Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 108-117
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.9778

Klasifikasi Jenis Kelamin Berbasis Citra Mata Menggunakan Vision Transformer ViT dengan Strategi Discriminative Fine-Tuning

Gde Made Hanura[*], Putu Hendra Suputra

Fakultas Teknik dan Kejuruan, Program Studi Ilmu Komputer, Universitas Pendidikan Ganesha, Singaraja, Indonesia
Email: [1,*]gde@student.undiksha.ac.id, [2]hendra.suputra@undiksha.ac.id
Email Penulis Korespondensi: gde@student.undiksha.ac.id
Submitted: 28/04/2026; Accepted: 26/05/2026; Published: 26/05/2026

Abstrak–Sistem identifikasi biometrik berbasis wajah memiliki keterbatasan signifikan ketika wajah subjek tertutup, baik akibat penggunaan masker pasca pandemi COVID-19 maupun cadar karena alasan budaya dan agama. Kondisi ini menciptakan celah keamanan nyata, sebagaimana terbukti pada kasus penyusupan berbasis penyamaran gender di Masjid Jannatul Firdaus Makassar. Dalam situasi tersebut, mata menjadi satu-satunya fitur biometrik yang konsisten terekspos. Penelitian ini mengusulkan penerapan Vision Transformer (ViT-B/16) yang dipretrain pada ImageNet-21K dengan strategi fine tuning bertahap berbasis prinsip discriminative learning rate untuk mengklasifikasikan jenis kelamin dari citra mata. Dataset *Female and Male Eyes* dari Kaggle terdiri dari 11.525 citra mata yang dibagi menjadi data latih (64%), validasi (16%), dan pengujian (20%). Eksperimen dilakukan dalam dua seri: Seri B menguji variasi jumlah transformer block yang di-*unfreeze* (0–6 block), dan Seri C menguji variasi rasio discriminative learning rate antara classifier dan encoder (5:1, 10:1, 3:1). Konfigurasi optimal dengan 6 block di-unfreeze dan rasio 3:1 mencapai akurasi 95,70%, precision 97,67%, recall 92,69%, dan weighted F1-score 0,9569, melampaui MobileNet (93,90%) dan K-Nearest Neighbor (68,81%). Hasil ini menunjukkan bahwa ViT dengan strategi discriminative fine tuning efektif untuk klasifikasi jenis kelamin berbasis citra mata dan berpotensi diterapkan pada sistem keamanan biometrik.

Kata Kunci: Vision Transformer; Klasifikasi Jenis Kelamin; Citra Mata; Discriminative Learning Rate; Fine Tuning

Abstract–Face-based biometric identification systems have significant limitations when a subject's face is covered, whether due to mask usage after the COVID-19 pandemic or face veils for cultural and religious reasons. This creates real security gaps, as evidenced by the gender-disguise infiltration incident at Masjid Jannatul Firdaus in Makassar. In such situations, the eyes remain the only consistently exposed biometric feature. This study proposes the application of Vision Transformer (ViT-B/16) pretrained on ImageNet-21K with a progressive fine-tuning strategy based on the discriminative learning rate principle to classify gender from eye images. The *Female and Male Eyes* dataset from Kaggle consists of 11,525 eye images divided into training (64%), validation (16%), and testing (20%) sets. Experiments were conducted in two series: Series B tested variations in the number of unfrozen transformer blocks (0–6), and Series C tested discriminative learning rate ratios between the classifier and encoder (5:1, 10:1, 3:1). The optimal configuration with 6 unfrozen blocks and a 3:1 ratio achieved 95.70% accuracy, 97.67% precision, 92.69% recall, and 0.9569 weighted F1-score, surpassing MobileNet (93.90%) and K-Nearest Neighbor (68.81%). These results indicate that ViT with discriminative fine-tuning is effective for gender classification from eye images and has potential for biometric security applications.

Keywords: Vision Transformer; Gender Classification; Eye Image; Discriminative Learning Rate; Fine Tuning

1. PENDAHULUAN

Penggunaan masker, cadar, niqab, dan helm yang lazim dalam berbagai konteks seperti pandemi, keagamaan, budaya, dan keselamatan telah menciptakan tantangan struktural bagi sistem identifikasi otomatis berbasis wajah. Studi dari NIST menunjukkan bahwa tingkat kesalahan pengenalan wajah dapat meningkat hingga 50 kali lipat ketika subjek mengenakan masker [1]. Selain itu, kasus nyata seperti penyusupan berbasis penyamaran wajah di fasilitas umum [2] menegaskan bahwa keterbatasan ini tidak hanya bersifat teknis, tetapi juga menjadi permasalahan keamanan yang mendesak. Kondisi ini mendorong terjadinya pergeseran paradigma dalam biometrik, dari pemanfaatan fitur wajah secara keseluruhan menuju fitur yang tetap terekspos meskipun wajah tertutup, yaitu area mata.

Area mata menyimpan informasi biometrik yang kaya dan unik, mencakup pola iris, tekstur sklera, morfologi pupil, serta karakteristik periorbital seperti kontur kelopak mata, arkus supraorbital, distribusi bulu mata, dan geometri sudut kantus [3,4]. Berbagai penelitian menunjukkan bahwa karakteristik tersebut mengandung informasi diskriminatif yang cukup untuk membedakan jenis kelamin [5–9]. Namun demikian, klasifikasi jenis kelamin berbasis citra mata menghadirkan tantangan representasi yang khas. Fitur-fitur diskriminatif tersebut tidak terlokalisasi pada satu area tertentu, melainkan tersebar dan bersifat relasional. Perbedaan antara mata pria dan wanita sering kali terletak pada hubungan spasial antar area yang tidak berdekatan secara geometris dalam citra, seperti proporsi antara lebar arkus alis terhadap apertura kelopak, maupun kurva bulu mata terhadap kontur kantus lateral.

