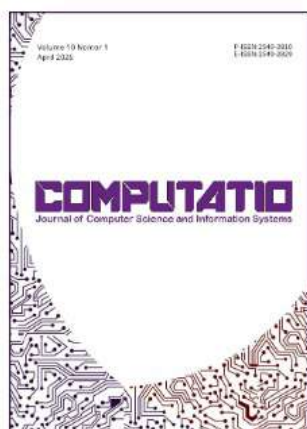


Prediksi Emosi Manusia Menggunakan Swin Transformer Berdasarkan Ekspresi Wajah



pdf

Published: Jul 23, 2026

DOI:
<https://doi.org/10.24912/computatio.v10i1.38116>

Keywords:
Deep Learning, Ekspresi Wajah, Haar Cascade, Prediksi Emosi, Swin Transformer

Dimensions



Juventia Agnesia

Chairisni Lubis

Abstract

Perkembangan pesat teknologi Artificial Intelligence (AI) telah membawa pengaruh besar dalam bidang pengolahan citra, khususnya pada sistem analisis wajah untuk mengenali emosi manusia melalui ekspresi wajah. Penelitian ini merupakan sebuah sistem pengenalan emosi wajah dengan memanfaatkan metode deep learning berbasis arsitektur Swin Transformer, yang mengandalkan mekanisme shifted window self-attention untuk mengekstraksi informasi fitur secara efektif baik pada skala lokal maupun global. Program aplikasi yang dikembangkan menerima masukan berupa citra wajah dalam bentuk foto dan dirancang agar dapat dioperasikan secara real time. Sebelum memasuki tahap klasifikasi emosi, proses pendeteksian wajah dilakukan terlebih dahulu menggunakan metode Haar Cascade, kemudian dilanjutkan dengan tahapan pra-pemrosesan untuk menyesuaikan data input dengan kebutuhan model. Fitur-fitur yang dihasilkan oleh Swin Transformer selanjutnya diproses menggunakan fully connected layer dengan fungsi aktivasi softmax untuk menghasilkan keluaran berupa label emosi beserta nilai confidence score. Berdasarkan hasil pengujian, sistem pengenalan emosi menggunakan Swin Transformer memperoleh rata-rata akurasi sebesar 87,54% dengan akurasi validasi mencapai 74,82%. Sementara itu, metode Haar Cascade menunjukkan performa pendeteksian wajah dengan rata-rata akurasi sebesar 98,44%, meskipun masih ditemukan kesalahan deteksi pada objek non-wajah dalam kondisi tertentu. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa Swin Transformer memiliki potensi yang sangat baik untuk dikembangkan lebih lanjut dalam sistem prediksi emosi berbasis ekspresi wajah di masa mendatang

ISSUE

Vol. 10 No. 1 (2026): Computatio: Journal of Computer Science and Information Systems

SECTION

Sinta Accreditation



PAPER



Download Template Here

Submit Paper Here

NEW AUTHOR ?

Temporarily, the registration will be handled via google form. Please provide valid google email to fill the form.

Please click "register button" below

Register

ABOUT

::: Contact :::
::: Editorial Team :::
::: Reviewer :::
::: Focus and Scope :::
::: Publication Ethics :::
::: Indexing :::

Keywords

PREDIKSI EMOSI MANUSIA MENGUNAKAN SWIN TRANSFORMER BERDASARKAN EKSPRESI WAJAH

Juventia Agnecia¹, Chairisni Lubis²

^{1,2} Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia

E-mail: ¹juventia.535220237@stu.untar.ac.id, Corresponding Author : chairisnil@fti.untar.ac.id

Abstrak

Perkembangan pesat teknologi *Artificial Intelligence* (AI) telah membawa pengaruh besar dalam bidang pengolahan citra, khususnya pada sistem analisis wajah untuk mengenali emosi manusia melalui ekspresi wajah. Penelitian ini merupakan sebuah sistem pengenalan emosi wajah dengan memanfaatkan metode *deep learning* berbasis arsitektur *Swin Transformer*, yang mengandalkan mekanisme *shifted window self-attention* untuk mengekstraksi informasi fitur secara efektif baik pada skala lokal maupun global. Program aplikasi yang dikembangkan menerima masukan berupa citra wajah dalam bentuk foto dan dirancang agar dapat dioperasikan secara real time. Sebelum memasuki tahap klasifikasi emosi, proses pendeteksian wajah dilakukan terlebih dahulu menggunakan metode *Haar Cascade*, kemudian dilanjutkan dengan tahapan pra-pemrosesan untuk menyesuaikan data input dengan kebutuhan model. Fitur-fitur yang dihasilkan oleh Swin Transformer selanjutnya diproses menggunakan *fully connected layer* dengan fungsi aktivasi softmax untuk menghasilkan keluaran berupa label emosi beserta nilai *confidence score*. Berdasarkan hasil pengujian, sistem pengenalan emosi menggunakan *Swin Transformer* memperoleh rata-rata akurasi sebesar 87,54% dengan akurasi validasi mencapai 74,82%. Sementara itu, metode *Haar Cascade* menunjukkan performa pendeteksian wajah dengan rata-rata akurasi sebesar 98,44%, meskipun masih ditemukan kesalahan deteksi pada objek non-wajah dalam kondisi tertentu. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa Swin Transformer memiliki potensi yang sangat baik untuk dikembangkan lebih lanjut dalam sistem prediksi emosi berbasis ekspresi wajah di masa mendatang.

