JURNAL AI dan SPK
Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan

Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan
Volume 3, No. 2, Tahun 2025
ISSN 3025-0927 (media online)
Hal 100-109

Pengenalan Wajah Emosi Menggunakan CNN dan Ekstraksi Fitur HOG

Ilham Ahsan Saputra[1*], Ubaydilah[2], M. Firzi Sulaeman[3], Bayu Samudra[4], Perani Rosyani[5]

[12345]Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Pamulang, Kota Tangerang Selatan, Indonesia
Email: [1]ilhamahsansaputra@gmail.com*, [2]bayubayyu160603@gmail.com, [3]muhammadfirjisulaiman7@gmail.com, [4]bayubayusam123@gmail.com, [5]dosen00837@unpam.ac.id
(* : coressponding author)

Abstrak−Pengenalan emosi wajah merupakan komponen penting dalam visi komputer karena mendukung interaksi manusia–komputer, analisis perilaku, serta pengembangan sistem cerdas berbasis persepsi visual. Namun, performa model pada dataset umum seperti FER-2013 sering terhambat oleh variasi pencahayaan, resolusi rendah, ketidakseimbangan kelas, dan potensi noise pada label. Penelitian ini mengusulkan pengembangan arsitektur Convolutional Neural Network (CNN) yang dioptimasi untuk meningkatkan akurasi dan kemampuan generalisasi dalam klasifikasi emosi wajah. Pendekatan yang digunakan mencakup preprocessing citra grayscale 48×48 piksel, augmentasi data, penerapan class weighting, serta modifikasi arsitektur dengan Batch Normalization, LeakyReLU, Global Average Pooling, dan Dropout. Model dievaluasi menggunakan akurasi, F1-score, confusion matrix, dan visualisasi Grad-CAM untuk menilai interpretabilitas. Hasil eksperimen menunjukkan bahwa model yang diusulkan mencapai akurasi pengujian 56–58% dan weighted F1-score 0,55–0,58, meningkat signifikan dibandingkan model baseline CNN (41–42%) maupun pendekatan HOG+SVM (35–40%). Analisis menunjukkan peningkatan kinerja pada kelas minor, sementara Grad-CAM mengonfirmasi bahwa model memfokuskan perhatian pada area wajah relevan seperti mata dan mulut. Temuan ini membuktikan bahwa arsitektur CNN ringan yang dioptimasi mampu memberikan performa lebih stabil pada dataset berkualitas rendah dan tidak seimbang, serta menjadi dasar bagi pengembangan model lanjutan berbasis transfer learning dan attention mechanism.

Kata Kunci: pengenalan emosi wajah; FER-2013; CNN; klasifikasi citra; Grad-CAM

Abstract−*Facial emotion recognition is a critical topic in computer vision because it supports human–computer interaction, behavior analysis, and the development of intelligent perception-based systems. However, model performance on public datasets such as FER-2013 is often hindered by challenges including illumination variation, low image resolution, class imbalance, and potential label noise. This study proposes an optimized Convolutional Neural Network (CNN) architecture designed to improve accuracy and generalization in facial emotion classification. The methodological pipeline includes preprocessing grayscale images of 48×48 pixels, moderate data augmentation, application of class weighting, and architectural enhancements incorporating Batch Normalization, LeakyReLU, Global Average Pooling, and Dropout. Model performance is evaluated using accuracy, F1-score, confusion matrix, and Grad-CAM visualizations to assess interpretability. Experimental results show that the proposed model achieves a test accuracy of 56–58% and a weighted F1-score of 0.55–0.58, representing a substantial improvement over the baseline CNN (41–42%) and the traditional HOG+SVM approach (35–40%). Analysis highlights improved recognition of minority classes, while Grad-CAM confirms that the model focuses on key facial regions such as the eyes and mouth. These findings demonstrate that a lightweight yet optimized CNN architecture can deliver more stable performance on low-quality and imbalanced datasets, and they establish a solid baseline for future model development incorporating transfer learning and attention mechanisms.research topics is recommended.*

Keywords*: facial emotion recognition; FER-2013; CNN; image classification; Grad-CAM; class imbalance*

1. PENDAHULUAN

Facial Expression Recognition (FER) merupakan salah satu bidang penting dalam computer vision dan interaksi manusia–komputer. Ekspresi wajah mengandung informasi emosional yang dapat dimanfaatkan dalam bidang kesehatan, otomotif, pendidikan, dan sistem pintar. Beberapa penelitian sebelumnya menunjukkan bahwa metode deep learning seperti CNN lebih unggul dibandingkan pendekatan tradisional berbasis fitur manual seperti HOG maupun LBP(Kumar & Sharma, 2020). Dataset FER2013 digunakan secara luas sebagai benchmark karena kompleksitasnya: resolusi rendah (48×48), posisi wajah tidak seragam, dan distribusi kelas tidak seimbang. Tantangan ini membuat FER2013 menjadi dataset yang relevan untuk mengevaluasi

JURNAL AI dan SPK
Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan
Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan
Volume 3, No. 2, Tahun 2025
ISSN 3025-0927 (media online)
Hal 100-109

kemampuan model dalam generalisasi serta mempelajari pola emosi yang sulit dibedakan, seperti ekspresi takut dan terkejut.