Sifat relasional dan global dari fitur periorbital tersebut menuntut pendekatan pemodelan yang mampu menangkap long-range spatial dependency secara langsung, yaitu kemampuan untuk memodelkan hubungan antar area dalam citra tanpa dibatasi oleh jarak spasial. Vision Transformer (ViT) menawarkan kemampuan tersebut melalui mekanisme multi-head self-attention [10,18]. Berbeda dengan pendekatan berbasis konvolusi yang mengekstraksi fitur secara lokal dan bertahap, ViT membagi citra menjadi sejumlah patch dan memprosesnya secara paralel.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 108-117
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.9778

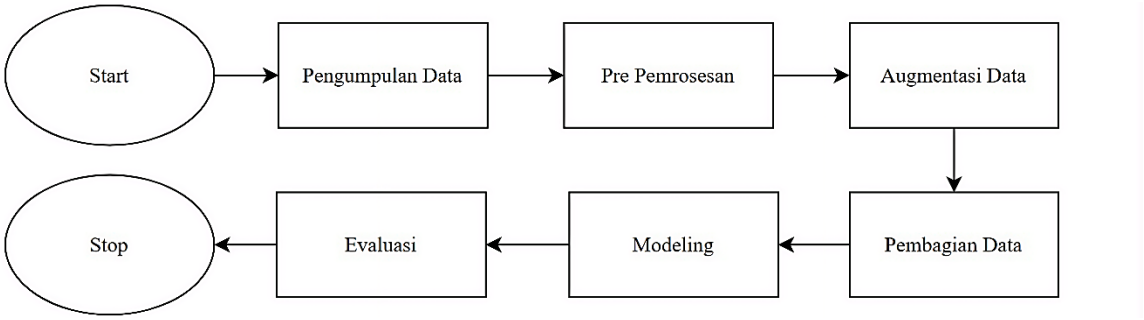
Meskipun demikian, penerapan ViT secara efektif pada domain biometrik mata memerlukan strategi pelatihan yang cermat. Model ViT yang dipralatih pada dataset skala besar seperti ImageNet-21K menghasilkan representasi visual yang bersifat umum. Namun, distribusi citra mata memiliki karakteristik yang berbeda secara signifikan, sehingga proses fine-tuning yang dilakukan secara naif berpotensi merusak representasi awal (pretrained representation) atau menghasilkan adaptasi yang kurang optimal [11]. Untuk mengatasi hal tersebut, digunakan strategi discriminative fine-tuning, yaitu pemberian laju pembelajaran (learning rate) yang berbeda pada setiap lapisan model. Lapisan awal yang menangkap fitur dasar dipertahankan dengan laju pembelajaran kecil, sedangkan lapisan akhir yang berperan dalam pembentukan fitur semantik diperbarui dengan laju yang lebih besar agar mampu beradaptasi dengan karakteristik domain citra mata.

Sejumlah penelitian sebelumnya telah mengeksplorasi klasifikasi jenis kelamin berbasis citra mata dengan berbagai pendekatan [13–15]. Namun demikian, hingga saat ini belum ditemukan penelitian yang secara spesifik mengintegrasikan arsitektur Vision Transformer dengan strategi discriminative fine-tuning untuk tugas tersebut. Oleh karena itu, penelitian ini mengusulkan penggunaan ViT-B/16 yang dipralatih pada ImageNet-21K dengan eksplorasi sistematis terhadap: (1) jumlah transformer block yang di-unfreeze beserta analisis diminishing returns-nya, serta (2) rasio discriminative learning rate yang optimal untuk domain biometrik mata. Pendekatan ini diharapkan mampu menyeimbangkan antara adaptasi terhadap domain baru dan preservasi representasi awal yang telah dipelajari model.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Untuk mendukung pemahaman terhadap alur penelitian, Gambar 1 menyajikan diagram tahapan metode penelitian yang menggambarkan secara sistematis proses klasifikasi jenis kelamin berbasis citra mata manusia dengan pendekatan Vision Transformer (ViT). Alur penelitian dimulai dari tahap pengumpulan data, *pre-processing*, augmentasi data, pembagian data, *modeling*, hingga evaluasi.



Gambar 1 Tahapan Penelitian

Berdasarkan Gambar 1, berikut penjelasan setiap tahapan penelitian:

a. Pengumpulan Data.
   Tahap ini merupakan titik awal (*start*) dari seluruh alur penelitian. Data citra mata dikumpulkan dari dataset publik *Female and Male Eyes* yang diperoleh dari platform Kaggle, terdiri dari 11.525 citra mata yang terbagi ke dalam dua kelas, yaitu pria dan wanita.
b. Pre-Pemrosesan.
   Seluruh citra yang telah dikumpulkan kemudian diproses untuk memastikan konsistensi dan kualitas input. Tahap ini mencakup resize citra dari resolusi awal 60×60 piksel menjadi 224×224 piksel sesuai standar input ViT-B/16, konversi ruang warna RGB ke grayscale sebagai strategi mitigasi brightness bias, serta normalisasi menggunakan parameter ImageNet.
c. Augmentasi Data.
   Untuk meningkatkan keberagaman sampel data latih dan mengurangi risiko overfitting, diterapkan teknik augmentasi yang meliputi ColorJitter dengan intensitas 0,3, rotasi acak ±10 derajat, dan horizontal flip. Augmentasi hanya diterapkan pada data latih, sedangkan data validasi dan pengujian tidak mengalami augmentasi guna menjaga objektivitas evaluasi.
d. Pembagian Data.
   Dataset dibagi menggunakan skema 80:20, di mana 80% digunakan sebagai data pelatihan dan 20% sebagai data pengujian. Dari data pelatihan, dilakukan pembagian ulang dengan rasio 80:20 untuk memperoleh data validasi, sehingga menghasilkan tiga subset: data latih (64%), data validasi (16%), dan data uji (20%).
e. Modeling.
   Pada tahap ini, model ViT-B/16 yang telah dipretrain pada ImageNet-21K dilatih menggunakan data latih dengan strategi fine tuning bertahap berbasis prinsip discriminative learning rate. Eksperimen dilakukan

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 108-117
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.9778

dalam dua seri, yaitu Seri B untuk menentukan jumlah transformer block optimal yang di-unfreeze, dan Seri C untuk mengoptimalkan rasio discriminative learning rate antara classifier head dan encoder.

f. Evaluasi.
Tahap akhir sebelum stop adalah pengukuran performa model menggunakan data pengujian. Evaluasi dilakukan menggunakan Confusion Matrix dan metrik Accuracy, Precision, Recall, serta F1-Score untuk mengukur kemampuan model dalam mengklasifikasikan jenis kelamin secara menyeluruh dan per kelas.