Kata kunci— *Deep Learning, Ekspresi Wajah, Haar Cascade, Prediksi Emosi, Swin Transformer*

Abstract

The rapid advancement of *Artificial Intelligence* (AI) technology has significantly influenced the field of image processing, particularly in facial analysis systems aimed at recognizing human emotions through facial expressions. This study proposes the development of a facial emotion recognition system utilizing a deep learning approach based on the *Swin Transformer* architecture, which employs a *shifted window self-attention* mechanism to effectively capture both local and global facial features. The proposed system accepts facial image inputs in the form of photographs and is designed to operate in real-time environments. Prior to emotion classification, facial regions are detected using the *Haar Cascade* method, followed by a data preprocessing stage to ensure compatibility with the model's input requirements. Feature representations extracted by the *Swin Transformer* are subsequently processed through a *fully connected layer* with a softmax activation function to generate emotion labels along with their corresponding confidence scores. Experimental results indicate that the *Swin Transformer*-based emotion recognition system achieves an average accuracy of 87.54%, with a validation accuracy of 74.82%. Meanwhile, the *Haar Cascade* face detection method demonstrates a high averagedetection accuracy of 98.44%, although misdetections of non-facial objects are

observed under certain conditions. Overall, the findings of this study suggest that the Swin Transformer architecture holds strong potential for further development and optimization in future facial expression-based emotion prediction systems.

Keywords— *Deep Learning, Facial Expressions, Haar Cascade, Emotion Prediction, Swin Transformer*

1. PENDAHULUAN

Perkembangan teknologi kecerdasan buatan atau *Artificial Intelligence* (AI) dalam beberapa tahun terakhir ini telah memberikan kontribusi besar di berbagai bidang, salah satunya adalah untuk analisis wajah. Deteksi ekspresi wajah manusia untuk mengidentifikasi emosi yang sedang dirasakan menjadi salah satu cabang yang menarik perhatian karena emosi adalah bagian penting dari kondisi psikologis manusia yang berperan untuk memunculkan suatu tindakan sebagai respons terhadap stimulus tertentu [1][2]. Emosi tidak hanya mencerminkan perasaan, tetapi juga mempengaruhi cara seseorang berinteraksi dengan lingkungan sekitarnya. Menurut Garcia et al. (2017), emosi dapat diidentifikasi melalui enam pendekatan utama, yaitu analisis percakapan, teks, gerakan tubuh dan gestur, kondisi fisiologis, serta ekspresi wajah [3]. Di antara semua pendekatan tersebut, ekspresi wajahlah yang dianggap sebagai indikator paling dominan karena dapat diamati secara langsung. Namun, terdapat beberapa tantangan yang mungkin akan dihadapi seperti adanya variasi citra karena foto diambil dari sudut pandang yang berbeda, pengekspresian yang tergantung masing-masing pribadi, dan ekspresi netral serta sedih yang cenderung mirip sehingga tidak menutup kemungkinan jika model sulit membedakan emosi tersebut [4].

Metode berbasis *Convolutional Neural Network* (CNN) selama ini telah menunjukkan perkembangan yang cukup pesat dan mampu memberikan performa yang cukup baik dalam pengenalan citra. Meskipun demikian, pendekatan ini masih memiliki keterbatasan dalam menangkap hubungan spasial yang kompleks, khususnya dalam membedakan variasi ekspresi wajah yang memiliki karakteristik visual yang mirip, misalnya emosi sedih dan netral [5]. Keterbatasan tersebut menimbulkan pertanyaan terkait kemampuan metode konvensional dalam menangani kompleksitas struktur citra wajah secara menyeluruh.

Perancangan ini mengusulkan penggunaan metode Transformer dengan arsitekturnya yaitu *Swin Transformer* sebagai metode utama dalam sistem pengenalan emosi berdasarkan ekspresi wajah. Model ini bekerja dengan memanfaatkan mekanisme *shifted window self-attention*, di mana citra dibagi ke dalam beberapa *patch* atau jendela kecil yang diproses secara bertahap dan digeser untuk menangkap informasi global dari seluruh area wajah [6]. Keunggulan tersebut memungkinkan model untuk memahami pola visual yang kompleks, termasuk variasi sudut pengambilan gambar dan kondisi pencahayaan.

Tahapan perancangan meliputi pengumpulan dataset citra wajah manusia dengan empat kelas emosi yang telah diberikan label, yaitu senang (*happy*), sedih (*sad*), netral (*neutral*), dan marah (*angry*). Selanjutnya citra akan melalui proses Haar Cascades untuk deteksi wajah serta pra-pemrosesan data untuk meningkatkan kualitas data agar sesuai dengan kebutuhan model. Model Swin Transformer kemudian diterapkan, dilatih, divalidasi, dan dievaluasi menggunakan *confusion matrix* dengan berbagai metriknya yaitu akurasi, presisi, *recall*, dan *f1-score* guna memastikan sistem mampu mengenali emosi wajah secara akurat. Melalui tahapan ini, perancangan diharapkan dapat berkontribusi dalam pengembangan sistem analisis wajah yang lebih baik di masa mendatang.

2. METODE PENELITIAN

2.1 Data

Data yang digunakan pada penelitian ini diperoleh dari *website* Kaggle yang merupakan salah satu *platform* penyedia dataset terbesar [7]. Dataset yang akan digunakan merupakan data citra *Emotion Detection* berupa citra wajah manusia yang terdiri dari 35.685 citra yang terbagi menjadi tujuh kelas yaitu *angry*, *disgusted*, *fearful*, *happy*, *neutral*, *sad*, dan *surprised*. Akan tetapi, jumlah dataset yang akan digunakan adalah sebanyak 26.217 citra dengan empat kelas yaitu *happy*, *angry*, *sad*, dan *neutral*. Hal ini dilakukan karena ketiga kelas lainnya memiliki jumlah citra yang jauh lebih sedikit yang menyebabkan data tidak seimbang sehingga memungkinkan model bias terhadap kelas yang jumlah datanya lebih banyak. Detail jumlah sampel data tersebut dapat dilihat pada Tabel 1 dan Tabel 2 di bawah ini.

