SMARTICS Journal, Vol.12 No.02 2026. p155-164
ISSN online: 2476-9754, ISSN print: 2623-0429
DOI : https://doi.org/10.21067/smartics.v12i2.14778

Evaluasi Pengenalan Wajah Wisudawan *Real-time* Berbasis *Visual Geometry Group-Face* dan *Cosine Distance*

Abbu Solihin Alhakim[a*], Rudi Heriansyah[b], Zaid Romegar Mair[c]
[abc]Universitas Indo Global Mandiri, Indonesia
[*]correspondence email : alhakimabusolihin@gmail.com

Abstract*—Manually identifying graduates can slow down the procession and increase the risk of misidentification. This study evaluates a real-time facial recognition system using VGG-Face for feature extraction and Cosine Distance for matching. Faces are detected using OpenCV's Haar Cascade and represented as 4096-dimensional embeddings via the VGG-Face model (accessed through DeepFace). Identity is determined based on the minimum Cosine Distance value using a 0.6 threshold and validated by consistent detection across three consecutive frames. Testing involved 20 individuals (10 registered graduates and 10 unregistered individuals) across four conditions: distances of 50 cm, 100 cm, and 100–150 cm, as well as a 50 cm distance with the subject wearing glasses. Out of 885 recorded recognition events, the results yielded 416 True Positive s, 440 True Negatives, 22 false Positive s, and 7 false Negatives. Evaluation metrics showed an accuracy of 96.7232%, precision of 94.9772%, recall of 98.3452%, and an F1-score of 96.6318%. The average system latency was 0.6877 seconds. For the registered group, varying the distance up to 150 cm resulted in minimal changes to the recognition rate, whereas wearing glasses reduced the recognition rate to 95%. The findings indicate that the system supports rapid graduate identification in semi-controlled environments, however, the rejection of unregistered faces and robustness against occlusion require further improvement.

***Index Terms*—Cosine Distance; DeepFace; face recognition; real-time system; VGG-Face.**

Abstrak*–Identifikasi wisudawan secara manual berpotensi memperlambat alur prosesi dan meningkatkan risiko kesalahan dalam penyebutan identitas. Penelitian ini mengevaluasi sistem pengenalan wajah real-time menggunakan VGG-Face sebagai ekstraktor fitur dan Cosine Distance sebagai metode pencocokan. Wajah dideteksi menggunakan Haar Cascade pada OpenCV, kemudian direpresentasikan menjadi embedding berdimensi 4096 melalui model VGG-Face yang diakses menggunakan DeepFace. Identitas ditentukan berdasarkan nilai Cosine Distance minimum dengan threshold 0,6 dan validasi kemunculan identitas selama tiga frame berturut-turut. Pengujian melibatkan 20 individu yang terdiri atas 10 wisudawan terdaftar dan 10 individu tidak terdaftar pada empat kondisi, yaitu jarak 50 cm, 100 cm, lebih dari 100 hingga 150 cm, serta 50 cm dengan kacamata. Dari 885 event recognition yang tercatat, diperoleh 416 True Positive, 440 True Negative, 22 false Positive, dan 7 false Negative. Hasil evaluasi menunjukkan accuracy sebesar 96,7232%, precision sebesar 94,9772%, recall sebesar 98,3452%, dan F1-score sebesar 96,6318%. Rata-rata latency sistem sebesar 0,6877 detik. Pada kelompok terdaftar, variasi jarak hingga 150 cm memberikan perubahan recognition rate yang terbatas, sedangkan penggunaan kacamata menurunkan recognition rate menjadi 95%. Hasil penelitian menunjukkan bahwa sistem mampu mendukung identifikasi wisudawan secara cepat pada kondisi semi-terkontrol, tetapi penolakan wajah tidak terdaftar dan ketahanan terhadap oklusi masih perlu ditingkatkan.

***Kata Kunci*—Cosine Distance; DeepFace; pengenalan wajah; sistem real-time; VGG-Face.**

I. PENDAHULUAN

Pengenalan wajah merupakan salah satu penerapan *computer vision* yang banyak digunakan untuk melakukan verifikasi dan identifikasi individu berdasarkan representasi karakteristik wajah. Perkembangan tekno*log*i *deep learning* telah mendorong penggunaan representasi fitur berdimensi tinggi dalam sistem pengenalan wajah [1], [2]. Li et al. membahas perkembangan tekno*log*i pengenalan wajah dan penerapan metode berbasis pembelajaran mendalam untuk meningkatkan kemampuan sistem dalam merepresentasikan karakteristik wajah [2]. Representasi tersebut dapat digunakan untuk membandingkan kesamaan antara citra masukan dan data referensi [1], [3].

VGG-*Face* merupakan model berbasis *Convolutional Neural Network* (CNN) yang digunakan untuk menghasilkan representasi fitur wajah [4]. CNN memiliki kemampuan mengekstraksi fitur visual secara bertingkat sehingga dapat digunakan sebagai dasar pemrosesan citra wajah. Pada penelitian ini, VGG-

Face digunakan sebagai model *pre-trained* pada tahap *inference* sehingga penelitian berfokus pada integrasi dan evaluasi sistem, bukan pengembangan arsitektur CNN baru.

Penerapan CNN pada pengolahan citra menunjukkan kemampuan metode tersebut dalam mengekstraksi fitur dan mengenali pola visual. Mair dan Irfani menerapkan CNN untuk pengenalan pola citra, sedangkan Nurlita et al. menggunakan pendekatan CNN dalam mendeteksi kemiringan tulisan tangan dengan mempertimbangkan karakteristik dan orientasi citra [5], [6]. Temuan tersebut mendukung penggunaan representasi berbasis CNN sebagai dasar ekstraksi fitur pada penelitian ini.

Pada sistem pengenalan wajah berbasis *embedding*, hasil ekstraksi fitur perlu dibandingkan dengan representasi wajah referensi untuk menentukan identitas yang paling sesuai. *Cosine Similarity* dapat digunakan untuk mengukur kesamaan dua vektor, sedangkan *Cosine Distance* digunakan dalam penelitian ini sebagai ukuran jarak dengan nilai yang lebih kecil menunjukkan representasi yang semakin dekat [7]–[8]. Pendekatan berbasis *cosine* telah digunakan pada berbagai sistem pengenalan dan verifikasi wajah untuk membandingkan representasi fitur.

