KLASIFIKASI EKSPRESI WAJAH MENGGUNAKAN *CONVOLUTIONAL NEURAL NETWORK (CNN) EFFICIENTNETB0*

SKRIPSI

Oleh:

SAKILA AULIA MAHARANI
NIM. 220605110055

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025

KLASIFIKASI EKSPRESI WAJAH MENGGUNAKAN *CONVOLUTIONAL NEURAL NETWORK (CNN) EFFICIENTNETB0*

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh:
SAKILA AULIA MAHARANI
NIM. 220605110055

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2025

HALAMAN PERSETUJUAN

KLASIFIKASI EKSPRESI WAJAH MENGGUNAKAN *CONVOLUTIONAL NEURAL NETWORK (CNN) EFFICIENTNETB0*

SKRIPSI

Oleh:
SAKILA AULIA MAHARANI
NIM. 220605110055

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 09 Desember 2025

Pembimbing I,	Pembimbing II,
Tri Mukti Lestari, M.Kom	Okta Qomaruddin Aziz, M.Kom
NIP. 199111082020122005	NIP. 19911019 201903 1 013

Mengetahui
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Supriyono, M.Kom
NIP. 19841010 201903 1 012

HALAMAN PENGESAHAN

KLASIFIKASI EKSPRESI WAJAH MENGGUNAKAN *CONVOLUTIONAL NEURAL NETWORK (CNN) EFFICIENTNETB0*

SKRIPSI

Oleh :
SAKILA AULIA MAHARANI
NIM. 220605110055

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 29 Desember 2025

Susunan Dewan Penguji

Ketua Penguji	: Dr. M. Amin Hariyadi, M.T NIP. 19670118 200501 1 001	( )
Anggota Penguji I	: Khadijah Fahmi Hayati Holle, M.Kom NIP. 19900626 202203 2 002	( )
Anggota Penguji II	: Tri Mukti Lestari, M.Kom NIP. 199111082020122005	( )
Anggota Penguji III	: Okta Qomaruddin Aziz, M.Kom NIP. 19911019 201903 1 013	( )

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Supriyono, M.Kom
NIP. 19841010 201903 1 012

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama	: Sakila Aulia Maharani
NIM	: 220605110055
Fakultas / Program Studi	: Sains dan Teknologi / Teknik Informatika
Judul Skripsi	: Klasifikasi Ekspresi Wajah Menggunakan *Convolutional Neural Network (CNN) EffectientNetB0*

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 22 Januari 2025
Yang membuat pernyataan,

Sakila Aulia Maharani
NIM.220605110055

MOTTO

"Inner calm emerges from a mind disciplined to accept reality rationally"

"Ketenangan lahir dari pikiran yang terlatih untuk menerima realitas secara rasional"

HALAMAN PERSEMBAHAN

Puji syukur kehadirat ALLAH SWT yang telah memberikan rahmat dan
Hidayah-Nya sehingga penulis diberi kemudahan dalam
Menyelesaikan penulisan skripsi ini dengan baik.
Saya persembahkan karya ini kepada:

Bunda tercinta, Balkis
Yang selalu mendampingi dengan sabar dan tak kenal lelah
Serta dukungan, dan doa tulus yang tiada henti

Ayah tercinta, alm. Irwan
Semoga Allah SWT memberikan tempat terbaik untuk Ayah di sisi-Nya, .

Adikku, Muhammad Farel Akmal
Yang sudah selalu mendukung dan memberi semangat

Segenap keluarga besar A.M. Yusuf
Yang juga mendukung dan selalu memberikan doa – doa tiada henti

KATA PENGANTAR

Assalamualaikum wr wb.

Alhamdulillah segala puji dan Syukur senantiasa penulis panjatkan pada Allah subhanahu wa ta'ala atas berkat Rahmat, serta hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul "Klasifikasi Ekspresi Wajah Menggunakan *Convolutional Neural Network (CNN) EffectientNetB0*". Sholawat serta salam tetap tercurahkan kepada Nabi Muhammad SAW. Dan semoga kita semua mendapat syafaatnya di hari akhir kelak, Aamiin.