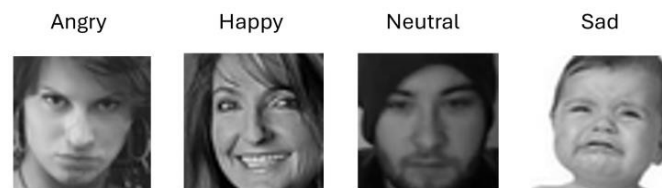
Tabel 1. Jumlah Data Latih untuk Proses Pelatihan

No	Kelas	Jumlah Data
1	<i>Angry</i> (Marah)	3.995
2	<i>Happy</i> (Senang)	7.215
3	<i>Neutral</i> (Netral)	4.965
4	<i>Sad</i> (Sedih)	4.830
Jumlah Data Latih		21.005

Tabel 2 Jumlah Data Uji untuk Proses Pelatihan

No	Kelas	Jumlah Data
1	<i>Angry</i> (Marah)	958
2	<i>Happy</i> (Senang)	1.774
3	<i>Neutral</i> (Netral)	1.233
4	<i>Sad</i> (Sedih)	1.247
Jumlah Data Latih		5.212

Contoh citra dari setiap kelas pada dataset *Emotion Detection* tersebut dapat dilihat pada Gambar 1 berikut. Data citra tersebut memiliki ukuran 48 x 48 piksel dan *grayscale*.



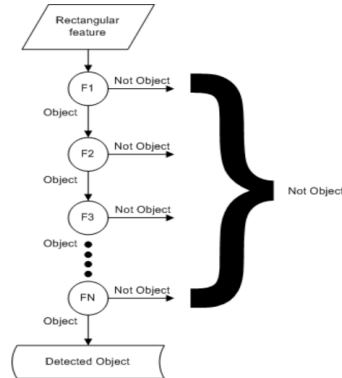
Gambar 1. Contoh Citra Kelas Emosi

2.2 Haar Cascade

Metode *Haar Cascade Classifier* merupakan teknik yang umum dipakai untuk mendeteksi objek, khususnya wajah pada citra digital. Algoritma ini memanfaatkan fitur *Haar* untuk mengenali pola visual tertentu pada gambar, yang kemudian dipilih melalui proses pembelajaran dari kumpulan data wajah dan area selain wajah [8]. Metode ini bekerja dengan menggabungkan banyak *classifier* sederhana dalam bentuk *cascade* sehingga bisa mendeteksi bagian wajah secara bertahap dan cepat.

Proses pendeteksian wajah dengan *Haar Cascade Classifier* diawali dengan pengambilan citra input, yang kemudian diubah menjadi citra *grayscale* untuk memudahkan ekstraksi fitur pixel. Selanjutnya *classifier* dijalankan pada citra tersebut dengan parameter seperti *scaleFactor* dan *minNeighbors*, yang masing-masing berfungsi untuk menentukan skala deteksi dan jumlah tetangga yang diperlukan agar suatu area dianggap sebagai wajah [9]. Setelah area

lolos melalui seluruh tahap *classifier*, sistem akan menggambar *bounding box* di sekitar lokasi wajah yang terdeteksi. Studi penerapan metode ini menunjukkan bahwa teknik ini cukup baik dalam menangani deteksi wajah dalam berbagai kondisi latar belakang secara cepat dan efisien. Proses dari *haar cascade classifier* dapat dilihat pada Gambar 2 di bawah ini.



Gambar 2. Proses Deteksi Objek Haar Cascade

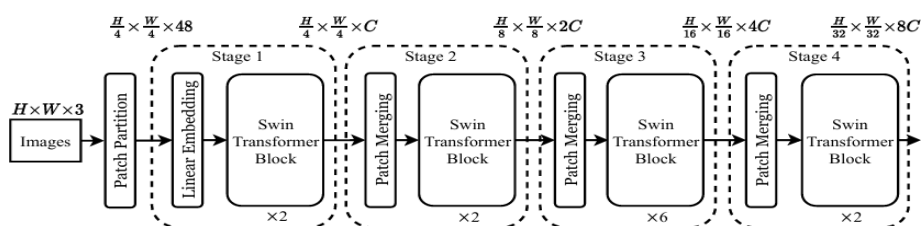
Metode *Haar Cascade* sering dimanfaatkan karena kecepatan komputasinya dan dukungan implementasi pada pustaka populer seperti OpenCV [8]. Keunggulan utama metode ini adalah kemampuannya untuk memproses citra secara cepat bahkan sehingga dapat diterapkan pada perangkat dengan keterbatasan sumber daya komputasi.

2.3 Swin Transformer

Swin Transformer merupakan model pengembangan dari metode *Vision Transformer* (ViT) yang berfungsi untuk melakukan tugas klasifikasi citra dengan tujuan meningkatkan efisiensi dalam memproses citra beresolusi tinggi. *Swin Transformer* ini pertama kali dikemukakan oleh Ze Liu et al (2022) yang menghadirkan mekanisme *shifted window* dimana citra dibagi ke dalam jendela lokal, kemudian digeser agar model mempelajari hubungan antar jendela tersebut [6]. Selain itu, *Swin Transformer* juga dirancang dengan representasi hierarkis yaitu model ini tidak hanya bekerja pada *fixed size patch*, tetapi melakukannya dengan proses bertahap dari resolusi paling rendah ke resolusi yang lebih tinggi [10].