Penelitian Suhardi et al. menggabungkan YOLOv8n, VGG-*Face*, dan *Cosine Similarity* untuk sistem presensi perkuliahan dan melaporkan *accuracy* sekitar 90% tanpa augmentasi serta meningkat hingga sekitar 97% setelah augmentasi [9]. Meskipun memiliki kesamaan pada penggunaan VGG-*Face* dan pencocokan berbasis *cosine*, konteks penelitian tersebut berbeda dengan identifikasi wisudawan secara *real-time*. Sistem pada penelitian ini juga harus menghadapi wajah yang tidak terdaftar, mempertahankan keputusan pada aliran video, serta mengevaluasi variasi jarak dan penggunaan kacamata.

Penelitian berbasis video *real-time* juga perlu mempertimbangkan karakteristik pemrosesan streaming dan waktu *respons* sistem. Rizaldi et al. menggunakan OpenCV dalam sistem pengenalan gerakan berbasis video *real-time* dan melakukan pengujian terhadap variasi jarak, sedangkan Heriansyah dan Utomo mengevaluasi performa pengolahan citra dan pemrosesan video dengan mempertimbangkan efisiensi komputasi [10], [11]. Aspek tersebut relevan dengan evaluasi *latency* dan variasi jarak pada penelitian ini.

Keberadaan individu yang tidak terdapat dalam basis data merupakan karakteristik *open-set face recognition*, ketika sistem harus mengenali identitas yang diketahui sekaligus menolak individu yang tidak diketahui [12], [13]. Penggunaan *threshold* yang terlalu longgar dapat meningkatkan *false acceptance*, sedangkan *threshold* yang terlalu ketat dapat meningkatkan *false rejection*. Selain itu, kualitas citra dan oklusi dapat memengaruhi kemampuan sistem dalam membentuk representasi wajah yang konsisten [14], [15]. Dengan demikian, evaluasi sistem tidak hanya perlu mempertimbangkan keberhasilan pengenalan, tetapi juga pola kesalahan dan kondisi yang memengaruhi kualitas representasi.

Berdasarkan permasalahan tersebut, penelitian ini mengevaluasi sistem pengenalan wajah *real-time* untuk identifikasi wisudawan menggunakan VGG-*Face* melalui *framework DeepFace* dan *Cosine Distance*. Sistem mengintegrasikan deteksi wajah menggunakan *Haar Cascade*, enam *embedding* referensi untuk setiap wisudawan, pencocokan berdasarkan *Cosine Distance* minimum, *threshold* 0,6, dan validasi identitas selama tiga *frame* berturut-turut. Evaluasi dilakukan terhadap 10 wisudawan terdaftar dan 10 individu tidak terdaftar pada empat kondisi pengujian, yaitu jarak 50 cm, 100 cm, lebih dari 100 hingga 150 cm, serta 50 cm dengan kacamata. Penelitian ini bertujuan menganalisis performa klasifikasi, *latency*, pengaruh jarak dan oklusi ringan terhadap pengenalan, serta pola *false acceptance*.

II. TINJAUAN PUSTAKA

Pengenalan wajah merupakan bidang *computer vision* yang bertujuan mengenali atau memverifikasi identitas individu berdasarkan karakteristik visual wajah. Perkembangan *deep learning* memungkinkan wajah direpresentasikan ke dalam *embedding* berdimensi tinggi yang kemudian dibandingkan dengan data referensi [1]–[3].

A. *Convolutional Neural Network*

Convolutional Neural Network (CNN) merupakan arsitektur *deep learning* yang banyak digunakan dalam pengolahan citra karena mampu mengekstraksi fitur visual secara otomatis melalui proses konvolusi [3]. Mair dan Irfani menunjukkan penerapan CNN untuk pengenalan pola citra, sedangkan Nurlita et al. menerapkan CNN dalam pengolahan citra dengan mempertimbangkan variasi orientasi objek [5], [6].

VGG-*Face* merupakan model pengenalan wajah berbasis CNN yang digunakan untuk menghasilkan representasi fitur wajah [4]. Pada penelitian ini, VGG-*Face* digunakan sebagai model *pre-trained* pada

tahap *inference* untuk menghasilkan *embedding* wajah tanpa retraining maupun *fine-tuning*.

B. *DeepFace* dan *Face Embedding*

DeepFace merupakan sistem pengenalan wajah berbasis *deep learning* yang menghasilkan representasi fitur wajah untuk proses verifikasi dan identifikasi [16]. Dalam implementasi penelitian ini, model VGG-*Face* diakses melalui *DeepFace* untuk menghasilkan *face embedding*.

Setiap wajah direpresentasikan menjadi *embedding* berdimensi 4096. *Embedding* tersebut digunakan sebagai data referensi dan dibandingkan dengan *embedding* dari wajah masukan. Penggunaan beberapa *embedding* untuk satu identitas memungkinkan sistem mempertahankan variasi karakteristik wajah dalam proses pencocokan.

C. *Cosine Distance*

Cosine Similarity digunakan untuk mengukur tingkat kesamaan dua vektor berdasarkan orientasinya, sedangkan *Cosine Distance* digunakan sebagai ukuran jarak dengan nilai yang lebih kecil menunjukkan representasi yang semakin dekat [7], [8]. Cahyono dan Imelda menerapkan CNN *MobileFaceNet* yang dikombinasikan dengan *Cosine Similarity* untuk verifikasi wajah, menunjukkan bahwa perbandingan representasi berbasis *cosine* dapat digunakan sebagai dasar pengambilan keputusan pada sistem verifikasi wajah [8]. Pendekatan berbasis *cosine* tersebut relevan dengan penelitian ini karena *embedding* wajah digunakan sebagai representasi yang dibandingkan terhadap data referensi.

$$S_{cos}(\mathbf{x},\mathbf{y}) = \frac{\sum_{i=1}^{n} x_i\, y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}} \qquad (1)$$

Pada penelitian ini, *embedding* wajah masukan dibandingkan dengan seluruh *embedding* referensi dan nilai *Cosine Distance* minimum digunakan untuk menentukan kandidat identitas. Setiap wisudawan memiliki enam *embedding* referensi yang disimpan secara terpisah.

$$D_{cos}(\mathbf{x},\mathbf{y}) = 1 - S_{cos}(\mathbf{x},\mathbf{y}) \qquad (2)$$

dengan x dan y merupakan dua *embedding* wajah yang dibandingkan, x_i dan y_i merupakan elemen ke-*i* dari masing-masing *embedding*, *n* merupakan jumlah dimensi *embedding*, S_{cos} merupakan *Cosine Similarity*, dan *Dcos* merupakan *Cosine Distance*.

