×32 piksel, yang terbagi menjadi 50.000 data pelatihan dan 10.000 data pengujian. Seluruh citra dikelompokkan ke dalam 100 kelas objek, dengan masing-masing kelas memiliki 500 citra latih dan 100 citra uji. Kompleksitas jumlah kelas yang tinggi menjadikan CIFAR-100 sebagai salah satu *benchmark* dataset yang banyak digunakan untuk mengevaluasi kemampuan generalisasi model klasifikasi citra.

Tahap *data preparation* diawali dengan memuat dataset CIFAR-100 kemudian melakukan konversi tipe data citra dari uint8 menjadi float32. Selanjutnya dilakukan *resize data* sesuai kebutuhan masukan model, normalisasi nilai piksel ke rentang 0–1, serta *one-hot encoding* terhadap *label* kelas agar sesuai dengan fungsi aktivasi *Softmax* pada lapisan keluaran. Untuk meningkatkan kemampuan generalisasi model, diterapkan teknik *data augmentation* berupa rotasi, translasi, *zoom*, dan *horizontal flip*. Tahapan tersebut menghasilkan variasi data pelatihan sehingga mampu meningkatkan kemampuan generalisasi model sekaligus mengurangi risiko *overfitting* selama proses pelatihan (Budi et al., 2024; Nurdyansyah et al., 2024).

Tahap *modeling* menggunakan pendekatan *Transfer Learning*, yaitu memanfaatkan bobot awal (*pre-trained weights*) hasil pelatihan pada dataset *ImageNet*, kemudian melakukan *fine-tuning parsial* pada lapisan klasifikasi agar sesuai dengan karakteristik dataset CIFAR-100. Pendekatan ini dipilih karena mampu mempercepat proses pelatihan sekaligus meningkatkan performa klasifikasi ketika jumlah data pelatihan terbatas (Alzubaidi et al., 2021; Wijaya et al., 2021).

Penelitian membandingkan dua arsitektur *Convolutional Neural Network*, yaitu *MobileNetV2* dan *ResNet50*. *MobileNetV2* dikenal sebagai arsitektur ringan yang dirancang untuk menghasilkan efisiensi komputasi tinggi dengan jumlah parameter yang relatif kecil sehingga sesuai untuk implementasi pada perangkat dengan sumber daya terbatas (Hermanto et al., 2024; Yong et al., 2023). Sebaliknya, *ResNet50* memiliki kedalaman jaringan yang lebih tinggi sehingga mampu mengekstraksi fitur visual yang lebih kompleks dan menghasilkan performa klasifikasi yang lebih baik pada berbagai kasus klasifikasi citra (Berliani et al., 2023; Utami et al., 2025). Konfigurasi utama kedua model ditunjukkan pada Tabel 1.

Tabel 1. Konfigurasi Arsitektur *MobileNetV2* dan *ResNet50*

Parameter	MobileNetV2	ResNet50
<i>Input Size</i>	32 × 32 × 3	32 × 32 × 3
<i>Pre-trained Weight</i>	<i>ImageNet</i>	<i>ImageNet</i>
<i>Include Top</i>	<i>False</i>	<i>False</i>
<i>Global Average Pooling</i>	Digunakan	Digunakan
<i>Dense Layer</i>	256 Unit (ReLU)	256 Unit (ReLU)
<i>Dropout</i>	0,5	0,5
<i>Output Layer</i>	Dense 100 (Softmax)	Dense 100 (Softmax)

Sumber: Hasil Penelitian (2026)

Pada Tabel 1 menjelaskan kedua model menggunakan konfigurasi lapisan klasifikasi yang identik sehingga proses perbandingan dapat dilakukan secara objektif. Setelah proses *fine-tuning parsial*, kedua model dikompilasi menggunakan parameter pelatihan yang sama, kemudian dilakukan proses pelatihan (*training*) dengan dukungan *callback* berupa *EarlyStopping*, *ModelCheckpoint*, dan *ReduceLROnPlateau*. Selama proses pelatihan, bobot model terbaik

disimpan (*save best model*) sehingga model dengan performa validasi terbaik digunakan pada tahap evaluasi.