Dalam penulisan skripsi ini, penulis menyadari banyak pihak yang terlibat baik dalam proses membimbing penulisan dan juga memberikan semangat dan dukungan moril atau materiil. Untuk itu, penulis ingin menyampaikan terima kasih kepada:

1. Prof. Dr. HJ. Ilfi Nur Diana, M.SI., CAHRM., CRMP., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Agus Mulyono, M.Kes., selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Supriyono, M.Kom., selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang
4. Ibu Tri Mukti Lestari, M.Kom selaku Dosen Pembimbing satu, yang dapat memposisikan diri sebagai sahabat bahkan keluarga yang telah memberikan banyak arahan saran serta masukan di jam dan hari kapan pun sehingga penulis dapat menyelesaikan skripsi ini dengan baik dan tanpa adanya tekanan

5. Bapak Okta Qomaruddin Aziz, M.Kom selaku dosen pembimbing dua yang telah memberikan bimbingan, masukan, dan arahan yang sangat berharga selama proses penyusunan skripsi ini.
6. Bapak Dr. M. Amin Hariyadi, M.T dan Ibu Khadijah Fahmi Hayati Holle, M.Kom selaku dosen Penguji I dan dosen Penguji II yang telah menguji serta memberikan banyak saran untuk menyelesaikan skripsi ini dengan baik.
7. Nia Faricha S, Si., selaku admin Program Studi Teknik Informatika yang selalu sabar memberikan informasi, membantu, dan memberikan arahan selama perkuliahan dan proses penulisan skripsi ini.
8. Seluruh Dosen, Admin, Laboran dan jajaran Staf Program Studi Teknik Informatika yang telah memberikan banyak bantuan selama studi ini.
9. Bunda tercinta, Bunda Balkis, yang senantiasa mendampingi penulis dengan penuh kesabaran, kasih sayang, dan ketulusan. Doa serta dukungan Bunda yang tiada henti menjadi sumber kekuatan utama bagi penulis dalam menghadapi setiap proses dan tantangan selama perkuliahan.
10. Ayah tercinta, alm. Irwan, yang meskipun telah berpulang, setiap nasihat, doa, dan nilai kehidupan yang Ayah tanamkan senantiasa menjadi pegangan dan penyemangat penulis. Semoga Allah SWT memberikan tempat terbaik di sisi-Nya dan menerima seluruh amal ibadah Ayah.
11. Adik tercinta, Muhammad Farel Akmal, yang selalu memberikan dukungan, semangat, dan doa kepada penulis sehingga mampu terus melangkah dan menyelesaikan skripsi ini.

12. Segenap keluarga besar A.M. Yusuf, yang senantiasa memberikan doa, perhatian, dan dukungan moril kepada penulis selama menempuh pendidikan hingga penyusunan skripsi ini dapat diselesaikan.
13. Keluarga sepupu “Mongers”, yang telah tumbuh dan berproses bersama penulis, serta setia memberikan dukungan, semangat, dan kebersamaan yang berarti dalam setiap fase perjalanan hidup penulis.
14. Yoza Setya Febriyanti, gadis Tulungagung yang dengan penuh ketulusan telah menemani penulis ke mana pun melangkah, membantu menghadapi rasa cemas dan anxiety, serta menjadi sosok pendukung penting bagi penulis untuk bertahan dan beradaptasi dalam kehidupan perantauan.
15. Gavrila Pandita, gadis Malang yang telah berkenan menjadi teman dan membantu penulis mengenal serta menyesuaikan diri dengan Kota Malang selama menjalani kehidupan sebagai mahasiswa Rantau.

Penulis sadar bahwa skripsi ini masih sangat jauh dari kata sempurna dan mungkin terdapat kesalahan di dalamnya. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun untuk mengembangkan skripsi ini agar lebih bermanfaat bagi dirinya dan pembaca pada umumnya.

Wassalamualaikum Warahmatullahi Wabarakatuh.

Malang, 22 Desember 2024

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

ABSTRAK

Maharani, Sakila Aulia. 2025. **Klasifikasi Ekspresi Wajah Menggunakan *Convolutional Neural Network (CNN) EfficientNetB0***. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Tri Mukti Lestari, M.Kom (II) Okta Qomaruddin Aziz, M.Kom

.

