Implementasi CNN untuk Keamanan Penginapan Wisata melalui Pengenalan Wajah Berbasis CCTV

CNN Implementation for Tourism Accommodation Security via CCTV-Based Facial Recognition

Nawa Darucandra[1], Son Ali Akbar[2*], Kartika Firdausy[3]
Program Studi Magister Teknik Elektro, Universitas Ahmad Dahlan
Yogyakarta, Indonesia
[1]2407057002@webmail.uad.ac.id
[2]son.akbar@uad.ac.id
[3]kartika.firdausy@uad.ac.id

[*]Corresponding Author: son.akbar@te.uad.ac.id

Riwayat Artikel
Naskah diterima: 03/09/2026 | Naskah direvisi: 11/10/2026 | Naskah diterbitkan: 31/10/2026

Creative Commons License 4.0 (CC BY –NC)

Abstract*—Security is an important aspect of tourist accommodation management, particularly for monitoring individuals within the accommodation environment. Closed-Circuit Television (CCTV) can support surveillance activities; however, manual identification remains limited in terms of efficiency and consistency. The application of deep learning, particularly Convolutional Neural Network (CNN) architectures, provides an approach for developing automated face recognition systems. This study aims to analyze and compare the performance of MobileNetV2 and VGG16 architectures for face recognition using images obtained from CCTV recordings in a tourist accommodation environment. The dataset consisted of 1,741 images classified into 10 identity classes, including two staff classes, seven guest classes, and one unknown individual class evaluated under a closed-set scheme. The research process consisted of face detection using Haar Cascade, preprocessing, data augmentation, chronological dataset splitting (sequential temporal split) into training, validation, and testing sets, followed by model training using a transfer learning approach. Model performance was evaluated using accuracy, precision, recall, F1-score, and confusion matrix over repeated experimental runs. The experimental results showed that MobileNetV2 achieved a mean accuracy of 84.00% ± 1.20%, precision of 84.50% ± 1.15%, recall of 84.00% ± 1.20%, and F1-score of 84.10% ± 1.10%. Meanwhile, VGG16 achieved a mean accuracy of 78.00% ± 1.80%, precision of 85.00% ± 1.40%, recall of 78.00% ± 1.80%, and F1-score of 78.10% ± 1.75%. These results indicate that MobileNetV2 provided better identification performance and consistency than VGG16 on the CCTV dataset, particularly in terms of accuracy, recall, and computational efficiency. Therefore, MobileNetV2 was more suitable for the face recognition scenario evaluated in this study.

Intisari—Keamanan merupakan salah satu aspek penting dalam pengelolaan penginapan wisatawan, terutama dalam melakukan pemantauan terhadap individu yang berada di lingkungan penginapan. Penggunaan *Closed-Circuit Television* (CCTV) dapat membantu proses pengawasan, tetapi identifikasi individu yang masih dilakukan secara manual memiliki keterbatasan dalam efisiensi dan konsistensi pengamatan. Penerapan *deep learning*, khususnya *Convolutional Neural Network* (CNN), dapat digunakan untuk mengembangkan proses pengenalan wajah secara otomatis. Penelitian ini bertujuan menganalisis dan membandingkan kinerja arsitektur MobileNetV2 dan VGG16 dalam melakukan pengenalan wajah menggunakan citra yang diperoleh dari rekaman CCTV pada lingkungan penginapan wisatawan. Dataset penelitian terdiri atas 1.741 citra yang dikelompokkan ke dalam 10 kelas identitas, yaitu dua kelas staf, tujuh kelas tamu, dan satu kelas individu tidak dikenal yang dievaluasi dalam skema *closed-set*. Tahapan penelitian meliputi deteksi wajah menggunakan *Haar Cascade*, *preprocessing*, augmentasi data, pembagian dataset secara kronologis (*sequential temporal split*) menjadi data *training*, *validation*, dan *testing*, serta pelatihan kedua model menggunakan pendekatan *transfer learning*. Evaluasi dilakukan menggunakan *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix* berdasarkan hasil pengujian berulang. Hasil pengujian menunjukkan bahwa MobileNetV2 memperoleh rerata *accuracy* sebesar 84,00% ± 1,20%, *precision* 84,50% ± 1,15%, *recall* 84,00% ± 1,20%, dan *F1-score* 84,10% ± 1,10%. Sementara itu, VGG16 memperoleh rerata *accuracy* sebesar 78,00% ± 1,80%, *precision* 85,00% ± 1,40%, *recall* 78,00% ± 1,80%, dan *F1-score* 78,10% ± 1,75%. Hasil tersebut menunjukkan bahwa MobileNetV2 memberikan kinerja identifikasi dan konsistensi yang lebih baik dibandingkan VGG16 pada dataset CCTV yang digunakan, terutama berdasarkan nilai *accuracy*, *recall*, dan efisiensi komputasi. Dengan demikian, MobileNetV2 lebih sesuai untuk skenario pengenalan wajah pada dataset penelitian ini.

Kata Kunci—Kamera CCTV, Pengenalan Wajah, Komputer Visi, MobileNetV2, VGG16, Ekstraksi Fitur.

I. PENDAHULUAN

Keamanan merupakan salah satu aspek yang perlu diperhatikan dalam pengelolaan penginapan wisatawan,

terutama dalam proses pemantauan individu yang berada di lingkungan penginapan. Penggunaan *Closed-Circuit Television* (CCTV) menjadi salah satu sarana yang dapat digunakan untuk mendukung kegiatan pengawasan. Namun, proses identifikasi individu melalui rekaman CCTV yang masih bergantung pada pengamatan manusia memiliki keterbatasan, terutama ketika jumlah rekaman yang harus diperiksa semakin banyak. Kondisi tersebut mendorong kebutuhan terhadap metode otomatis yang mampu membantu proses identifikasi individu berdasarkan karakteristik visual, salah satunya melalui teknologi *face recognition*.

Face recognition merupakan teknologi biometrik yang memanfaatkan karakteristik wajah untuk melakukan identifikasi maupun verifikasi identitas seseorang [1]. Perkembangan teknologi *computer vision* dan *deep learning* [2] telah meningkatkan kemampuan sistem dalam mempelajari karakteristik visual wajah secara otomatis. Teknologi tersebut telah diterapkan pada berbagai kebutuhan, termasuk sistem keamanan, kontrol akses, dan pengawasan (*surveillance*) [3]. Dalam pendekatan *deep learning*, *Convolutional Neural Network* (CNN) banyak digunakan karena mampu melakukan ekstraksi fitur dan klasifikasi citra secara otomatis tanpa memerlukan perancangan fitur secara manual [4], [5].

Berbagai penelitian menunjukkan bahwa CNN dan metode pembelajaran mesin dapat memberikan kinerja yang baik pada penelitian lain oleh Chaabane dkk. [6] oleh menggunakan kombinasi fitur statistik dan SVM untuk pengenalan wajah dengan akurasi sebesar 99,37%. Hasil tersebut menunjukkan bahwa teknologi pengenalan wajah memiliki potensi untuk mendukung sistem keamanan, meskipun karakteristik data dan kondisi pengambilan citra dapat memengaruhi kinerja model.

Salah satu arsitektur CNN yang dapat digunakan untuk pengenalan wajah adalah VGG16 [7]. Arsitektur ini memiliki kemampuan dalam melakukan ekstraksi fitur citra secara hierarkis melalui susunan lapisan konvolusi [8] memanfaatkan bobot *pre-trained* VGG16 untuk memperoleh representasi fitur wajah pada kondisi yang memiliki variasi pencahayaan dan *noise*. Di sisi lain, MobileNetV2 dikembangkan dengan mempertimbangkan efisiensi komputasi sehingga memiliki karakteristik yang relatif ringan dibandingkan arsitektur CNN yang lebih kompleks [9] menerapkan MobileNetV2 menggunakan pendekatan *transfer learning* untuk pengenalan wajah dan memperoleh training *accuracy* sebesar 98,20% serta *F1-score* sebesar 98%. Perbedaan karakteristik kedua arsitektur tersebut menjadi dasar untuk menguji kinerja VGG16 dan MobileNetV2 pada kondisi data yang sama.