Cara kerja Swin Transformer adalah dengan menggunakan mekanisme *shifted window self-attention* yang memungkinkan pemrosesan citra dengan membaginya menjadi *patch-patch* (jendela) kecil dan menggeser *patch* untuk menangkap fitur global sekaligus lokal [6]. Terdapat beberapa keunggulan yang dimiliki oleh metode Swin Transformer ini di antaranya yaitu membutuhkan komputasi yang lebih rendah dari metode Transformer lainnya karena adanya *window-based self-attention*, menghasilkan performa yang cukup baik untuk dataset yang besar, serta dapat menangani resolusi citra yang cukup kompleks [11].

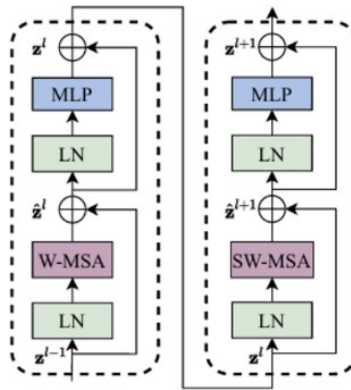
Dalam arsitektur Swin Transformer pada Gambar 3 [6], dapat dilihat bahwa model ini terdiri dari beberapa komponen utama yang bekerja secara sekuensial (berurutan), mulai dari *patch partition*, *linear embedding*, *Swin transformer block*, sampai ke *patch merging* [6].



Gambar 3. Arsitektur Swin Transformer

Swin Transformer dimulai dari tahap *patch partition* yaitu citra dipotong menjadi *patch* kecil berukuran tetap, dimana setiap *patch* tersebut direpresentasikan sebagai vektor satu dimensi sehingga total *patch* yang diperoleh berjumlah $\frac{H}{P} \times \frac{W}{P}$, dengan P adalah ukuran *patch*. Proses ini

membuat citra yang kompleks menjadi lebih sederhana agar mudah diproses. Setelah itu, vektor hasil *patch partition* tersebut diubah ke ruang representasi baru berdimensi C dengan *linear projection*, yang memungkinkan model merepresentasikan setiap *patch* ke dalam *embedding* yang lebih informatif. Tahapan inilah yang disebut *linear embedding*. Kemudian, output tersebut akan dimasukkan ke *Swin Transformer block* yang terdiri dari beberapa *layer* sekuensial untuk mengekstraksi fitur. Isi dari *swin transformer block* dapat dilihat pada Gambar 4 [6].



Gambar 4. *Swin Transformer Block*

Pada Gambar 5 tersebut, dapat terlihat bahwa isi dari *swin transformer block* adalah *layer normalization* (LN), *Window Multi-Head Self Attention* (W-MSA), *Residual Connection* (bentuk lingkaran), *Multi Layer Perceptron* (MLP), dan *Shifted Window Multi-Head Self Attention* (SW-MSA). Komponen utama dari *swin transformer block* tersebut ada di W-MSA yang berfungsi untuk mencari informasi pada setiap *patch* dan SW-MSA yang memanfaatkan pergeseran jendela (*window*) agar informasi global pada citra terjaga. *Self Attention* pada W-MSA dan SW-MSA dapat didefinisikan dengan rumus (1) berikut [6].

$$Attention(Q, K, V) = Softmax_{\frac{QK^T}{\sqrt{d}}} V \quad (1)$$

Di rumus tersebut, Q merupakan *query matrix* (vektor pernyataan), K adalah *key matrix* (vektor kunci), V yaitu *value matrix* (vektor nilai), K^T yang merupakan transpos dari *key matrix*, serta faktor skala normalisasi yang didefinisikan dengan $\sqrt{d}$. Setelah melewati *stage 1* seperti yang ada di Gambar 4, fitur yang dihasilkan akan melalui tahap *patch merging* dimana *patch* kecil digabungkan menjadi *patch* besar guna untuk menurunkan resolusi spasial dan meningkatkan kedalaman fitur [11].

2.4 Metode Evaluasi

Metode evaluasi yang digunakan pada penelitian ini adalah *Confusion Matrix* yang umum digunakan untuk menganalisis performa model klasifikasi dengan melakukan perbandingan prediksi model terhadap data sebenarnya [12]. *Confusion matrix* memiliki beberapa metrik evaluasi yang digunakan untuk menilai kinerja model klasifikasi. Metrik yang digunakan dalam penelitian ini meliputi akurasi, *precision*, *recall*, dan *f1-score*. Keempat metrik tersebut digunakan untuk memberikan gambaran menyeluruh terhadap kemampuan model dalam melakukan prediksi emosi.

2.4.1 Akurasi

Akurasi merupakan rasio hasil yang diprediksi benar oleh model dari total keseluruhan prediksi [13]. Akurasi mengukur seberapa sering model untuk menghasilkan prediksi yang tepat secara keseluruhan dengan rumus (2) [13].

$$Akurasi = \frac{Jumlah\ prediksi\ benar}{Total\ data} \times 100\% \quad (2)$$

2.4.2 Precision

Precision atau yang disebut juga presisi berfungsi untuk mengukur keakuratan model dalam memprediksi setiap kelas. Rumus *precision* dapat dilihat pada rumus (3) dimana TP adalah *True Positive* dan FP adalah *False Positive* [14].

$$Precision\ x = \frac{TP\ x}{TP\ x + FP\ x} \quad (3)$$

2.4.3 Recall

Recall atau sensitivitas dimanfaatkan untuk mengukur banyaknya data positif dari suatu kelas yang diprediksi secara benar. Secara matematis, rumus *recall* dapat dilihat pada rumus (4) yaitu TP adalah *True Positive* dan FP adalah *False Positive* [14].

