JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 159-168
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Penerapan Transfer Learning MobileNetV2 untuk Sistem Pengenalan Gerakan Tangan Interaktif

Marcellino Andelta Pinem[1], Fathony Mursyid[2], Eldika Rubiana[3*], Abraham Imanuel Sinaga[4], Haikal Ryan Saputra[5]

[1,2,3,4,5]Informatika, Universitas Bina Sarana Informatika, Depok, Jawa Barat
Email: [1]marcellpinem@gmail.com, [2]gusfathon.fm@gmail.com, [3*]eldikarubiana@gmail.com, [4]abrm.imnl@gmail.com, [5]haikalryan04@gmail.com
Email Penulis Korespondensi: [3]eldikarubiana@gmail.com

Abstrak − Pengenalan gerakan tangan merupakan komponen penting dalam interaksi manusia-komputer yang memerlukan akurasi tinggi untuk aplikasi praktis. Penelitian ini menerapkan transfer learning MobileNetV2 dengan strategi two-phase fine-tuning untuk meningkatkan akurasi pengenalan tujuh gerakan tangan pada dataset HaGRID. Dataset terdiri dari 175.000 gambar yang terbagi menjadi 140.000 data latih, 17.500 data validasi, dan 17.500 data uji. Metode two-phase meliputi Phase 1 dengan frozen base layers menghasilkan akurasi 75,83%, dan Phase 2 dengan fine-tuning selective layers meningkatkan akurasi menjadi 98,88% pada data validasi dan 98,86% pada data uji. Peningkatan signifikan sebesar 23,05% berhasil dicapai hanya dalam 10 epochs total dengan durasi training 6,5 jam. Model berhasil mengeliminasi seluruh confusion pairs yang sebelumnya mencapai 18,64% pada Phase 1 menjadi 0% confusion di Phase 2. Kontribusi utama penelitian ini adalah demonstrasi strategi two-phase fine-tuning yang efisien untuk model lightweight dengan akurasi setara arsitektur kompleks, memberikan solusi praktis untuk implementasi sistem pengenalan gerakan real-time pada perangkat mobile dan embedded system tanpa mengorbankan performa.

Kata Kunci: Hand Gesture Recognition, Transfer Learning, MobileNetV2, Deep Learning, HaGRID Dataset, Two-phase Fine-Tuning

Abstract − Hand gesture recognition is a critical component in human-computer interaction that requires high accuracy for practical applications. This study applies MobileNetV2 transfer learning with a two-phase fine-tuning strategy to improve the recognition accuracy of seven hand gestures on the HaGRID dataset. The dataset consists of 175,000 images divided into 140,000 training data, 17,500 validation data, and 17,500 test data. The two-phase method includes Phase 1 with frozen base layers achieving 75.83% accuracy, and Phase 2 with selective layer fine-tuning improving accuracy to 98.88% on validation data and 98.86% on test data. A significant improvement of 23.05% was achieved in only 10 total epochs with a training duration of 6.5 hours. The model successfully eliminated all confusion pairs that previously reached 18.64% in Phase 1 to 0% confusion in Phase 2. The main contribution of this research is the demonstration of an efficient two-phase fine-tuning strategy for lightweight models with accuracy comparable to complex architectures, providing practical solutions for implementing real-time gesture recognition systems on mobile and embedded system devices without sacrificing performance.

Keywords: Hand Gesture Recognition, Transfer Learning, MobileNetV2, Deep Learning, HaGRID Dataset, Two-phase Fine-Tuning

1. PENDAHULUAN

Teknologi pengenalan gerakan tangan (hand gesture recognition) menjadi topik riset penting dalam pengembangan sistem interaksi manusia-komputer dan bidang visi komputer. Kehadiran teknologi ini membuka peluang komunikasi yang lebih natural antara pengguna dan perangkat digital tanpa bergantung pada peranti masukan konvensional. Implementasi sistem pengenalan gerakan tangan dapat ditemukan pada berbagai sektor, mulai dari sistem komunikasi bahasa isyarat bagi penyandang disabilitas, pengendali perangkat rumah pintar, permainan interaktif, sistem robotik, hingga lingkungan realitas virtual. Urgensi pengembangan teknologi ini kian meningkat seiring maraknya perangkat bergerak dan komputasi tepi yang menuntut solusi pengenalan yang cepat dan presisi tinggi.

Perkembangan terkini menunjukkan bahwa metode deep learning memberikan kinerja unggul dibandingkan pendekatan tradisional dalam mengenali gerakan tangan. Berbagai penelitian mengimplementasikan arsitektur kompleks untuk mencapai akurasi tinggi: Aurangzeb dkk. [1] mengembangkan sistem berbasis deep learning untuk membantu komunikasi penyandang tuna rungu-wicara serta aplikasi bidang kesehatan, Mujahid dkk. [2] berhasil membangun model deteksi real-time menggunakan arsitektur YOLOv3, Haq dkk. [3] mengimplementasikan mekanisme self-attention untuk meningkatkan ketepatan prediksi, Rahim dkk.[4] mengembangkan arsitektur hibrida three-stream untuk menangkap gerakan dinamis, Yaseen dkk. [5] mengkombinasikan MediaPipe dengan Inception-v3 dan LSTM dalam kerangka kerja terintegrasi, Zerrouki dkk. [6] mengaplikasikan sistem deep learning pada lingkungan museum virtual dengan sensor pakai, dan Faisal dkk.[7] memanfaatkan teknik transformasi domain menggunakan dataglove berbiaya terjangkau. Terlepas dari capaian akurasi yang tinggi, mayoritas arsitektur deep learning yang kompleks menuntut kapasitas komputasi yang substansial serta durasi training yang panjang, menghambat implementasi pada perangkat dengan daya komputasi terbatas seperti smartphone dan embedded system.

Sebagai alternatif, arsitektur lightweight MobileNetV2 menawarkan kompromi optimal antara tingkat akurasi dan efisiensi pemrosesan melalui pendekatan depthwise separable convolutions yang mereduksi parameter dan operasi matematis. Implementasi MobileNetV2 yang dikombinasikan dengan transfer learning menunjukkan efektivitas pada berbagai ranah aplikasi: Gulzar [8] memperoleh akurasi tinggi dalam mengklasifikasikan citra buah-buahan menggunakan arsitektur ini, Banoth dan Murthy [9] mengadopsi transfer learning MobileNetV2 untuk kategorisasi citra tanah, Barman

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted:**13/11/2025**; Accepted: 27**/12/2025**; Published: **15/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 159-168
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

dan Susan [10] menggunakannya pada klasifikasi multi-label remote sensing, dan Xiang dkk. [11] menerapkannya untuk pengelompokan citra buah, yang semuanya memvalidasi adaptabilitas arsitektur tersebut. Meskipun demikian, aplikasi MobileNetV2 untuk mengenali gerakan tangan pada dataset berskala besar seperti HaGRID masih membutuhkan optimalisasi lebih lanjut guna memaksimalkan performa. Dataset HaGRID (HAnd Gesture Recognition Image Dataset) yang dipublikasikan oleh Alexander dkk. [12] termasuk dalam dataset berukuran besar untuk pengenalan gerakan tangan statis dengan koleksi melebihi 550.000 citra mencakup 18 kategori gerakan dengan keragaman kondisi seperti intensitas cahaya, latar belakang, dan situasi pengambilan gambar yang bervariasi. Pengembangan lebih lanjut dilakukan oleh Nuzhdin dkk. [13] melalui HaGRIDv2 yang menampung 1 juta citra untuk mengenali gerakan statis maupun dinamis.