Keputusan penerimaan identitas menggunakan *threshold Cosine Distance* sebesar 0,6. Nilai yang memenuhi batas penerimaan dapat dipertimbangkan sebagai identitas yang sesuai, sedangkan wajah yang tidak memenuhi *threshold* dikategorikan sebagai tidak dikenali.

D. *Open-set Face recognition*

Pada penerapan pengenalan wajah secara nyata, tidak seluruh wajah yang ditangkap kamera terdapat dalam basis data. Kondisi tersebut berkaitan dengan *open-set recognition*, yaitu kemampuan sistem untuk mengenali identitas yang diketahui sekaligus menolak identitas yang tidak tersedia pada data referensi [12], [13].

Threshold menjadi komponen penting dalam kondisi tersebut. *Threshold* yang terlalu longgar dapat meningkatkan *false acceptance* terhadap wajah asing, sedangkan *threshold* yang terlalu ketat dapat meningkatkan *false rejection* pada wajah terdaftar. Oleh karena itu, penelitian ini tidak hanya mengevaluasi kemampuan mengenali wisudawan terdaftar, tetapi juga kemampuan menolak individu yang tidak terdaftar.

Selain *threshold*, penelitian menerapkan validasi kemunculan identitas selama tiga *frame* berturut-turut untuk membantu menstabilkan keputusan pada aliran video *real-time*.

E. Evaluasi Kinerja

Evaluasi sistem dilakukan menggunakan confusion matrix yang terdiri atas *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Keempat komponen tersebut digunakan untuk menghitung *accuracy*, *precision*, *recall*, dan *F1-score* sebagai ukuran kinerja klasifikasi.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \qquad (3)$$

$$Precision = \frac{TP}{TP + FP} \qquad (4)$$

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

$$F_1 = 2\frac{P \times R}{P + R} \quad (6)$$

Selain kinerja klasifikasi, *latency* digunakan untuk mengevaluasi waktu pemrosesan sistem pada setiap *event recognition*. Rizaldi et al. menunjukkan relevansi pemrosesan video *real-time* menggunakan OpenCV, sedangkan Heriansyah dan Utomo membahas evaluasi performa pengolahan citra dan video dengan mempertimbangkan aspek komputasi dan waktu *respons* [10], [11].

III. METODE

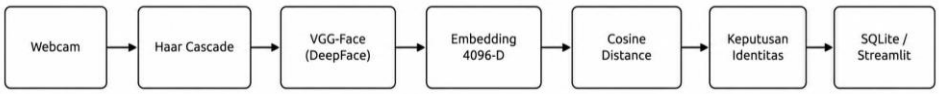
Penelitian ini menggunakan pendekatan kuantitatif melalui eksperimen komputasional untuk mengevaluasi sistem pengenalan wajah *real-time*. Sistem menerima video dari *webcam*, mendeteksi wajah menggunakan *Haar Cascade* pada OpenCV, menghasilkan *embedding* menggunakan model *pre-trained* VGG-*Face* melalui *DeepFace*, kemudian membandingkan *embedding* masukan dengan *embedding* referensi menggunakan *Cosine Distance*. Keputusan identitas ditentukan berdasarkan *threshold* dan distabilkan melalui kemunculan identitas pada tiga *frame* berturut-turut.

A. Arsitektur Sistem dan Lingkungan Implementasi

Alur pemrosesan sistem dimulai dari pengambilan *frame* video melalui *webcam*. Setiap *frame* dikonversi ke *grayscale* dan diproses menggunakan *Haar Cascade Classifier* untuk menentukan lokasi wajah. Parameter deteksi yang digunakan adalah *scaleFactor* 1,3 dan *minNeighbors* 5. Area wajah yang terdeteksi kemudian dipotong dan diteruskan ke fungsi *DeepFace*.represent() untuk menghasilkan *embedding*.

Karena proses deteksi wajah telah dilakukan sebelumnya oleh OpenCV, parameter *enforce_detection=False* digunakan agar *DeepFace* tidak melakukan validasi deteksi ulang. Konfigurasi alignment mengikuti perilaku default *DeepFace*. Model VGG-*Face* digunakan dalam kondisi *pre-trained* pada tahap *inference* tanpa retraining maupun *fine-tuning*. Dengan demikian, parameter seperti *epoch*, *optimizer*, *learning rate*, *batch size*, dan *loss function* tidak menjadi variabel penelitian.

Setiap wajah direpresentasikan menjadi *embedding* berdimensi 4096. *Embedding* masukan dibandingkan dengan seluruh *embedding* referensi yang tersimpan pada basis data. Sistem memilih nilai *Cosine Distance* terkecil sebagai kandidat identitas, kemudian menerapkan *threshold* 0,6 dan mekanisme validasi tiga *frame* berturut-turut sebelum identitas dinyatakan stabil.



Gambar 1. Arsitektur pemrosesan sistem pengenalan wajah real-time

Lingkungan implementasi penelitian ditunjukkan pada Tabel 1.