Kata kunci: Ekspresi Wajah*, Convolutional Neural Network, EfficientNetB0*, FER-13*, MixUp, Label Smoothing*

Perkembangan teknologi kecerdasan buatan, khususnya di bidang *computer vision*, memungkinkan pengolahan dan analisis citra wajah secara otomatis untuk mengenali ekspresi emosi manusia. Klasifikasi ekspresi wajah memiliki peranan penting dalam berbagai aplikasi, seperti *human–computer interaction*, sistem pemantauan emosi, dan analisis perilaku. Penelitian ini bertujuan untuk melakukan klasifikasi ekspresi wajah menggunakan *Convolutional Neural Network (CNN*) dengan arsitektur *EfficientNetB0* pada dataset FER-13 yang terdiri dari tujuh kelas emosi, yaitu *angry, disgust, fear, happy, neutral, sad,* dan *surprise*. Penelitian ini menerapkan empat skenario pelatihan yang berbeda untuk menganalisis pengaruh teknik augmentasi data, *MixUp*, dan *label smoothing* terhadap performa model. Seluruh citra diproses dengan ukuran masukan 96×96 piksel dan dilatih menggunakan optimizer Adam dengan batch size 10. Augmentasi citra konvensional diterapkan secara *on-the-fly*, sementara MixUp digunakan sebagai teknik regularisasi berbasis interpolasi linear antar sampel. Evaluasi performa model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil pengujian menunjukkan bahwa penggunaan strategi pelatihan yang tepat berpengaruh signifikan terhadap kinerja model. Skenario terbaik diperoleh pada kombinasi augmentasi *MixUp* dan *label smoothing*, dengan nilai akurasi sebesar 79% dan F1-score sebesar 0.78. Model menunjukkan performa yang baik pada ekspresi dengan ciri visual yang kuat seperti *happy* dan *surprise*, namun masih menghadapi tantangan pada ekspresi yang memiliki kemiripan fitur visual seperti *fear, sad,* dan *neutral*. Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa EfficientNetB0 efektif digunakan untuk klasifikasi ekspresi wajah, terutama ketika dikombinasikan dengan teknik augmentasi dan regularisasi yang sesuai. Penelitian ini diharapkan dapat menjadi dasar untuk pengembangan lebih lanjut, seperti penggunaan dataset yang lebih besar atau penerapan arsitektur CNN yang lebih kompleks.

ABSTRACT

Maharani, Sakila Aulia. 2025. **Facial Expression Classification Using *Convolutional Neural Network (CNN) EfficientNetB0***. Undergraduate Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Supervisor: (I) Tri Mukti Lestari, M.Kom (II) Okta Qomaruddin Aziz, M.Kom

.

Keywords: Facial Expression, *Convolutional Neural Network, EfficientNetB0*, FER-13, *MixUp, Label Smoothing*

The rapid development of artificial intelligence, particularly in the field of computer vision, has enabled automatic processing and analysis of facial images for recognizing human emotional expressions. Facial expression classification plays an important role in various applications, such as human–computer interaction, emotion monitoring systems, and behavioral analysis. This study aims to classify facial expressions using a Convolutional Neural Network (CNN) with the EfficientNetB0 architecture on the FER-13 dataset, which consists of seven emotion classes, namely angry, disgust, fear, happy, neutral, sad, and surprise. This research applies four different training scenarios to analyze the effects of data augmentation techniques, MixUp, and label smoothing on model performance. All images are processed with an input size of 96×96 pixels and trained using the Adam optimizer with a batch size of 10. Conventional image augmentation is applied on-the-fly, while MixUp is employed as a regularization technique based on linear interpolation between samples. Model performance is evaluated using accuracy, precision, recall, and F1-score metrics. The experimental results indicate that appropriate training strategies significantly affect the model's performance. The best performance is achieved by combining MixUp augmentation and label smoothing, resulting in an accuracy of 79% and an F1-score of 0.78. The model performs well on expressions with strong visual characteristics, such as happy and surprise, but still faces challenges in distinguishing expressions with similar visual features, such as fear, sad, and neutral. Based on these results, it can be concluded that EfficientNetB0 is effective for facial expression classification, particularly when combined with suitable augmentation and regularization techniques. This study is expected to serve as a foundation for further research, such as utilizing larger datasets or exploring more complex CNN architectures.