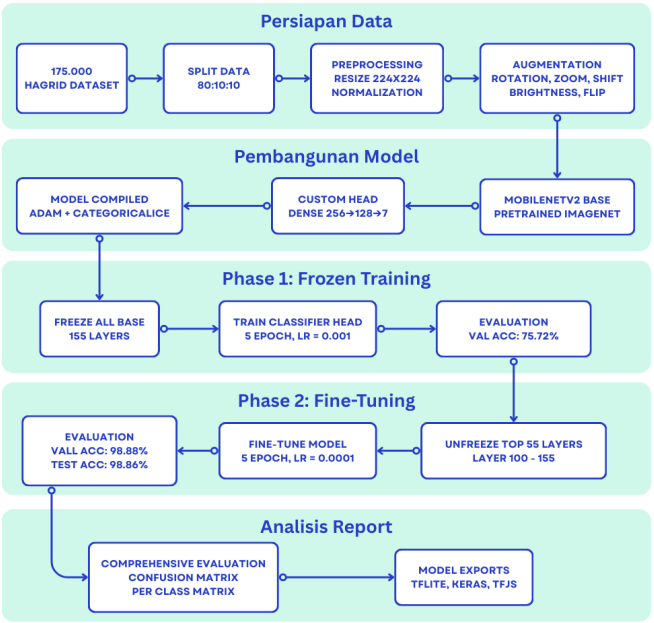
Permasalahan fundamental dalam pelatihan jaringan deep learning adalah penanganan class imbalance dan confusion antar kategori yang memiliki kemiripan visual. ValizadehAslani dkk. [14], [15] memperkenalkan pendekatan two-phase fine-tuning untuk mengatasi data tidak seimbang. Metodologi ini menggunakan fase awal dengan frozen base layers untuk ekstraksi fitur generik, dilanjutkan fase kedua berupa fine-tuning pada layer terpilih untuk adaptasi domain khusus. Teknik ini terbukti ampuh meningkatkan akurasi serta memitigasi overfitting, namun implementasinya pada pengenalan gerakan tangan dengan dataset bervolume besar masih jarang dilakukan. Tinjauan terhadap riset terdahulu mengidentifikasi empat celah penelitian yang signifikan: pertama, arsitektur deep learning kompleks menuntut kapasitas komputasi tinggi sehingga tidak efisien untuk deployment pada perangkat mobile; kedua, pemanfaatan MobileNetV2 pada pengenalan gerakan tangan dengan dataset berskala masif belum mencapai hasil optimal; ketiga, isu confusion pairs antar kelas dengan karakteristik visual serupa belum dikaji secara komprehensif; keempat, strategi two-phase transfer learning yang efektif menangani class imbalance jarang diterapkan pada domain ini.

Riset ini berupaya mengisi kekosongan tersebut dengan mengimplementasikan strategi two-phase transfer learning pada arsitektur MobileNetV2 untuk mengenali gerakan tangan menggunakan subset HaGRID. Kontribusi penelitian meliputi: (1) penerapan strategi two-phase fine-tuning efisien yang meningkatkan akurasi MobileNetV2 dari 75,83% ke 98,88% hanya dengan 10 epochs; (2) penghapusan total confusion pairs yang sebelumnya mencapai 18,64% menjadi nihil; (3) pembuktian bahwa model lightweight MobileNetV2 mampu menandingi performa arsitektur kompleks pada dataset 175.000 citra; dan (4) analisis menyeluruh efektivitas two-phase transfer learning dalam menyelesaikan masalah confusion pada klasifikasi gerakan tangan. Tujuan spesifik penelitian adalah membangun sistem pengenalan gerakan tangan yang akurat dan efisien menggunakan MobileNetV2 dengan two-phase transfer learning pada subset HaGRID mencakup 7 kategori gerakan (ok, one, palm, peace, rock, thumb_down, thumb_up) dengan total 175.000 gambar. Sistem yang dikembangkan ditargetkan mencapai akurasi tinggi dengan mempertahankan efisiensi komputasi untuk implementasi pada perangkat berdaya terbatas, sekaligus mengeliminasi confusion antar kelas secara efektif.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini mengimplementasikan pendekatan *two-phase transfer learning* pada arsitektur MobileNetV2 untuk sistem pengenalan gerakan tangan. Metodologi yang dirancang mencakup persiapan dataset, konstruksi arsitektur model, pelatihan bertahap dua Phase, dan evaluasi komprehensif performa model. Tahapan penelitian secara keseluruhan divisualisasikan pada Gambar 1 yang menunjukkan alur kerja dari persiapan data hingga eksport model final.



Gambar 1. Flowchart Tahapan Penelitian

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted:**13/11/2025**; Accepted: 27**/12/2025**; Published: **15/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 159-168
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

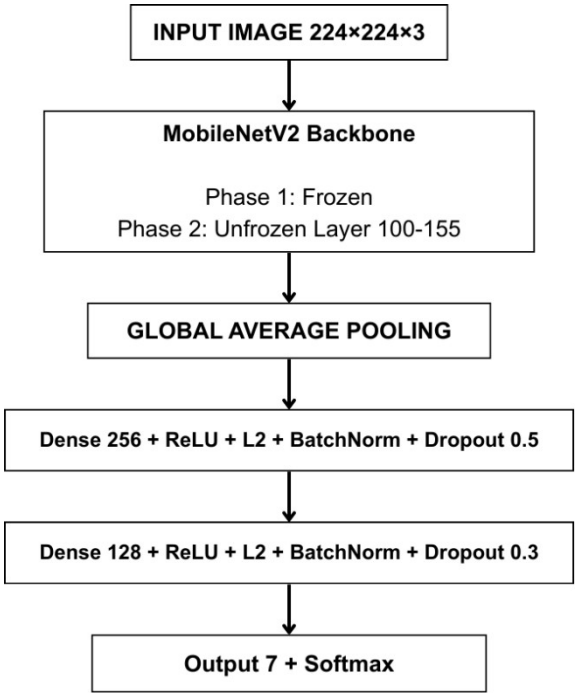
2.2 Dataset dan Preprocessing

Dataset yang digunakan merupakan subset dari HaGRID (*HAnd Gesture* Recognition *Image Dataset*) [12] terdiri dari 7 kategori gerakan tangan: *ok*, *one*, *palm*, *peace*, *rock*, *thumb_down*, *thumb_up*. Setiap kategori mengandung 25.000 citra menghasilkan 175.000 gambar yang dibagi menjadi data latih (140.000 citra), data validasi (17.500 citra), dan data uji (17.500 citra) dengan rasio 80:10:10. Distribusi kelas bersifat seimbang dengan jumlah sampel identik pada setiap partisi untuk menghindari bias klasifikasi.