Seluruh eksperimen dilakukan menggunakan *TensorFlow* dan *Keras* pada lingkungan *Google Colab* dengan dukungan GPU. Kedua model dilatih menggunakan parameter yang sama agar proses perbandingan berlangsung secara adil. Konfigurasi pelatihan yang digunakan disajikan pada Tabel 2.

Tabel 2. Parameter Pelatihan Model

Parameter	Nilai
<i>Optimizer</i>	<i>Adam</i>
<i>Learning Rate</i>	0,0001
<i>Batch Size</i>	32
Epoch Maksimum	50
<i>Loss Function</i>	<i>Categorical Crossentropy</i>
<i>Callback</i>	<i>EarlyStopping, ModelCheckpoint, ReduceLROnPlateau</i>

Sumber: Hasil Penelitian (2026)

Pada Tabel 2 menjelaskan kedua model dilatih menggunakan konfigurasi yang sama sehingga menghasilkan proses evaluasi yang lebih adil (*fair comparison*). *Optimizer Adam* dipilih karena memiliki kemampuan konvergensi yang cepat dan stabil pada proses pelatihan jaringan saraf dalam. Sementara itu, *EarlyStopping* digunakan untuk menghentikan pelatihan apabila tidak terjadi peningkatan performa validasi dalam beberapa *epoch* berturut-turut, *ModelCheckpoint* digunakan untuk menyimpan bobot model terbaik, dan *ReduceLROnPlateau* menyesuaikan nilai *learning rate* secara otomatis ketika proses pembelajaran mengalami stagnasi.

Tahap *evaluation* dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*. Selain mengevaluasi performa klasifikasi, penelitian juga mengukur efisiensi komputasi berdasarkan waktu pelatihan (*training time*), waktu inferensi (*inference time*), jumlah parameter, dan ukuran model. Seluruh hasil evaluasi kemudian dianalisis pada tahap analisis perbandingan untuk mengidentifikasi kelebihan dan kekurangan masing-masing arsitektur sehingga dapat ditentukan model yang memberikan keseimbangan terbaik antara performa klasifikasi dan efisiensi komputasi.

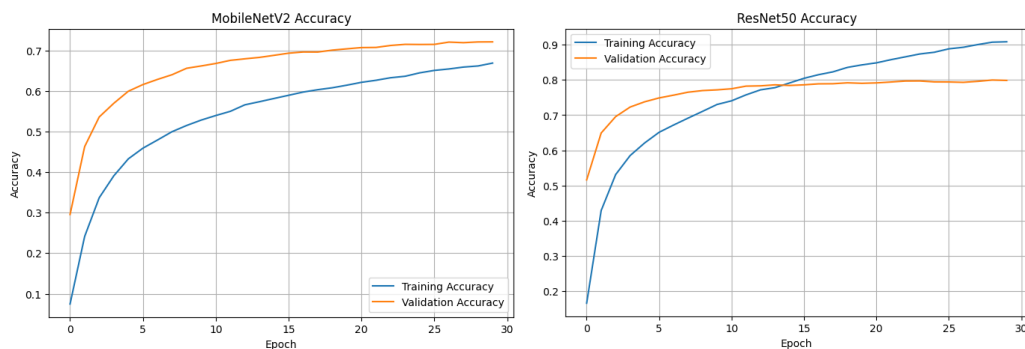
Tahap terakhir adalah *deployment*, yaitu mengimplementasikan model terbaik hasil analisis ke dalam aplikasi berbasis web menggunakan *Streamlit*. Aplikasi memungkinkan pengguna mengunggah citra eksternal, kemudian sistem melakukan klasifikasi objek secara otomatis menggunakan model hasil pelatihan. Hasil prediksi ditampilkan dalam bentuk nama objek beserta tingkat keyakinan (*confidence score*) dan selanjutnya dikonversi menjadi keluaran suara menggunakan *Google Text-to-Speech (gTTS)*. Implementasi ini berfungsi sebagai *proof of concept* bahwa model klasifikasi citra berbasis Transfer Learning dapat diintegrasikan ke dalam teknologi asistif berbasis kecerdasan buatan yang mampu menyampaikan informasi visual dalam bentuk audio bagi pengguna.