Tabel 1. Lingkungan Implementasi Sistem

Komponen	Spesifikasi
Perangkat	Lenovo LOQ 15IRX9
Prosesor	Intel Core i5 generasi ke-13
Memori	12 GB RAM
GPU	NVIDIA GeForce RTX 4050
Sistem Operasi	Windows 11
Kamera	Webcam internal HD 720p
Python	3.10.0
DeepFace	0.0.99
OpenCV-Python	4.13.0.92
TensorFlow / Keras	2.21.0 / 3.12.1
Streamlit	1.56.0
Basis Data	SQLite

Seluruh eksperimen dijalankan secara lokal pada perangkat dan posisi kamera yang sama. Penggunaan *CPU* atau *GPU* oleh *TensorFlow* tidak diverifikasi sebagai variabel eksperimen, sehingga *latency* diperlakukan sebagai performa sistem pada konfigurasi perangkat yang digunakan.

B. Data Referensi, Subjek, dan Protokol Pengujian

Subjek penelitian terdiri atas 20 individu yang dibagi menjadi 10 wisudawan terdaftar dan 10 individu tidak terdaftar. Pada tahap registrasi, setiap wisudawan direkam dalam enam variasi wajah, yaitu frontal-netral, tersenyum, melihat ke kiri, melihat ke kanan, jarak dekat, dan jarak lebih jauh. Setiap citra menghasilkan satu *embedding* yang disimpan secara terpisah dan dikaitkan dengan identitas wisudawan yang sama. Keenam *embedding* tidak dirata-ratakan sehingga satu identitas mempunyai beberapa representasi yang dapat digunakan untuk menghadapi variasi pose, ekspresi, dan skala wajah.

Citra registrasi berbeda dari *frame* yang digunakan pada pengujian. Seluruh subjek mengetahui penggunaan data wajah dalam penelitian dan menyatakan kesediaannya mengikuti pengujian. Identitas individu tidak ditampilkan dalam artikel dan analisis per subjek menggunakan kode anonim.

Pengujian dilakukan pada pencahayaan normal dengan posisi kamera dipertahankan. Empat kondisi digunakan untuk mengevaluasi variasi jarak dan oklusi ringan berupa penggunaan kacamata. Distribusi *event recognition* ditunjukkan pada Tabel 2.

Tabel 2. Kondisi Pengujian dan Distribusi Event Recognition

Kondisi	Tujuan	Terdaftar	Tidak Terdaftar	Total
50 cm normal	Kinerja pada jarak dekat	110	109	219
100 cm normal	Kinerja pada jarak menengah	90	99	189
>100–150 cm	Toleransi pada jarak lebih jauh	143	138	281
50 cm + kacamata	Pengaruh oklusi sebagian wajah	80	116	196
Total		423	462	885

Pengujian dilakukan secara berulang pada setiap subjek dan kondisi. Jumlah *event* tidak diperlakukan sebagai kuota tetap karena dapat berubah akibat wajah yang tidak terdeteksi, penyesuaian jarak, atau identitas yang belum stabil selama tiga *frame*. Wajah yang sama sekali tidak terdeteksi tidak menghasilkan *log*, sedangkan wajah yang berhasil dideteksi tetapi tidak memenuhi kriteria pengenalan dicatat sebagai Tidak Dikenali.

Dengan demikian, 885 *event* merupakan data yang telah mencapai tahap *recognition* dan pencatatan, bukan jumlah seluruh paparan wajah terhadap kamera. Setiap *event* disimpan pada tabel *log_recognition* yang memuat identitas, waktu, *confidence*, *Cosine Distance*, *latency*, ukuran wajah, status, dan estimasi jarak. Seluruh *event* pada periode eksperimen, termasuk kegagalan identifikasi, dipertahankan untuk analisis.

C. Representasi Wajah dan Aturan Keputusan

Untuk setiap wajah masukan, VGG-*Face* melalui *DeepFace* menghasilkan *embedding* berdimensi 4096. *Embedding* tersebut dibandingkan dengan seluruh *embedding* referensi yang tersedia pada basis data. Nilai *Cosine Distance* terkecil dari seluruh pasangan *embedding* digunakan sebagai jarak minimum, sedangkan identitas pemilik *embedding* tersebut menjadi kandidat identitas.

$$d_{min} = \min_{j} \min_{k\in\{1,\ldots,6\}} D_{cos}(\mathbf{e}, \mathbf{r}_{j,k}) \quad (7)$$

dengan (e) merupakan *embedding* wajah masukan, ($r_{j,k}$) merupakan *embedding* referensi ke-(k) dari identitas ke-(j), (D_{cos}) merupakan *Cosine Distance*, dan (d_{min}) merupakan nilai jarak minimum dari seluruh pasangan *embedding*.

Kandidat identitas diterima apabila nilai (d_{min}) memenuhi *threshold* sebesar 0,6. *Threshold* ditetapkan sebelum eksperimen dan tidak dioptimasi menggunakan 885 *event* pengujian sehingga data uji tidak digunakan untuk menentukan batas penerimaan.

Untuk menstabilkan keputusan pada video *real-time*, identitas kandidat harus muncul pada tiga *frame* berturut-turut. Apabila kandidat berubah, penghitung kestabilan diulangi. Mekanisme ini digunakan untuk mengurangi keputusan sesaat akibat perubahan hasil pencocokan antar*frame*.

Pada *event* dari kelompok terdaftar yang berhasil dikenali, sistem menghasilkan nilai *confidence* berdasarkan transformasi internal terhadap *Cosine Distance*. Nilai tersebut digunakan sebagai indikator internal pencocokan dan tidak diperlakukan sebagai probabilitas yang terkalibrasi secara statistik.

D. Teknik Evaluasi

Evaluasi dilakukan dengan menempatkan wisudawan terdaftar sebagai kelas positif dan individu tidak terdaftar sebagai kelas negatif. Seluruh 885 *event recognition* yang tercatat digunakan dalam evaluasi, yang kemudian dikelompokkan menjadi *True Positive*, *True Negative*, *False Positive*, dan *False Negative*.

Accuracy, *precision*, *recall*, dan *F1-score* dihitung berdasarkan confusion matrix sesuai definisi pada Tinjauan Pustaka. *Recognition rate* dihitung pada kelompok aktual terdaftar untuk masing-masing kondisi pengujian.

$$RR_c = \frac{TP_c}{TP_c + FN_c} \quad (8)$$

dengan dan masing-masing merupakan *True Positive* dan *false Negative* pada kondisi pengujian ke-, sedangkan merupakan *recognition rate* pada kondisi tersebut.