Tahap *preprocessing* melibatkan transformasi seluruh citra ke resolusi 224x224 piksel menggunakan interpolasi bilinear. Normalisasi piksel dilakukan dengan fungsi *preprocess_input* MobileNetV2 yang mengubah nilai dari rentang [0, 255] menjadi [-1, 1] sesuai skema ImageNet. Augmentasi data diterapkan eksklusif pada data latih meliputi rotasi acak hingga 20 derajat, pergeseran horizontal dan vertikal 15%, *zoom* dengan faktor 0,15, manipulasi *brightness* dalam rentang [0,85, 1,15], dan *horizontal flip* secara acak untuk meningkatkan kemampuan generalisasi model.

2.3 Arsitektur Model

Arsitektur yang dikembangkan memanfaatkan MobileNetV2 [8] sebagai *backbone* ekstraksi fitur dengan bobot *pretrained* ImageNet. MobileNetV2 dipilih karena efisiensinya melalui *inverted residual blocks* dan *depthwise separable convolutions* menghasilkan model *lightweight* dengan parameter minimal namun representasi fitur kuat. Konfigurasi mengekstrak fitur dari layer terakhir sebelum klasifikasi menggunakan *global average pooling* untuk representas vektor satu dimensi.



Gambar 2. Arsitektur Model MobileNetV2 dengan Classification Head Kustom

Diagram menunjukkan alur data dari input image (224×224×3) melalui MobileNetV2 backbone yang menggunakan pretrained weights ImageNet, dilanjutkan dengan global average pooling dan classification head tiga layer. Phase 1 membekukan seluruh backbone, sementara Phase 2 melakukan unfreezing pada 55 layer terakhir (layer 100-155) untuk fine-tuning adaptif.

Bagian *classification head* dirancang dengan arsitektur kustom: layer pertama *fully connected* 256 neuron dengan aktivasi ReLU, regularisasi L2 (λ = 0,0001), *batch normalization*, dan *droptout* 0,5; layer kedua 128 neuron dengan konfigurasi serupa namun *dropout* 0,3; layer output 7 neuron dengan aktivasi *softmax* untuk distribusi probabilitas tujuh kategori gesture.

2.4 Strategi Pelatihan Two-phase

Proses pelatihan dibagi menjadi dua Phase dengan karakteristik berbeda. Phase 1 membekukan (*freeze*) seluruh layer *backbone* MobileNetV2 sehingga hanya *classification head* yang dilatih untuk mempelajari representasi fitur spesifik domain gerakan tangan. Phase ini menggunakan *learning rate* 0,001 dengan optimizer Adam dan dilatih 5 *epochs*. Fungsi *loss categorical crossentropy* dengan *label smoothing* 0,1 digunakan untuk mengurangi *overfitting* dan meningkatkan generalisasi. *Batch size* yang digunakan 128 sampel per iterasi.

Phase 2 melakukan *unfreezing* pada layer MobileNetV2 dimulai dari layer ke-100 hingga terakhir untuk adaptasi lebih dalam, sementara layer awal tetap dibekukan mempertahankan fitur genetik tingkat rendah. *Learning rate*

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted:**13/11/2025**; Accepted: 27**/12/2025**; Published: **15/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 159-168
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

diturunkan menjadi 0,0001 mencegah perubahan bobot drastis dan menjaga stabilitas. Phase ini juga dilatih 5 *epochs* dengan konfigurasi *loss*, *batch size*, dan regularisasi identik dengan Phase pertama.

2.5 Evaluasi dan Metrik Performa

Mekanisme *callback* diimplementasikan meliputi *ModelCheckpoint* untuk menyimpan bobot terbaik berdasarkan akurasi validasi tertinggi, *ReduceLROnPlateau* yang menurunkan *learning rate* 50% jika *validation loss* tidak membaik selama 2 *epochs*, dan *EarlyStopping* yang menghentikan pelatihan jika akurasi validasi stagnan selama 4-5 *epochs* untuk efisiensi komputasi. Evaluasi performa dilakukan pada data validasi dan data uji untuk menggunakan metrik akurasi klasifikasi, *precision*, *recall*, dan *F1-Score* per-kelas untuk analisis performa individual setiap kategori gesture. *Confusion matrix* divisualisasikan mengidentifikasi pola kesalahan dan *confusion pairs* antar kelas dengan kemiripan visual. Pasangan dengan tingkat *confusion* di atas 5% pada Phase 1 dan 2% pada Phase 2 dikategorikan problematik memerlukan perhatian khusus. Perbandingan performa Phase 1 dan Phase 2 mengevaluasi efektivitas strategi *two-phase fine-tuning* dalam meningkatkan akurasi dan mengurangi *confusion*. Visualisasi kurva pelatihan untuk akurasi dan *loss* menganalisis konvergensi model dan mendeteksi indikasi *overfitting* atau *underfitting*. Keseluruhan eksperimen dilakukan dengan menggunakan *framework* TensorFlow dengan akselerasi GPU untuk mempercepat komputasi.

2.6 Environment Komputasi

Keseluruhan eksperimen dilakukan menggunakan Google Colaboratory dengan spesifikasi hardware yang ditampilkan pada Tabel 1. Framework TensorFlow 2.15 dengan backend Keras digunakan untuk implementasi model, dilengkapi dengan akselerasi GPU NVIDIA Tesla T4 untuk mempercepat proses pelatihan dan inferensi.