2.2 Metode Klasifikasi Citra

Metode klasifikasi citra dalam penelitian ini mengombinasikan arsitektur Vision Transformer (ViT-B/16) sebagai model utama dengan strategi *fine tuning* bertahap berbasis prinsip *discriminative learning rate* sebagai teknik optimasi. ViT-B/16 digunakan untuk mengekstraksi dan memproses informasi spasial dari setiap citra mata melalui mekanisme *self-attention*, sehingga relasi global antar *patch* citra yang merepresentasikan fitur-fitur diskriminatif jenis kelamin, seperti kontur periorbital, tekstur iris, dan distribusi bulu mata, dapat dimodelkan secara komprehensif dalam proses klasifikasi [10][23]. Pemilihan arsitektur ViT-B/16 didasarkan pada kemampuannya dalam menangkap dependensi spasial jarak jauh (*long-range spatial dependency*) yang tidak dapat dimodelkan secara efektif oleh CNN, karena ViT memproses seluruh *patch* citra secara paralel dan mempertimbangkan hubungan antar *patch* di seluruh area citra sekaligus, sehingga lebih *robust* terhadap variasi spasial global [10][18]. Model ViT-B/16 yang digunakan telah dipretrain pada dataset ImageNet-21K, sehingga bobot awalnya telah mengandung representasi fitur visual yang kaya dan dapat diadaptasi ke domain klasifikasi citra mata melalui proses *fine tuning* [16].

Fine tuning dilakukan dengan menerapkan prinsip *discriminative learning rate* [11][26] yang menetapkan *learning rate* berbeda untuk setiap bagian model guna menciptakan keseimbangan antara adaptasi terhadap domain baru dan preservasi representasi *pretrained*. *Layer* yang berada dekat dengan *output* mendapatkan *learning rate* lebih besar untuk adaptasi yang lebih agresif, sedangkan *layer* yang berada dekat dengan *input* mendapatkan *learning rate* lebih kecil untuk menjaga keutuhan representasi yang telah dipelajari. Proses *fine tuning* dilakukan dalam dua seri eksperimen sistematis menggunakan Optimizer AdamW dengan *weight decay* 1×10^{-4} dan mekanisme *early stopping* (*patience*=4) [17]. Seri B menentukan jumlah *transformer block* optimal yang di-*unfreeze* dari total 12 *block*, dengan variasi 0, 2, 4, dan 6 *block* yang di-*unfreeze* dari bagian atas encoder [27]. Seri C selanjutnya mengoptimalkan rasio *discriminative learning rate* antara *classifier head* dan *encoder* menggunakan konfigurasi terbaik Seri B sebagai fondasi, dengan menguji variasi rasio 5:1, 10:1, dan 3:1. Lapisan *output* asli model diganti dengan *classifier head* berupa lapisan Dense 2 neuron beraktivasi Softmax untuk menghasilkan prediksi probabilistik pada kelas Pria dan Wanita [16]. Hasil prediksi model kemudian dibandingkan dengan label aktual pada data uji untuk menghitung performa klasifikasi. Evaluasi akhir dilakukan menggunakan *confusion matrix* yang menjadi dasar perhitungan metrik evaluasi *accuracy*, *precision*, *recall*, dan *F1-score* [14][15][28][29].

3. HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil penelitian serta pembahasan yang diperoleh dari penerapan metode klasifikasi jenis kelamin berbasis citra mata manusia menggunakan arsitektur Vision Transformer (ViT-B/16) dengan strategi fine tuning bertahap berbasis prinsip discriminative learning rate. Pembahasan mencakup evaluasi model baseline sebagai acuan performa awal, hasil eksperimen Seri B terkait variasi jumlah transformer block yang di-unfreeze, hasil eksperimen Seri C terkait optimasi rasio discriminative learning rate, evaluasi model optimal pada data pengujian, serta perbandingan performa dengan penelitian terdahulu dalam domain yang sama. Seluruh hasil yang disajikan pada bagian ini bertujuan untuk menjawab rumusan masalah penelitian dan mendukung tujuan yang telah ditetapkan, khususnya dalam membuktikan efektivitas ViT-B/16 dengan strategi discriminative fine tuning sebagai solusi klasifikasi jenis kelamin andal dalam kondisi wajah tertutup.

3.1 Deskripsi dan Karakteristik Data

Penelitian ini menggunakan dataset publik Female and Male Eyes yang diperoleh dari Kaggle [12], dapat dilihat pada Gambar 2.

Gambar 2. Data Citra Mata Female and Male Eyes

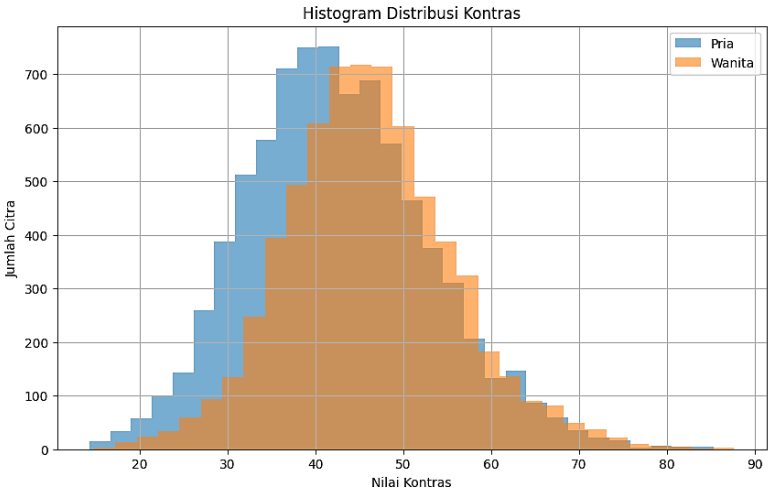
Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 108-117
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.9778

Dataset ini terdiri dari 11.525 citra mata dengan resolusi awal 60x60 piksel, terbagi menjadi 6.323 citra mata pria dan 5.202 citra mata wanita. Dataset dibagi menggunakan skema 80:20, di mana 80% (9.220 citra) digunakan sebagai data pelatihan dan 20% (2.305 citra) sebagai data pengujian. Dari data pelatihan tersebut, dilakukan pembagian ulang dengan rasio 80:20 untuk memperoleh data validasi. Distribusi lengkap dataset disajikan pada Tabel 1.

Tabel 1 Distribusi Dataset Sebelum Augmentasi

Split	Total	Pria	Wanita	Persentase
Training	7.377	4.047	3.330	64%
Validation	1.843	1.011	832	16%
Testing	2.305	1.265	1.040	20%
Total	11.525	6.323	5.202	100%

Berdasarkan Gambar 3, 1analisis distribusi kontras menunjukkan adanya perbedaan karakteristik visual antar kelas. Citra wanita memiliki rata-rata kontras sebesar 46,13 ± 9,58, sedangkan citra pria sebesar 42,60 ± 10,44. Histogram distribusi kontras menunjukkan bahwa kedua kelas memiliki area distribusi yang saling tumpang tindih, namun citra wanita cenderung memiliki nilai kontras yang sedikit lebih tinggi dibandingkan citra pria. Perbedaan ini mengindikasikan adanya variasi intensitas lokal dan tekstur pada area mata antar kelas.