$$Recall\ x = \frac{TP\ x}{TP\ x + FP\ x} \quad (4)$$

2.4.4 F1-Score

F1-Score merupakan hasil harmonisasi dari *precision* dan *recall* untuk mencapai keseimbangan serta secara matematis dapat ditulis dengan rumus (5) [14].

$$F1\ Score\ x = 2 \times \frac{Presisi\ x \times Sensitivitas\ x}{Presisi\ x + Sensitivitas\ x} \quad (5)$$

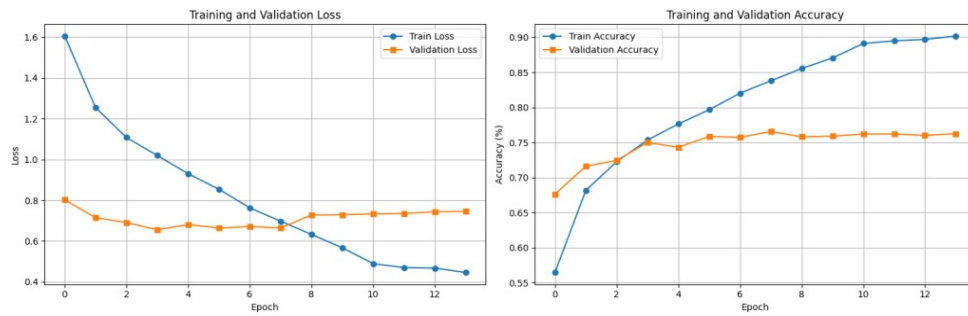
3. HASIL DAN PEMBAHASAN

3.1 Pra-Pemrosesan Data

Sebelum diterapkan ke dalam perancangan, data terlebih dahulu melewati tahap *preprocessing data* atau pra-pemrosesan data untuk membantu metode yang akan digunakan agar dapat menghasilkan akurasi yang lebih baik. Pra-pemrosesan data yang dilakukan adalah dengan melakukan *resize* ukuran citra menjadi 224 x 224 *pixel* agar dapat diproses model nantinya, normalisasi, serta melakukan *undersampling* dan *oversampling* sekaligus augmentasi data (*flipping*, *zoom*, rotasi, dan *brightness adjustment*) pada skenario yang berbeda yang bertujuan untuk meningkatkan kualitas data.

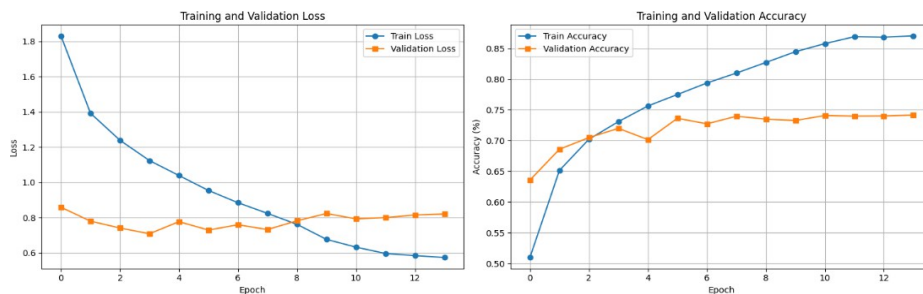
3.2 Skenario Eksperimen

Eksperimen dilakukan sebanyak dua kali percobaan dengan menggunakan pengolahan data latih yang berbeda dimana pada percobaan pertama, data latih pada kelas *angry*, *sad*, dan *neutral* akan diperbanyak melalui proses *oversampling* sekaligus augmentasi untuk menangani ketidakseimbangan jumlah data sehingga data latih pada kelas *angry* menjadi 7.110 data, kelas *sad* menjadi 7.115 data, dan kelas *neutral* menjadi 7.035 data. Sedangkan pada percobaan kedua, seluruh data latih akan dilakukan *undersampling* terlebih dahulu menjadi 3.000 data, baru setelah itu dilakukan proses augmentasi sekaligus *oversampling* menjadi 5.000 data agar jumlah data latih menjadi lebih seimbang. Proses pelatihan dilakukan dengan 100 *epoch* menggunakan *earlystopping* dimana proses *training* akan dihentikan secara otomatis jika *validation loss* tidak mengalami penurunan setelah 10 *epoch* berturut-turut. Grafik *loss* dan akurasi pada percobaan pertama dapat dilihat pada Gambar 5 berikut.



Gambar 5. Grafik Loss dan Akurasi Data *Oversampling* + Augmentasi

Gambar 5 tersebut menunjukkan bahwa *training* berhenti di *epoch* ke-14 dimana *validation loss* (validasi loss) terendah terdapat pada *epoch* ke-4 dan *validation accuracy* (akurasi validasi) tertinggi ada di *epoch* ke-8. Kemudian, grafik pelatihan *loss* dan akurasi percobaan kedua yang menggunakan data latih *undersampling* + *oversampling* + augmentasi dapat dilihat pada Gambar 6.



Gambar 6. Grafik Loss dan Akurasi Data *Undersampling* + *Oversampling* + Augmentasi

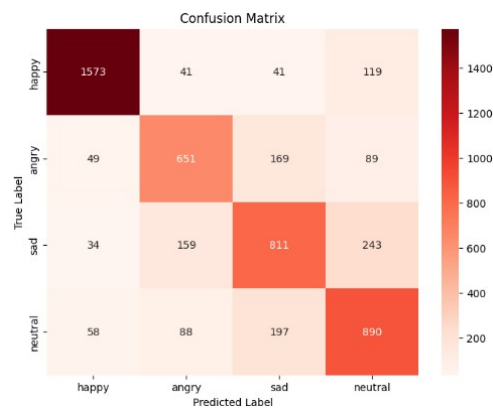
Pada percobaan kedua yang menggunakan data latih yang sudah di *undersampling* + *oversampling* + augmentasi, pelatihan juga berhenti di *epoch* ke-14 dimana *validation loss* terendah terdapat pada *epoch* ke-4 dan *validation accuracy* tertinggi terdapat pada *epoch* ke-14.