مستخلص البحث

ماهاراني، سكيلا أوليا. ٢٠٢٥ .**تصنيف تعبيرات الوجه باستخدام الشبكة العصبية التلافيفية (CNN) EfficientNetB0**.
أطروحة البكالوريوس. قسم الهندسة المعلوماتية، كلية العلوم والتكنولوجيا,الجامعة الإسلامية نيجيري مولانا مالك إبراهيم,
.مالانج. لمشرف الأول: تري موكتي ليستاري، الماجستير. المشرف الثاني: أوكتا قمر الدين عزيز، الماجستير

الكلمات الرئيسية:تعابير الوجه، الشبكات العصبية الالتفافية FER-13, EfficientNetB0,
Label Smoothing, MixUp,

أتاح التطور السريع للذكاء الاصطناعي، ولا سيما في مجال رؤية الحاسوب، المعالجة والتحليل الآليين لصور الوجه بهدف ،التعرف على تعابير الوجه العاطفية. ويلعب تصنيف تعابير الوجه دورًا هامًا في تطبيقات متنوعة، مثل التفاعل بين الإنسان والحاسوب ذات (CNN) وأنظمة مراقبة المشاعر، وتحليل السلوك. تهدف هذه الدراسة إلى تصنيف تعابير الوجه باستخدام شبكة عصبية التفافية ،التي تتكون من سبع فئات عاطفية، وهي: الغضب، والاشمئزاز، والخوف ،FER-13 على مجموعة بيانات EfficientNetB0 بنية ،والسعادة، والحياد، والحزن، والمفاجأة. يطبق هذا البحث أربعة سيناريوهات تدريب مختلفة لتحليل تأثير تقنيات زيادة البيانات وتنعيم التصنيفات على أداء النموذج. تتم معالجة جميع الصور بحجم إدخال 96×96 بكسل، ويتم تدريبها باستخدام ،MixUp كتقنية تنظيم تعتمد MixUp بحجم دفعة 32. يتم تطبيق زيادة الصور التقليدية بشكل فوري، بينما يُستخدم *Adam* مُحسِّن تشير النتائج .F1 على الاستيفاء الخطي بين العينات. يُقيّم أداء النموذج باستخدام مقاييس الدقة، والضبط، والاستدعاء، ومقياس التجريبية إلى أن استراتيجيات التدريب المناسبة تؤثر بشكل كبير على أداء النموذج. وقد تحقق أفضل أداء من خلال الجمع بين تقنية قدره 0.78. يُظهر النموذج أداءً جيدًا F1 لزيادة البيانات وتنعيم التصنيفات، مما أدى إلى دقة بلغت 79% ومقياس MixUp على التعبيرات ذات الخصائص البصرية القوية، مثل السعادة والمفاجأة، ولكنه لا يزال يواجه تحديات في تمييز التعبيرات ذات السمات فعال في تصنيف EfficientNetB0 البصرية المتشابهة، مثل الخوف والحزن والحياد. بناءً على هذه النتائج، يمكن الاستنتاج أن ،تعابير الوجه، خاصةً عند دمجه مع تقنيات زيادة البيانات والتنظيم المناسبة. من المتوقع أن تُشكل هذه الدراسة أساسًا لمزيد من الأبحاث .مثل استخدام مجموعات بيانات أكبر أو استكشاف بنى شبكات عصبية تلافيفية أكثر تعقيدًا.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi kecerdasan buatan *Artificial Intelligence* (AI) telah mendorong kebutuhan sistem yang mampu memahami kondisi emosional manusia. Salah satu pendekatan penting adalah *Facial Expression Recognition* (FER), karena ekspresi wajah merupakan indikator utama dalam komunikasi non-verbal yang dapat dimanfaatkan di berbagai bidang modern, mulai dari keamanan, interaksi manusia– komputer, pendidikan daring, hingga kesehatan mental (Huang, 2024; Aly, 2025). Urgensi penelitian FER semakin tinggi seiring meningkatnya aplikasi berbasis kecerdasan buatan yang menuntut interaksi lebih alami dan adaptif.

Dalam konteks ini, kemampuan memahami ekspresi dan emosi manusia juga sejalan dengan ajaran Islam yang menekankan pentingnya empati dan kepekaan terhadap perasaan orang lain. Rasulullah ﷺ bersabda:

لَا يُؤْمِنُ أَحَدُكُمْ حَتَّى يُحِبَّ لِأَخِيهِ مَا يُحِبُّ لِنَفْسِهِ

"Tidak beriman salah seorang di antara kamu hingga ia mencintai saudaranya sebagaimana ia mencintai dirinya sendiri." (HR. Bukhari dan Muslim).