Gambar 3. Perbedaan Distribusi Brightness Antar Kelas

Perbedaan distribusi kontras tersebut menunjukkan adanya variasi karakteristik tekstur dan intensitas lokal pada area mata antar kelas yang berpotensi memengaruhi proses pembelajaran model. Untuk menyederhanakan representasi visual dan mengurangi ketergantungan model terhadap informasi warna yang tidak relevan, seluruh citra dikonversi dari format RGB ke grayscale. Konversi ini dilakukan agar model lebih berfokus pada pola struktur dan tekstur area mata selama proses pembelajaran. Selanjutnya, citra grayscale direplikasi menjadi tiga channel identik agar tetap kompatibel dengan bobot pretrained ViT-B/16 yang mengharapkan input tiga channel.

3.2 Tahapan Pra-Pemrosesan Data

Tahap pra-pemrosesan data dilakukan untuk memastikan konsistensi dan kualitas citra sebelum diproses pada tahap ekstraksi fitur dan pelatihan model. Citra mata dari dataset *Female and Male Eyes* memiliki resolusi awal yang kecil (60×60 piksel) dan mengandung potensi *bias* kecerahan antar kelas yang perlu ditangani secara sistematis. Tahapan yang dilakukan adalah sebagai berikut.

a. Resize (Penyamaan Ukuran).
   Seluruh citra di-resize dari resolusi awal 60×60 piksel menjadi 224×224 piksel menggunakan fungsi pengolahan citra pada library Python. Penyamaan ukuran ini dilakukan untuk menyesuaikan format input yang dibutuhkan oleh arsitektur ViT-B/16, di mana model tersebut dirancang untuk memproses citra berukuran 224×224 piksel yang kemudian dibagi menjadi 196 patch berukuran 16×16 piksel.
b. Konversi RGB ke Grayscale.
   Analisis distribusi kontras dataset menunjukkan adanya perbedaan karakteristik visual antar kelas, di mana citra wanita memiliki rata-rata kontras sebesar 46,13 ± 9,58 sedangkan citra pria sebesar 42,60 ± 10,44. Perbedaan distribusi kontras tersebut menunjukkan adanya variasi intensitas lokal dan karakteristik tekstur pada area mata antar kelas yang berpotensi memengaruhi proses pembelajaran model. Untuk menyederhanakan representasi visual dan mengurangi ketergantungan model terhadap informasi warna yang tidak relevan, seluruh citra dikonversi dari format RGB ke grayscale. Konversi ini dilakukan agar model

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 108-117
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.9778

lebih berfokus pada pola struktur dan tekstur area mata selama proses pembelajaran. Selanjutnya, citra grayscale direplikasi menjadi tiga channel identik agar tetap kompatibel dengan bobot pretrained ViT-B/16 yang mengharapkan input tiga channel.

c. Normalisasi.
Seluruh citra dinormalisasi menggunakan parameter ImageNet dengan nilai mean = [0.485, 0.456, 0.406] dan standard deviation = [0.229, 0.224, 0.225]. Normalisasi ini dilakukan untuk menyelaraskan distribusi nilai piksel citra dengan distribusi data yang digunakan selama proses pretraining ViT-B/16 pada ImageNet-21K, sehingga bobot pretrained dapat dimanfaatkan secara optimal selama proses fine tuning.

d. Augmentasi Data.
Augmentasi diterapkan khusus pada data latih untuk meningkatkan keberagaman sampel dan mengurangi risiko overfitting. Teknik augmentasi yang diterapkan meliputi ColorJitter dengan intensitas 0,3, rotasi acak ±10 derajat, dan horizontal flip. Augmentasi rotasi menghasilkan penambahan 7.376 citra baru sehingga total data latih setelah augmentasi meningkat dari 7.377 menjadi 14.754 citra. Data validasi dan pengujian tidak mengalami augmentasi guna menjaga objektivitas evaluasi model.

e. Pembagian Dataset.
Dataset dibagi menggunakan skema 80:20, di mana 80% (9.220 citra) digunakan sebagai data pelatihan dan 20% (2.305 citra) sebagai data pengujian. Dari data pelatihan, dilakukan pembagian ulang dengan rasio 80:20 untuk memperoleh data validasi, sehingga menghasilkan tiga subset akhir: data latih sebanyak 7.377 citra (64%), data validasi sebanyak 1.843 citra (16%), dan data uji sebanyak 2.305 citra (20%).

3.3 Implementasi dan Pengujian Model

Tahap implementasi dilakukan dengan menerapkan arsitektur ViT-B/16 yang telah dipretrain pada ImageNet-21K sebagai model klasifikasi jenis kelamin berbasis citra mata, dengan strategi *fine tuning* bertahap berbasis prinsip *discriminative learning rate* sebagai teknik optimasi. Pengujian dilakukan dalam dua seri eksperimen sistematis untuk mengevaluasi pengaruh jumlah *transformer block* yang di-*unfreeze* dan rasio *discriminative learning rate* terhadap performa model, serta melalui analisis *confusion matrix* untuk memahami pola kesalahan klasifikasi secara lebih rinci.

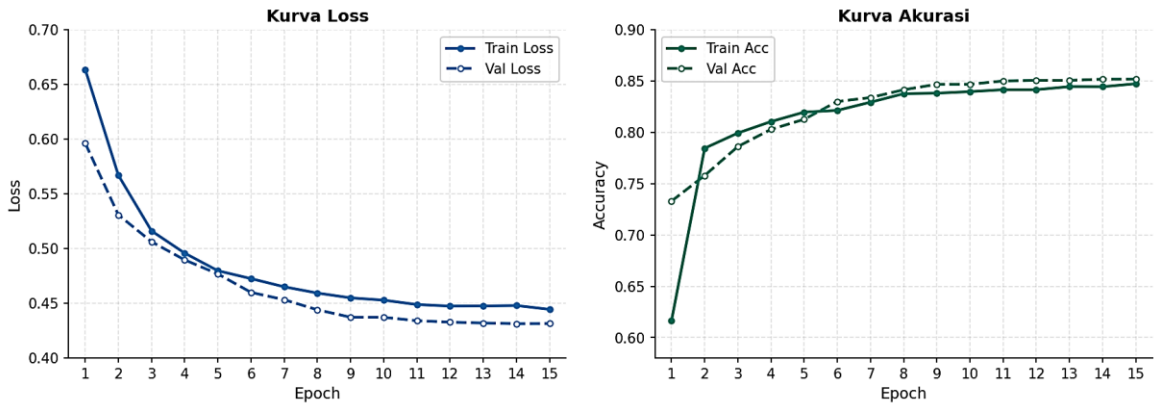
3.4 Evaluasi Model Baseline

Eksperimen dimulai dengan penetapan model *baseline* menggunakan konfigurasi *frozen backbone* — seluruh 12 *encoder block* dibekukan sehingga hanya *classifier head* yang dilatih, menghasilkan 0,1 juta *trainable parameter*. Rincian konfigurasi model baseline disajikan pada Tabel 2.