3. Hasil dan Pembahasan

Bagian ini menyajikan hasil penelitian berdasarkan tahapan pada metodologi *CRISP-DM*, meliputi hasil pelatihan model, evaluasi performa dan efisiensi komputasi, analisis perbandingan *MobileNetV2* dan *ResNet50*, pemilihan model terbaik, serta implementasi model terpilih pada *Web App* yang terintegrasi dengan fitur *Google Text-to-Speech (gTTS)*.

3.1. Hasil Pelatihan Model

Tahap pertama evaluasi dilakukan dengan mengamati proses pembelajaran kedua model selama pelatihan. Kinerja *MobileNetV2* dan *ResNet50* dianalisis berdasarkan perkembangan nilai *training accuracy*, *validation accuracy*, *training loss*, dan *validation loss* pada setiap *epoch*. Visualisasi perubahan nilai akurasi kedua model selama proses pelatihan ditunjukkan pada Gambar 2.

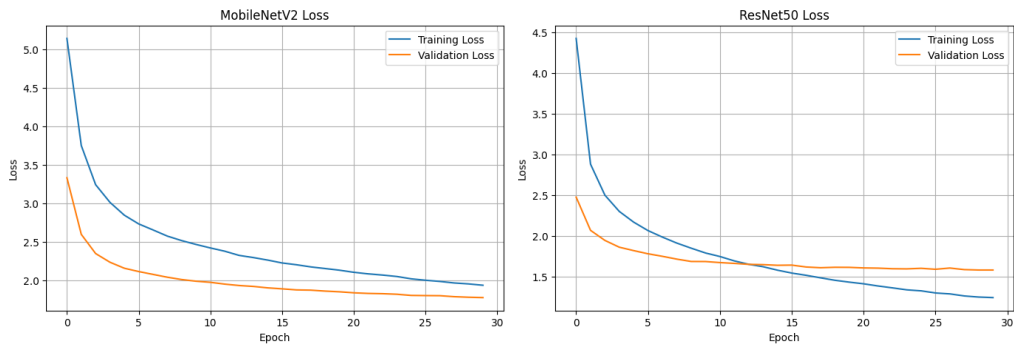


Sumber: Hasil Penelitian (2026)

Gambar 2. Grafik Accuracy *MobileNetV2* dan *ResNet50*

Pada Gambar 2 menunjukkan terlihat bahwa kedua model mengalami peningkatan akurasi secara bertahap seiring bertambahnya *epoch*. *MobileNetV2* memperlihatkan proses konvergensi yang relatif cepat, namun nilai akurasi validasi cenderung berhenti pada kisaran 72%. Sebaliknya, *ResNet50* menunjukkan peningkatan akurasi yang lebih stabil hingga mencapai hampir 80%. Kondisi tersebut menunjukkan bahwa arsitektur *ResNet50* memiliki kemampuan ekstraksi fitur yang lebih baik sehingga mampu menghasilkan kemampuan generalisasi yang lebih tinggi terhadap data uji.

Selain akurasi, kualitas proses pembelajaran juga diamati melalui perubahan nilai *loss* selama pelatihan. Nilai *loss* menggambarkan besar kesalahan prediksi model terhadap data pelatihan maupun data validasi.



Sumber: Hasil Penelitian (2026)

Gambar 3. Grafik Loss MobileNetV2 dan ResNet50

Pada Gambar 3 menunjukkan grafik nilai *training loss* dan *validation loss* pada kedua model terus mengalami penurunan hingga mencapai kondisi yang relatif stabil. Penurunan *loss* pada *ResNet50* berlangsung lebih konsisten dibandingkan *MobileNetV2*, sehingga menghasilkan nilai akhir yang lebih kecil. Hal tersebut mengindikasikan bahwa *ResNet50* mampu mempelajari representasi fitur yang lebih baik tanpa menunjukkan gejala *overfitting* yang signifikan. Untuk memberikan gambaran yang lebih ringkas mengenai proses pelatihan kedua model, hasil akhir pelatihan dirangkum pada Tabel 3.