Tabel 1. Spesifikasi Hardware

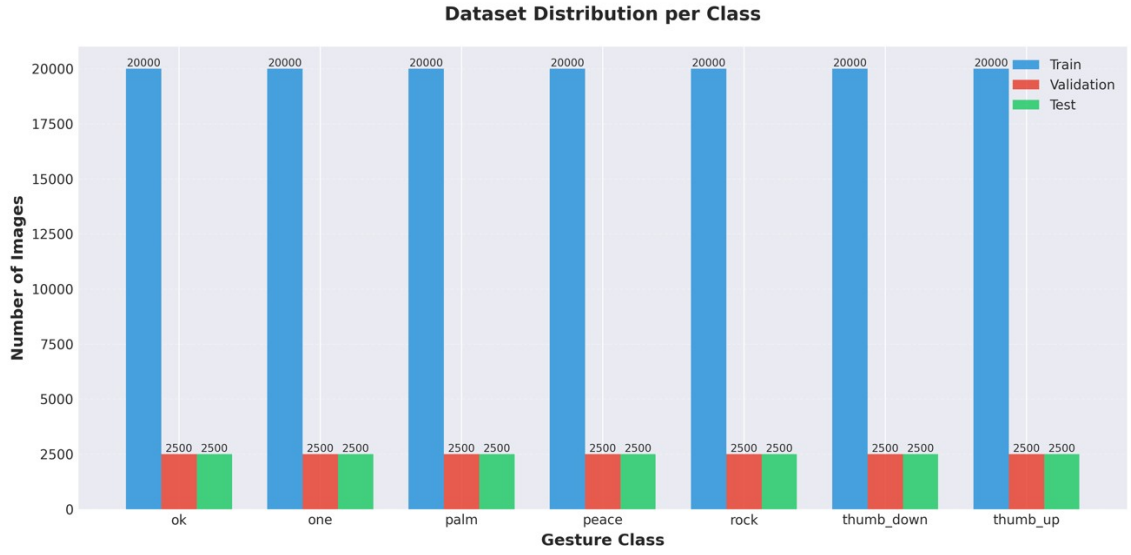
Komponen	Spesifikasi
Platform	Google Colaboratory (Free Tier)
GPU	NVIDIA Tesla T4 (16GB VRAM)
System RAM	12,7 GB
CPU	Intel Xeon (2 vCPU @ 2.20GHz)
Storage	Google Drive

Penggunaan GPU Tesla T4 dengan 16GB VRAM memungkinkan pelatihan batch size 128 sampel tanpa kendala memory overflow. Total durasi pelatihan 6.5 jam untuk 10 epochs menunjukkan efisiensi komputasi yang tinggi untuk dataset berukuran 175.000 gambar.

3. HASIL DAN PEMBAHASAN

3.1 Distribusi Dataset

Dataset yang digunakan dalam penelitian ini merupakan subset dari HaGRID yang telah diorganisir dengan distribusi seimbang untuk memastikan model tidak mengalami bias terhadap kelas tertentu. Distribusi dataset pada setiap partisi ditampilkan pada Gambar 3 yang memvisualisasikan jumlah sampel untuk tujuh kategori gesture pada data latih, validasi, dan uji.



This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted:**13/11/2025**; Accepted: 27**/12/2025**; Published: **15/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 159-168
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

Gambar 3. Distribusi Dataset per Kelas

Dari Gambar 3 terlihat bahwa setiap kategori gesture memiliki jumlah sampel yang identik pada masing-masing partisi, dengan total 20.000 citra per kelas pada data latih, 2.500 citra pada data validasi, dan 2.500 citra pada data uji. Distribusi seimbang ini penting untuk mencegah model mengalami bias klasifikasi terhadap kelas mayoritas dan memastikan evaluasi yang objektif pada setiap kategori gesture. Total dataset mencapai 175.000 gambar yang terdistribusi merata, memberikan representasi yang memadai untuk melatih arsitektur *deep learning* dengan kapasitas parameter yang cukup besar.

3.2 Hasil Pelatihan Phase 1: Frozen Base Training

Phase 1 merupakan tahap pelatihan awal dengan strategi pembekuan seluruh layer pada *backbone* MobileNetV2, sehingga hanya *classification head* yang mengalami pembaruan bobot selama proses pembelajaran. Strategi ini memungkinkan model untuk mempelajari pemetaan fitur dari representasi ImageNet ke domain gerakan tangan tanpa mengubah ekstraksi fitur tingkat rendah yang telah dipelajari pada dataset skala besar. Proses pelatihan Phase 1 dilakukan selama 5 *epochs* dengan *learning rate* 0,001 menggunakan optimizer Adam.



Gambar 4. Training Curves Phase 1

Gambar 4 menunjukkan kurva pelatihan Phase 1 untuk metrik akurasi dan *loss* pada data latih dan validasi. Terlihat bahwa akurasi validasi mengalami peningkatan dari 74,14% pada *epoch* pertama menjadi 75,83% pada *epoch* ketiga, yang merupakan performa terbaik pada Phase ini. Akurasi pelatihan menunjukkan tren peningkatan yang stabil dari 63,45% hingga 71,81%, mengindikasikan bahwa *classification head* berhasil mempelajari pemetaan fitur dengan baik. Gap antara akurasi latih dan validasi yang relatif kecil menunjukkan bahwa model tidak mengalami *overfitting* yang signifikan pada Phase ini. Nilai *loss* pada data validasi menurun secara konsisten dari 1,059 menjadi 1,002, menandakan konvergensi yang baik meskipun terdapat ruang untuk peningkatan performa.

Pola konvergensi yang stabil tanpa fluktuasi signifikan mengindikasikan bahwa kombinasi *label smoothing* sebesar 0.1 dan regularisasi L2 (λ=0.0001) berhasil mencegah model dari *overfitting* prematur pada data latih. Karakteristik kurva *loss* yang menurun secara monoton tanpa adanya peningkatan kembali (*rebound*) menunjukkan bahwa *learning rate* 0.001 yang dipilih sudah optimal untuk Phase pembelajaran awal. Efektivitas strategi pembekuan *backbone* juga terlihat dari cepatnya konvergensi dalam 5 *epochs*, dimana model mampu mencapai plateau performa tanpa memerlukan iterasi tambahan yang akan memboroskan waktu komputasi.

Analisis per-kelas pada Phase 1 menunjukkan variasi performa yang cukup signifikan antar kategori gesture. Kelas *thumb_up* mencapai akurasi tertinggi sebesar 90,44% diikuti oleh *thumb_down* dengan 85,80%, menunjukkan bahwa kedua gesture ini memiliki karakteristik visual yang distinktif dan mudah dibedakan oleh model. Sebaliknya kelas *peace* mengalami performa terburuk dengan akurasi hanya 52,84%, mengindikasikan adanya kesulitan dalam membedakan gesture ini dari kategori lain. Kelas *rock* juga menunjukkan akurasi yang relatif rendah sebesar 64,00%, sementara kelas *one* mencapai 80,76%. Performa tinggi pada kelas *thumb_up* dan *thumb_down* dapat dijelaskan oleh karakteristik visual yang sangat distinktif dimana orientasi ibu jari memberikan fitur yang mudah diidentifikasi oleh *convolutional layers*. Sebaliknya, kesulitan model dalam mengklasifikasikan *peace*, *rock*, dan *one* mencerminkan tantangan fundamental dalam membedakan gesture yang memiliki konfigurasi jari serupa namun dengan perbedaan subtle pada jumlah dan posisi jari yang terangkat. Fenomena ini konsisten dengan penelitian sebelumnya yang menggunakan arsitektur *deep learning* untuk klasifikasi gesture, dimana kemiripan intra-class variance yang tinggi menjadi bottleneck utama dalam mencapai akurasi optimal. Hasil ini mengungkapkan bahwa terdapat *confusion* yang substansial pada beberapa pasangan kelas yang memiliki kemiripan visual, khususnya antara gesture yang melibatkan jari-jari terangkat seperti *peace*, *rock*, dan *one*.