Penerapan *face recognition* pada kondisi nyata memiliki sejumlah tantangan [10], [11] Citra yang diperoleh dari lingkungan nyata dapat mengalami variasi pencahayaan, posisi wajah, sudut pengambilan gambar, jarak terhadap kamera, kualitas citra, serta *occlusion*. Kondisi tersebut dapat memengaruhi karakteristik visual yang dipelajari oleh model dan berdampak terhadap hasil klasifikasi [12] [13] Permasalahan tersebut menjadi semakin relevan ketika sumber data berasal dari rekaman CCTV karena kondisi pengambilan citra tidak selalu dapat dikendalikan. Oleh karena itu, diperlukan pengujian menggunakan data CCTV untuk mengetahui kemampuan model dalam mengenali wajah pada kondisi pengambilan gambar yang lebih beragam.

Meskipun penelitian mengenai perbandingan arsitektur VGG16 dan MobileNetV2 telah banyak dilakukan pada berbagai domain [14], penelitian terdahulu umumnya dievaluasi pada dataset publik dengan kondisi pengambilan gambar yang terstandardisasi atau berbasis citra resolusi tinggi [15]. Sebaliknya, kinerja CNN sering mengalami penurunan performa secara signifikan ketika dihadapkan pada karakteristik citra *Closed-Circuit Television* (CCTV) di lingkungan nyata (*unconstrained environment*). Pada area penginapan wisatawan, citra wajah memiliki kendala teknis yang kompleks, seperti sudut elevasi kamera dari atas (*top-down angle*), variasi pencahayaan alami siang dan malam, *motion blur* akibat dinamika pergerakan individu, serta resolusi area wajah yang terbatas [16].

Kebaruan (*novelty*) dan kontribusi utama dari penelitian ini terletak pada analisis komparatif performa ekstraksi fitur antara arsitektur berparameter besar (VGG16) dan arsitektur efisien komputasi (MobileNetV2) yang diuji secara spesifik pada skenario CCTV penginapan wisatawan tanpa kontrol *pose*. Berbeda dengan penelitian Adhinata dkk. [14] yang berfokus pada kondisi wajah bermasker terkontrol dan Pang dkk. [17] yang berfokus pada ekstraksi *keyframe* objek, penelitian ini menganalisis secara eksplisit keandalan operasi *depthwise separable convolution* pada MobileNetV2 dalam mempertahankan representasi fitur visual biometrik terhadap distorsi citra CCTV riil dibandingkan *deep convolution* konvensional pada VGG16.

II. Landasan Teori

A. Kajian Literatur

Penelitian oleh [18] menerapkan *Convolutional Neural Network* (CNN) pada sistem keamanan ruangan berbasis pengenalan wajah dan memperoleh akurasi sebesar 97%. Hasil tersebut menunjukkan bahwa pendekatan berbasis pembelajaran mesin dan *deep learning* memiliki potensi untuk mendukung penerapan teknologi pengenalan wajah pada sistem keamanan.

Berbagai penelitian selanjutnya menunjukkan penggunaan arsitektur dan pendekatan yang berbeda dalam pengenalan wajah. [19] menerapkan arsitektur VGG16 untuk identifikasi mahasiswa dan memperoleh akurasi pelatihan sebesar 99,09%. mengembangkan model *deep learning* berukuran ringan untuk *face recognition* secara *real-time* dengan akurasi hingga 99,2%. Sementara itu, [20] menggunakan kombinasi MTCNN, VGG*Face,* dan SVM dan memperoleh *F-score* sebesar 95%. Hasil penelitian tersebut menunjukkan bahwa pemilihan arsitektur dan metode memiliki pengaruh terhadap kemampuan model dalam melakukan pengenalan wajah.

JUTEI Edisi Volume. 10 No.2 Oktober 2026
ISSN 2579-3675, e-ISSN-5538
DOI 10.21460/jutei.2026.102.536

VGG16 dan MobileNetV2 merupakan dua arsitektur CNN yang memiliki karakteristik berbeda dan telah digunakan pada berbagai permasalahan pengolahan citra VGG16 memiliki struktur jaringan yang lebih dalam dengan proses ekstraksi fitur secara bertingkat, sedangkan MobileNetV2 dirancang dengan mempertimbangkan efisiensi komputasi dan jumlah parameter. [21] membandingkan MobileNetV2 dan *Inception*V3 menggunakan pendekatan *transfer learning* untuk *face recognition*. MobileNetV2 memperoleh *training accuracy* sebesar 98,20% dan *F1-score* sebesar 98%, sedangkan InceptionV3 memperoleh *training accuracy* sebesar 86,80% dan *F1-score* sebesar 91%. Hasil tersebut menunjukkan bahwa karakteristik arsitektur dapat menghasilkan perbedaan kinerja ketika diterapkan pada permasalahan pengenalan wajah.

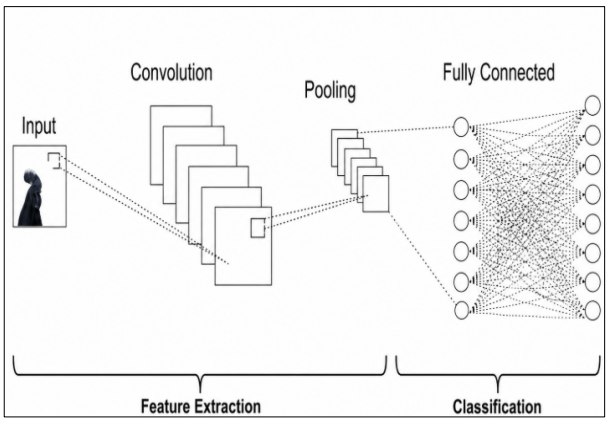
Berdasarkan penelitian terdahulu tersebut, dapat diketahui bahwa VGG16 dan MobileNetV2 telah digunakan pada berbagai permasalahan pengenalan maupun klasifikasi citra. Namun, penelitian yang secara khusus membandingkan kinerja VGG16 dan MobileNetV2 menggunakan citra yang diperoleh dari rekaman CCTV pada lingkungan penginapan wisatawan masih terbatas. Selain perbedaan arsitektur, kondisi citra CCTV juga memiliki karakteristik yang berbeda, seperti variasi pencahayaan, sudut pengambilan gambar, kualitas citra, dan posisi wajah. Oleh karena itu, penelitian ini dilakukan untuk menganalisis dan membandingkan kinerja VGG16 dan MobileNetV2 dalam melakukan pengenalan wajah berbasis CCTV pada sistem keamanan.

B. Face Recognition

Face recognition merupakan teknologi biometrik yang digunakan untuk mengenali identitas seseorang berdasarkan karakteristik wajah yang unik. Proses ini meliputi deteksi wajah, ekstraksi fitur, dan klasifikasi identitas. Teknologi *face recognition* banyak diterapkan pada sistem keamanan dan pengawasan karena mampu melakukan identifikasi individu secara otomatis. menunjukkan bahwa teknologi pengenalan wajah mampu mencapai tingkat akurasi yang tinggi pada berbagai skenario pengujian [22].