Tabel 2. Konfigutasi Model Baseline

Parameter	Nilai
Classifier Learning Rate	5e-5
Encoder Learning Rate	1e-5
Augmentasi Data	Grayscale + ColorJitter(0,3) + Rotasi 10*
Backbone	Frozen (12 encoder block dibekukan)
Trainable Parameters	0,1 juta
Jumlah Epoch	15

Selama 15 epoch pelatihan, model menunjukkan konvergensi yang stabil sebagaimana terlihat pada Gambar 1.



Gambar 4. Kurva Konvergensi Training dan Validasi Model Baseline

Berdasarkan Gambar 4, kurva train loss dan val loss menunjukkan penurunan yang konsisten sepanjang 15 epoch, dari masing-masing 0,6639 dan 0,5964 pada epoch pertama menjadi 0,4445 dan 0,4315 pada epoch

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 108-117
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.9778

terakhir. Kurva akurasi menunjukkan tren meningkat yang stabil, dengan val accuracy mencapai 0,8519 pada epoch ke-14. Tidak terdapat indikasi overfitting yang signifikan, ditandai oleh val loss yang terus menurun dan val accuracy yang secara konsisten berada di atas train accuracy kondisi ini mengindikasikan bahwa augmentasi data memberikan efek regularisasi yang efektif selama pelatihan. Hasil evaluasi model baseline pada data pengujian disajikan pada Tabel 2.

Tabel 2. Classification Report Evaluasi Model Baseline

Kelas	Precision	Recall	F1-Score	Support
Pria	0,9086	0,9668	0,9368	1.265
Wanita	0,9562	0,8817	0,9175	1.040
Accuracy	---	---	0,9284	2.305
Macro Avg	0,9324	0,9243	0,9271	2.305

Model baseline mencapai akurasi 92,84% dengan weighted F1-score 0,9281. Kelas Pria menunjukkan recall tinggi (96,68%) dengan precision lebih rendah (90,86%), sedangkan kelas Wanita menunjukkan precision tinggi (95,62%) namun recall lebih rendah (88,17%). Pola ini menjadi dasar pertimbangan untuk melakukan fine tuning lebih lanjut guna meningkatkan performa, khususnya pada kelas Wanita.

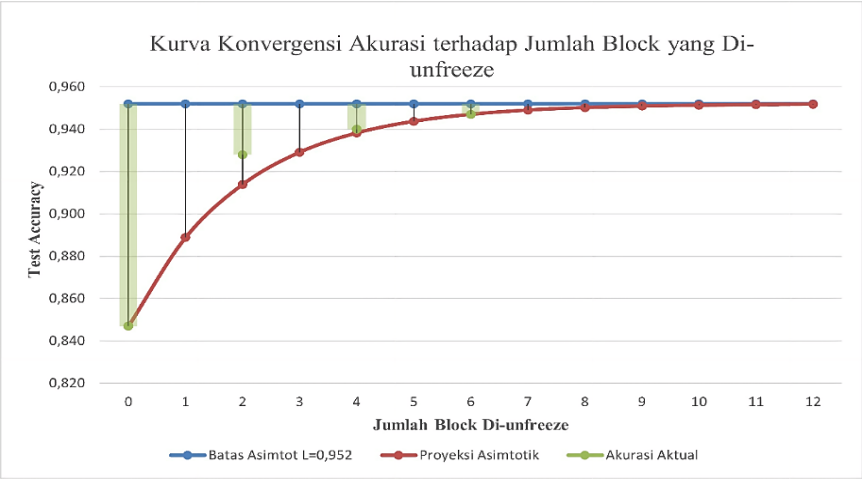
3.5 Eksperimen Seri B — Variasi Jumlah Transformer Block yang Di-unfreeze

Tabel 3 menyajikan hasil eksperimen variasi jumlah *transformer block* yang di-*unfreeze*. Terdapat peningkatan performa yang konsisten seiring bertambahnya *block* yang di-*unfreeze*, namun laju peningkatannya tidak linear.

Tabel 3. Hasil Eksperimen Jumlah Transformer Block yang Di-unfreeze (Seri B)

Kode	Block	Params (Jt)	Best Epoch	Val Acc (%)
B0	0	0,1	14	85,1
B1	2	14	13	94,1
B2	4	28	12	95,8
B3	6	42	7	96,7

Konfigurasi B0 (*backbone frozen*) menghasilkan akurasi 0,847. Membuka 2 *block* (B1) menghasilkan lonjakan signifikan ke akurasi 0,928. Kenaikan berlanjut pada B2 (4 *block*) menjadi 0,940. Konfigurasi B3 dengan 6 *block* dan 42 juta *trainable parameter* dipilih sebagai terbaik seri B dengan akurasi 0,947. Gambar 2 memperlihatkan pola *diminishing returns* yang mengkonfirmasi B3 sebagai titik optimal.



Gambar 5. Kurva Konvergensi Test Accuracy Terhadap Jumlah Block yang Di-Unfreeze

Berdasarkan Gambar 5, terlihat bahwa setiap penambahan jumlah transformer block yang di-unfreeze berkontribusi terhadap peningkatan performa model, terutama pada transisi dari B0 ke B1 yang menunjukkan lonjakan paling signifikan. Hal ini mengindikasikan bahwa fine-tuning pada sebagian layer atas sudah cukup untuk menyesuaikan representasi fitur pretrained dengan karakteristik dataset. Namun, setelah titik tersebut, kurva mulai menunjukkan kecenderungan melandai pada konfigurasi B2 dan B3. Pola ini menegaskan adanya fenomena diminishing returns, di mana peningkatan jumlah parameter yang dilatih tidak lagi diikuti oleh kenaikan akurasi yang proporsional.

Selain itu, kurva konvergensi juga menunjukkan bahwa konfigurasi dengan jumlah block yang lebih banyak cenderung mencapai performa optimal dalam jumlah epoch yang lebih sedikit. Hal ini menunjukkan bahwa model menjadi lebih fleksibel dalam beradaptasi terhadap data ketika lebih banyak lapisan dilibatkan dalam proses pembelajaran. Meskipun demikian, peningkatan kompleksitas model juga perlu dipertimbangkan,

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 108-117
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.9778

sehingga pemilihan konfigurasi B3 sebagai titik optimal didasarkan pada kombinasi akurasi validasi tertinggi (96,7%) dan kecepatan konvergensi yang paling cepat (best epoch ke-7) di antara seluruh konfigurasi yang diuji.