Investigasi lebih lanjut mengidentifikasi 11 pasangan kelas (*confusion pairs*) dengan tingkat kebingungan di atas 5%. Pasangan terproblematik adalah *peace* yang diprediksi sebagai *rock* dengan tingkat *confusion* 18,64% diikuti oleh *rock* yang diprediksi sebagai *one* (16,40%). Konfusi tinggi ini disebabkan oleh kemiripan pose tangan pada ketiga gesture tersebut yang sama-sama menampilkan jari terangkat dengan variasi jumlah dan posisi. Pasangan problematik lainnya

This is an open access article under the CC–BY-SA license

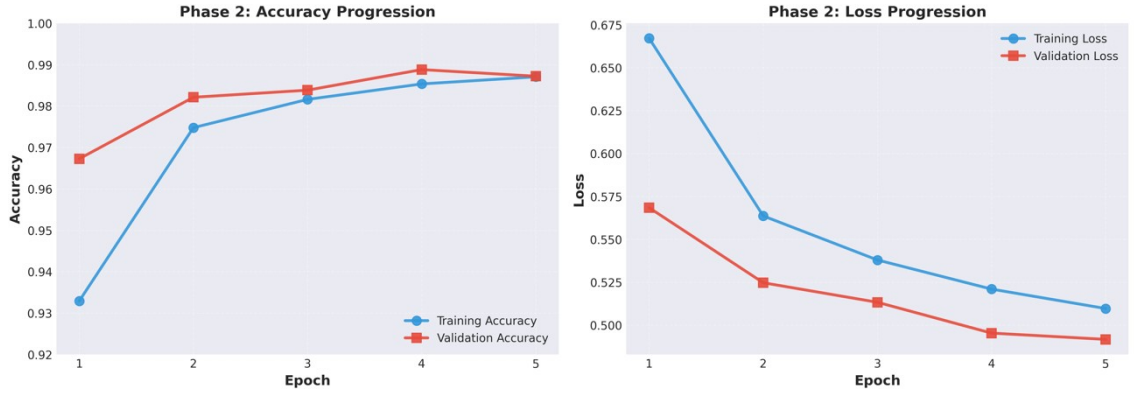
Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted:**13/11/2025**; Accepted: 27**/12/2025**; Published: **15/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 159-168
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

termasuk *palm* dengan *ok* (10,64%), *thumb_down* dengan *thumb_up* (10,64%), dan beberapa pasangan lain dengan tingkat *confusion* antara 5-8%. Temuan ini menjadi dasar justifikasi untuk melakukan *fine-tuning* pada Phase 2 guna meningkatkan kemampuan diskriminasi model terhadap kelas-kelas yang memiliki karakteristik visual serupa.

3.3 Hasil Pelatihan Phase 2: Fine Tuning

Phase 2 mengimplementasikan strategi *fine-tuning* dengan melakukan *unfreezing* pada 55 layer terakhir dari *backbone* MobileNetV2 (layer 100-155) untuk memungkinan adaptasi lebih mendalam terhadap karakteristik spesifik dataset gerakan tangan. *Learning rate* diturunkan secara drastis menjadi 0,0001 untuk mencegah perubahan bobot yang terlalu agresif yang dapat merusak representasi fitur yang telah dipelajari. Strategi ini bertujuan untuk mempertahankan fitur generik tingkat rendah sambil mengoptimalkan fitur tingkat tinggi yang lebih spesifk untuk domain gerakan tangan.



Gambar 5. Training Curves Phase 2 (Accuracy dan Loss)

Gambar 5 memperlihatkan kurva pelatihan Phase 2 yang menunjukkan performa yang dramatis dibandingkan Phase 1. Akurasi validasi melonjak dari 96,73% pada *epoch* pertama menjadi 98,88% pada epoch ketiga, mencapai target akurasi tinggi yang diharapkan. Akurasi pelatihan juga meningkat dari 93,29% hingga 98,71%, dengan gap yang sangat kecil antara akurasi latih dan validasi (sekitar 0,01%), mengindikasikan bahwa model mencapai generalisasi yang sangat baik tanpa *overfitting*. Konvergensi yang cepat ini menunjukkan efektivitas strategi *two-phase training* pada6 Phase 2. Nilai *loss* menurun secara signifikan dari 0,667 menjadi 0,492 pada data validasi, menandakan bahwa model ini berhasil meminimalkan kesalahan prediksi dengan sangat baik.

Perbandingan komprehensif antara Phase 1 dan Phase 2 mengungkapkan peningkatan performa yang luar biasa pada semua metrik evaluasi. Akurasi keseluruhan meningkat sebesar 23,05 poin persentase dari 75,83% pada Phase 1 menjadi 98,88% pada Phase 2, melampaui target yang ditetapkan. Analisis per-kelas menunjukkan bahwa peningkatan paling dramatis terjadi pada kelas *peace* yang melonjak 44,36% dari akurasi 52,84% menjadi 97,20%, mengubah kelas dengan performa terburuk menjadi salah satu yang paling akurat. Kelas *rock* juga mengalami transformasi signifikan dengan peningkatan 34,12% dari 64,00% menjadi 98,12%, mendemonstrasikan efektivitas *fine-tuning* dalam mengatasi kesulitan membedakan gesture dengan karakteristik visual serupa.

Kelas-kelas yang sudah memiliki performa baik pada Phase 1 tetap mengalami peningkatan substansial: *ok* meningkat 23,84% menjadi 98,96%, *one* naik 18,64% menjadi 99,40%, *palm* bertambah 17,36% menjadi 99,20%, *thumb_down* meningkat 14,00% mencapai 99,80%, dan *thumb_up* naik 9,04% menjadi 99,48%, semuanya mendekati akurasi sempurna. Metrik agregat menunjukkan konsistensi peningkatan dimana *average precision* naik 22,43% dari 76,46% ke 98,89%, *average recall* meningkat 23,05% dari 75,83% ke 98,88%, dan *average F1-score* bertambah 23,28% dari 75,60% ke 98,88%. Aspek paling mengesankan adalah eliminasi total dari 11 *problem pairs* (pasangan kelas dengan *confusion* >5%) yang ada pada Phase 1 menjadi tidak ada satupun pasangan dengan *confusion* >2% pada Phase 2, dengan penurunan *worst confusion* dari 18,64% menjadi di bawah 2%.

Aspek paling mengesankan dari hasil Phase 2 adalah eliminasi total dari seluruh *confusion pairs* yang sebelumnya bermasalah. Pada Phase 1 terdapat 11 pasangan kelas dengan tingkat kebingungan di atas 5%, dengan *confusion* terburuk mencapai 18,64%. Setelah *fine-tuning* pada Phase 2, tidak ada satupun pasangan kelas yang memiliki tingkat *confusion* di atas 2%, menandakan bahwa model berhasil mempelajari fitur-fitur distinktif yang membedakan setiap gesture dengan sangat baik. Eliminasi *confusion pairs* ini sangat krusial untuk aplikasi praktis sistem pengenalan gerakan tangan dimana kesalahan klasifikasi dapat mengganggu pengalaman pengguna atau bahkan menyebabkan kesalahan perintah pada sistem kontrol.