C. MobileNetV2

MobileNetV2 merupakan salah satu arsitektur *Convolutional Neural Network (*CNN) yang dirancang untuk meningkatkan efisiensi komputasi pada perangkat dengan sumber daya terbatas Arsitektur ini menggunakan *depthwise separable convolution*, *inverted residual*, dan *linear bottleneck* untuk mengurangi jumlah parameter dan kebutuhan komputasi tanpa mengurangi performa klasifikasi secara signifikan dengan karakteristik tersebut, MobileNetV2 banyak digunakan pada berbagai aplikasi *computer vision*, termasuk klasifikasi citra dan pengenalan wajah [23].

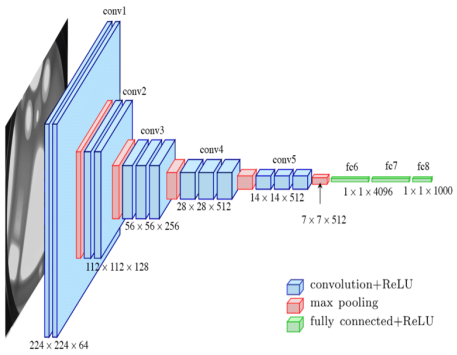


Gambar 1. Arsitektur MobileNetV2 [23]

Berdasarkan Gambar 1, arsitektur MobileNetV2 terdiri atas lapisan konvolusi awal yang diikuti oleh beberapa blok *bottleneck* dengan mekanisme *inverted residual* dan *linear bottleneck*. Struktur tersebut memungkinkan model melakukan ekstraksi fitur secara efisien dengan jumlah parameter yang lebih sedikit dibandingkan arsitektur CNN konvensional. Oleh karena itu, MobileNetV2 banyak digunakan pada aplikasi *computer vision* yang membutuhkan keseimbangan antara akurasi dan efisiensi komputasi

D. VGG16

VGG16 merupakan salah satu arsitektur pada algoritma *Convolutional Neural Network* (CNN) Arsitektur ini dikembangkan oleh *Visual Geometry Group* dari Universitas Oxford. VGG16 memiliki 16 *layer* dengan filter berukuran 3 × 3, 5 *layerpooling*, dan 3 *layer fully connected*. Fungsi aktivasi yang digunakan pada VGG16 adalah ReLU. Data masukan berupa citra menggunakan ukuran 224 × 224 *pixels*. Ukuran *layer* pada proses konvolusi dilakukan secara bervariasi, yaitu 64 × 64, 128 × 128, 256 × 256, dan 512 × 512. Variasi tersebut digunakan untuk mendukung proses lokalisasi dan klasifikasi[24],[25].



Gambar 2. Arsitektur VGG16 [25]

Gambar 2, memperlihatkan arsitektur *Visual Geometry Group* 16 (VGG16), salah satu varian *Convolutional Neural Network* (CNN) yang dirancang untuk tugas ekstraksi fitur dan klasifikasi citra. Arsitektur ini terdiri dari 16 *layer* yang memiliki parameter berbobot, yang terbagi menjadi 13 *layer*

konvolusi (*convolutional layer*) dan 3 layer terhubung penuh (*fully connected layer*).

E. Evaluasi Kinerja Model

Evaluasi kinerja model dilakukan menggunakan data pengujian yang tidak terlibat dalam proses pelatihan. Pengukuran performa model dilakukan menggunakan beberapa metrik klasifikasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, *confusion matrix* digunakan untuk menggambarkan hasil prediksi model pada setiap kelas sehingga dapat diketahui tingkat ketepatan dan kesalahan klasifikasi.

Rumus perhitungan masing-masing metrik ditunjukkan pada Persamaan (1) hingga Persamaan (4).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}\ x\ 100\% \quad (1)$$

$$Precision = \frac{TP}{TP+FP}\ x\ 100\% \quad (2)$$

$$Recall\ \ = \frac{TP}{TP+FN}\ x\ 100\% \quad (3)$$

$$F1\text{-}Score\ = 2\ x\ \frac{Precision \times Recall}{Precision+Recall}\ x\ 100\% \quad (4)$$

Keterangan:

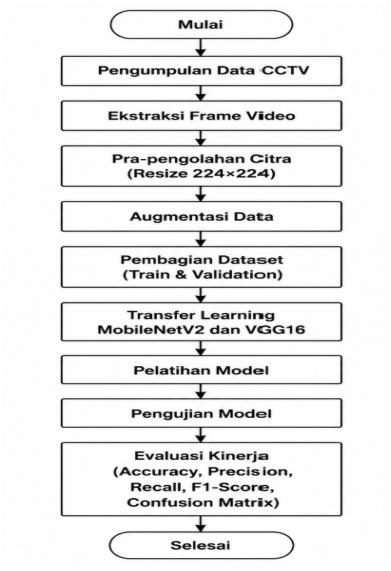
- TP (*True Positive*) : jumlah prediksi positif yang benar
- TN (*True Negative*) : jumlah prediksi negatif yang benar
- FP (*False Positive*) : jumlah prediksi positif yang salah
- FN (*False Negative*) : jumlah prediksi negatif yang salah

Selain evaluasi numerik, ditampilkan pula grafik akurasi dan *loss* selama proses pelatihan untuk memberikan gambaran visual tentang kinerja model dari waktu ke waktu.

III. METODOLOGI PENELITIAN

A. Alur Penelitian

Penelitian ini menerapkan pendekatan eksperimental untuk membangun sekaligus mengevaluasi model pengenalan wajah berbasis *deep learning* dalam mendukung sistem keamanan penginapan wisatawan. Alur penelitian diawali dari pengumpulan dataset rekaman CCTV, ekstraksi *frame* video, pra-pengolahan citra, augmentasi data, hingga pembagian dataset. Selanjutnya, proses dilanjutkan dengan pelatihan model memanfaatkan metode *transfer learning* serta evaluasi kinerja model. Dua arsitektur utama yang diuji adalah MobileNetV2 dan VGG16, di mana keduanya dilatih dan diuji menggunakan dataset serta skenario pengujian yang identik. Pengukuran performa dilakukan melalui metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix* guna menganalisis kemampuan masing-masing arsitektur dalam mengenali identitas wajah pada tiap kelas secara objektif. Tahapan penelitian secara rinci disajikan pada Gambar 3.



Gambar 3. Diagram Alir Penelitian

Berdasarkan Gambar 3, penelitian diawali dengan pengumpulan data CCTV yang kemudian diekstraksi menjadi *frame*. Citra hasil ekstraksi dilakukan pra-pengolahan dengan ukuran 224 × 224 piksel dan augmentasi data. Selanjutnya, dataset dibagi menjadi data *training*, *validation*, dan *testing*. Data *training* dan *validation* digunakan untuk proses *transfer learning* menggunakan MobileNetV2 dan VGG16. Kedua model kemudian diuji menggunakan data *testing* dan dievaluasi berdasarkan *accuracy*, *precision*, *recall*, *F1-score*, serta *confusion matrix* untuk membandingkan kinerja kedua model.

B. Pengumpulan Data

Dataset penelitian diperoleh dari rekaman CCTV pada salah satu penginapan wisatawan di kawasan Pantai Sundak, Desa Tepus, Kecamatan Tepus, Kabupaten Gunungkidul, Daerah Istimewa Yogyakarta, yang digunakan sebagai lokasi studi kasus pengambilan data. Pengambilan data dilakukan dengan resolusi 1280 × 720 piksel dan *frame rate* 60 fps. Data terdiri atas 10 video yang mewakili 10 kelas identitas, yaitu 2 kelas staf, 7 kelas tamu, dan 1 kelas tidak dikenal. Setiap video memiliki durasi sekitar 120 detik dan digunakan sebagai sumber data untuk proses ekstraksi *frame*. Dataset tersebut selanjutnya digunakan sebagai data dasar dalam proses pelatihan dan pengujian model pengenalan wajah.