Latency dianalisis pada seluruh *event recognition* menggunakan nilai rata-rata sebagai indikator utama, sedangkan median dan persentil ke-95 (P95) digunakan untuk menggambarkan distribusi waktu pemrosesan.

$$\bar{L}_c = \frac{1}{N_c}\sum_{i=1}^{N_c} L_i \quad (9)$$

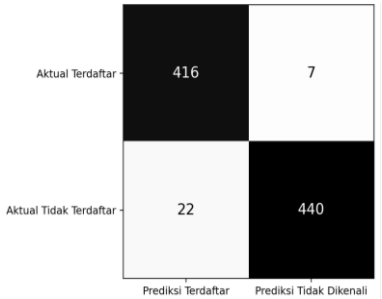
dengan merupakan *latency event* ke-, merupakan jumlah *event* pada kondisi pengujian ke-, dan merupakan rata-rata *latency* pada kondisi tersebut.

Untuk analisis jarak dan penggunaan kacamata, *confidence*, *Cosine Distance*, dan *face width* dihitung hanya pada *event* terdaftar yang berhasil dikenali dengan benar agar analisis kualitas pencocokan tidak tercampur dengan *event* Tidak Dikenali atau *false acceptance*.

IV. HASIL DAN PEMBAHASAN

A. Performa Pengenalan Keseluruhan

Eksperimen menghasilkan 885 *event recognition* yang terdiri atas 423 *event* dari kelompok wisudawan terdaftar dan 462 *event* dari kelompok individu tidak terdaftar. Pada kelompok terdaftar, sebanyak 416 *event* berhasil dikenali dan 7 *event* tidak berhasil dikenali. Sementara itu, pada kelompok tidak terdaftar, 440 *event* berhasil ditolak dan 22 *event* salah diterima sebagai identitas yang tersedia pada basis data. Dengan demikian, sistem menghasilkan 856 keputusan benar dan 29 keputusan salah.



Gambar 2. Confusion matrix sistem pengenalan wajah

Ringkasan hasil evaluasi ditunjukkan pada Tabel 3.

Tabel 3. Ringkasan Performa Keseluruhan

Komponen	Nilai
True Positive (TP)	416
True Negative (TN)	440
False Positive (FP)	22
False Negative (FN)	7
Accuracy	96,7232%
Precision	94,9772%
Recall	98,3452%
F1-score	96,6318%

Hasil tersebut menunjukkan bahwa sistem memiliki *recall* sebesar 98,3452%, yang merupakan nilai tertinggi di antara metrik evaluasi. Nilai tersebut menunjukkan bahwa sebagian besar *event* dari kelompok wisudawan terdaftar berhasil dikenali. *False Negative* hanya berjumlah 7 *event* atau 1,65% dari 423 *event* terdaftar. Sebaliknya, *precision* sebesar 94,9772% dipengaruhi oleh 22 *false Positive* yang terjadi pada kelompok individu tidak terdaftar. Dari 462 *event* wajah asing, 22 *event* atau 4,76% salah diterima sebagai identitas yang tersedia pada basis data.

Perbedaan antara *recall* dan *precision* tersebut menunjukkan karakteristik *threshold* 0,6 yang relatif

mempertahankan sensitivitas terhadap wajah terdaftar, tetapi masih memungkinkan sebagian wajah asing memiliki jarak *embedding* yang berada pada wilayah penerimaan. Kondisi ini penting pada sistem identifikasi wisudawan karena kesalahan menerima wajah asing berpotensi menghasilkan identitas yang salah, meskipun *accuracy* agregat tetap tinggi.

Pada tingkat subjek, lima dari sepuluh wisudawan memperoleh *recognition rate* sebesar 100%, sedangkan lima subjek lainnya berada pada rentang 95,12%–98,31%. Tujuh *false Negative* tersebar pada lima subjek sehingga kegagalan pengenalan tidak terkonsentrasi pada satu identitas tertentu. Pola tersebut menunjukkan bahwa variasi keberhasilan lebih berkaitan dengan kondisi pengambilan wajah dan kualitas representasi pada *event* tertentu daripada kegagalan sistem terhadap satu identitas secara konsisten.

Hasil keseluruhan menunjukkan bahwa pendekatan VGG-*Face* dan pencocokan berbasis *Cosine Distance* mampu memberikan performa klasifikasi yang tinggi pada kondisi pengujian yang digunakan. Hasil ini sebanding dengan penelitian Suhardi et al. yang menggunakan VGG-*Face* dalam sistem presensi dan melaporkan *accuracy* hingga sekitar 97% setelah augmentasi [9]. Perbandingan tersebut tidak dimaksudkan untuk menyatakan keunggulan langsung karena protokol pengujian berbeda. Penelitian Suhardi et al. berfokus pada presensi, sedangkan penelitian ini menggunakan aliran video *real-time*, melibatkan individu tidak terdaftar, menggunakan enam *embedding* referensi untuk setiap wisudawan, serta mengevaluasi jarak, kacamata, *false acceptance*, dan *latency*.

B. Kecepatan Identifikasi

Latency keseluruhan berada pada rentang 0,1524–2,0015 detik dengan rata-rata 0,6877 detik. Statistik *latency* untuk setiap kondisi pengujian disajikan pada Tabel 4.

Tabel 4. Statistik Latency Berdasarkan Kondisi Pengujian

Kondisi	Event	Mean(s)	Median(s)	P95(s)	Rentang(s)
50 cm normal	219	0,6905	0,7556	1,3366	0,1606–2,0015
100 cm normal	189	0,7021	0,7691	1,1174	0,1610–1,4028
>100–150 cm	281	0,6196	0,7401	1,1321	0,1524–1,3875
50 cm + kacamata	196	0,7683	0,8330	1,3718	0,1639–1,7736
Keseluruhan	885	0,6877	0,7747	1,2481	0,1524–2,0015

Rata-rata *latency* tertinggi terdapat pada kondisi 50 cm dengan kacamata, yaitu 0,7683 detik, sedangkan rata-rata terendah terdapat pada kondisi >100–150 cm, yaitu 0,6196 detik. Hasil tersebut menunjukkan bahwa peningkatan jarak tidak secara langsung meningkatkan waktu pemrosesan. Setelah area wajah diperoleh, pipeline tetap menjalankan proses representasi dan pencocokan *embedding* sehingga ukuran wajah pada *frame* tidak secara sederhana berbanding lurus dengan waktu *inference*.