Tabel 3. Ringkasan Hasil Pelatihan MobileNetV2 dan ResNet50

Parameter	MobileNetV2	ResNet50
Jumlah data latih	50.000 citra	50.000 citra
Jumlah data validasi	10.000 citra	10.000 citra
Jumlah epoch	30	30
Batch size	32	32
Training accuracy akhir	66,87%	90,81%
Validation accuracy terbaik	72,10%	79,95% (Epoch 29)
Training loss akhir	1,9322	1,2424
Validation loss terbaik	1,7717	1,5808 (Epoch 29)
Learning rate akhir	1×10^{-5}	2×10^{-6}
Waktu pelatihan	1.886,44 detik (± 31 menit 26 detik)	4.500,05 detik (± 75 menit)
Model terbaik yang disimpan	mobilenetv2_best.keras	resnet50_best.keras

Sumber: Hasil Penelitian (2026)

Pada Tabel 3 menjelaskan bahwa ResNet50 menghasilkan nilai *validation accuracy* yang lebih tinggi dibandingkan *MobileNetV2*. Sementara itu, *MobileNetV2* menunjukkan proses pelatihan yang lebih cepat karena jumlah parameter yang lebih sedikit. Hasil pelatihan menunjukkan bahwa kedua model memiliki karakteristik yang berbeda, sehingga diperlukan

evaluasi lebih lanjut terhadap performa klasifikasi dan efisiensi komputasi sebelum menentukan model terbaik.

3.2. Evaluasi Model

Evaluasi performa klasifikasi dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Keempat metrik tersebut digunakan untuk memberikan gambaran yang lebih komprehensif mengenai kemampuan model dalam mengklasifikasikan citra pada seluruh kelas dataset CIFAR-100. Hasil evaluasi kedua model disajikan pada Tabel 4.

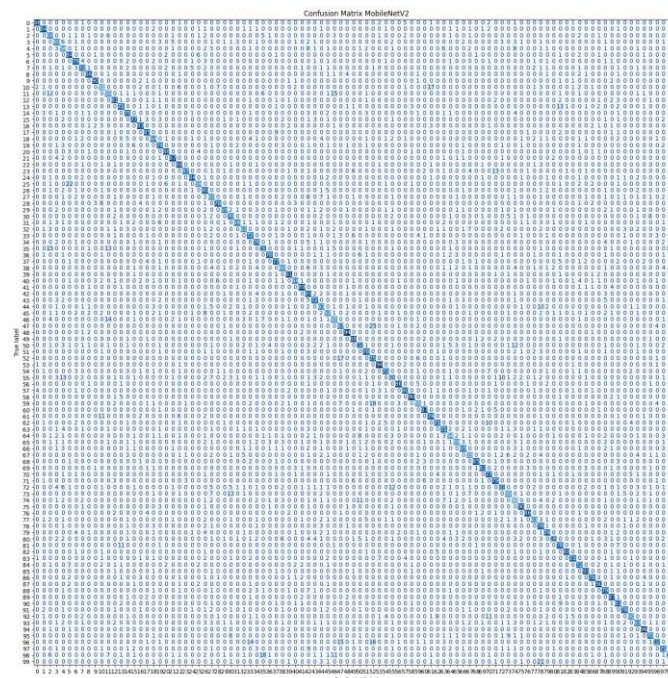
Tabel 4. Hasil Evaluasi *MobileNetV2* dan *ResNet50*

<i>Model</i>	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F1-score</i>
<i>MobileNetV2</i>	0,7210	0,7272	0,7210	0,7192
<i>ResNet50</i>	0,7995	0,8020	0,7995	0,7989

Sumber: Hasil Penelitian (2026)

Pada Tabel 4 menjelaskan *ResNet50* memperoleh *accuracy* sebesar 79,95%, lebih tinggi dibandingkan *MobileNetV2* sebesar 72,10%. Peningkatan performa tersebut juga tercermin pada nilai *precision*, *recall*, dan *F1-score*. Hasil tersebut menunjukkan bahwa *ResNet50* mampu mengenali pola visual yang lebih kompleks sehingga menghasilkan prediksi yang lebih akurat pada dataset multikelas.