Hadis ini menunjukkan bahwa memahami ekspresi dan emosi sesama merupakan bagian dari membangun hubungan sosial yang harmonis nilai yang juga menjadi dasar pengembangan sistem pengenalan ekspresi wajah dalam membantu interaksi manusia dan teknologi agar lebih manusiawi.

Secara teoritis, Paul Ekman mengklasifikasikan tujuh ekspresi dasar universal *angry, disgust, fear, happy, neutral, sad,* dan *surprise* yang dapat dikenali lintas budaya (Zeng et al., 2024). Fakta ini menjadi landasan penting dalam pembangunan sistem otomatis untuk pengenalan emosi. FER sendiri bertujuan membuat mesin dapat mengenali ekspresi wajah secara otomatis melalui pengolahan citra, dan telah banyak diteliti dengan penerapan nyata di berbagai sektor kehidupan, mulai dari interaksi manusia–komputer hingga bidang kesehatan untuk memantau kondisi psikologis pasien. Dengan cakupan yang luas, FER menjadi bidang penelitian yang terus berkembang dan memiliki nilai praktis yang tinggi.

Meskipun memiliki potensi besar, *Facial Expression Recognition (FER)* menghadapi tantangan kompleks di dunia nyata. Dataset FER-13 bersifat *in the wild*, sehingga ekspresi wajah sangat dipengaruhi oleh pencahayaan, sudut kamera, posisi kepala, hingga adanya occlusion seperti masker atau kacamata (Duan, C. 2023). Selain itu, karakteristik individu seperti usia, gender, dan etnis menyebabkan ekspresi wajah tidak selalu ditampilkan secara konsisten. Distribusi kelas yang tidak seimbang, misalnya ekspresi “happy” yang lebih banyak dibanding “disgust” atau “fear”, juga dapat menimbulkan bias pada model (Mejia-Escobar, L., et al. 2023).

Berbagai pendekatan telah dilakukan untuk mengatasi masalah FER. Pada masa awal, metode berbasis *handcrafted features* seperti *Local Binary Pattern (LBP)* dan *Histogram of Oriented Gradients (HOG)* dikombinasikan dengan algoritma *machine learning* klasik seperti *Support Vector Machine (SVM)* atau *K-*

Nearest Neighbor (KNN) (Liao, J., Lin, Y., Ma, T., et al. 2023). Walaupun metode ini relatif sederhana dan efektif dalam kondisi terkendali, hasilnya menurun drastis ketika diterapkan pada dataset *in the wild*.

Saat ini, pendekatan berbasis *Convolutional Neural Network (CNN)* dengan *transfer learning*, seperti EfficientNetB0, banyak digunakan. Metode ini mampu mengekstraksi fitur kompleks dari citra wajah secara otomatis dan lebih tahan terhadap variasi pencahayaan, pose, dan occlusion, sehingga performanya lebih baik pada dataset FER-13 yang *in the wild*.

Perkembangan *deep learning* membawa perubahan besar dalam bidang *Facial Expression Recognition (FER). Convolutional Neural Networks (CNN)* menjadi metode utama karena mampu mengekstraksi fitur visual secara otomatis dari citra wajah (Zhao et al., 2024). CNN memanfaatkan lapisan konvolusi untuk mendeteksi pola spasial seperti tepi, tekstur, dan bentuk, yang kemudian digabungkan menjadi representasi fitur yang kaya. Arsitektur CNN modern seperti VGGNet, ResNet, Inception, dan EfficientNet terbukti memberikan akurasi tinggi dalam berbagai kompetisi pengenalan citra (Insani & Santoso, 2024). EfficientNet memperkenalkan konsep *compound scaling*, yang menyeimbangkan kedalaman, lebar, dan resolusi jaringan untuk mencapai performa optimal dengan jumlah parameter lebih efisien (Anthony Tan Zhen Ren et al., 2021).