Median keseluruhan sebesar 0,7747 detik dan P95 sebesar 1,2481 detik menunjukkan bahwa distribusi *latency* memiliki ekor yang lebih panjang dibandingkan nilai rata-ratanya. Oleh karena itu, istilah *real-time* dalam penelitian ini lebih tepat dimaknai sebagai pemrosesan video kontinu dengan *respons* rata-rata di bawah satu detik, bukan sebagai jaminan bahwa setiap *event* selesai diproses dalam waktu kurang dari satu detik. Nilai maksimum 2,0015 detik menunjukkan terdapat *event* tertentu yang membutuhkan waktu pemrosesan lebih panjang.

Nilai *latency* pada penelitian ini berasal dari field *latency*_sec pada *log recognition* dan merepresentasikan durasi pemrosesan pengenalan pada pipeline yang diuji. Nilai tersebut tidak mencakup waktu operator, antrean fisik, proses tampilan antarmuka, maupun keluaran suara. Pembatasan definisi ini diperlukan agar klaim mengenai kecepatan sistem tetap sesuai dengan metrik yang benar-benar dicatat.

Hasil *latency* juga memperlihatkan bahwa pengukuran *respons* merupakan aspek penting pada sistem pengolahan video *real-time*. Hal tersebut sejalan dengan penelitian Rizaldi et al. yang menggunakan OpenCV pada pipeline video *real-time* serta melakukan evaluasi pada variasi jarak, dan penelitian Heriansyah dan Utomo yang menekankan evaluasi performa pengolahan citra melalui aspek waktu *respons* dan efisiensi komputasi [10], [11].

C. Pengaruh Jarak dan Penggunaan Kacamata

Evaluasi berdasarkan kondisi menunjukkan bahwa *accuracy* sistem tetap berada di atas 95% pada seluruh skenario pengujian. Kondisi 50 cm normal memberikan hasil terbaik dengan *accuracy* 98,17% dan *recall* 100%, sedangkan kondisi 50 cm dengan kacamata menghasilkan *recall* terendah sebesar 95%. Pada kondisi tersebut terdapat empat *event* wisudawan terdaftar yang tidak berhasil dikenali.

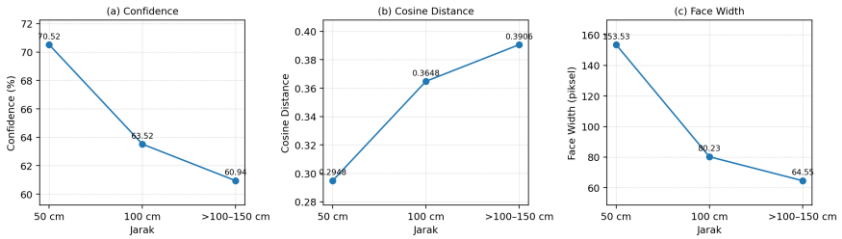
Tabel 5. Performa Pengenalan Berdasarkan Kondisi Pengujian

Kondisi	Recognition Rate	Keterangan
50 cm normal	100%	Performa terbaik
100 cm normal	97,78%	Penurunan terbatas
>100–150 cm	99,30%	Tetap tinggi
50 cm + kacamata	95,00%	Terendah

Pada kelompok aktual terdaftar, *recognition rate* sebesar 100% diperoleh pada jarak 50 cm, 97,78% pada 100 cm, dan 99,30% pada >100–150 cm. Nilai tersebut tidak membentuk pola penurunan linear terhadap jarak. Oleh karena itu, pengaruh jarak lebih tepat diamati melalui perubahan kualitas representasi wajah dan ukuran wajah yang terdeteksi daripada hanya melalui perubahan *recognition rate*.

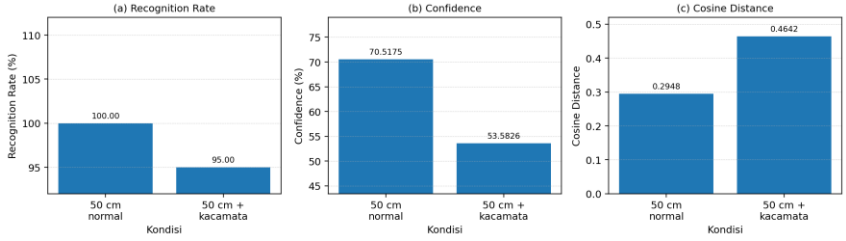
Pada *event* terdaftar yang dikenali dengan benar, *confidence* rata-rata turun dari 70,52% pada 50 cm menjadi 63,52% pada 100 cm dan 60,94% pada >100–150 cm. Pada saat yang sama, rata-rata *Cosine Distance* meningkat dari 0,2948 menjadi 0,3648 dan 0,3906. Perubahan tersebut menunjukkan bahwa bertambahnya jarak cenderung menghasilkan representasi yang lebih jauh dari *embedding* referensi, meskipun perubahan tersebut belum cukup untuk menurunkan *recognition rate* secara konsisten.

Face width menunjukkan perubahan yang paling konsisten, yaitu dari rata-rata 153,53 piksel pada 50 cm menjadi 80,23 piksel pada 100 cm dan 64,55 piksel pada >100–150 cm. Berkurangnya ukuran wajah berarti area yang tersedia untuk membentuk representasi visual menjadi lebih kecil. Dengan demikian, jarak memberikan pengaruh yang lebih jelas terhadap karakteristik *embedding* daripada terhadap keputusan akhir pada rentang jarak yang diuji. Temuan ini menunjukkan bahwa *threshold* dan kemampuan model dalam mempertahankan fitur diskriminatif berperan dalam menentukan apakah perubahan kualitas representasi akhirnya menyebabkan kesalahan pengenalan.