3.6 Eksperimen Seri C — Variasi Rasio Discriminative Learning Rate

Menggunakan konfigurasi B3 sebagai fondasi, Seri C menguji empat variasi rasio *learning rate*. Rancangan variasi disajikan pada Tabel 4

Tabel 4. Rancangan Variasi Rasio Discriminative Learning Rate (Seri C)

Kode	Variasi	Cls LR	Enc LR	Rasio	Dasar Pemilihan
C0	Baseline (5:1)	5×10^{-5}	1×10^{-5}	05:01	Rasio acuan dari konfigurasi B3
C1	Cls agresif (10:1)	1×10^{-4}	1×10^{-5}	10:01	Classifier belajar lebih cepat
C2	Cls konservatif (3:1)	3×10^{-5}	1×10^{-5}	03:01	Update lebih merata dan seimbang
C3	Enc konservatif (10:1)	5×10^{-5}	5×10^{-6}	10:01	Cegah catastrophic forgetting

Hasil pengujian disajikan pada Tabel 5. Seluruh variasi seri C melampaui referensi C0. C1 (rasio 10:1 via *Classifier LR* agresif 1×10^{-4}) mencapai akurasi 0,956. Sebaliknya, C3 dengan rasio 10:1 yang dicapai melalui pengecilan *Encoder LR* hanya menghasilkan akurasi 0,940. Temuan ini membuktikan bahwa cara mencapai rasio lebih penting dari rasio itu sendiri: *encoder* yang terlalu dibatasi tidak dapat mengoptimalkan 42 juta parameternya meski secara teknis tidak dibekukan.

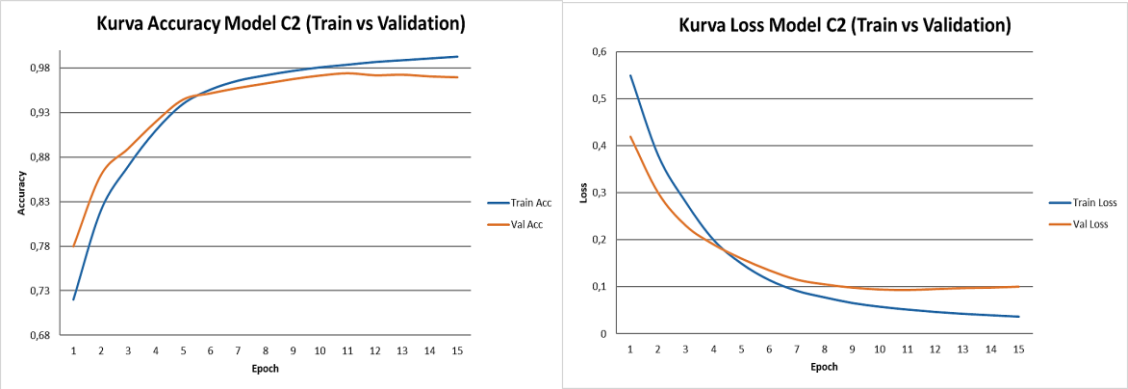
Tabel 5. Hasil Eksperimen Variasi Rasio Discriminative Learning Rate (Seri C)

Kode	Rasio	Cls LR	Enc LR	Best Ep.	Val Acc
C0	05:01	5×10^{-5}	1×10^{-5}	7	96,7
C1	10:01	1×10^{-4}	1×10^{-5}	8	97
C2	03:01	3×10^{-5}	1×10^{-5}	11	97,2
C3	10:01	5×10^{-5}	5×10^{-6}	12	96,4

Konfigurasi C2 dengan rasio 3:1 terbukti paling optimal dengan akurasi 0,957. Rasio yang lebih kecil menciptakan keseimbangan *update* yang tepat: *classifier* memimpin adaptasi terhadap domain baru sementara *encoder* mendapat kebebasan yang cukup untuk menyesuaikan representasinya secara presisi.

3.7 Evaluasi Model Optimal

Model C2 dilatih menggunakan mekanisme *early stopping* (patience=4). Val accuracy mencapai puncak pada epoch ke-11 sebesar 0,9745. Setelah epoch ke-11, val accuracy berfluktuasi sementara *train accuracy* terus naik, mengindikasikan *overfitting* ringan yang tertangani oleh *early stopping* pada epoch 15. Bobot epoch ke-11 disimpan sebagai model final.



Gambar 6. Kurva Loss (Kiri) Dan Accuracy (Kanan)

Evaluasi model final pada data pengujian disajikan pada Tabel 6. Model mencapai akurasi 95,70% dengan weighted F1-score 0,9569. Terdapat asimetri performa: kelas Pria memiliki *recall* tinggi (98,18%) namun *precision* lebih rendah (94,23%), sedangkan kelas Wanita menunjukkan *precision* tinggi (97,67%) namun *recall* lebih rendah (92,69%).

Tabel 6. Classification Report Evaluasi Model Optimal pada Data Pengujian

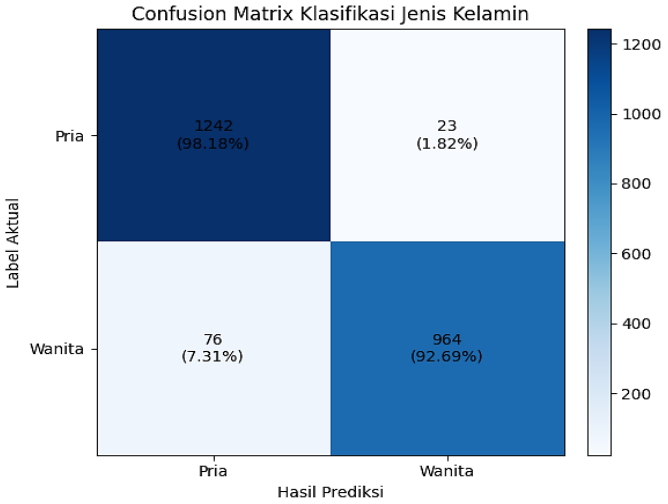
Kelas	Precision	Recall	F1-Score	Support
Pria	0,9423	0,9818	0,9617	1.265
Wanita	0,9767	0,9269	0,9512	1.040
Accuracy	---	---	0,957	2.305
Macro Avg	0,9595	0,9544	0,9564	2.305

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 108-117
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.9778

Kelas	Precision	Recall	F1-Score	Support
Weighted Avg	0,9578	0,957	0,9569	2.305

Keterangan: Total Sampel Pengujian = 2.305 Citra

Pola ini menunjukkan adanya kecenderungan model untuk lebih sensitif terhadap fitur yang dominan pada kelas Pria, yang kemungkinan dipengaruhi oleh distribusi data maupun karakteristik visual citra. Dari total 1.040 sampel Wanita, sebanyak 76 citra diprediksi sebagai Pria, sebagaimana terlihat pada Gambar 7 hasil analisis *confusion matrix*.