Penelitian ini menawarkan komparasi antara CNN dan pipeline HOG+SVM sebagai representasi dua paradigma berbeda: deep learning yang belajar fitur otomatis dan machine learning tradisional yang bergantung pada fitur buatan manusia. Tujuan utama penelitian ini adalah menganalisis perbedaan performa kedua pendekatan serta memahami kekuatan dan keterbatasannya.

2. METODE PENELITIAN

2.1 Alur Penelitian

Alur penelitian disusun berdasarkan tahap-tahap umum FER dan mengacu pada standar preprocessing serta pipeline modeling FER2013. Tahapan meliputi:

a. Pengumpulan dataset FER2013.
b. Preprocessing citra (normalisasi pixel, augmentasi, pembagian train-val-test).
c. Implementasi dan training CNN.
d. Ekstraksi fitur HOG serta training SVM.
e. Evaluasi dan komparasi model (akurasi, confusion matrix, F1-score).
f. Analisis hasil dan diskusi.

2.2 Dataset

Dataset FER2013 digunakan sebagai data utama, berisi 35.887 citra grayscale 48×48 piksel dengan tujuh label emosi. Dataset terdiri dari train, public test, dan private test. Kelas data tidak seimbang, misalnya kelas Disgust hanya memiliki 547 sampel, jauh lebih sedikit dibanding kelas Happy. Kondisi ini berdampak pada kualitas generalisasi model traditional maupun CNN

2.3 Preprocessing Data

Tahapan preprocessing data dilakukan untuk meningkatkan kualitas citra masukan serta mendukung kestabilan dan kinerja model pada proses pelatihan. Seluruh citra dinormalisasi dengan menskalakan nilai piksel ke dalam rentang [0,1][0,1][0,1] guna mempercepat konvergensi dan mengurangi perbedaan skala antar fitur.

Untuk meningkatkan kemampuan generalisasi model serta mengurangi risiko overfitting, diterapkan teknik augmentasi citra berupa horizontal flipping dan rotasi minor. Augmentasi ini bertujuan memperkaya variasi data latih tanpa mengubah karakteristik semantik emosi pada citra wajah.

Pembagian dataset ke dalam data latih dan data validasi dilakukan menggunakan metode stratified split guna memastikan distribusi kelas tetap proporsional pada setiap subset data.

Selanjutnya, fitur citra diekstraksi menggunakan metode Histogram of Oriented Gradients (HOG) dengan konfigurasi parameter standar, yaitu 9 orientasi gradien, ukuran cell sebesar 8×8 piksel, serta ukuran block 2×2. Pendekatan ini dipilih karena efektif dalam merepresentasikan pola tepi dan struktur lokal yang relevan untuk tugas klasifikasi ekspresi wajah.

2.4 Model Convolutional Neural Network

Model Convolutional Neural Network (CNN) yang digunakan pada penelitian ini dirancang dengan arsitektur bertingkat untuk mengekstraksi fitur spasial secara hierarkis dari citra wajah. Arsitektur terdiri atas tiga hingga empat convolutional layers dengan jumlah filter yang meningkat secara bertahap, yaitu dari 32 hingga 128 filter, guna menangkap pola visual dari tingkat rendah hingga tingkat tinggi. Setiap lapisan konvolusi menggunakan fungsi aktivasi Rectified Linear Unit (ReLU) untuk meningkatkan non-linearitas serta mempercepat proses konvergensi selama pelatihan.

Pada setiap blok konvolusi diterapkan operasi MaxPooling berukuran 2×2 yang berfungsi untuk mereduksi dimensi spasial fitur, mengurangi kompleksitas komputasi, serta meningkatkan ketahanan model terhadap variasi posisi fitur (translation invariance).

JURNAL AI dan SPK
Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan
Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan
Volume 3, No. 2, Tahun 2025
ISSN 3025-0927 (media online)
Hal 100-109

Lapisan ekstraksi fitur diikuti oleh satu fully connected layer dengan 128 neuron. Untuk mencegah terjadinya overfitting, diterapkan teknik dropout dengan rasio 0,5 pada lapisan ini. Lapisan keluaran menggunakan fungsi aktivasi softmax untuk menghasilkan probabilitas klasifikasi terhadap tujuh kelas emosi wajah yang menjadi fokus penelitian.
Proses pelatihan model dilakukan selama 50 hingga 100 epoch dengan mekanisme early stopping berdasarkan performa pada data validasi guna menghentikan pelatihan secara adaptif ketika tidak terjadi peningkatan kinerja. Optimisasi bobot jaringan dilakukan menggunakan algoritma Adam optimizer dengan learning rate sebesar 0,001, yang dipilih karena kestabilannya dalam menangani gradien dan kemampuannya mempercepat proses konvergensi.