Gambar 3. Perubahan *confidence*, *Cosine Distance*, dan *face width* berdasarkan jarak pada *event* terdaftar yang dikenali benar

Penggunaan kacamata menghasilkan perubahan yang lebih jelas dibandingkan variasi jarak. Pada jarak 50 cm, *recognition rate* turun dari 100% pada kondisi normal menjadi 95% ketika subjek menggunakan kacamata. *Confidence* rata-rata pada *event* yang dikenali benar turun dari 70,5175% menjadi 53,5826%, sedangkan *Cosine Distance* meningkat dari 0,2948 menjadi 0,4642. Karena jarak dipertahankan sama, perubahan tersebut lebih konsisten dengan pengaruh oklusi ringan dan perubahan pola visual pada area sekitar mata daripada perubahan skala wajah.



Gambar 3. Perbandingan kondisi 50 cm normal dan 50 cm menggunakan kacamata pada kelompok terdaftar

Kacamata tidak menutupi seluruh wajah, tetapi dapat mengubah pola lokal dan informasi visual di sekitar mata yang digunakan dalam pembentukan *embedding*. Hasil ini searah dengan kajian mengenai *face recognition under occlusion* yang menunjukkan bahwa gangguan pada area wajah dapat mengubah representasi fitur dan menurunkan kemampuan pencocokan [14], [15]. Meskipun demikian, *recognition rate* sebesar 95% menunjukkan bahwa model masih dapat memanfaatkan informasi dari bagian wajah lainnya sehingga sebagian besar *event* tetap dikenali.

Perlu diperhatikan bahwa *recognition rate* pada penelitian ini dihitung terhadap *event* yang telah mencapai tahap *recognition*. Wajah yang sama sekali tidak terdeteksi oleh *Haar Cascade* tidak masuk ke

log_recognition. Oleh karena itu, *recognition rate* tidak dapat ditafsirkan sebagai probabilitas keberhasilan dari seluruh paparan fisik wajah terhadap kamera. Pemisahan antara keberhasilan deteksi dan keberhasilan *recognition* merupakan batas interpretasi penting pada evaluasi sistem ini.

D. *False acceptance*, *Open-set Recognition*, dan Perbandingan Hasil

Sebanyak 22 *false Positive* ditemukan pada kelompok individu tidak terdaftar. Kesalahan tersebut tidak tersebar merata pada sepuluh subjek. Delapan subjek tidak menghasilkan *false acceptance* sama sekali, sedangkan dua subjek menyumbang seluruh *false acceptance*. Satu subjek menghasilkan 13 *false acceptance* dari 59 *event*, sedangkan subjek lainnya menghasilkan 9 *false acceptance* dari 56 *event*.

Tabel 5. Performa Pengenalan Berdasarkan Kondisi Pengujian

Kelompok Subject	Event	False Positive	False Acceptance Rate
Sebagian besar subjek	-	0	0%
Subjek asing A	56	9	16,07%
Subjek asing B	59	13	22,03%
Total dua subjek	115	22	19,13%

Konsentrasi kesalahan pada dua subjek menunjukkan bahwa *false acceptance* tidak tersebar merata. Pemeriksaan *log* menunjukkan bahwa 13 *false acceptance* pada satu subjek terbagi menjadi delapan dan lima *event* terhadap dua identitas referensi, sedangkan sembilan *false acceptance* pada subjek lainnya terdiri atas delapan *event* terhadap satu identitas dan satu *event* terhadap identitas lain. Pola tersebut menunjukkan bahwa sebagian *embedding* wajah asing memiliki kedekatan relatif terhadap *embedding* referensi tertentu sehingga nilai (d_{min}) berada dalam wilayah penerimaan *threshold* 0,6. Berdasarkan kondisi pengujian, *false Positive* masing-masing berjumlah 4 *event* pada 50 cm, 5 *event* pada 100 cm, 8 *event* pada >100–150 cm, dan 5 *event* pada kondisi kacamata. Karena jumlah *event* wajah asing berbeda pada setiap kondisi, perbandingan tidak dilakukan hanya berdasarkan jumlah absolut. *Precision* terendah diperoleh pada kondisi kacamata sebesar 93,83%, sedangkan kondisi 50 cm normal mencapai 96,49%.

Dalam perspektif *open-set face recognition*, hasil tersebut menunjukkan bahwa *threshold* 0,6 masih memberikan ruang bagi sebagian wajah asing untuk diterima [12], [13]. Oleh karena itu, peningkatan sistem tidak hanya perlu diarahkan pada *accuracy* agregat, tetapi juga pada pengurangan *false acceptance* dengan tetap mempertahankan *recall*. Dalam konteks prosesi wisuda, *false Negative* dapat ditindaklanjuti melalui verifikasi ulang, sedangkan *false Positive* berpotensi menghasilkan identitas yang salah. Pengembangan selanjutnya dapat diarahkan pada evaluasi *threshold* yang lebih sistematis atau strategi pencocokan yang lebih adaptif.

Penelitian ini memiliki beberapa keterbatasan. Pengujian hanya melibatkan 20 individu, sedangkan 885 *event* merupakan repeated observations dan bukan 885 subjek independen. Pengujian dilakukan pada pencahayaan normal dan lingkungan semi-terkontrol sehingga hasil belum dapat digeneralisasikan pada pencahayaan ekstrem, pose besar, lingkungan outdoor, atau oklusi berat. Selain itu, VGG-*Face* tidak dilakukan *fine-tuning*, *threshold* 0,6 bersifat tetap, wajah yang gagal dideteksi tidak masuk confusion matrix, dan sistem belum memiliki liveness detection atau anti-spoofing. Dengan demikian, hasil penelitian lebih tepat diposisikan sebagai evaluasi prototipe pada kondisi pengujian yang ditetapkan, bukan sebagai validasi sistem biometrik siap produksi.