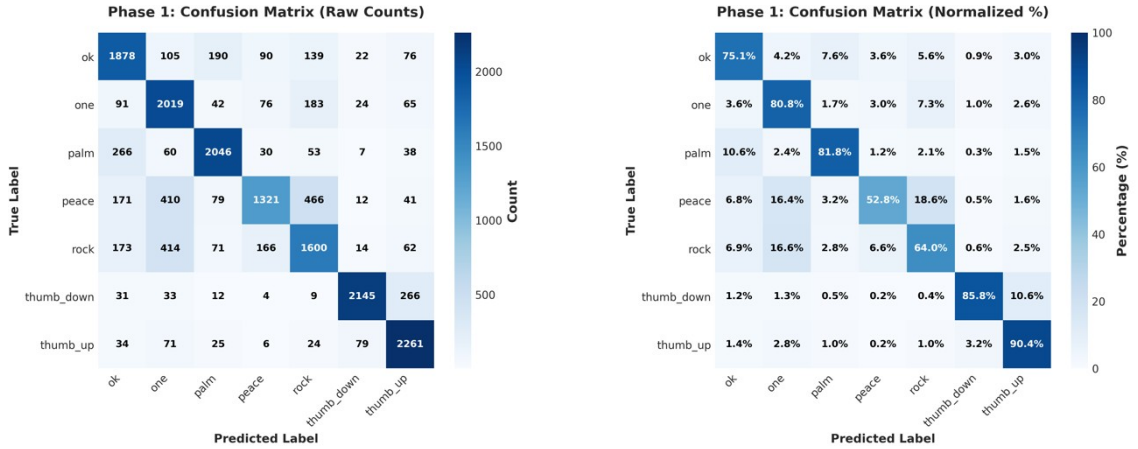
3.4 Analisis Confusion Matrix

Untuk memahami pola kesalahan klasifikasi secara mendalam, analisis *confusion matrix* dilakukan untuk membandingkan performa Phase 1 dan Phase 2 secara visual.

This is an open access article under the CC–BY-SA license

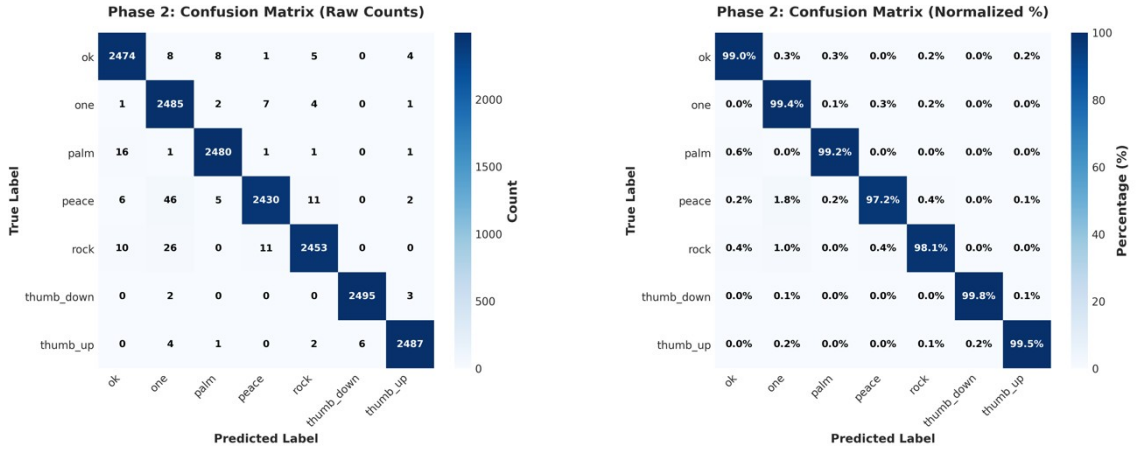
Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted:**13/11/2025**; Accepted: 27**/12/2025**; Published: **15/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 159-168
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin



Gambar 6. Confusion Matrix Phase 1 (Raw Counts dan Normalized)

Gambar 6 menampilkan *confusion matrix* Phase 1 dalam bentuk *raw counts* dan *normalized percentage*. Visualiasi ini mengkonfirmasi temuan sebelumnya bahwa terdapat confusion yang substansial pada beberapa pasangan kelas. Diagonal matriks yang mempresentasikan prediksi benar menunjukkan intensitas yang bervariasi, dengan kelas *thumb_up* dan *thumb_down* memiliki nilai diagonal tertinggi, sementara *peace* dan *rock* menunjukkan nilai yang lebih rendah. *Off-diagonal* elemen dengan intensitas tinggi mengindikasikan *confusion pairs* yang signifikan, khususnya pada sel (*peace, rock*), (*rock, one*), dan (*peace, one*). Pola *confusion* ini mencerminkan tantangan dalam membedakan gesture yang melibatkan konfigurasi jari yang serupa, dimana perbedaan antara dua jari terangkat (*peace*), tiga jari terangkat dengan variasi (*rock*), dan satu jari terangkat (*one*) memerlukan pembelajaran fitur yang lebih halus.



Gambar 7. Confusion Matrix Phase 2 (Raw Counts dan Normalized)

Gambar 7 menunjukkan *confusion matrix* Phase 2 yang menampilkan perbaikan dramatis dibandingkan Phase 1. Diagonal matriks memperlihatkan intensitas yang sangat tinggi dan hampir seragam untuk semua kelas, mengindikasikan bahwa model dapat mengklasifikasikan hampir semua sampel dengan benar. Elemen *off-diagonal* menunjukkan intensitas yang sangat rendah dengan warna yang hampir putih, menandakan bahwa kesalahan klasifikasi telah diminimalkan secara drastis. Bahkan pasangan kelas yang sebelumnya bermasalah seperti *peace-rock*, *rock-one*, dan *peace-one* kini menunjukkan tingkat confusion yang *negligible* di bawah 2%. Transformasi visual yang jelas antara Gambar 6 dan Gambar 7 secara efektif mengilustrasikan keberhasilan strategi *two-phase fine-tuning* dalam mengatasi masalah *confusion* antar kelas.