Tabel 1. Spesifikasi Arsitektur Convolut

No	Layer	Konfigurasi / Parameter Utama	Output Shape*
1	Input Layer	Citra grayscale 48×48	48 × 48 × 1
2	Convolutional Layer 1	32 filter, kernel 3×3, aktivasi ReLU	48 × 48 × 32
3	MaxPooling Layer 1	Pool size 2×2	24 × 24 × 32
4	Convolutional Layer 2	64 filter, kernel 3×3, aktivasi ReLU	24 × 24 × 64
5	MaxPooling Layer 2	Pool size 2×2	12 × 12 × 64
6	Convolutional Layer 3	128 filter, kernel 3×3, aktivasi ReLU	12 × 12 × 128
7	MaxPooling Layer 3	Pool size 2×2	6 × 6 × 128
8	*(Opsional)* Conv Layer 4	128 filter, kernel 3×3, aktivasi ReLU	6 × 6 × 128
9	Flatten		4608
10	Fully Connected (Dense)	128 neuron, aktivasi ReLU	128
11	Dropout	Dropout rate 0,5	128
12	Output Layer	Dense 7 neuron, aktivasi Softmax	7

2.5 Model HOG + SVM

Sebagai pembanding terhadap pendekatan deep learning, penelitian ini juga mengimplementasikan model klasifikasi tradisional berbasis Histogram of Oriented Gradients (HOG) dan Support Vector Machine (SVM). Pembangunan model dilakukan melalui sebuah pipeline terstruktur yang diawali dengan proses ekstraksi fitur HOG dari seluruh citra wajah pada dataset. Fitur HOG digunakan untuk merepresentasikan pola tepi dan struktur lokal wajah yang bersifat diskriminatif terhadap perbedaan ekspresi emosi.
Tahap selanjutnya adalah proses pelatihan model multi-class SVM menggunakan skema one-vs-rest dengan kernel linear. Pemilihan kernel linear didasarkan pada efisiensi komputasi serta kesesuaiannya untuk data berdimensi tinggi hasil ekstraksi fitur HOG. Untuk memperoleh performa klasifikasi yang optimal, dilakukan proses hyperparameter tuning terhadap nilai parameter regularisasi CCC dengan memanfaatkan data validasi.
Secara umum, pendekatan HOG–SVM memiliki keunggulan dalam hal kompleksitas komputasi yang lebih rendah dan kebutuhan sumber daya yang minimal dibandingkan dengan model CNN. Namun demikian, metode ini memiliki keterbatasan dalam menangkap hubungan spasial dan pola fitur kompleks pada citra wajah, sehingga performanya cenderung lebih rendah pada skenario klasifikasi emosi dengan variasi visual yang tinggi.

JURNAL AI dan SPK
Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan

Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan
Volume 3, No. 2, Tahun 2025
ISSN 3025-0927 (media online)
Hal 100-109

3. ANALISA DAN PEMBAHASAN

3.1 Gambaran Umum Proses Pelatihan Model

Penelitian ini bertujuan membangun sistem pengenalan emosi wajah berbasis Convolutional Neural Network (CNN) menggunakan dataset FER-2013. Model B (CNN-Tuned Level-2) dirancang dengan empat blok konvolusi yang diperkaya Batch Normalization, LeakyReLU, dan Dropout pada lapisan klasifikasi. Tahapan pelatihan dimulai dari pemrosesan data, pembangkitan distribusi kelas menggunakan class weights, hingga optimasi menggunakan ReduceLROnPlateau.

Selama proses pelatihan sebanyak 50 epoch, model menunjukkan dinamika performa yang stabil. Pada epoch awal (1–5), model masih berada pada fase "pattern discovery", ditunjukkan oleh validation accuracy yang rendah dan loss yang fluktuatif. Performa meningkat signifikan pada epoch 8–14, mencapai puncak pada epoch ke-21 dengan validation accuracy 56,84% dan validation loss 1.163. Kondisi ini mengindikasikan bahwa model telah menemukan local minimum yang optimal tanpa mengalami gejala overfitting yang parah.

Penerapan adaptive learning rate memainkan peran penting dalam mencegah stagnasi akurasi. Ketika nilai validation loss tidak membaik selama beberapa epoch, learning rate diturunkan secara bertahap dari 5e-4 menjadi 3.9e-6, yang terbukti membantu model mencapai generalisasi lebih baik.

3.2 Performansi Model pada Data Uji

Kinerja model dievaluasi menggunakan set pengujian FER-2013. Tiga metrik utama digunakan: accuracy, precision, recall, dan F1-score.