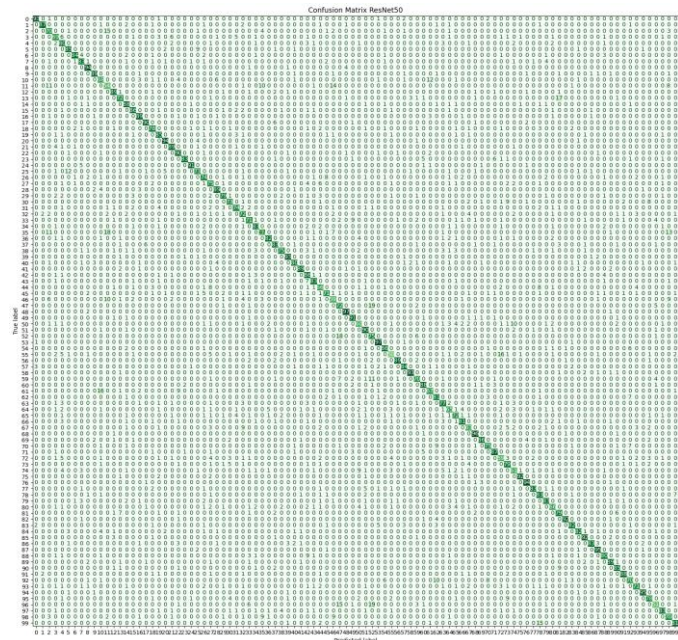
Selain menggunakan metrik numerik, evaluasi model juga dilakukan melalui *confusion matrix* untuk mengetahui distribusi prediksi benar dan kesalahan klasifikasi pada setiap kelas. Hasil *confusion matrix MobileNetV2* ditunjukkan pada Gambar 4.



Sumber: Hasil Penelitian (2026)

Gambar 4. *Confusion Matrix MobileNetV2*

Pada Gambar 4 menunjukkan *confusion matrix MobileNetV2* memperlihatkan dominasi nilai pada diagonal utama yang menunjukkan sebagian besar citra berhasil diklasifikasikan dengan benar. Namun demikian, masih terlihat sejumlah prediksi yang tersebar di luar diagonal, yang mengindikasikan adanya kesalahan klasifikasi pada beberapa kelas yang memiliki karakteristik visual serupa. Sebagai pembandingan, *confusion matrix ResNet50*.



Sumber: Hasil Penelitian (2026)

Gambar 5. *Confusion Matrix ResNet50*

Pada Gambar 5 menunjukkan dibandingkan *MobileNetV2*, *confusion matrix ResNet50* menunjukkan distribusi warna yang lebih dominan pada diagonal utama dan lebih sedikit penyebaran pada elemen di luar diagonal. Hal tersebut menunjukkan bahwa *ResNet50* mampu mengurangi jumlah kesalahan klasifikasi sehingga menghasilkan performa yang lebih baik pada sebagian besar kelas.

3.3. Analisis Perbandingan *MobileNetV2* dan *ResNet50*

Selain performa klasifikasi, penelitian ini juga membandingkan efisiensi komputasi kedua model. Parameter yang dievaluasi meliputi waktu pelatihan, waktu inferensi, jumlah parameter, dan ukuran model. Hasil perbandingan tersebut disajikan pada Tabel 5.

Tabel 5. Perbandingan *MobileNetV2* dan *ResNet50*

Model	Accuracy	Training Time (Second)	Inference Time (Second)	Parameter	Model Size
MobileNetV2	72.10%	1886.44	14.09	2,391,204	22.36
ResNet50	79.95%	4500.05	41.28	23,800,804	213.85

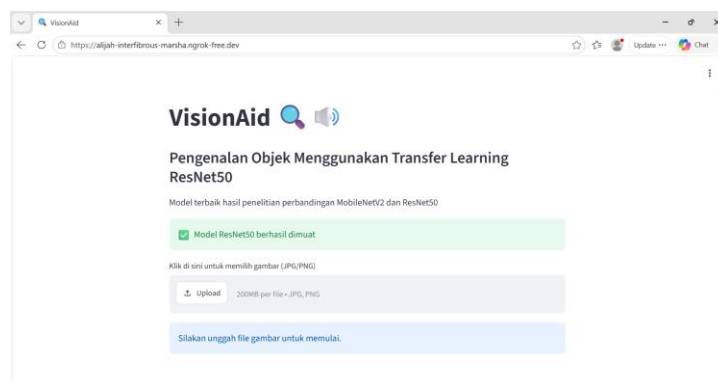
Sumber: Hasil Penelitian (2026)