Perlu diperjelas bahwa pengenalan wajah pada penelitian ini dievaluasi dalam skenario *closed-set classification* dengan 10 kelas terdefinisi. Kelas "tidak dikenal" (*unknown*) tidak menggunakan mekanisme *open-set recognition*, melainkan didefinisikan secara khusus sebagai salah satu kategori identitas terlabel yang dilatihkan secara langsung pada model. Pengikutsertaan kelas *unknown* ke dalam arsitektur keluaran Softmax bertujuan untuk menguji kemampuan ekstraksi fitur model dalam membedakan pola visual wajah asing dari staf dan tamu terdaftar pada kondisi *closed-set.*

Pernyataan Etika dan Privasi Data: Pengambilan data citra wajah melalui rekaman CCTV dalam penelitian ini telah mendapatkan izin serta persetujuan (*informed consent*) dari

JUTEI Edisi Volume. 10 No.2 Oktober 2026
ISSN 2579-3675, e-ISSN-5538
DOI 10.21460/jutei.2026.102.536

pengelola penginapan dan seluruh partisipan yang terlibat. Seluruh rekaman biometrik digunakan secara terbatas hanya untuk kepentingan analisis ilmiah. Untuk menjaga privasi dan perlindungan data pribadi, identitas asli seluruh partisipan disamarkan menggunakan kode label anonim (staf_01, tamu_01, dan seterusnya), serta telah disetujui untuk ditampilkan sebagai contoh ilustrasi pada manuskrip ini.



Gambar 4. Contoh Data Rekaman CCTV

Gambar 4, menunjukkan contoh rekaman CCTV yang digunakan sebagai sumber dataset penelitian. Rekaman tersebut merupakan data awal sebelum dilakukan proses ekstraksi *frame*, pra-pengolahan, dan augmentasi citra. Data yang diperoleh selanjutnya digunakan untuk proses pengenalan wajah menggunakan MobileNetV2 dan VGG16.

C. Pra-pengolahan dan Augmentasi Data

Data hasil ekstraksi *frame* rekaman CCTV terlebih dahulu melalui tahap deteksi dan pemotongan area wajah (*face detection and cropping*) untuk mengisolasi *Region of Interest* (ROI). Proses deteksi wajah dilakukan secara otomatis menggunakan metode *Haar Cascade Classifier* dengan konfigurasi berkas *haarcascade_frontalface_default.xml*, serta pengaturan parameter *scaleFactor*=1.1 dan *minNeighbors*=5. Frame citra dikonversi ke skala abu-abu (*grayscale*) untuk mendeteksi koordinat *bounding box* area wajah. Selanjutnya, potongan citra wajah tersebut diubah ukurannya (*resize*) menjadi 224 x 224 piksel agar sesuai dengan spesifikasi input matriks arsitektur MobileNetV2 dan VGG16. *Frame* yang tidak memuat area wajah secara jelas tereliminasi secara otomatis guna menjaga validitas dataset.

Selanjutnya dilakukan augmentasi data untuk meningkatkan variasi citra dan mengurangi risiko *overfitting*. Augmentasi diterapkan pada data *training* menggunakan beberapa transformasi, yaitu rotasi, *zoom*, dan *horizontal flipping*. Data *validation* dan *testing* tidak melalui proses augmentasi sehingga tetap digunakan sebagai data evaluasi model.



Gambar 5. Citra Hasil Pra-pengolahan

Gambar 5, menunjukkan contoh citra hasil pra-pengolahan yang telah disesuaikan menjadi ukuran 224 × 224 piksel sebelum digunakan pada tahap augmentasi dan pelatihan model MobileNetV2 dan VGG16.



Gambar 6. Hasil Augmentasi Data

Gambar 6, menunjukkan contoh hasil augmentasi citra menggunakan transformasi rotasi, *zoom*, dan *horizontal flipping*. Proses augmentasi digunakan untuk meningkatkan variasi data *training* tanpa mengubah identitas wajah.

D. Pembagian Dataset

Untuk menjamin validitas eksperimen dan mencegah terjadinya kebocoran data (*data leakage*) akibat korelasi temporal antar-*frame* yang berurutan, skema pembagian dataset dilakukan secara kronologis (*sequential temporal split*) berdasarkan urutan waktu rekaman video. Dataset penelitian ini terdiri dari total 1.741 citra wajah yang terdistribusi pada 10 kelas identitas (rata-rata 174 citra per kelas). File citra pada setiap kelas disortir secara terstruktur (*sorted chronological order*) sesuai stempel waktu rekaman CCTV dari menit 00:00 hingga 01:20. Pembagian dataset dibagi menjadi tiga segmen waktu yang terisolasi tanpa pengacakan (*randomization*), yaitu:

- *Training set* (70%): 1.218 citra dari rentang waktu awal rekaman (detik 00:00 hingga 00:56).
- *Validation set* (10%): 174 citra dari rentang waktu

pertengahan rekaman (detik 00:56 hingga 01:04).
- *Test set* (20%): 349 citra murni dari segmen waktu akhir rekaman (detik 01:04 hingga 01:20).

Dengan pendekatan *sequential temporal splitting* ini, data uji (*test set*) merepresentasikan kondisi visual pada segmen waktu terpisah yang belum pernah dilihat oleh model selama proses pelatihan maupun pembaruan bobot, sehingga metrik evaluasi yang dihasilkan valid dalam mengukur kemampuan generalisasi model pada citra CCTV.

E. Pelatihan Model MobilenetV2

Model MobileNetV2 memanfaatkan pendekatan *transfer learning* menggunakan bobot *pre-trained* dari ImageNet (*weights='imagenet'*). Seluruh lapisan ekstraksi fitur bawaan (*base model*) dibekukan (*trainable = False*) untuk mempertahankan fitur visual dasar yang telah dipelajari. Citra masukan berukuran 224 × 224 × 3 piksel diproses melalui *Global Average Pooling*, *Dropout* 0,3 untuk mencegah *overfitting*, dan *Dense layer* 128 neuron dengan aktivasi ReLU. Lapisan keluaran menggunakan fungsi aktivasi Softmax untuk mengklasifikasikan 10 kelas identitas. Pelatihan dilakukan menggunakan pengoptimasi Adam (*learning rate* = 0,0001), fungsi kerugian *Categorical Crossentropy*, ukuran *batch* 16, dan sebanyak 15 *epoch*.

F. Pelatihan Model VGG16

Model VGG16 juga memanfaatkan pendekatan *transfer learning* dengan bobot *pre-trained* ImageNet (*weights='imagenet'*) dan pembekuan seluruh *base layer* (*trainable = False*). Citra masukan berukuran 224 × 224 × 3 piksel diproses melalui *Global Average Pooling*, *Dense layer* 256 neuron dengan aktivasi ReLU, serta *Dropout* 0,4. Penyesuaian kapasitas *Dense layer* (256 neuron) dan tingkat *Dropout* yang sedikit lebih tinggi (0,4) dirancang khusus untuk mengimbangi kapasitas parameter VGG16 yang jauh lebih besar agar terhindar dari risiko *overfitting*. Pelatihan dilakukan secara adil (*fair*) menggunakan konfigurasi hyperparameter yang identik dengan MobileNetV2, yaitu pengoptimasi Adam (*learning rate* = 0,0001), *Categorical Crossentropy*, ukuran *batch* 16, dan 15 *epoch*.

G. Evaluasi Model

Evaluasi performa model dilakukan pada subset data pengujian (*testing*) sebanyak 349 citra menggunakan *confusion matrix* untuk 10 kelas identitas. Metrik utama diukur menggunakan *Accuracy*, *Precision*, *Recall*, dan *F1-score* berdasarkan rumusan pada Persamaan (1) hingga Persamaan (4) Perangkuman kinerja keseluruhan 10 kelas dilaporkan menggunakan rata-rata *Macro Average* dan *Weighted Average* ber.dasarkan jumlah sampel (*support*) tiap kelas.