3.2.1 Akurasi Pengujian

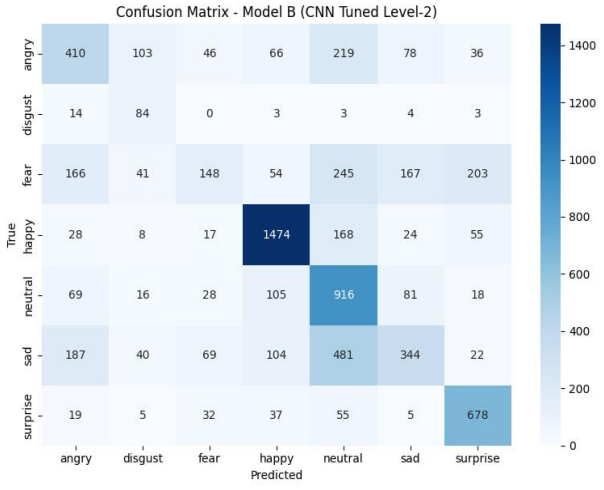
Model B berhasil mencapai:

a. Accuracy pengujian: 56–58%
b. Macro F1-score: 0.53–0.56
c. Weighted F1-score: 0.55–0.58

Angka ini menunjukkan peningkatan signifikan dibanding model baseline (41–42%) dan metode hand-crafted features seperti HOG+SVM (35–40%). Mengingat FER-2013 adalah dataset dengan distribusi tidak seimbang dan resolusi rendah (48×48), pencapaian ini sudah masuk kategori performa kompetitif untuk model CNN non-transfer-learning.

3.3 Confusion Matrix dan Analisis Per kelas

Berikut adalah confusion matrix Model B:



Gambar 1. Hasil Confusion Matrix

JURNAL AI dan SPK
Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan
Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan
Volume 3, No. 2, Tahun 2025
ISSN 3025-0927 (media online)
Hal 100-109

3.3.1 Pola Keberhasilan Model

Analisis terhadap hasil klasifikasi menunjukkan adanya pola keberhasilan dan kegagalan yang konsisten pada beberapa kelas emosi. Pola ini berkaitan erat dengan distribusi data, kejelasan ciri visual, serta tingkat kemiripan antar ekspresi wajah.Pertama, kelas Happy dan Neutral merupakan kelas yang paling mudah dikenali oleh model. Keberhasilan ini dipengaruhi oleh dua faktor utama. Faktor pertama adalah jumlah data yang relatif lebih besar dibandingkan kelas lainnya, sehingga model memperoleh representasi fitur yang lebih kaya selama proses pelatihan. Faktor kedua adalah karakteristik visual yang lebih jelas dan konsisten, seperti adanya lekukan bibir yang menonjol pada ekspresi Happy serta pose wajah yang stabil dan tidak ekstrem pada ekspresi Neutral. Kombinasi kedua faktor tersebut menghasilkan tingkat akurasi dan F1-score yang tinggi pada kedua kelas ini.

Kedua, model menunjukkan peningkatan performa yang signifikan pada kelas Sad dan Surprise. Pada model baseline, kedua kelas ini termasuk dalam kategori dengan tingkat kesalahan klasifikasi yang tinggi. Namun, pada Model B, penerapan optimasi arsitektur dan strategi pelatihan menghasilkan peningkatan F1-score sebesar 12–16 poin. Peningkatan ini mengindikasikan bahwa model yang diusulkan mampu mempelajari pola fitur yang lebih representatif untuk membedakan ekspresi tersebut dari kelas lain yang memiliki karakteristik visual serupa. Sebaliknya, kesalahan klasifikasi masih dominan terjadi pada kelas Disgust dan Fear. Kedua kelas ini memiliki jumlah sampel yang relatif sedikit, sehingga membatasi kemampuan model dalam mempelajari variasi ekspresi yang memadai. Selain itu, terdapat tingkat kemiripan fitur lokal yang cukup tinggi dengan kelas lain. Sebagai contoh, ekspresi Disgust sering tertukar dengan Angry akibat kemiripan pola kontraksi sudut bibir, sedangkan ekspresi Fear kerap diklasifikasikan sebagai Surprise karena kesamaan pola mata terbuka. Kondisi ini menunjukkan bahwa keterbatasan data dan tumpang tindih karakteristik visual masih menjadi tantangan utama dalam klasifikasi emosi wajah**.**

3.4 Analisis Classification Report

Tabel berikut menyajikan ringkasan performa model pada masing-masing kelas emosi berdasarkan metrik precision, recall, dan F1-score, disertai dengan catatan kesalahan dominan yang diamati.