V. SIMPULAN

Penelitian ini menunjukkan bahwa sistem pengenalan wajah *real-time* menggunakan VGG-*Face* melalui *DeepFace* dan *Cosine Distance* mampu melakukan identifikasi wisudawan pada kondisi semi-terkontrol dengan performa yang tinggi. Dari 885 *event recognition* yang tercatat, sistem memperoleh *accuracy* sebesar 96,7232%, *precision* sebesar 94,9772%, *recall* sebesar 98,3452%, dan *F1-score* sebesar 96,6318%, dengan rata-rata *latency* 0,6877 detik. Hasil pengujian berdasarkan jarak menunjukkan bahwa perubahan hingga >100–150 cm belum menyebabkan penurunan *recognition rate* yang konsisten, sedangkan penggunaan kacamata pada jarak 50 cm menurunkan *recognition rate* menjadi 95% dan meningkatkan *Cosine Distance* serta menurunkan *confidence* pada *event* yang dikenali dengan benar.

Meskipun demikian, sistem masih memiliki keterbatasan pada kemampuan menolak wajah tidak terdaftar. Sebanyak 22 *false Positive* ditemukan pada kelompok individu tidak terdaftar dan terkonsentrasi pada dua subjek, sehingga pengembangan selanjutnya perlu diarahkan pada pengendalian *false acceptance* dan evaluasi *threshold* yang lebih sistematis. Dengan mempertimbangkan hasil pengujian dan batasan lingkungan eksperimen, sistem yang dikembangkan dapat digunakan sebagai prototipe identifikasi wisudawan secara *real-time* pada kondisi semi-terkontrol, tetapi masih memerlukan

pengujian yang lebih beragam dan penguatan mekanisme *open-set* serta *anti-spoofing* sebelum diterapkan pada lingkungan operasional yang lebih luas.

VI. DAFTAR PUSTAKA

[1] M. Wang and W. Deng, "Deep Face Recognition: A Survey," Neurocomputing, vol. 429, pp. 215–244, Mar. 2021, doi: 10.1016/j.neucom.2020.10.081.
[2] L. Li, X. Mu, S. Li, and H. Peng, "A Review Of Face Recognition Technology," IEEE Access, vol. 8, pp. 139110–139120, 2020, doi: 10.1109/ACCESS.2020.3011028.
[3] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed, "Past, Present, and Future of Face recognition: A Review," Electronics, vol. 9, no. 8, p. 1188, Jul. 2020, doi: 10.3390/electronics9081188.
[4] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face recognition," Procedings Br. Mach. Vis. Conf. 2015, pp. 41.1-41.12, 2015, doi: 10.5244/C.29.41.
[5] Z. R. Mair and M. H. Irfani, "Permainan INGBAS (Gunting, Batu, Kertas) Menggunakan Arsitektur Convolutional Neural Network," J. Tek. Inform. dan Sist. Inf., vol. 10, no. 1, pp. 1019–1026, 2023, doi: DOI: 10.35957/jatisi.v10i1.2891.
[6] A. Nurlita, M. Haviz Irfani, and Z. Romegar Mair, "Pendekatan Convolutional Neural Network dalam Mendeteksi Kemiringan Tulisan Tangan Menggunakan Framework YOLO," J. FASILKOM, vol. 15, no. 3, pp. 640–647, Jan. 2026, doi: 10.37859/jf.v15i3.10406.
[7] M. Siddik and A. Sirait, "Implementasi Algoritma Cosine Similarity Dalam Pengenalan Wajah Untuk Aplikasi Absensi," JOISIE (Journal Inf. Syst. Informatics Eng., vol. 9, no. 1, p. 1, Feb. 2025, doi: 10.35145/joisie.v9i1.4772.
[8] J. Zhang and N. Hu, "Cross-age Face Image Similarity Measurement Based on Deep Learning Algorithms," Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 5, 2023, doi: 10.14569/IJACSA.2023.01405123.
[9] C. Listya Suhardi, C. M. R. Eislam, and F. A. Wicaksono, "Implementasi Sistem Verifikasi dan Pengenalan Wajah Berbasis YOLOv8n dan VGG-Face untuk Absensi Perkuliahan," Pseudocode, vol. 13, no. 1, pp. 44–53, Feb. 2026, doi: 10.33369/pseudocode.13.1.44-53.
[10] Muhammad Rizaldi, Rudi Heriansyah, and Nazori Suhandi, "Pengenalan Gerakan Tangan untuk Kontrol Slide Presentasi Menggunakan Framework Mediapipe, OpenCV, dan Model LSTM," J. Publ. Tek. Inform., vol. 4, no. 3, pp. 98–111, Sep. 2025, doi: 10.55606/jupti.v4i3.5332.
[11] R. Heriansyah and W. M. Utomo, "Performance Evaluation of Digital Image Processing by Using Scilab," JUITA J. Inform., vol. 9, no. 2, p. 239, Nov. 2021, doi: 10.30595/juita.v9i2.8434.
[12] E. Lopez-Lopez, X. M. Pardo, and C. V. Regueiro, "Incremental Learning from Low-labelled Stream Data in Open-set Video Face recognition," Pattern Recognit., vol. 131, p. 108885, Nov. 2022, doi: 10.1016/j.patcog.2022.108885.
[13] R. H. Vareto, Y. Linghu, T. E. Boult, W. R. Schwartz, and M. Günther, "Open-set face recognition with maximal entropy and Objectosphere loss," Image Vis. Comput., vol. 141, p. 104862, Jan. 2024, doi: 10.1016/j.imavis.2023.104862.
[14] D. Zeng, R. Veldhuis, and L. Spreeuwers, "A survey of face recognition techniques under occlusion," IET Biometrics, vol. 10, no. 6, pp. 581–606, Nov. 2021, doi: 10.1049/bme2.12029.
[15] G. Jeevan, G. C. Zacharias, M. S. Nair, and J. Rajan, "An empirical study of the impact of masks on face recognition," Pattern Recognit., vol. 122, p. 108308, Feb. 2022, doi: 10.1016/j.patcog.2021.108308.
[16] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in 2014 IEEE Conference on Computer vision and Pattern Recognition, IEEE, Jun. 2014, pp. 1701–1708. doi: 10.1109/CVPR.2014.220.