Gambar 7. Hasil analisis confusion matrix

Temuan ini mengindikasikan adanya kecenderungan ketidakseimbangan sensitivitas model antar kelas yang kemungkinan dipengaruhi oleh distribusi data maupun kemiripan karakteristik visual pada beberapa citra area mata.

3.8 Perbandingan dengan Penelitian Terdahulu

Tabel 7 menyajikan perbandingan performa model yang diusulkan terhadap penelitian terdahulu dalam domain klasifikasi jenis kelamin berbasis citra mata. Model ViT-B/16 dengan discriminative fine tuning menghasilkan akurasi 95,70%, melampaui seluruh pendekatan sebelumnya pada dataset yang sama.

Tabel 7. Perbandingan Performa dengan Penelitian Terdahulu

Penelitian	Metode	Dataset	Akurasi
Aini et al. [14]	CNN	11.525 citra mata	90,18%
Aini et al. [14]	MobileNetV2	11.525 citra mata	93,90%
Pradana & Wijiyanto [15]	CNN + Haar Cascade	11.525 citra mata	92,00%
Penelitian ini	ViT-B/16 + Discriminative LR	11.525 citra mata	95,70%

Peningkatan performa sebesar 1,8 percentage point dibandingkan MobileNetV2 menunjukkan bahwa pendekatan berbasis Vision Transformer ViT memiliki kemampuan representasi fitur yang baik untuk klasifikasi biometrik berbasis area mata. Selain itu, penerapan discriminative fine tuning dengan rasio learning rate 3:1 memungkinkan proses adaptasi parameter berlangsung lebih seimbang antara classifier head dan encoder pretrained, sehingga memberikan performa yang lebih optimal dibandingkan pendekatan pada penelitian sebelumnya.

4. KESIMPULAN

Penelitian ini berhasil menerapkan Vision Transformer ViT-B/16 dengan strategi discriminative fine tuning untuk klasifikasi jenis kelamin berbasis citra mata manusia. Pendekatan ini dilakukan untuk memanfaatkan kemampuan Vision Transformer dalam menangkap representasi global citra melalui mekanisme self-attention, sehingga model diharapkan mampu mengenali karakteristik biometrik area mata secara lebih efektif dibandingkan pendekatan konvensional berbasis Convolutional Neural Network (CNN). Hasil eksperimen menunjukkan bahwa konfigurasi unfreeze 6 dari 12 transformer block memberikan performa paling optimal karena mampu menjaga keseimbangan antara kemampuan adaptasi domain dan stabilitas representasi pretrained yang telah dipelajari sebelumnya. Selain itu, penerapan rasio discriminative learning rate sebesar 3:1 antara classifier head dan encoder terbukti menghasilkan proses pembelajaran yang lebih efektif dibandingkan

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 108-117
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.9778

konfigurasi lainnya, karena lapisan classifier dapat beradaptasi lebih cepat tanpa merusak representasi fitur dasar pada encoder. Model terbaik yang dihasilkan pada penelitian ini mampu mencapai akurasi sebesar 95,70% pada data pengujian dan menunjukkan performa yang lebih baik dibandingkan beberapa pendekatan berbasis CNN maupun metode klasik pada dataset yang sama. Hasil tersebut mengindikasikan bahwa pendekatan berbasis Vision Transformer memiliki kemampuan representasi fitur yang baik untuk klasifikasi biometrik area mata manusia. Meskipun demikian, penelitian ini masih memiliki beberapa keterbatasan. Analisis interpretabilitas model seperti attention map atau Grad-CAM belum dilakukan sehingga mekanisme perhatian spasial yang dipelajari model belum dapat divisualisasikan secara langsung. Selain itu, penggunaan dataset beresolusi rendah yang kemudian di-upscale menjadi 224×224 piksel berpotensi memengaruhi kualitas informasi visual yang diterima model selama proses pelatihan. Oleh karena itu, penelitian selanjutnya disarankan untuk mengeksplorasi metode interpretabilitas model, menggunakan dataset dengan variasi kondisi yang lebih kompleks, serta mengembangkan arsitektur yang lebih ringan dan efisien agar dapat diimplementasikan secara real-time pada perangkat edge atau sistem biometrik berbasis mobile.

REFERENCES

[1] M. Ngan, P. Grother, dan K. Hanaoka, "Ongoing Face Recognition Vendor Test (FRVT) Part 6A: Face Recognition Accuracy with Masks Using Pre-COVID-19 Algorithms," NIST Interagency Report 8311, National Institute of Standards and Technology, Gaithersburg, MD, USA, Jul. 2020. DOI: 10.6028/NIST.IR.8311

[2] Detik News, "Pria Bercadar Menyusup ke Jemaah Wanita di Masjid Makassar Diamankan," *Detik.com*, 2024. [Online]. Available: https://news.detik.com/berita/d-7259609. [Accessed: Apr. 20, 2026]

[3] K. Nguyen, H. Proença, dan F. Alonso-Fernandez, "Deep Learning for Iris Recognition: A Survey," *ACM Computing Surveys*, vol. 56, no. 9, Art. no. 223, 2024. DOI: 10.1145/3637525

[4] S. Minaee, A. Abdolrashidi, H. Su, M. Bennamoun, dan D. Zhang, "Biometrics Recognition Using Deep Learning: A Survey," *Artificial Intelligence Review*, vol. 56, no. 8, hlm. 8647–8695, 2023. DOI: 10.1007/s10462-022-10237-x

[5] D. Kwasny dan D. Hemmerling, "Gender and Age Estimation Methods Based on Speech Using Deep Neural Networks," *Sensors*, vol. 21, no. 14, Art. no. 4785, Jul. 2021. DOI: 10.3390/s21144785

[6] S. Haseena et al., "Prediction of the Age and Gender Based on Human Face Images Based on Deep Learning Algorithm," *Computational Intelligence and Neuroscience*, vol. 2022, Art. no. 1413597, 2022. DOI: 10.1155/2022/1413597