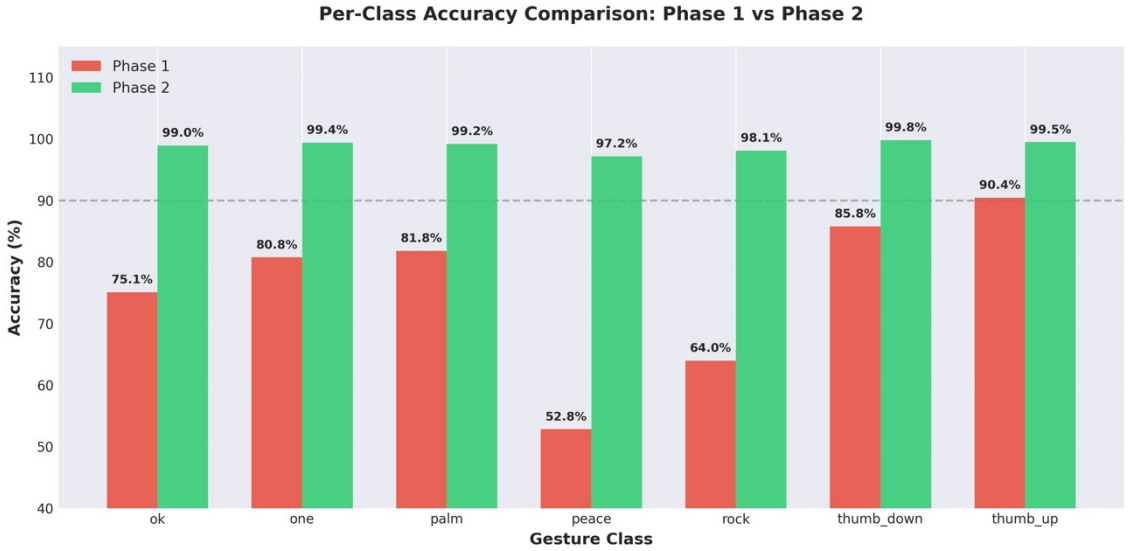
3.5 Perbandingan dan Evaluasi Final

Untuk memberikan perspektif yang lebih komprehensif tentang peningkatan performa, Gambar 8 menyajikan perbandingan visual akurasi per-kelas antara Phase 1 dan Phase 2.

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted:**13/11/2025**; Accepted: 27**/12/2025**; Published: **15/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 159-168
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin



Gambar 8. Perbandingan Per-Class Accuracy Phase 1 vs Phase 2

Gambar 8 memvisualiasasikan dengan jelas bahwa semua kategori gesture mengalami peningkatan performa yang substansial setelah *fine-tuning*. Bar chart menunjukkan bahwa pada Phase 1, hanya dua kelas (*thumb_up* dan *thumb_down*) yang melampaui threshold 85%, sementara *peace* dan *rock* berada jauh di bawah 70%. Setelah Phase 2, seluruh kelas melampaui 97%, mendemonstrasikan konsistensi performa yang tinggi. Garis referensi pada 90% memperjelas bahwa target akurasi tinggi telah tercapai dan bahkan terlampaui dengan margin yang signifikan untuk semua kategori.

Evaluasi final dilakukan pada data uji yang salah sekali tidak terlihat oleh model selama proses pelatihan untuk memvalidasi kemampuan generalisasi. Hasil evaluasi pada *test set* disajikan pada Tabel 2.

Tabel 2. Hasil Evaluasi Final pada Test Set

Gesture	Precision	Recall	F1-Score	Support
ok	98,88%	99,08%	98,98%	2500
one	96,10%	99,44%	97,74%	2500
palm	99,64%	99,12%	97,04%	2500
peace	98,98%	97,04%	98,00%	2500
rock	99,03%	97,84%	98,43%	2500
thumb_down	99,68%	99,96%	99,82%	2500
thumb_up	99,84%	99,56%	99,70%	2500
Macro Average	98,88%	98,86%	98,86%	17500

Tabel 2 menunjukkan bahwa model mencapai performa yang sangat konsisten pada *test set* dengan akurasi keseluruhan 98,86%, hanya berbeda 0,02% dari akurasi validasi (98,88%). Konsistensi ini mengkonfirmasi bahwa model memiliki kemampuan generalisasi *overfitting* terhadap data validasi. Nilai *precision* dan recall yang tinggi dan seimbang untuk semua kelas menunjukkan bahwa model tidak hanya akurat dalam mengidentifikasi kelas positif (*recall*) tetapi juga meminimalkan *false positive* (*precision*). Kelas *thumb_down* mencapai performa terbaik dengan *F1-Score* 99,82%, sementara *peace* memiliki *F1-Score* terendah namun tetap sangat tinggi pada 98,00%. Hasil ini memvalidasi bahwa strategi *two-phase transfer learning* tidak hanya meningkatkan akurasi pada data validasi tetapi juga menghasilkan model yang robust untuk data baru yang belum pernah ditemui sebelumnya.

Dibandingkan dengan penelitian sebelumnya yang menggunakan MobileNetV2 untuk tugas klasifikasi visual [8], [9], [10], [11], hasil yang dicapai dalam penelitian ini berada pada tingkat yang kompetitif bahkan dengan dataset yang lebih besar dan kompleks. Keberhasilan eliminasi *confusion pairs* merupakan kontribusi yang sangat relevan untuk implementasi praktis sistem pengenalan gerakan tangan, dimana kesalahan klasifikasi pada gesture yang mirip sering menjadi hambatan utama dalam aplikasi *real-world*. Efisiensi pelatihan yang dicapai dengan hanya 10 *epochs* total (5+5) juga menunjukkan keunggulan strategi *two-phase* dibandingkan pelatihan konvensional yang memerlukan puluhan hingga ratusan *epochs* untuk mencapai konvergensi optimal.

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted:**13/11/2025**; Accepted: 27**/12/2025**; Published: **15/04/2026**

JURNAL MEDIA INFORMATIKA [JUMIN]
Volume 7 Nomor 2 April 2026, Page 159-168
ISSN 2808-005X (media online)
Available Online at http://ejournal.sisfokomtek.org/index.php/jumin

4. KESIMPULAN

Penelitian Penelitian ini berhasil mengimplementasikan strategi two-phase transfer learning pada arsitektur MobileNetV2 untuk sistem pengenalan gerakan tangan menggunakan subset dataset HaGRID dengan 175.000 gambar mencakup tujuh kategori gesture. Pendekatan dua tahap yang terdiri dari fase pembekuan base layers (Phase 1) dan fase fine-tuning selektif (Phase 2) terbukti sangat efektif dalam meningkatkan performa model secara signifikan. Phase 1 dengan frozen backbone mencapai akurasi 75,83%, kemudian Phase 2 dengan unfreezing 55 layer terakhir meningkatkan akurasi menjadi 98,88% pada data validasi dan 98,86% pada data uji, mencapai peningkatan sebesar 23,05 poin persentase hanya dalam 10 epochs total dengan durasi pelatihan 6,5 jam.

Kontribusi utama penelitian ini adalah keberhasilan mengeliminasi seluruh confusion pairs yang sebelumnya mencapai 11 pasangan dengan tingkat kebingungan tertinggi 18,64% menjadi nihil dengan semua pasangan kelas memiliki confusion di bawah 2%. Peningkatan paling dramatis terjadi pada kelas peace yang melonjak dari 52,84% menjadi 97,20%, dan kelas rock dari 64,00% menjadi 98,12%, mendemonstrasikan efektivitas fine-tuning dalam mengatasi kesulitan membedakan gesture dengan karakteristik visual serupa. Konsistensi performa antara data validasi dan uji mengkonfirmasi kemampuan generalisasi yang sangat baik tanpa indikasi overfitting, dengan semua kelas mencapai precision, recall, dan F1-Score di atas 96%.