3.3 Evaluasi Model

Evaluasi model dilakukan dengan dua cara yaitu dengan melihat *confusion matrix* hasil pelatihan dan pengujian pada program secara langsung. Pada percobaan pertama yang menggunakan data latih *oversampling* + augmentasi, model disimpan ke dalam dua jenis *file* yaitu *model_best_loss* (model dengan *validation loss* terendah sebesar) dan *model_best_acc* (model dengan *validation accuracy* tertinggi). *Confusion matrix* dari *model_best_loss* dan *model_best_acc* dapat dilihat pada Gambar 7 dan Gambar 8 berikut.

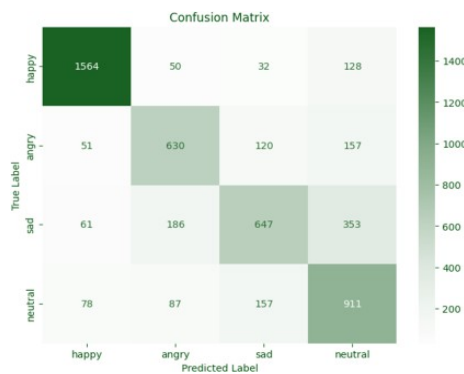


Gambar 7. *Confusion Matrix* Model *Validation Loss* Terendah *Oversampling* + Augmentasi (*model_best_loss*)

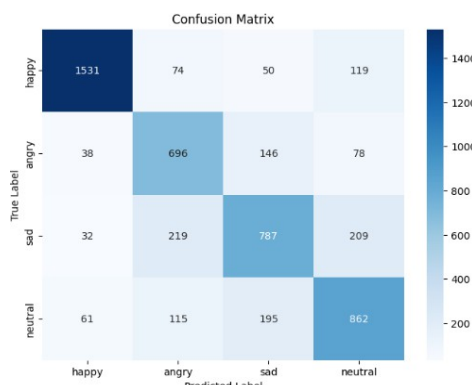


Gambar 8. *Confusion Matrix Model Validation Accuracy Tertinggi Data Oversampling + Augmentasi (model_best_acc)*

Kemudian, percobaan kedua dilakukan dengan menggunakan data latih yang di *undersampling* terlebih dahulu, baru di *oversampling* sekaligus augmentasi. Model pada percobaan kedua ini juga disimpan ke dalam dua file yang berbeda yaitu `model_best_loss_pt2` (model dengan *validation loss* terendah sebesar) dan `model_best_acc_pt2` (model dengan *validation accuracy* tertinggi). Gambar 9 dan Gambar 10 berikut menunjukkan *confusion matrix* dari `model_best_loss_pt2` dan `model_best_acc_pt2`.



Gambar 9. *Confusion Matrix Model Validation Loss Terendah Undersampling + Oversampling + Augmentasi (model_best_loss_pt2)*



Gambar 10. *Confusion Matrix Model Validation Accuracy Tertinggi Undersampling + Oversampling + Augmentasi (model_best_acc_pt2)*

Detail nilai akurasi validasi dan akurasi pelatihan (*training*) pada kedua percobaan yang telah dilakukan dengan masing-masing dua model yang tersimpan dapat dilihat pada Tabel 3 berikut.

Tabel 3. Rekapitulasi Hasil Akurasi Kedua Eksperimen

		Akurasi Validasi	Akurasi <i>Training</i>
Data Latih <i>Oversampling</i> + Augmentasi	Model_best_loss	74.98%	75.33%
	Model_best_acc	76.55%	83.78%
Data Latih <i>Undersampling</i> + <i>Oversampling</i> + Augmentasi	Model_best_loss_pt2	71.99%	73.08%
	Model_best_acc_pt2	74.16%	87.05%

Berdasarkan Tabel 3 tersebut, dapat terlihat bahwa model yang memperoleh akurasi validasi tertinggi adalah model_best_loss dan model_best_acc yang didapatkan dari percobaan pertama yang menggunakan data latih *oversampling* + augmentasi (pada kelas *angry*, *sad*, dan *neutral*). Oleh karena itu, kedua model tersebut yang akan dilakukan pengujian secara manual pada program karena akurasi validasi berguna untuk memastikan model dapat mempelajari pola baru yang tidak ada di data latih sebelumnya.

Dataset yang akan digunakan untuk pengujian manual pada program terbagi menjadi 2 yaitu dataset pengujian 1 dan dataset pengujian 2. Dataset pengujian 1 terdiri dari 96 citra uji berupa citra wajah manusia yang terbagi menjadi 4 kelas dan diperoleh dari hasil pemotretan menggunakan kamera *handphone* terhadap 24 orang. Dataset pengujian 2 merupakan hasilprediksi secara langsung dari *webcam* perangkat (*real time*) yang terdiri dari 36 citra uji dari 12 orang dengan total 42 objek. Hasil prediksi dataset pengujian 1 pada model_best_loss dan model_best_acc dapat dilihat dari Tabel 4 dan Tabel 5 berikut.

Tabel 4. Prediksi Keseluruhan Dataset Pengujian 1 Pada model best_loss

Eksperimen Pengujian	Jumlah Benar	Persentase Akurasi
Jumlah area wajah yang terdeteksi	93/96	96.88%
Jumlah prediksi emosi yang benar	75/96	78.01%
Rata – rata		87.44%

Tabel 5. Prediksi Keseluruhan Dataset Pengujian 1 Pada model best_acc

Eksperimen Pengujian	Jumlah Benar	Persentase Akurasi
Jumlah area wajah yang terdeteksi	93/96	96.88%
Jumlah prediksi emosi yang benar	83/96	86.46%
Rata – rata		91.67%

Berdasarkan kedua tabel tersebut, dapat dibuatkan *confusion matrix* yang dapat dilihat pada Tabel 6 yang merupakan implementasi dari model_best_loss dan Tabel 7 adalah implementasi dari model_best_acc.