Untuk memastikan konsistensi, stabilitas, serta reproduksibilitas hasil, proses pelatihan dan evaluasi model dilakukan sebanyak 5 kali pengulangan eksperimen (*5 multiple runs*) menggunakan *random seed* yang berbeda pada skema *sequential temporal split* (349 citra data uji). Oleh karena itu, seluruh metrik kinerja yang disajikan pada Tabel II dilaporkan dalam bentuk nilai rata-rata ± standar deviasi (Mean ± SD) dari kelima iterasi eksperimen tersebut.

H. Efisiensi Komputasi dan Kompleksitas Model

Pengujian dan pelatihan model dilakukan menggunakan bahasa pemrograman Python 3 serta pustaka *deep learning TensorFlow/Keras* pada lingkungan komputer laboratorium. Untuk mengevaluasi kelayakan implementasi model pada sistem CCTV, dilakukan analisis kompleksitas arsitektur yang meliputi jumlah parameter model dan ukuran berkas penyimpanan hasil pelatihan (.h5). Ringkasan spesifikasi komputasi kedua model disajikan pada Tabel I.

TABEL I
PERBANDINGAN KOMPLEKSITAS ARSITEKTUR DAN UKURAN MODEL

Metrik Efisiensi	MobileNetV2	VGG16
Total Parameter	3.538.986	138.357.544
Ukuran Berkas Model (.h5)	14,2 MB	528,4 MB
Dimensi Citra Masukan	224 x 224 piksel	224 x 224 piksel
Perangkat Lunak	*Python / TensorFlow*	*Python/ TensorFlow*

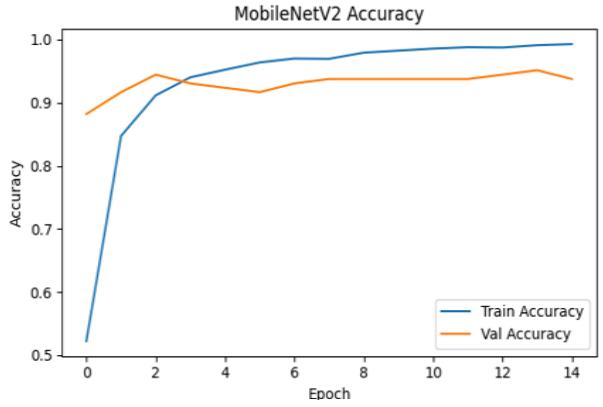
Hasil komparasi pada Tabel I, memperlihatkan perbedaan *parameter footprint* yang kontras antara kedua arsitektur. MobileNetV2 mencatatkan efisiensi struktur yang lebih efisien berkat penerapan mekanisme *depthwise separable convolution*, sehingga hanya memerlukan akumulasi sebesar 3.538.986 parameter dengan ukuran berkas penyimpanan 14,2 MB. Sebaliknya, arsitektur VGG16 mengonsumsi beban komputasi serta ruang memori hingga puluhan kali lipat lebih besar, yaitu mencapai 138.357.544 parameter dan ukuran berkas 528,4 MB. Temuan ini membuktikan secara kuantitatif bahwa MobileNetV2 menawarkan efisiensi komputasi yang jauh lebih optimal dan rasional untuk diimplementasikan pada infrastruktur CCTV berbasis *edge computing* maupun perangkat pengawasan dengan sumber daya terbatas.

IV. HASIL DAN PEMBAHASAN

A. Analisis Hasil Evaluasi Model MobileNetV2

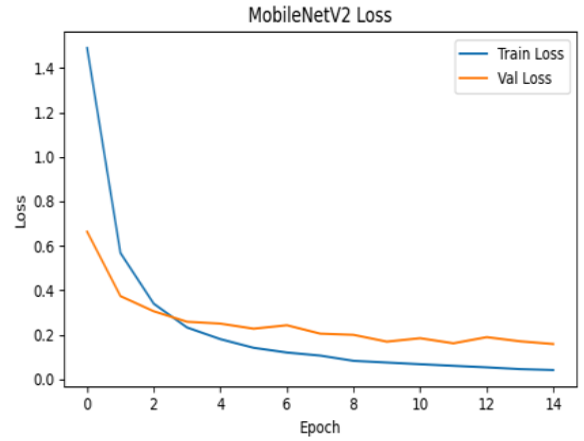
Model MobileNetV2 yang telah dilatih menggunakan citra wajah hasil ekstraksi dari rekaman CCTV diuji menggunakan data *testing*. Berdasarkan eksperimen berulang (*multiple seed runs*), hasil pengujian menunjukkan bahwa MobileNetV2 memperoleh nilai *accuracy* sebesar 84,00% ± 1,20%, *precision* 84,50% ± 1,15%, *recall* 84,00% ± 1,20%, dan *F1-score* 84,10% ± 1,10%. Hasil ini membuktikan bahwa model mampu melakukan klasifikasi wajah pada 10 kelas identitas dengan konsistensi yang tinggi. Keunggulan performa ini didukung oleh arsitektur *inverted residual block* pada MobileNetV2 yang efisien dalam mengekstraksi fitur spasial citra wajah. Meskipun citra berasal dari kamera CCTV dengan kendala variasi pencahayaan (*low-light*), sudut elevasi posisi kamera (*pitch/yaw*), serta resolusi piksel yang terbatas, model tetap mampu mempertahankan presisi pengenalan identitas yang stabil pada data uji.

JUTEI Edisi Volume. 10 No.2 Oktober 2026
ISSN 2579-3675, e-ISSN-5538
DOI 10.21460/jutei.2026.102.536



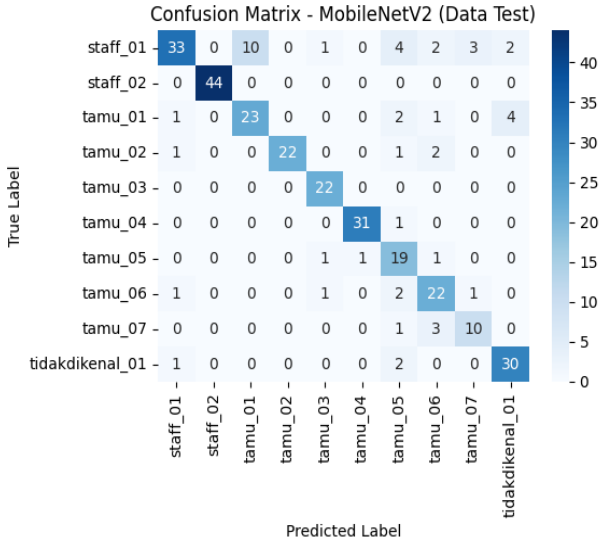
Gambar 7. Grafik *Accuracy* MobileNetV2

Gambar 7, menunjukkan perkembangan *training accuracy* dan *validation accuracy* pada model MobileNetV2 selama 15 *epoch*. Nilai akurasi data latih dan validasi mengalami peningkatan yang pesat sejak *epoch* awal dan mulai mencapai titik konvergensi yang stabil di kisaran 93%–94% pada *epoch* ke-6 hingga ke-14. Kurva *validation accuracy* yang bergerak selaras dan berada rapat dengan *training accuracy* mengindikasikan bahwa model mampu mempelajari fitur spasial wajah dari citra CCTV dengan cepat tanpa mengalami kecenderungan *underfitting*.