Namun, CNN dengan arsitektur besar menghadapi kendala praktis terkait efisiensi komputasi dan ketersediaan sumber daya. Model seperti VGG16, VGG19, atau ResNet50 memiliki jumlah parameter sangat besar sehingga memerlukan kapasitas memori tinggi, waktu pelatihan lama, dan GPU berperforma tinggi agar

dapat berjalan optimal. Pada dataset dengan jumlah sampel terbatas seperti FER-13, penggunaan arsitektur kompleks juga berisiko menyebabkan *overfitting*, di mana model terlalu menyesuaikan diri pada data latih sehingga performanya menurun pada data uji (Parmonangan et al., 2023).

Untuk mengatasi keterbatasan ini, penelitian terbaru memanfaatkan CNN *pretrained* dan efisien*,* seperti MobileNetV2*,* MobileNetV3*,* dan EfficientNetB0. EfficientNetB0 memiliki jumlah parameter lebih sedikit dibanding ResNet50, namun tetap mampu mengekstraksi fitur kompleks dari citra wajah, serta lebih efisien dalam komputasi dan penggunaan memori (Zeng, J., Li, Y., Xu, M., & Deng, W., 2024). Dengan demikian, CNN pretrained seperti EfficientNetB0 sangat cocok untuk penelitian akademis yang memiliki keterbatasan perangkat keras, karena mampu memberikan hasil kompetitif dengan biaya komputasi lebih rendah.

Selain efisiensi arsitektur, aspek penting lain dalam penerapan model *pretrained* adalah penyesuaian format citra *input*. Model yang dilatih dengan ImageNet umumnya membutuhkan citra berformat RGB (3-*channel*), sehingga dataset *grayscale* seperti FER-13 harus direplikasi menjadi tiga kanal. Strategi ini telah digunakan secara luas dalam berbagai penelitian implementasi *transfer learning* pada citra medis yang juga berformat grayscale.

(Gu et al., 2024) menunjukkan bahwa pada klasifikasi pneumonia berbasis X-ray, citra grayscale direplikasi menjadi tiga channel untuk menyesuaikan input model *pretrained* ImageNet. Mereka menjelaskan bahwa meskipun informasi pada tiap channel sama, replikasi 3-channel membuat pola intensitas grayscale dapat

diproses oleh *feature extractor* pretrained tanpa perlu mengubah arsitektur model. Pendekatan ini terbukti stabil dan tetap mempertahankan performa model.

Penelitian lain oleh (George et al. 2022) pada klasifikasi COVID-19 juga menegaskan bahwa reproduksi citra satu channel menjadi tiga channel merupakan metode standar untuk memungkinkan penggunaan CNN *pretrained* pada dataset X-ray yang *grayscale*. Mereka menyebutkan bahwa metode ini sederhana, kompatibel dengan arsitektur pretrained, dan tidak menurunkan kualitas representasi fitur pada tahap ekstraksi.

Dengan dasar temuan ilmiah tersebut, pendekatan replikasi 3-channel pada FER-13 menjadi relevan dan sesuai standar praktik penelitian modern, sehingga model *EfficientNetB0* dapat memanfaatkan bobot awal *pretrained* secara optimal meskipun dataset asli tidak memiliki format RGB.

Evaluasi performa dilakukan menggunakan metrik standar klasifikasi *multiclass*, termasuk *confusion matrix, accuracy, precision, recall, dan F1-score. Confusion matrix* memberikan informasi detail tentang kesalahan klasifikasi, misalnya ekspresi “fear” yang sering salah dikenali sebagai “surprise”. Dengan kombinasi metrik ini, evaluasi model dapat dilakukan secara komprehensif.

Penelitian ini diharapkan untuk mengimplementasikan metode klasifikasi ekspresi wajah menggunakan CNN *pretrained* EfficientNetB0pada dataset FER-13. Dengan pendekatan ini, diharapkan sistem mampu mencapai akurasi tinggi, efisien dalam penggunaan sumber daya komputasi, dan relevan dengan kondisi nyata. Penelitian ini diharapkan dapat menghasilkan sistem FER yang adaptif, efisien, dan bermanfaat luas, serta memungkinkan lebih banyak pihak

mengembangkan sistem pengenalan ekspresi wajah meskipun memiliki keterbatasan perangkat keras (Zhang et al., 2022; Khan et al., 2024).