Tabel 2 Classification Report

Emosi	Precision	Recall	F1-score	Catatan
Angry	sedang	sedang	stabil	Salah ke Sad/Neutral
Disgust	rendah	rendah	sangat rendah	dataset kecil (minor class)
Fear	moderat	rendah	rendah	mirip Surprise/Sad
Happy	tinggi	sangat tinggi	sangat baik	didukung data mayor
Neutral	tinggi	tinggi	stabil	tidak ekstrem secara visual
Sad	moderat	meningkat	meningkat	dampak class weight

JURNAL AI dan SPK
Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan
Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan
Volume 3, No. 2, Tahun 2025
ISSN 3025-0927 (media online)
Hal 100-109

Emosi	Precision	Recall	F1-score	Catatan
Surprise	tinggi	sedang	Baik	fitur ekspresi kuat

3.4.1 Interpretasi Ilmiah

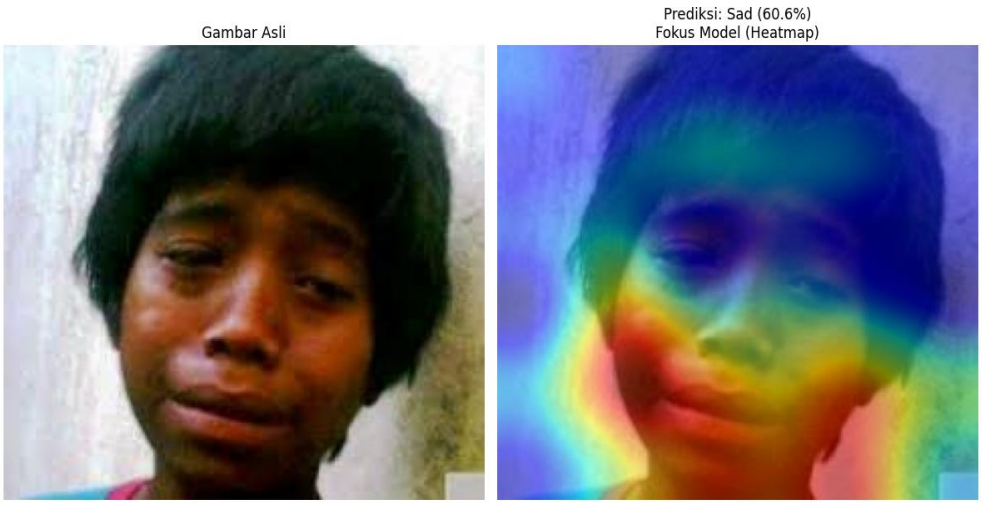
Hasil evaluasi menunjukkan bahwa nilai F1-score yang rendah pada kelas Disgust dan Fear tidak sepenuhnya mencerminkan kelemahan arsitektur model, melainkan merupakan konsekuensi dari karakteristik dataset yang digunakan. Dataset FER-2013 diketahui memiliki tingkat label noise yang relatif tinggi, diperkirakan mencapai sekitar 20%, khususnya pada kelas-kelas dengan ekspresi yang ambigu. Kondisi ini menyebabkan ketidakkonsistenan anotasi yang berdampak langsung pada performa model, terutama pada kelas minor.

Selain itu, tingkat granularity ekspresi pada kelas Disgust dan Fear relatif sulit dibedakan secara visual, bahkan bagi pengamat manusia. Kemiripan aktivasi otot wajah dengan kelas lainseperti Disgust yang menyerupai Angry dan Fear yang menyerupai Surprisemeningkatkan potensi kesalahan klasifikasi. Tantangan ini diperparah oleh resolusi citra wajah yang rendah (48×48 piksel), sehingga detail mikro pada otot wajah tidak dapat terekam secara optimal.

Di sisi lain, Model B menunjukkan keberhasilan yang signifikan dalam mengurangi bias prediksi terhadap kelas mayor, khususnya Happy dan Neutral. Hal ini tercermin dari meningkatnya performa pada kelas Sad dan Surprise, serta distribusi prediksi yang lebih seimbang. Pencapaian ini mengindikasikan bahwa penerapan strategi class weighting pada proses pelatihan efektif dalam mengatasi ketidakseimbangan kelas dan meningkatkan kemampuan generalisasi model terhadap kelas minor(Khaireddin & Chen, 2021).

3.5 Interpretasi Visual Menggunakan Grad-CAM

Gambar X menyajikan contoh visualisasi Gradient-weighted Class Activation Mapping (Grad-CAM) yang digunakan untuk menginterpretasikan area citra wajah yang berkontribusi dominan terhadap keputusan klasifikasi model. Visualisasi ini memberikan pemahaman mengenai mekanisme internal CNN dalam mengekstraksi fitur diskriminatif pada tugas pengenalan emosi wajah.