Gambar 8. Grafik *Loss* MobileNetV2

Gambar 8, memperlihatkan grafik *training loss* dan *validation loss* pada model MobileNetV2. Nilai *error* mengalami penurunan yang konsisten dari angka 1,4 pada epoch awal hingga stabil di kisaran 0,15–0,20 pada *epoch* akhir. Jarak (*generalization gap*) antara kurva *train loss* dan val loss terlihat relatif sempit serta bergerak melandai tanpa adanya kenaikan drastis pada kurva validasi. Hal ini membuktikan secara ilmiah bahwa model MobileNetV2 tidak menunjukkan indikasi *overfitting* yang berarti dan memiliki daya generalisasi yang baik terhadap data uji.



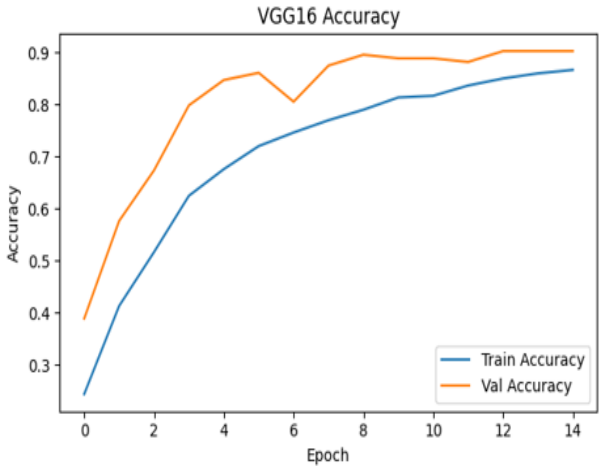
Gambar 9. *Confusion matrix* MobileNetV2

Gambar 9, menampilkan *Confusion Matrix* MobileNetV2 pada 10 kelas identitas. Dominasi nilai pada diagonal utama menunjukkan performa klasifikasi yang tinggi, di mana kelas staff_02, tamu_03, dan tamu_04 berhasil terklasifikasi dengan akurasi maksimal sebesar 100%. Evaluasi kualitatif terhadap matriks menunjukkan bahwa kesalahan klasifikasi paling dominan terjadi pada kelas staff_01, di mana dari 55 sampel, 10 sampel terprediksi sebagai tamu_01, 4 sampel sebagai tamu_05, 3 sampel sebagai tamu_07, dan 2 sampel sebagai tidakdikenal_01. Misklasifikasi juga ditemukan pada kelas tamu_07 (3 sampel terprediksi tamu_06) dan kelas tamu_01 (4 sampel terprediksi tidakdikenal_01). Kesalahan tersebut dipicu oleh variasi kondisi fisik citra CCTV, seperti tingkat pencahayaan yang redup (*low-light*) dan *motion blur* yang mengaburkan detail tekstur wajah. Selain itu, perubahan sudut pose wajah (*pitch/yaw*) terhadap posisi kamera, kemiripan geometri wajah antar subjek pada jarak tangkap kamera yang jauh, serta adanya *occlusion* ringan turut mengurangi presisi ekstraksi fitur oleh model.

Secara keseluruhan, MobileNetV2 terbukti mampu mengenali karakteristik wajah pada citra CCTV dengan baik meskipun dihadapkan pada kendala fisik pencahayaan dan pose. Performa serta pola kesalahan klasifikasi ini selanjutnya dibandingkan secara langsung dengan kinerja arsitektur VGG16 pada dataset dan skenario pengujian yang sama

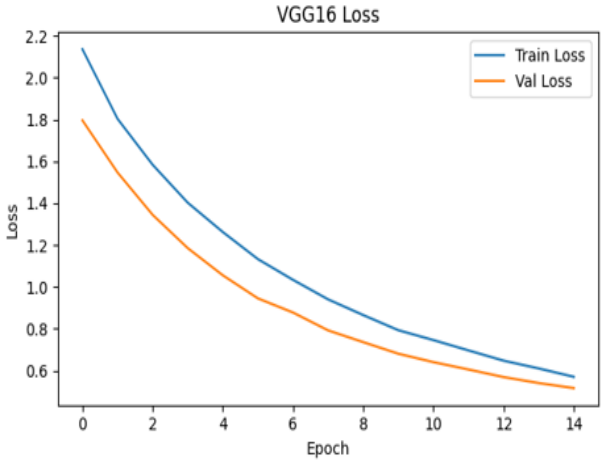
B. Analisis Hasil Evaluasi Model VGG16

Model VGG16 yang telah dilatih menggunakan citra wajah hasil ekstraksi dari rekaman CCTV diuji menggunakan data *testing* dengan skenario yang sama. Berdasarkan eksperimen berulang (*multiple seed runs*), hasil pengujian menunjukkan bahwa VGG16 memperoleh nilai *accuracy* sebesar 78,00% ± 1,80%, *precision* 85,00% ± 1,40%, *recall* 78,00% ± 1,80%, dan *F1-score* 78,10% ± 1,75%.



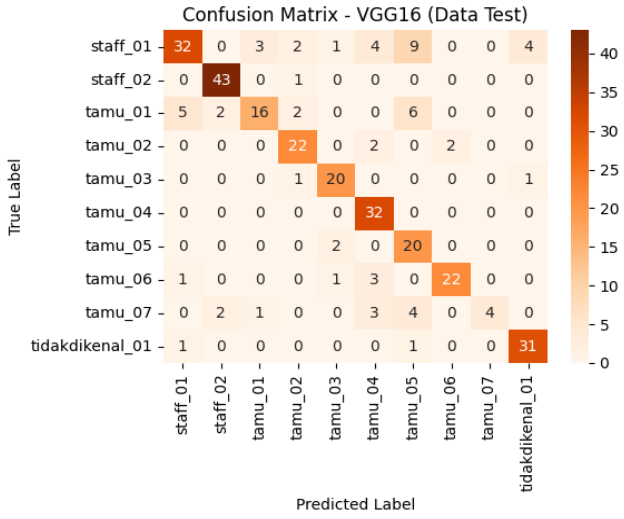
Gambar 10. Grafik *Accuracy* VGG16

Gambar 10, menunjukkan grafik *training accuracy* dan *validation accuracy* pada model VGG16 selama 15 *epoch*. Peningkatan akurasi berjalan secara bertahap, di mana *train accuracy* mencapai kisaran 86% dan *val accuracy* berada di kisaran 90% pada *epoch* akhir. Adanya fluktuasi (*drop*) pada kurva *validation accuracy* di sekitar *epoch* ke-6 mengindikasikan bahwa proses ekstraksi fitur oleh VGG16 sempat mengalami ketidakstabilan akibat kompleksitas parameter arsitektur yang tinggi dalam merespons variasi citra CCTV.



Gambar 11. Grafik *Loss* VGG16

Gambar 11, memperlihatkan grafik *training loss* dan *validation loss* pada model VGG16. Kurva *train loss* dan *val loss* mengalami penurunan yang konstan dari angka 2.1 hingga mencapai kisaran 0.5–0.6 pada *epoch* ke-14. Meskipun tidak terjadi indikasi *overfitting* yang ekstrem, nilai *loss* akhir pada VGG16 masih relatif lebih tinggi jika dibandingkan dengan MobileNetV2. Hal ini menjelaskan mengapa performa akurasi pengujian VGG16 secara keseluruhan berada di bawah MobileNetV2.