Tabel 6. *Confusion Matrix* untuk model best_loss pada Dataset Pengujian 1

		Label Prediksi			
Label Sebenarnya		Netral	Senang	Sedih	Marah
	Netral	22	1	1	0
	Senang	1	23	0	0
	Sedih	6	0	15	2
	Marah	4	2	2	15

Tabel 7. *Confusion Matrix* untuk model *best_acc* pada Dataset Pengujian 1

Label Sebenarnya	Label Prediksi				
		Netral	Senang	Sedih	Marah
	Netral	20	1	3	0
	Senang	0	24	0	0
	Sedih	1	0	22	0
	Marah	2	1	3	17

Dari *confusion matrix* tersebut, dapat dilihat bahwa kelas yang paling bagus untuk diprediksi dari dataset pengujian 1 adalah kelas Senang dimana pada model *best_loss* berhasil terprediksi 23/24 citra sehingga mendapatkan akurasi 95.38% dan model *best_acc* berhasil memprediksi 24/24 citra sehingga didapatkan akurasinya yaitu 100%. Selain itu, kelas yang paling kurang adalah kelas Marah yang pada model *best_loss* terprediksi 15/24 dengan akurasi 62.5% sedangkan pada model *best_acc* terprediksi 17/24 sehingga didapatkan akurasinya yaitu 70.38%. Selanjutnya pengujian dilakukan menggunakan dataset pengujian 2 yang merupakan hasil prediksi dari *webcam* perangkat secara langsung sebanyak 36 citra uji yang juga terbagi menjadi 4 kelas dengan total 42 objek. Hasil prediksi dari dataset pengujian 2 dari model *best_loss* dan model *best_acc* dapat dilihat pada Tabel 8 dan Tabel 9 berikut.

Tabel 8. Prediksi Keseluruhan Dataset Pengujian 2 Pada model *best_loss*

Eksperimen Pengujian	Jumlah Benar	Persentase Akurasi
Jumlah area wajah yang terdeteksi	42/42	100%
Jumlah prediksi emosi yang benar	40/42	95.24%
Rata – rata		97.62%

Tabel 9. Prediksi Keseluruhan Dataset Pengujian 2 Pada model *best_acc*

Eksperimen Pengujian	Jumlah Benar	Persentase Akurasi
Jumlah area wajah yang terdeteksi	42/42	100%
Jumlah prediksi emosi yang benar	38/42	90.47%
Rata – rata		95.24%

Confusion matrix juga dapat diterapkan berdasarkan hasil kedua tabel tersebut yang dapat dilihat pada Tabel 10 dan Tabel 11 dimana Tabel 10 adalah implementasi dari model *best_loss* dan Tabel 11 adalah implementasi dari model *best_acc*.

Tabel 10. *Confusion Matrix* untuk model *best_loss* pada Dataset Pengujian 2

Label Sebenarnya	Label Prediksi				
		Netral	Senang	Sedih	Marah
	Netral	10	0	0	0
	Senang	0	10	0	0
	Sedih	0	0	10	0
	Marah	1	0	1	10

Tabel 11. *Confusion Matrix* untuk model *best_acc* pada Dataset Pengujian 2

Label Sebenarnya	Label Prediksi				
		Netral	Senang	Sedih	Marah
	Netral	9	0	1	0
	Senang	1	9	0	0
	Sedih	1	0	9	0
	Marah	0	0	1	11

Berdasarkan kedua *confusion matrix* tersebut, kelas yang paling bagus untuk diprediksi dari dataset pengujian 2 adalah kelas Netral, Senang dan Sedih pada model *best_loss* dimana pada 10/10 objek sehingga mendapatkan akurasi 100% dan kelas Marah pada model *best_acc* yang berhasil memprediksi 11/12 citra sehingga didapatkan akurasinya yaitu 91.66%. Kemudian, dari *confusion matrix* yang diuji pada program (Tabel 6, 7, 10, dan 11), didapatkan *classification report* seperti Tabel 12 (dataset pengujian 1) dan Tabel 13 (dataset pengujian 2) berikut.

Tabel 12. *Classification Report* Pada Dataset Pengujian 1

	Akurasi	Presisi	Sensitivitas	F1-score
Model_best_loss	78.01%	81.67%	78.12%	78.44%
Model_best_acc	86.46%	89.46%	86.46%	87.17%

Tabel 13. *Classification Report* Pada Dataset Pengujian 2

	Akurasi	Presisi	Sensitivitas	F1-score
Model_best_loss	95.24%	95.45%	95.83%	95.32%
Model_best_acc	90.47%	90.9%	90.42%	90.44%

Secara keseluruhan, dapat disimpulkan bahwa model *best_loss* lebih unggul dalam memprediksi emosi pada data citra dinamis yang lebih kompleks sedangkan model *best_acc* menghasilkan performa yang lebih unggul ketika memprediksi data uji statis dimana kedua model tersebut menunjukkan keunggulan yang tergantung pada jenis data yang akan diuji. Rata-rata akurasi yang didapatkan oleh model *Swin Transformer* berdasarkan pengujian yang dilakukan secara manual pada program adalah sebesar 87.54, sedangkan rata-rata akurasi yang didapatkan pada metode *Haar Cascades* adalah sebesar 98.44%.