1.2 Pernyataan Masalah

Bagaimana performa klasifikasi ekspresi wajah pada dataset FER-13 menggunakan *Convolutional Neural Network* (CNN) berbasis *EfficientNetB0* dalam mengenali tujuh kelas ekspresi dasar: *Angry, Disgust, Fear, Happy, Neutral, Sad,* dan *Surprise* berdasarkan hasil evaluasi *multiclass confusion matrix*?

1.3 Batasan Masalah

1. Dataset yang digunakan adalah FER-13, yang berisi citra wajah dengan tujuh kelas ekspresi dasar (*Angry, Disgust, Fear, Happy, Neutral, Sad, dan Surprise).*
2. Evaluasi performa model dilakukan menggunakan *multiclass confusion matrix.*

1.4 Tujuan Penelitian

Mengukur performa *Convolutional Neural Network (CNN)* berbasis *EfficientNetB0* pada dataset FER-13 dalam mengenali tujuh kelas ekspresi dasar: *Angry, Disgust, Fear, Happy, Neutral, Sad,* dan *Surprise* berdasarkan hasil evaluasi *multiclass confusion matrix.*

1.5 Manfaat Penelitian

1. Penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan ilmu pengetahuan di bidang *Facial Expression Recognition (FER)* dengan

menerapkan metode *Convolutional Neural Network (CNN)* berbasis arsitektur *EfficientNetB0*, sehingga memperkaya referensi akademik mengenai penerapan model deep learning yang efisien dan akurat untuk pengenalan ekspresi wajah.

2. Secara praktis, penelitian ini dapat menjadi alternatif solusi yang efisien dan ringan secara komputasi untuk sistem pengenalan ekspresi wajah tanpa memerlukan perangkat keras berkapasitas tinggi, serta berpotensi diterapkan dalam berbagai bidang seperti keamanan, interaksi manusia–komputer, pendidikan, dan kesehatan**.**

.

BAB II
STUDI PUSTAKA

2.1 Penelitian Terdahulu

Penelitian pada pengenalan ekspresi wajah (*Facial Expression Recognition*/FER) dalam beberapa tahun terakhir diarahkan pada pengembangan model yang tahan terhadap kondisi *in-the-wild* sekaligus tetap efisien secara komputasi. Dua tantangan utama yang sering muncul adalah keterbatasan jumlah dan kualitas data beranotasi sehingga berisiko overfitting, serta variasi kondisi citra seperti pencahayaan, pose, *occlusion*, resolusi, dan perbedaan identitas yang menurunkan performa pada data nyata. Untuk mengatasi masalah tersebut, metode-modern menggabungkan strategi arsitektur ringan, teknik transfer learning, dan pendekatan *data-centric* seperti augmentasi dan rebalancing kelas (Pham, Duong, Ho, Lee, & Hong, 2023).

Dataset *in-the-wild* menjadi fokus penting karena karakternya lebih menantang dibanding dataset laboratorium. FER-13, yang bersifat *in-the-wild*, menghadirkan variasi pencahayaan, pose non-frontal, dan ekspresi wajah yang berbeda-beda, menjadikannya tolok uji bagi model *Facial Expression Recognition (FER)* (Duan, C., 2023). Selain itu, studi evaluasi menekankan pentingnya menjaga independensi subjek antar bagian pelatihan dan pengujian, serta memperhatikan distribusi kelas agar tidak bias ke kelas mayoritas (Mejia-Escobar, L., et al., 2023).

Seiring meningkatnya kebutuhan efisiensi dan akurasi, sejumlah penelitian mengadaptasi CNN *pretrained* sebagai *backbone* ekstraksi fitur. Arsitektur

modern seperti EfficientNetB0 menawarkan keseimbangan antara kedalaman, lebar, dan resolusi jaringan melalui *compound scaling*, sehingga mampu mengekstraksi fitur kompleks dengan jumlah parameter yang lebih efisien dibandingkan arsitektur besar seperti ResNet50 atau VGG16 (Anthony Tan Zhen Ren et al., 2021).

Dalam konteks efisiensi komputasi, penyesuaian resolusi citra masukan merupakan langkah *preprocessing* yang umum dilakukan pada penelitian klasifikasi ekspresi wajah. Proses *resize* bertujuan untuk menyesuaikan dimensi citra dengan kebutuhan input model CNN serta menurunkan beban komputasi tanpa menghilangkan informasi visual utama yang merepresentasikan ekspresi wajah. Pendekatan ini banyak digunakan dalam penelitian berbasis *transfer learning*, khususnya pada dataset FER yang memiliki keterbatasan resolusi dan jumlah data.