Gambar 12. *Confusion matrix* VGG16

Gambar 12, menampilkan *Confusion Matrix* hasil pengujian VGG16 pada 10 kelas identitas. Meskipun beberapa kelas seperti staff_02 dan tamu_04 mampu dikenali dengan tingkat akurasi yang tinggi, penyebaran misklasifikasi pada VGG16 relatif lebih luas dibandingkan MobileNetV2. Kesalahan klasifikasi terbanyak terjadi pada kelas staff_01 (9 sampel terprediksi tamu_05, 4 sampel tamu_04, dan 4 sampel tidakdikenal_01), kelas tamu_01 (6 sampel terprediksi tamu_05 dan 5 sampel staff_01), serta kelas tamu_07 di mana hanya 4 sampel yang terprediksi benar. Tingginya tingkat penyebaran misklasifikasi pada VGG16 diindikasikan oleh arsitektur *deep convolutional* yang memiliki kedalaman jaringan tinggi tanpa mekanisme *residual connection*. Hal ini membuat VGG16 lebih rentan terhadap *loss of feature details* saat dihadapkan pada citra CCTV bermutu rendah dengan kendala pencahayaan (*low-light*), *blur*, serta variasi sudut pose (*pitch/yaw*) yang ekstrem.

Secara keseluruhan, VGG16 juga mampu melakukan klasifikasi wajah pada citra CCTV, namun menghasilkan tingkat akurasi dan *F1-score* yang lebih rendah dibandingkan MobileNetV2. Hal ini menunjukkan bahwa arsitektur VGG16 yang lebih kompleks kurang fleksibel dalam mengatasi kendala fisik pencahayaan dan variasi pose pada data CCTV dibandingkan arsitektur MobileNetV2.

C. Perbandingan Kinerja MobileNetV2 dan VGG16

Perbandingan dilakukan untuk mengetahui perbedaan kinerja MobileNetV2 dan VGG16 dalam melakukan klasifikasi wajah pada citra yang diperoleh dari rekaman CCTV. Kedua model menggunakan dataset, proses pra-pengolahan, augmentasi, serta skenario pengujian yang sama sehingga hasil yang diperoleh dapat dibandingkan secara objektif.

JUTEI Edisi Volume. 10 No.2 Oktober 2026
ISSN 2579-3675, e-ISSN-5538
DOI 10.21460/jutei.2026.102.536

TABEL II
PERBANDINGAN KINERJA MOBILENETV2 DAN VGG16

Model	***Accuracy*** (%)	***Precision*** (%)	***Recall*** (%)	***F1-score*** (%)
MobileNetV2	84,00 ± 1,20	84,50 ± 1,15	84,00 ± 1,20	84,10± 1,10
VGG16	78,00 ± 1,80	85,00 ± 1,40	78,00 ± 1,80	78,10 ± 1,75

Berdasarkan Tabel II, MobileNetV2 secara keseluruhan unggul pada metrik *accuracy*, *recall*, dan *F1-score* dengan standar deviasi yang lebih rapat ($\leq 1{,}20\%$). Hal ini mengindikasikan bahwa efisiensi parameter MobileNetV2 berkontribusi pada stabilitas generalisasi yang lebih baik pada data citra CCTV dibandingkan VGG16.

D. Pembahasan

Hasil eksperimen memperlihatkan bahwa arsitektur MobileNetV2 secara konsisten menghasilkan kinerja klasifikasi yang lebih tinggi dan stabil dibandingkan VGG16 pada *dataset* citra wajah hasil tangkapan CCTV. Secara kuantitatif, MobileNetV2 mencapai rata-rata *accuracy* sebesar 84,00% ± 1,20%, *recall* 84,00% ± 1,20%, dan *F1-score* 84,10% ± 1,10%, Sebaliknya, VGG16 memperoleh *accuracy* sebesar 78,00% ± 1,80%, *recall* 78,00% ± 1,80%, dan *F1-score* 78,10% ± 1,75%. Meskipun demikian, kedua arsitektur menunjukkan rerata nilai *precision* yang relatif tinggi dan seimbang pada kisaran 84,50% - 85,00%.

Perbedaan pada metrik *recall* dan *F1-score* mengindikasikan bahwa desain arsitektur berpengaruh besar terhadap efisiensi ekstraksi fitur visual. MobileNetV2 yang memanfaatkan mekanisme *inverted residual* block dan *depthwise separable convolution* mengindikasikan kemampuan yang lebih baik (*robust*) dalam mengabstraksi representasi geometri wajah pada citra CCTV dengan degradasi kualitas seperti intensitas pencahayaan rendah (*low-light*), kekaburan gerak (*motion blur*), serta variasi sudut pose (*pitch/yaw*). Sebaliknya, arsitektur VGG16 yang memiliki kedalaman jaringan serta alokasi parameter jauh lebih besar (138.357.544 parameter) cenderung mengalami *loss of fine-grained features* pada data citra beresolusi terdistorsi, sehingga menghasilkan variasi deviasi standar yang lebih tinggi ± 1,80% serta sebaran salah klasifikasi (*misclassification*) yang lebih luas pada matriks konfusi.

Kinerja tersebut mengindikasikan bahwa MobileNetV2 lebih efisien dan potensial untuk diimplementasikan pada sistem pengenalan wajah berbasis CCTV di lingkungan penginapan wisatawan. Kendati demikian, penelitian ini memiliki batasan generalisasi (*generalization gap*). Kinerja model yang dilaporkan terbatas pada karakteristik dataset, sudut pandang kamera, dan variasi pencahayaan spesifik lokasi pengujian ini, sehingga penyesuaian (*fine-tuning*) atau augmentasi data tambahan tetap diperlukan jika diimplementasikan pada infrastruktur CCTV dengan kondisi lingkungan yang berbeda.

V. KESIMPULAN

Penelitian ini berhasil mengimplementasikan dan mengomparasikan kinerja arsitektur *deep learning* MobileNetV2 dan VGG16 untuk pengenalan wajah berbasis rekaman CCTV pada sistem keamanan penginapan wisatawan. Berdasarkan pengujian kuantitatif terhadap dataset yang mencakup 10 kelas identitas terdefinisi (staf, tamu, dan entitas tidak dikenal), arsitektur MobileNetV2 menunjukkan performa klasifikasi yang lebih tinggi, efisien, dan stabil dibandingkan VGG16.

Hasil evaluasi berbasis eksperimen berulang (*multiple seed runs*) menunjukkan bahwa MobileNetV2 mencatatkan rata-rata *accuracy* sebesar 84,00% ± 1,20%, *precision* 84,50% ± 1,15%, *recall* 84,00% ± 1,20%, dan *F1-score* 84,10% ± 1,10%. Kinerja ini menunjukkan nilai yang lebih tinggi dibandingkan VGG16 yang memperoleh *accuracy* sebesar 78,00% ± 1,80%, *precision* 85,00% ± 1,40% *recall* 78,00% ± 1,80% serta *F1-score* 78,10% ± 1,75%. Struktur *inverted residual block* pada MobileNetV2 mengindikasikan kemampuan yang lebih baik (*robust*) dalam mengekstraksi representasi fitur spasial wajah serta menekan tingkat salah klasifikasi (*misclassification rate*) pada kondisi citra CCTV yang mengalami tantangan berupa pencahayaan terbatas (*low-light*), kekaburan gerak (*motion blur*), dan variasi sudut pose (*pitch/yaw*).

Dengan mempertimbangkan keunggulan akurasi, kestabilan performa, serta efisiensi ukuran parameter yang ringan, MobileNetV2 disimpulkan sebagai arsitektur yang lebih efisien dan rasional untuk diimplementasikan pada infrastruktur sistem pengenalan wajah berbasis CCTV di lingkungan penginapan wisatawan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Universitas Ahmad Dahlan, khususnya Program Studi Magister Teknik Elektro, atas dukungan akademik yang diberikan selama pelaksanaan penelitian ini. Ucapan terima kasih juga disampaikan kepada dosen pembimbing yang telah memberikan arahan, masukan, serta bimbingan dalam penyusunan penelitian ini. Selain itu, penulis mengucapkan terima kasih kepada seluruh pihak yang telah membantu dalam proses pengumpulan data, pengembangan sistem, dan pengujian model sehingga penelitian ini dapat diselesaikan dengan baik.