4. KESIMPULAN

Berdasarkan hasil evaluasi dan pengujian yang telah dilakukan, dapat disimpulkan bahwa penggunaan metode *Swin Transformer* yang dikombinasikan dengan *Haar Cascade* memberikan kinerja yang cukup baik dalam melakukan prediksi emosi. Secara umum, beberapa poin kesimpulan utama yang dapat ditarik dari penelitian ini adalah sebagai berikut:

1. Penerapan teknik augmentasi dan *oversampling* pada eksperimen pertama terbukti mampu menjaga keberagaman data tanpa menghilangkan informasi penting, sehingga model dapat mempelajari pola secara lebih optimal dibandingkan yang dilakukan *undersampling* terlebih dahulu pada eksperimen kedua. Hal ini ditunjukkan oleh rata-rata akurasi validasi yang didapatkan eksperimen pertama sebesar 75,76%, sedangkan pada eksperimen kedua diperoleh rata-rata akurasi validasi sebesar 73,89%.
2. Model *best_acc*, yaitu model dengan nilai akurasi validasi tertinggi, menunjukkan kinerja yang unggul pada pengujian data statis dengan akurasi mencapai 91,67%. Sementara itu, model *best_loss*, yaitu model dengan nilai *validation loss* terendah, menunjukkan performa yang lebih unggul pada pengujian data dinamis (*real time*) dengan tingkat akurasi sebesar 97,62%.
3. Berdasarkan proses pengujian yang dilakukan pada program, model *Swin Transformer* yang dikembangkan mampu mencapai tingkat akurasi keseluruhan sebesar 87,54% dalam memprediksi emosi berdasarkan ekspresi wajah.

DAFTAR PUSTAKA

- [1] R. Ramadani, Indriyati, and B. Priyambadha, "Deteksi Emosi Berdasarkan Ekspresi Wajah Menggunakan Metode Mini-Xception dengan Image Augmentation," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 9, Sep. 2025.
- [2] K. I. K. Jajan and Prof. Dr. Eng. A. M. Abdulazeez, "Facial Expression Recognition Based on Deep Learning: A Review," *Indonesian Journal of Computer Science*, vol. 13, no. 1, pp. 183–204, Feb. 2024, doi: 10.33022/ijcs.v13i1.3705.
- [3] J. M. Garcia-Garcia, V. M. R. Penichet, and M. D. Lozano, "Emotion detection : a technology review," in *Proceedings of the XVIII International Conference on Human Computer Interaction*, New York, NY, USA: ACM, Sep. 2017, pp. 1–8. doi: 10.1145/3123818.3123852.
- [4] C. Bialek, A. M. Grega, and M. Grega, "An Efficient Approach to Face Emotion Recognition with Convolutional Neural Networks," *Electronics (Basel)*, vol. 12, no. 12, pp. 1–21, Jun. 2023, doi: 10.3390/electronics12122707.
- [5] N. Yalçın and M. Alisawi, "Introducing a novel dataset for facial emotion recognition and demonstrating significant enhancements in deep learning performance through pre- processing techniques," *Heliyon*, vol. 10, no. 20, pp. 1–22, Oct. 2024, doi: 10.1016/j.heliyon.2024.e38913.
- [6] Z. Liu et al., "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Oct. 2021, pp. 10012–10022. doi: 10.1109/ICCV48922.2021.00986.
- [7] W. Chow, "A Pedagogy that Uses a Kaggle Competition for Teaching Machine Learning: an Experience Sharing," *2019 IEEE International Conference on Engineering, Technology and Education (TALE)*, pp. 1–5, Dec. 2019, doi: 10.1109/TALE48000.2019.9226005.
- [8] D. R. Rochmawati, "Deteksi Wajah Dengan Metode Haar Cascade Menggunakan Opencv (Face Detection With Haar Cascade Method Using Opencv)," *Jurnal Teknologi Komputer dan Informatika*, vol. 3, no. 1, Sep. 2024, doi: 10.59820/tekomin.v3i1.293.
- [9] C. A. Antipona, R. R. Magsino, R. M. Dioses, and K. E. Mata, "An Enhancement of Haar Cascade Algorithm Applied to Face Recognition for Gate Pass Security," *arXiv preprint*, Nov. 2024.
- [10] A. Nur, K. Nurhanafi, and E. R. Putri, "Brain Tumor Segmentation in MR Images Using Swin Transformer," *Atom Indonesia*, vol. 51, no. 2, pp. 97–108, May 2025, doi: 10.55981/aij.2025.1580.
- [11] D. Yao and Y. Shao, "A data efficient transformer based on Swin Transformer," *Vis Comput*, vol. 40, no. 4, pp. 2589–2598, Apr. 2024, doi: 10.1007/s00371-023-02939-2.
- [12] E. Elvin and C. Lubis, "Klasifikasi Citra Ikan Menggunakan convolutional Neural Network," *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 10, no. 1, Mar. 2022, doi: 10.24912/jiksi.v10i1.17827.
- [13] M. Y. Petra and L. Lina, "Sistem Pendeteksi Kantuk pada Mata Pengendara Roda Empat dengan Algoritma Convolutional Neural Network," *Computatio : Journal of Computer Science and Information Systems*, vol. 8, no. 1, pp. 20–28, Apr. 2024, doi: 10.24912/computatio.v8i1.16220.
- [14] H. Wisianto, H. R. Anglie, and A. Effendi, "Klasifikasi Daun Tebu Menggunakan Pendekatan Berbasis Grayscale, PCA DAN SVM," *Jurnal Ilmu Komputer dan Sistem Informasi*, vol. 13, no. 2, pp. 1–7, Sep. 2025, doi: <https://doi.org/10.24912/jiksi.v13i2.35138>.