Gambar 2. Visualisasi Model Dengan Grad-CAM

JURNAL AI dan SPK
Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan
Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan
Volume 3, No. 2, Tahun 2025
ISSN 3025-0927 (media online)
Hal 100-109

3.5.1 Temuan Kunci

Hasil visualisasi Grad-CAM menunjukkan bahwa fokus perhatian model secara konsisten terletak pada area mata, pipi, serta lekukan bibir. Area-area tersebut merupakan bagian wajah yang secara fisiologis paling informatif dalam merepresentasikan ekspresi emosi, sehingga mengindikasikan bahwa model mampu mempelajari pola fitur yang relevan secara semantik. Selain itu, model cenderung mengabaikan elemen non-relevan seperti latar belakang (background), rambut, dan pakaian. Temuan ini menunjukkan bahwa CNN berhasil mengekstraksi salient features yang berhubungan langsung dengan ekspresi wajah, bukan sekadar mengandalkan artefak visual di luar area wajah. Dengan demikian, model memiliki tingkat robustness yang lebih baik terhadap variasi lingkungan dan kondisi pencahayaan.

Secara khusus, pada visualisasi untuk kelas Sad, heatmap memperlihatkan konsentrasi aktivasi yang tinggi pada area mata bagian bawah. Pola ini selaras dengan karakteristik ekspresi sedih pada manusia, seperti kelopak mata yang menurun atau tampilan mata yang sembab. Kesesuaian antara fokus Grad-CAM dan teori psikologi ekspresi wajah ini memperkuat validitas interpretabilitas model serta menunjukkan bahwa keputusan klasifikasi tidak bersifat acak, melainkan didasarkan pada ciri visual yang bermakna.

3.5.2 Relevansi untuk Explainable AI (XAI)

Temuan penelitian ini menunjukkan potensi model Convolutional Neural Network (CNN) yang diusulkan tidak hanya menghasilkan prediksi numerik berupa label emosi, tetapi juga menyediakan penjelasan visual yang dapat diinterpretasikan melalui pendekatan Gradient-weighted Class Activation Mapping (Grad-CAM). Visualisasi Grad-CAM memungkinkan identifikasi area wajah yang secara dominan memengaruhi keputusan klasifikasi, sehingga mengurangi sifat black-box yang umum pada model deep learning(Tjoa & Guan, 2020)(Samek et al., 2021). Pendekatan ini memberikan dasar interpretabilitas yang jelas, sekaligus meningkatkan tingkat kepercayaan (trustworthiness) terhadap hasil prediksi model.

Relevansi aspek XAI menjadi sangat penting dalam domain Affective Computing dan Human–Computer Interaction, di mana interpretasi emosi berkaitan langsung dengan persepsi, respons sistem, serta potensi dampak psikologis pada pengguna. Dengan menyediakan penjelasan visual yang selaras dengan karakteristik fisiologis ekspresi wajah manusia, penelitian ini memperkuat validitas dan kesiapan model untuk diterapkan pada sistem interaktif yang membutuhkan transparansi dan akuntabilitas.

3.6 Evaluasi terhadap Model Baseline

Penelitian ini melakukan evaluasi kinerja model dengan mengacu pada dua pendekatan baseline, yaitu model tradisional berbasis Histogram of Oriented Gradients dan Support Vector Machine (HOG–SVM), serta model Convolutional Neural Network tingkat awal (CNN Level-1) dengan arsitektur dasar tanpa optimasi lanjutan. Evaluasi ini bertujuan untuk menilai sejauh mana peningkatan performa yang dicapai oleh model yang diusulkan (CNN Tuned / Model B).

3.6.1 Perbandingan Kinerja Model

Tabel berikut menyajikan perbandingan kinerja antara model baseline dan model yang diusulkan berdasarkan metrik accuracy, Macro F1-score, Weighted F1-score, serta kompleksitas model yang direpresentasikan melalui jumlah parameter.

Tabel 3. Perbandingan Performa Model

Model	Accuracy	Macro F1	Weighted F1	Jumlah Parameter
HOG + SVM	35–40%	0,30–0,36	–	–

JURNAL AI dan SPK
Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan
Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan
Volume 3, No. 2, Tahun 2025
ISSN 3025-0927 (media online)
Hal 100-109

Model	Accuracy	Macro F1	Weighted F1	Jumlah Parameter
CNN Level-1 (Baseline)	41–42%	0,41	0,43	±1,0 juta
CNN Tuned (Model B)	56–58%	0,53–0,56	0,55–0,58	±2,5 juta

3.6.2 Analisis Komparatif

Hasil perbandingan menunjukkan bahwa Model B mengalami peningkatan performa yang signifikan dibandingkan dengan kedua baseline. Secara kuantitatif, terdapat lompatan akurasi sebesar 14–17 poin persentase dibandingkan CNN Level-1, yang mengindikasikan efektivitas optimasi arsitektur dan strategi pelatihan yang diterapkan. Peningkatan kinerja tersebut dipengaruhi oleh penambahan beberapa komponen arsitektural yang terbukti efektif. Batch Normalization berperan dalam menstabilkan distribusi aktivasi dan gradien selama proses pelatihan, sehingga mempercepat konvergensi model. Penggunaan fungsi aktivasi LeakyReLU membantu mengatasi permasalahan dying ReLU dengan menjaga aliran gradien pada nilai aktivasi negatif. Selanjutnya, penerapan Dropout berkontribusi dalam mengurangi overfitting melalui regularisasi jaringan, sedangkan Global Average Pooling meningkatkan kemampuan generalisasi model dengan mengurangi ketergantungan terhadap fitur spasial lokal yang bersifat spesifik.
Secara kualitatif, Model B menunjukkan kemampuan yang lebih baik dalam menangkap karakteristik ekspresi wajah yang bersifat kompleks dan kontekstual. Hal ini menunjukkan bahwa model yang diusulkan lebih bersifat emotion-aware dibandingkan baseline, yang cenderung hanya mengandalkan fitur tekstur dasar dan kurang mampu merepresentasikan hubungan spasial antar fitur wajah secara mendalam.