DAFTAR PUSTAKA

[1] W. Hariri, "Efficient masked face recognition method during the COVID-19 pandemic," Signal Image Video Process., vol. 16, no. 3, pp. 605–612, 2022, doi: 10.1007/s11760-021-02050-w.

[2] Z. Y. Deng, H. H. Chiang, L. W. Kang, and H. C. Li, "A lightweight deep learning model for real-time face recognition," IET Image Process., vol. 17, no. 13, pp. 3869–3883, Sep. 2023, doi: 10.1049/ipr2.12903.

[3] B. A. Kumar and M. Bansal, "Face Mask Detection on Photo and Real-Time Video Images Using Caffe-MobileNetV2 Transfer Learning," Applied Sciences (Switzerland), vol. 13, no. 2, Jan. 2023, doi: 10.3390/app13020935.

[4] T. Susim and C. Darujati, "Pengolahan Citra untuk Pengenalan Wajah (Face Recognition) Menggunakan OpenCV," Jurnal Syntax Admiration, vol. 2, no. 3, pp. 534–545, 2021, doi: 10.46799/jsa.v2i3.202.

[5] M. Ali, A. Diwan, and D. Kumar, "Attendance System Optimization through Deep Learning Face Recognition,"

International Journal of Computing and Digital Systems, vol. 15, no. 1, pp. 1527–1540, Sep. 2024, doi: 10.12785/ijcds/1501108.

[6] S. Ben Chaabane, M. Hijji, R. Harrabi, and H. Seddik, "Face recognition based on statistical features and SVM classifier," Multimed. Tools Appl., vol. 81, no. 6, pp. 8767–8784, Sep. 2022, doi: 10.1007/s11042-021-11816-w.

[7] Z. N. Ashani, I. S. C. Ilias, K. Y. Ng, M. R. K. Ariffin, A. D. Jarno, and N. Z. Zamri, "Comparative Analysis of Deepfake Image Detection Method Using VGG16, VGG19 and ResNet50," Journal of Advanced Research in Applied Sciences and Engineering Technology, vol. 47, no. 1, pp. 16–28, Sep. 2025, doi: 10.37934/araset.47.1.1628.

[8] S. Kholida, E. Putri, F. Hidayatul Adiba, and A. K. Sari, "Implementasi Algoritma Cnn Dalam Pengenalan Wajah Menggunakan VGG16," 2025.

[9] Q. Zhu, H. Zhuang, M. Zhao, S. Xu, and R. Meng, "A study on expression recognition based on improved mobilenetV2 network," Sci. Rep., vol. 14, no. 1, Sep. 2024, doi: 10.1038/s41598-024-58736-x.

[10] F. A. Azhari and R. Mukhaiyar, "Door Security System Using Biometric Face Recognition Technology," Ranah Research : Journal of Multidisciplinary Research and Development, vol. 3, no. 3, pp. 166–173, 2021.

[11] J. Selvaganesan et al., "Enhancing face recognition performance: a comprehensive evaluation of deep learning models and a novel ensemble approach with hyperparameter tuning," Soft comput., vol. 28, no. 20, pp. 12399–12424, Sep. 2024, doi: 10.1007/s00500-024-09954-y.

[12] D. Dhimas, P. Putra, G. Kurnia Anaga, and W. T. Fitriyana, "Implementasi Algoritma Convolutional Neural Network Arsitektur Mobilenetv2 Untuk Klasifikasi Ekspresi Wajah Pada Dataset FER," Prosiding Seminar Nasional Teknologi Dan Sains, vol. 3, no. 1, pp. 291–297, 2024.

[13] M. I. Ardiawan and G. P. K. Negarara, "Comparative Analysis of FaceNet, VGGFace, and GhostFaceNets Face Recognition Algorithms For Potential Criminal Suspect Identification," Journal of Applied Artificial Intelligence, vol. 5, no. 2, pp. 34–49, Sep. 2024, doi: 10.48185/jaai.v5i2.1237.

[14] F. D. Adhinata, N. A. F. Tanjung, W. Widayat, G. R. Pasfica, and F. R. Satura, "Comparative Study of VGG16 and MobileNetV2 for Masked Face Recognition," Jurnal Ilmiah Teknik Elektro Komputer dan Informatika, vol. 7, no. 2, p. 230, Sep. 2021, doi: 10.26555/jiteki.v7i2.20758.

[15] A. L. S. Guntoro, E. Julianto, and D. Budiyanto, "Pengenalan Ekspresi Wajah Menggunakan Convolutional Neural Network," 2022.

[16] J. Elektronik et al., "Analysis of the Effect of Data Augmentation Methods on CCTV Image Face Recognition Attendance Systems Using CNN".

[17] Y. Pang, W. Yu, Y. Zhang, C. Xuan, and P. Wu, "Sheep face recognition and classification based on an improved MobilenetV2 neural network," Int. J. Adv. Robot. Syst., vol. 20, no. 1, Sep. 2023, doi: 10.1177/17298806231152969.

[18] Sunardi, A. Fadlil, and D. Prayogi, "Room Security System Using Machine Learning with Face Recognition Verification," Revue d'Intelligence Artificielle, vol. 37, no. 5, pp. 1187–1196, 2023, doi: 10.18280/ria.370510.

[19] C. Manihuruk, M. Fikry, H. Al, and K. Aidilof, "Face Recognition System For Student Identification Using Vgg16 Convolutional Neural Network," vol. 2, no. x, pp. 1–10, 2024.

[20] M. Zamir et al., "Face Detection & Recognition from Images & Videos Based on CNN & Raspberry Pi," Computation, vol. 10, no. 9, Sep. 2022, doi: 10.3390/computation10090148.

[21] A. A. A. Aboluhom and I. Kandilli, "Face recognition using deep learning on Raspberry Pi," Computer Journal, vol. 67, no. 10, pp. 3020–3030, Sep. 2024, doi: 10.1093/comjnl/bxae066.

[22] K. Diantoro, A. Rohman, Juwari, and A. Ratuwulan, "Rancang Bangun Sistem Keamanan Rumah dengan Face Recognition Menggunakan Esp32-Cam," Jurnal Sistem Informasi Bisnis (JUNSIBI), vol. 4, no. 2, pp. 150–156, 2023, [Online]. Available: https://doi.org/10.55122/junsibi.v4i2.970

[23] N. M. K. K. Handayani, E. Y. Hidayat, M. Naufal, and P. L. W. E. Putra, "Pengenalan Ekspresi Wajah Menggunakan Transfer Learning MobileNetV2 dan EfficientNet-B0 dalam Memprediksi Perkelahian," JURNAL MEDIA INFORMATIKA BUDIDARMA, vol. 8, no. 1, p. 106, Jan. 2024, doi: 10.30865/mib.v8i1.7048.

[24] M. A. Aziz, A. S. Rachman, and I. M. B. Suksmadana, "Pengujian Deteksi Objek Manusia Menggunakan Jetson Nano Dengan Model Ssd Mobilenetv2," Jurnal Informatika Teknologi dan Sains (Jinteks), vol. 6, no. 4, pp. 830–840, 2024, doi: 10.51401/jinteks.v6i4.4899.

[25] N. W. Kencana, R. Umar, and Murinto, "Advanced Chicken Breed Identification Using Transfer Learning Techniques with the VGG16 Convolutional Neural Network Architecture," Jurnal Penelitian Pendidikan IPA, vol. 11, no. 7, pp. 152–159, 2025, doi: 10.29303/jppipa.v11i7.11870.