Pada Tabel 5 menjelaskan bahwa *ResNet50* menghasilkan performa klasifikasi terbaik dengan *accuracy* sebesar 79,95%, tetapi membutuhkan waktu pelatihan sekitar 4.500 detik, waktu inferensi 4,128 ms per citra, serta memiliki lebih dari 23 juta parameter dengan ukuran model mencapai 213,85 MB. Sebaliknya, *MobileNetV2* memiliki *accuracy* yang lebih rendah, tetapi hanya memerlukan sekitar 2,39 juta parameter, ukuran model 22,36 MB, waktu pelatihan lebih singkat, serta waktu inferensi hanya 1,409 ms per citra. Temuan ini menunjukkan adanya *trade-off* yang jelas antara akurasi dan efisiensi komputasi. Apabila sistem diimplementasikan pada perangkat dengan sumber daya terbatas, *MobileNetV2* menjadi alternatif yang lebih ringan, sedangkan *ResNet50* lebih sesuai apabila prioritas utama adalah ketepatan klasifikasi.

Hasil evaluasi menampilkan bahwa *ResNet50* memperoleh *accuracy*, *precision*, *recall*, dan *F1-score* yang lebih tinggi dibandingkan *MobileNetV2* sehingga menunjukkan kemampuan klasifikasi yang lebih baik. Meskipun membutuhkan waktu pelatihan, waktu inferensi, jumlah parameter, dan ukuran model yang lebih besar, peningkatan performa klasifikasi yang dihasilkan menjadikan *ResNet50* dipilih sebagai model terbaik untuk tahap implementasi. Berdasarkan hasil analisis perbandingan, *ResNet50* dipilih sebagai model terbaik karena menghasilkan *accuracy* sebesar 79,95%, *precision* 80,20%, *recall* 79,95%, dan *F1-score* 79,89%, yang lebih tinggi dibandingkan *MobileNetV2*. Walaupun memiliki waktu pelatihan, waktu inferensi, jumlah parameter, dan ukuran model yang lebih besar, peningkatan performa klasifikasi dinilai lebih sesuai dengan tujuan penelitian sehingga model ini digunakan pada tahap *deployment*.

3.4. Implementasi

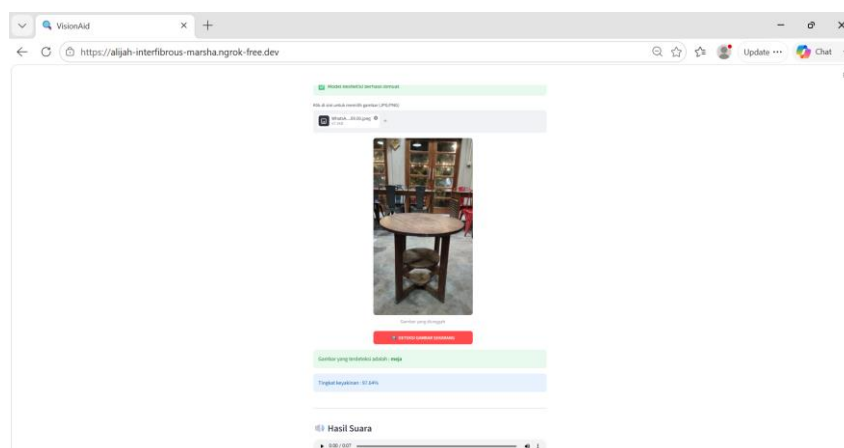
Tahap akhir penelitian adalah implementasi model terbaik (*ResNet50*) berbasis web menggunakan *Streamlit*. Selain menyediakan antarmuka untuk pengenalan objek, aplikasi juga diintegrasikan dengan *Google Text-to-Speech (gTTS)* sehingga hasil klasifikasi dapat disampaikan dalam bentuk suara. Web dikembangkan sebagai *proof of concept* untuk menunjukkan bahwa model hasil pelatihan dapat digunakan secara langsung dalam proses pengenalan objek. Tampilan utama aplikasi ditunjukkan pada Gambar 5.