Implikasi praktis dari hasil penelitian ini meliputi: (1) pengembangan sistem kontrol gesture untuk perangkat IoT dan smart home yang responsif dengan latensi rendah; (2) implementasi antarmuka komunikasi bahasa isyarat pada aplikasi mobile untuk meningkatkan aksesibilitas penyandang disabilitas; (3) deployment model pada edge devices tanpa memerlukan koneksi cloud, memastikan privasi data pengguna dan mengurangi ketergantungan bandwidth; serta (4) framework pelatihan yang dapat diadaptasi untuk domain klasifikasi visual lainnya dengan dataset berskala besar pada resource-constrained environments.

Hasil penelitian membuktikan bahwa model lightweight MobileNetV2 dengan strategi two-phase transfer learning mampu mencapai akurasi kompetitif dengan arsitektur kompleks sambil mempertahankan efisiensi komputasi yang sesuai untuk implementasi pada perangkat dengan sumber daya terbatas. Strategi ini menawarkan solusi praktis untuk pengembangan sistem pengenalan gerakan tangan yang akurat dan efisien, dengan potensi aplikasi pada berbagai domain seperti kontrol perangkat interaktif, bahasa isyarat, dan antarmuka human-computer interaction. Penelitian lanjutan dapat diarahkan pada ekspansi jumlah kelas gesture, implementasi untuk pengenalan gerakan dinamis, dan optimasi lebih lanjut untuk deployment pada perangkat edge dengan konversi model ke format TensorFlow Lite atau TensorFlow.js untuk aplikasi real-time.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada semua pihak yang telah mendukung terlaksananya penelitian ini. Terima kasih kepada pengembang dataset HaGRID yang telah menyediakan dataset berkualitas tinggi untuk komunitas riset pengenalan gerakan tangan. Apresiasi juga disampaikan kepada komunitas TensorFlow dan Keras atas penyediaan *framework* yang powerful untuk implementasi *deep learning*. Ucapan terima kasih juga ditujukan kepada penyedia layanan komputasi yang memfasilitasi eksperimen menggunakan akselerasi GPU.

REFERENCES

[1] K. Aurangzeb, K. Javeed, M. Alhussein, I. Rida, S. I. Haider, and A. Parashar, "Deep Learning Approach for Hand Gesture Recognition: Applications in Deaf Communication and Healthcare," *Computers, Materials & Continua*, vol. 78, no. 1, pp. 127–144, 2024, doi: 10.32604/cmc.2023.042886.

[2] A. Mujahid *et al.*, "Real-Time Hand Gesture Recognition Based on Deep Learning YOLOv3 Model," *Applied Sciences*, vol. 11, no. 9, p. 4164, May 2021, doi: 10.3390/app11094164.

[3] M. A. Haq, L. N. Q. Huy, M. Ridlwan, and I. Naila, "Leveraging Self-Attention Mechanism for Deep Learning in Hand-Gesture Recognition System," *E3S Web of Conferences*, vol. 500, p. 01009, Mar. 2024, doi: 10.1051/e3sconf/202450001009.

[4] M. Rahim, A. S. M. Miah, H. Akash, J. Shin, M. Hossain, and M. Hossain, *An Advanced Deep Learning Based Three-Stream Hybrid Model for Dynamic Hand Gesture Recognition*. 2024. doi: 10.48550/arXiv.2408.08035.

[5] Yaseen, O.-J. Kwon, J. Kim, S. Jamil, J. Lee, and F. Ullah, "Next-Gen Dynamic Hand Gesture Recognition: MediaPipe, Inception-v3 and LSTM-Based Enhanced Deep Learning Model," *Electronics (Basel)*, vol. 13, no. 16, p. 3233, Aug. 2024, doi: 10.3390/electronics13163233.

[6] N. Zerrouki *et al.*, "Deep Learning for Hand Gesture Recognition in Virtual Museum Using Wearable Vision Sensors," *IEEE Sens J*, vol. 24, no. 6, pp. 8857–8869, Mar. 2024, doi: 10.1109/JSEN.2024.3354784.

[7] Md. A. A. Faisal, F. F. Abir, M. U. Ahmed, and M. A. R. Ahad, "Exploiting domain transformation and deep learning for hand gesture recognition using a low-cost dataglove," *Sci Rep*, vol. 12, no. 1, p. 21446, Dec. 2022, doi: 10.1038/s41598-022-25108-2.

This is an open access article under the CC–BY-SA license

Terakreditasi SINTA 5 SK :72/E/KPT/2024
Submitted:**13/11/2025**; Accepted: 27**/12/2025**; Published: **15/04/2026**

[8] Y. Gulzar, "Fruit Image Classification Model Based on MobileNetV2 with Deep Transfer Learning Technique," *Sustainability*, vol. 15, no. 3, p. 1906, Jan. 2023, doi: 10.3390/su15031906.

[9] R. K. Banoth and B. V. R. Murthy, "Soil Image Classification Using Transfer Learning Approach: MobileNetV2 with CNN," *SN Comput Sci*, vol. 5, no. 1, p. 199, Jan. 2024, doi: 10.1007/s42979-023-02500-x.

[10] T. Barman and S. Susan, "Multi-Label Remote Sensing Image Classification using MobileNetV2," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, IEEE, Jun. 2024, pp. 1–4. doi: 10.1109/ICCCNT61001.2024.10725506.

[11] Q. Xiang, X. Wang, R. Li, G. Zhang, J. Lai, and Q. Hu, "Fruit Image Classification Based on MobileNetV2 with Transfer Learning Technique," in *Proceedings of the 3rd International Conference on Computer Science and Application Engineering*, New York, NY, USA: ACM, Oct. 2019, pp. 1–7. doi: 10.1145/3331453.3361658.

[12] K. Alexander, K. Karina, N. Alexander, K. Roman, and M. Andrei, "HaGRID – HAnd Gesture Recognition Image Dataset," in *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, IEEE, Jan. 2024, pp. 4560–4569. doi: 10.1109/WACV57701.2024.00451.

[13] A. Nuzhdin, A. Nagaev, A. Sautin, A. Kapitanov, and K. Kvanchiani, "HaGRIDv2: 1M Images for Static and Dynamic Hand Gesture Recognition," 2025. doi: 10.24132/CSRN.2025-1.

[14] T. ValizadehAslani *et al.*, "Two-stage fine-tuning with ChatGPT data augmentation for learning class-imbalanced data," *Neurocomputing*, vol. 592, p. 127801, Aug. 2024, doi: 10.1016/j.neucom.2024.127801.

[15] T. ValizadehAslani *et al.*, "Two-Stage Fine-Tuning: A Novel Strategy for Learning Class-Imbalanced Data," Jul. 2022.

This is an open access article under the CC–BY-SA license