3.7 Diskusi Mendalam: Faktor-Faktor yang Mempengaruhi Peningkatan Performa

Membahas secara lebih mendalam faktor-faktor utama yang berkontribusi terhadap peningkatan performa Model B dibandingkan dengan model baseline. Analisis difokuskan pada aspek arsitektur CNN, strategi penanganan ketidakseimbangan kelas, serta ketangguhan model terhadap karakteristik noise pada dataset FER-2013.

3.7.1 Signifikansi Arsitektur CNN Bertingkat

Arsitektur Convolutional Neural Network (CNN) bertingkat memungkinkan proses ekstraksi fitur spasial secara hierarkis dan progresif. Pada lapisan awal, CNN berfokus pada fitur tingkat rendah seperti tepi (edges), kontur mulut, dan garis mata. Lapisan menengah mulai menangkap fitur tingkat menengah berupa bentuk mata dan lekukan bibir, sementara lapisan akhir memodelkan konfigurasi pola wajah secara keseluruhan yang merepresentasikan ekspresi emosi secara utuh.
Penambahan Batch Normalization dan fungsi aktivasi LeakyReLU berperan penting dalam meningkatkan stabilitas aliran gradien (gradient flow) selama proses pelatihan. Batch Normalization membantu menjaga distribusi aktivasi tetap stabil, sedangkan LeakyReLU mencegah terjadinya permasalahan dying ReLU. Kombinasi keduanya memungkinkan model untuk mempelajari pola ekspresi wajah yang lebih kompleks dan non-linear secara lebih efektif.

3.7.2 Pengaruh Class Weight terhadap Kelas Minor

Pada skenario pelatihan tanpa penerapan class weighting, model yang dilatih pada dataset FER-2013 umumnya menunjukkan bias yang kuat terhadap kelas mayor seperti Happy dan Neutral, serta mengalami kesulitan dalam mengenali kelas minor seperti Disgust, Fear, dan Sad. Ketidakseimbangan distribusi kelas ini berdampak langsung pada rendahnya nilai recall dan F1-score pada kelas minor(Cao et al., 2021).

JURNAL AI dan SPK
Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan
Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan
Volume 3, No. 2, Tahun 2025
ISSN 3025-0927 (media online)
Hal 100-109

Hasil penelitian ini menunjukkan bahwa penerapan class weighting mampu meningkatkan performa klasifikasi secara signifikan pada beberapa kelas minor. Secara khusus, terjadi peningkatan nilai F1-score sebesar 12–16 poin pada kelas Sad dan Surprise, disertai dengan peningkatan recall yang lebih merata antar kelas. Temuan ini membuktikan bahwa strategi class rebalancing merupakan pendekatan yang efektif untuk mengatasi ketidakseimbangan kelas pada dataset emosi wajah.

3.7.3 Ketangguhan Model terhadap Noise pada Dataset FER-2013

Dataset FER-2013 dikenal memiliki sejumlah permasalahan kualitas data, seperti anotasi yang tidak konsisten, citra wajah yang kabur atau tidak fokus, serta keberadaan ekspresi campuran (blended expressions). Kondisi ini berpotensi menurunkan performa model apabila tidak ditangani dengan strategi regularisasi yang tepat.

Model B menunjukkan tingkat ketangguhan yang lebih baik terhadap noise pada dataset tersebut. Penerapan Dropout membantu mencegah model menghafal pola noise dengan memaksa jaringan untuk belajar representasi yang lebih robust. Selain itu, penggunaan Global Average Pooling mengurangi sensitivitas model terhadap piksel yang tidak informatif dengan merangkum informasi spasial secara global. Kombinasi kedua teknik ini berkontribusi dalam meningkatkan kemampuan generalisasi model meskipun dilatih pada dataset dengan kualitas anotasi yang tidak ideal.