Sumber: Hasil Penelitian (2026)

Gambar 6. Halaman Utama Web App

Pada Gambar 6 menunjukkan web menyediakan antarmuka sederhana yang memungkinkan pengguna mengunggah citra eksternal untuk dilakukan proses klasifikasi secara otomatis. Setelah gambar dipilih, sistem melakukan pra-pemrosesan dan inferensi menggunakan model *ResNet50*. Setelah proses inferensi selesai, aplikasi menampilkan hasil klasifikasi beserta tingkat keyakinan model dan mengubah hasil prediksi menjadi keluaran suara menggunakan *Google Text-to-Speech*. Hasil implementasi tersebut ditunjukkan pada Gambar 6.



Sumber: Hasil Penelitian (2026)

Gambar 7. Hasil Klasifikasi dan *Output Text-to-Speech*

Pada Gambar 7 menunjukkan hasil implementasi menunjukkan bahwa aplikasi berhasil mengenali objek pada citra eksternal serta menghasilkan keluaran suara yang sesuai dengan hasil prediksi. Integrasi antara model klasifikasi citra dan *Text-to-Speech* membuktikan bahwa sistem yang dikembangkan tidak hanya mampu melakukan pengenalan objek secara otomatis, tetapi juga berpotensi dimanfaatkan sebagai teknologi asistif untuk membantu penyandang disabilitas netra memperoleh informasi visual dalam bentuk audio. Hasil implementasi menunjukkan bahwa model terbaik yang diperoleh melalui proses evaluasi berhasil diterapkan pada *Web App* dan mampu menghasilkan keluaran *Text-to-Speech* sesuai alur *deployment* yang dirancang pada penelitian ini.

4. Kesimpulan

Penelitian ini berhasil menerapkan pendekatan *Transfer Learning* menggunakan arsitektur *MobileNetV2* dan *ResNet50* pada dataset CIFAR-100 melalui tahapan *data preparation*, pembangunan model, pelatihan, evaluasi, analisis perbandingan, pemilihan model terbaik, hingga implementasi (*deployment*). Berdasarkan hasil evaluasi, *ResNet50* menunjukkan performa klasifikasi yang lebih unggul dibandingkan *MobileNetV2*, ditunjukkan oleh nilai *accuracy*, *precision*, *recall*, dan *F1-score* yang lebih tinggi. Sebaliknya, *MobileNetV2* memiliki keunggulan pada aspek efisiensi komputasi dengan jumlah parameter yang lebih sedikit, ukuran model yang lebih kecil, serta waktu pelatihan yang lebih singkat. Hasil tersebut menunjukkan adanya *trade-off* antara performa klasifikasi dan efisiensi komputasi, sehingga pemilihan

arsitektur perlu disesuaikan dengan kebutuhan implementasi. Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa *ResNet50* unggul dalam kemampuan ekstraksi fitur yang kompleks, sedangkan *MobileNetV2* lebih sesuai untuk perangkat dengan sumber daya terbatas (Alzubaidi et al., 2021; Hermanto et al., 2024; Utami et al., 2025).

Berdasarkan hasil analisis perbandingan, *ResNet50* dipilih sebagai model terbaik dan berhasil diimplementasikan pada *Web App* berbasis Streamlit yang terintegrasi dengan fitur *Google Text-to-Speech (gTTS)* sebagai media penyampaian hasil klasifikasi dalam bentuk suara. Implementasi tersebut membuktikan bahwa model hasil pelatihan tidak hanya memiliki performa klasifikasi yang baik, tetapi juga dapat diwujudkan sebagai *proof of concept* teknologi asistif berbasis kecerdasan buatan yang mampu membantu penyampaian informasi visual dalam bentuk audio (Kalane & Mache, 2025; WHO, 2026). Penelitian selanjutnya dapat dikembangkan dengan memanfaatkan dataset yang lebih beragam, mengeksplorasi arsitektur *deep learning* yang lebih mutakhir, mengimplementasikan sistem pada perangkat *edge computing* atau *mobile*, serta mengintegrasikan sensor jarak seperti *Light Detection and Ranging (LiDAR)* atau sensor ultrasonik pada tingkat bantu agar sistem mampu memberikan informasi objek berdasarkan jarak secara *real-time* (Khatri et al., 2025; Mendez et al., 2021).