[7] C.-T. Hsiao, C.-Y. Lin, P.-S. Wang, dan Y.-T. Wu, "Application of Convolutional Neural Network for Fingerprint-Based Prediction of Gender, Finger Position, and Height," *Entropy*, vol. 24, no. 4, Art. no. 475, Mar. 2022. DOI: 10.3390/e24040475

[8] S. Zhang, X. Wang, A. Liu, C. Zhao, J. Wan, S. Escalera, H. Shi, Z. Wang, dan S. Z. Li, "A Dataset and Benchmark for Large-Scale Multi-Modal Face Anti-Spoofing," dalam *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, hlm. 919–928. DOI: 10.1109/CVPR.2019.00101

[9] B. M. S. Hasan dan R. J. Mstafa, "A Study of Gender Classification Techniques Based on Iris Images: A Deep Survey and Analysis," *Science Journal of University of Zakho*, vol. 10, no. 4, hlm. 222–234, 2022. DOI: 10.25271/sjuoz.2022.10.4.1039

[10] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, dan N. Houlsby, "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale," dalam *Proc. 9th International Conference on Learning Representations (ICLR 2021)*, May 2021. [Online]. Available: https://arxiv.org/abs/2010.11929

[11] J. Howard dan S. Ruder, "Universal Language Model Fine-tuning for Text Classification," dalam *Proc. 56th Annual Meeting of the Association for Computational Linguistics (ACL)*, Melbourne, Australia, 2018, hlm. 328–339. DOI: 10.18653/v1/P18-1031

[12] B. Pavel, "Female and Male Eyes," *Kaggle*, 2022. [Online]. Available: https://www.kaggle.com/datasets/burakbey0/female-and-male-eyes. [Accessed: Apr. 20, 2026]

[13] C. Kurniawan dan H. Irsyad, "Perbandingan Metode K-Nearest Neighbor Dan Naïve Bayes Untuk Klasifikasi Gender Berdasarkan Mata," *Jurnal Algoritme*, vol. 2, no. 2, hlm. 82–91, Apr. 2022. DOI: 10.35957/algoritme.v2i2.2358

[14] N. Aini dan D. Y. Liliana, "Prediksi Gender Berdasarkan Citra Mata Menggunakan Metode Convolutional Neural Network, Inception dan MobileNet," *Buletin Poltanesa*, vol. 23, no. 1, hlm. 226–232, Jun. 2022. DOI: 10.51967/tanesa.v23i1.1272

[15] A. I. Pradana dan W. Wijiyanto, "Identifikasi Jenis Kelamin Otomatis Berdasarkan Mata Manusia Menggunakan Convolutional Neural Network (CNN) dan Haar Cascade Classifier," *G-Tech: Jurnal Teknologi Terapan*, vol. 8, no. 1, hlm. 502–511, Jan. 2024. DOI: 10.33379/gtech.v8i1.3814

[16] H. Touvron *et al.*, "Training Data-Efficient Image Transformers & Distillation Through Attention," dalam *Proc. International Conference on Machine Learning (ICML)*, PMLR vol. 139, 2021, hlm. 10347–10357. [Online]. Available: https://arxiv.org/abs/2012.12877

[17] I. Loshchilov dan F. Hutter, "Decoupled Weight Decay Regularization," dalam Proc. International Conference on Learning Representations (ICLR), New Orleans, LA, USA, 2019. [Online]. Available: https://arxiv.org/abs/1711.05101

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, dan I. Polosukhin, "Attention Is All You Need," dalam *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017, hlm. 5998–6008. [Online]. Available: https://arxiv.org/abs/1706.03762

[19] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, dan B. Guo, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," dalam *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, hlm. 10012–10022. DOI: 10.1109/ICCV48922.2021.00986

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 7, No. 3, May 2026, Page 108-117
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v7i3.9778

[20] K. He, X. Chen, S. Xie, Y. Li, P. Dollar, dan R. Girshick, “Masked Autoencoders Are Scalable Vision Learners,” dalam *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, hlm. 16000–16009. DOI: 10.1109/CVPR52688.2022.01553

[21] M. Tan dan Q. V. Le, “EfficientNetV2: Smaller Models and Faster Training,” dalam *Proc. International Conference on Machine Learning (ICML)*, PMLR vol. 139, 2021, hlm. 10096–10106. [Online]. Available: https://arxiv.org/abs/2104.00298

[22] A. Radford *et al.*, “Learning Transferable Visual Models From Natural Language Supervision,” dalam *Proc. International Conference on Machine Learning (ICML)*, PMLR vol. 139, 2021, hlm. 8748–8763. [Online]. Available: https://arxiv.org/abs/2103.00020

[23] V. K. Suravarapu dan H. Y. Patil, "Performance Evaluation of Enhanced Deep Learning Classifiers for Person Identification and Gender Classification," *Scientific Reports*, vol. 15, Art. no. 28182, Aug. 2025. DOI: 10.1038/s41598-025-12474-w

[24] K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang, dan D. Tao, "A Survey on Vision Transformer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, hlm. 87–110, Jan. 2023. DOI: 10.1109/TPAMI.2022.3152247

[25] J. Yosinski, J. Clune, Y. Bengio, dan H. Lipson, "How Transferable Are Features in Deep Neural Networks?" dalam *Advances in Neural Information Processing Systems 27 (NIPS)*, 2014, hlm. 3320–3328. [Online]. Available: https://arxiv.org/abs/1411.1792

[26] B. J. Ferrell, "Fine-tuning Strategies for Classifying Community-Engaged Research Studies Using Transformer-Based Models: Algorithm Development and Improvement Study," *JMIR Formative Research*, vol. 7, Art. no. e41137, Feb. 2023. DOI: 10.2196/41137

[27] M. Raghu, T. Unterthiner, S. Kornblith, C. Zhang, dan A. Dosovitskiy, "Do Vision Transformers See Like Convolutional Neural Networks?" dalam Advances in Neural Information Processing Systems (NeurIPS), vol. 34, 2021, hlm. 12116–12128. [Online]. Available: https://arxiv.org/abs/2108.08810

[28] M. Hossin dan M. N. Sulaiman, "A Review on Evaluation Metrics for Data Classification Evaluations," International Journal of Data Mining & Knowledge Management Process, vol. 5, no. 2, hlm. 1–11, 2015. DOI: 10.5121/ijdkp.2015.5201

[29] C. Bisogni, L. Cascone, dan F. Narducci, "Periocular Data Fusion for Age and Gender Classification," *Journal of Imaging*, vol. 8, no. 11, Art. no. 307, Nov. 2022. DOI: 10.3390/jimaging8110307