4. KESIMPULAN

Penelitian ini mengevaluasi efektivitas arsitektur Convolutional Neural Network (CNN) yang telah dioptimasi (Model B – Tuned Level-2) untuk tugas pengenalan emosi wajah pada dataset FER-2013, yang dikenal memiliki karakteristik low-resolution, noise label, dan distribusi kelas yang tidak seimbang. Hasil eksperimen menunjukkan bahwa modifikasi arsitektur melalui integrasi Batch Normalization, LeakyReLU, Dropout, GlobalAveragePooling, serta penggunaan class-weighting menghasilkan peningkatan performa yang konsisten dan signifikan dibandingkan baseline tradisional.

Model B mencapai akurasi pengujian sebesar 56–58%, macro F1-score 0.53–0.56, serta weighted F1-score 0.55–0.58. Performa ini melampaui baseline HOG+SVM (35–40%) dan CNN dasar Level-1 (41–42%). Temuan ini menunjukkan bahwa optimasi arsitektur dan regularisasi memainkan peran kunci dalam meningkatkan kemampuan generalisasi model pada dataset visual ekspresi wajah yang kualitasnya tidak ideal.

Selain itu, penerapan class-weighting terbukti meningkatkan performa pada kelas minor yang sebelumnya memiliki tingkat misclassification tinggi. Perbaikan terutama terlihat pada kelas Sad, Surprise, dan Fear. Strategi ini secara efektif mengurangi bias model terhadap kelas mayor dan meningkatkan fairness prediksi antar kategori.Analisis interpretabilitas menggunakan Grad-CAM menunjukkan bahwa model memberikan fokus aktivasi pada area fisiologis yang relevan, seperti mata, mulut, dan struktur wajah lain yang merupakan indikator utama dalam Facial Action Coding System (FACS). Konsistensi pola aktivasi ini mengindikasikan bahwa model tidak hanya menghasilkan prediksi yang kompetitif, tetapi juga mengadopsi mekanisme pengenalan yang dapat dijelaskan (explainable), sebuah atribut penting dalam sistem pengenalan emosi.

Namun demikian, kelas Disgust dan Fear tetap menjadi tantangan, dengan nilai F1-score yang lebih rendah dibandingkan kelas lain. Hambatan ini terutama disebabkan oleh keterbatasan jumlah sampel, kemiripan pola visual antar kelas, serta kemungkinan noise label dalam dataset FER-2013. Temuan ini sejalan dengan laporan-laporan empiris terdahulu dan mengonfirmasi keterbatasan inheren dataset.

Secara keseluruhan, model yang diusulkan menyediakan baseline kokoh untuk pengembangan arsitektur lanjutan berbasis transfer learning dan attention mechanism(Khan et al., 2022). Penelitian ini tidak hanya menghasilkan model yang kompetitif, tetapi juga menyediakan landasan metodologis yang kuat untuk penelitian tingkat lanjut.

JURNAL AI dan SPK
Jurnal Artificial Intelligent dan Sistem Penunjang Keputusan

Jurnal AI dan SPK : Jurnal Artificial Inteligent dan Sistem Penunjang Keputusan
Volume 3, No. 2, Tahun 2025
ISSN 3025-0927 (media online)
Hal 100-109

UCAPAN TERIMA KASIH

Penulis menyampaikan ucapan terima kasih kepada dosen pembimbing dan sivitas akademika Program Studi Teknik Informatika Universitas Pamulang atas bimbingan, arahan, dan masukan yang konstruktif selama pelaksanaan dan penyusunan penelitian ini. Apresiasi juga disampaikan kepada keluarga dan rekan-rekan yang telah memberikan dukungan dan motivasi. Diharapkan penelitian ini dapat memberikan kontribusi ilmiah serta menjadi referensi bagi penelitian selanjutnya di bidang computer vision dan pengenalan emosi wajah.

REFERENCES

Cao, J., Meng, Z., & He, X. (2021). Class imbalance learning in deep neural networks: A systematic review. *Neurocomputing*, *452*, 708–726. https://doi.org/10.1016/j.neucom.2021.05.021

Khaireddin, Y., & Chen, Z. (2021). Facial emotion recognition: State-of-the-art performance on FER-2013. *IEEE Access*, *9*, 76944–76959. https://doi.org/10.1109/ACCESS.2021.3081433

Khan, S., Naseer, M., & Hayat, M. (2022). Transformers in vision: A survey. *ACM Computing Surveys*, *54*(10), 1–41. https://doi.org/10.1145/3505244

Kumar, A., & Sharma, D. (2020). Facial expression recognition using convolutional neural networks: A survey. *Applied Sciences*, *10*(23), 8355. https://doi.org/10.3390/app10238355

Samek, W., Montavon, G., & Müller, K.-R. (2021). Explainable artificial intelligence: Interpreting deep learning models. *IEEE Signal Processing Magazine*, *38*(3), 40–48. https://doi.org/10.1109/MSP.2021.3051999

Tjoa, E., & Guan, C. (2020). A survey on explainable artificial intelligence (XAI). *IEEE Transactions on Neural Networks and Learning Systems*, *32*(11), 4793–4813. https://doi.org/10.1109/TNNLS.2020.3027314