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Bhayangkara Jakarta Raya atas dukungan akademik yang diberikan selama pelaksanaan penelitian. Penulis juga menyampaikan penghargaan yang sebesar-besarnya kepada dosen pembimbing utama dan dosen pembimbing pendamping atas bimbingan, arahan, kritik, serta saran yang sangat berarti sehingga penelitian ini dapat diselesaikan dengan baik. Ucapan terima kasih turut disampaikan kepada seluruh dosen, keluarga, teman, dan semua pihak yang telah memberikan dukungan moral maupun motivasi selama proses penelitian. Penulis berharap hasil penelitian ini dapat memberikan manfaat bagi pengembangan teknologi klasifikasi citra berbasis *transfer learning* serta implementasi teknologi asistif berbasis *Google Text-to-Speech (gTTS)*.

Daftar Pustaka

- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. In *Journal of Big Data* (Vol. 8, Number 1). Springer International Publishing. <https://doi.org/10.1186/s40537-021-00444-8>
- Berliani, T., Rahardja, E., & Septiana, L. (2023). Perbandingan Kemampuan Klasifikasi Citra X-ray Paru-paru menggunakan Transfer Learning ResNet-50 dan VGG-16. *Journal of Medicine and Health*, 5(2), 123–135. <https://doi.org/10.28932/jmh.v5i2.6116>
- Budi, E. S., Chan, A. N., Alda, P. P., & Idris, M. A. F. (2024). Optimasi Model Machine Learning untuk Klasifikasi dan Prediksi Citra Menggunakan Algoritma Convolutional Neural Network.

- RESOLUSI: Rekayasa Teknik Informatika Dan Informasi*, 4(5), 502–509.
- Hermanto, A. R., Aziz, A., & Sudianto, S. (2024). Perbandingan Arsitektur MobileNetV2 dan ResNet50 untuk Klasifikasi Jenis Buah Kurma. *Jurnal Sistem Dan Teknologi Informasi SAN*, 12(4), 630–637. <https://doi.org/10.26418/justin.v12i4.80358>
- Kalane, A. S., & Mache, S. (2025). A Comprehensive Review on Text-To-Speech (TTS) Synthesis: Advances, Challenges, and Future Directions. *IJCRT: International Journal of Creative Research Thoughts*, 13(7). <https://doi.org/10.22214/ijraset.2025.71713>
- Khatri, M., Sahoo, M., Sayyad, S., & Sayyad, J. (2025). Adaptive and User-Friendly Framework for Image Classification with Transfer Learning Models. *Future Internet*, 17(8). <https://doi.org/10.3390/fi17080370>
- Krizhevsky, A. (2009). *Learning Multiple Layers of Features from Tiny Images*.
- Mendez, J., Bierzynski, K., Cuéllar, M. P., & Morales, D. P. (2021). Edge Intelligence: Concepts, Architectures, Applications, and Future Directions. *ACM Transactions on Embedded Computing Systems*, 21(5). <https://doi.org/10.1145/3486674>
- Nurdiyansyah, F., Kadir, S. F., Akbar, I., & Ursaputra, L. (2024). Penerapan Convolutional Neural Network Untuk Deteksi Kualitas Telur Ayam Ras Berdasarkan Warna Cangkang. *Jurnal MNEMONIC*, 7(1), 40–47.
- Tan, X., Qin, T., Soong, F., & Liu, T.-Y. (2021). *A Survey on Neural Speech Synthesis*.
- Utami, T. W., Novita, M., & Latifa, K. (2025). Implementation and Comparative Analysis of CNN and Transfer Learning Models (EfficientNetB0, MobileNetV2, and ResNet50) for Rice Leaf Disease Detection Based on Digital Images. *Journal of Applied Informatics and Computing*, 9(6), 3862–3870. <https://doi.org/10.30871/jaic.v9i6.11616>
- WHO. (2026). *Blindness and vision impairment*.
- Wijaya, A. E., Swastika, W., & Kelana, O. H. (2021). Implementasi Transfer Learning Pada Convolutional Neural Network Untuk Diagnosis Covid-19 Dan Pneumonia Pada Citra X-Ray. *SAINSBERTEK Jurnal Ilmiah Sains & Teknologi*, 2(1). <https://doi.org/10.33479/sb.v2i1.125>
- Yong, L., Ma, L., Sun, D., & Du, L. (2023). Application of MobileNetV2 to Waste Classification. *PLoS ONE*, 18(3), 1–16. <https://doi.org/10.1371/journal.pone.0282336>