Masalah *class imbalance* dan keterbatasan data tetap menjadi perhatian utama. Pendekatan yang umum digunakan meliputi *data augmentation, oversampling* kelas minoritas, dan teknik *rebalancing* lainnya. Strategi ini terbukti membantu model pretrained bersaing dalam performa tanpa harus menambah kompleksitas arsitektur (Pham et al., 2023).

Selain itu, aspek interpretabilitas semakin diperhatikan. Teknik seperti *Grad-CAM* digunakan untuk memvisualisasikan area wajah yang paling berpengaruh terhadap prediksi model, membantu memastikan model menggunakan fitur wajah yang relevan dan mengidentifikasi kelas yang sering tertukar (Obaid & Alrammahi, 2023).

Dari penelitian terdahulu, terlihat bahwa kombinasi strategi CNN *pretrained* sebagai ekstraktor, penggunaan *transfer learning*, augmentasi dan *rebalancing*, serta evaluasi interpretabilitas memberikan hasil paling menjanjikan pada dataset *in-the-wild*. Namun, terdapat gap penelitian relatif sedikit studi yang mengevaluasi *EfficientNetB0 pretrained* secara eksplisit pada dataset FER-13, dengan analisis mendalam terhadap masalah ketidakseimbangan kelas dan interpretabilitas model.

Tabel 2. 1 Penelitian Terdahulu

No	Nama Peneliti	Judul	Metode	Dataset	Hasil
1	Pham, Duong, Ho, Lee, & Hong (2023)	CNN-Based Facial Expression Recognition with Simultaneous Consideration of Inter-Class and Intra-Class Variations	CNN pretrained dengan perhatian terhadap variasi antar-kelas (inter-class) dan dalam kelas (intra-class)	FER-13, SFEW 2.0	Model mencapai akurasi 86,7% pada FER-13 dan 83,4% pada SFEW 2.0, meningkatkan ketahanan terhadap variasi pose dan pencahayaan serta menurunkan kesalahan antar kelas mirip
2	Shahzad, Bhatti, Jaffar, Akram, Alhajlah, & Mahmood (2023)	Balanced Evaluation Strategy for Facial Expression Recognition in the Wild	CNN-based FER dengan pembagian data berdasarkan independe	RAF-DB, FER-13	Model mencapai akurasi 85,9% pada RAF-DB dan 82,1% pada

			nsi subjek dan distribusi kelas yang seimbang		FER-13, membuktikan bahwa pembagian data seimbang meningkatkan reliabilitas dan menghindari bias kelas mayoritas
3	Bhosale & Chougule (2024)	Hybrid CNN-Random Forest Model for Robust Facial Expression Recognition	CNN untuk ekstraksi fitur dan Random Forest sebagai classifier	FER-13, JAFFE	Model hybrid CNN-RF menghasilkan akurasi 90,2% pada JAFFE dan 84,6% pada FER-13, lebih tinggi dibanding CNN tunggal, terutama pada citra in-the-wild
4	Obaid & Alrammahi (2023)	An Intelligent Facial Expression Recognition System Using a Hybrid Deep Convolutional Neural Network for Multimedia Applications	Hybrid Deep CNN dengan analisis interpretabilitas menggunakan Grad-CAM	FER-13, CK+	Model mencapai akurasi 92,8% pada CK+ dan 88,5% pada FER-13; Grad-CAM menunjukkan area wajah dominan yang

					memengaruhi prediksi, meningkatkan interpretabilitas model
5	Pham et al. (2023)	Data-Centric Enhancement in Lightweight CNN for FER	CNN ringan (MobileNetV2) + augmentasi data & class rebalancing	FER-13, SFEW 2.0	Model MobileNetV2 dengan augmentasi dan rebalancing mencapai akurasi 84,9%, membuktikan pendekatan data-centric lebih efektif dibanding menambah kedalaman jaringan CNN

Berdasarkan kajian literatur yang disajikan pada Tabel 2.1, dapat disimpulkan bahwa penelitian mengenai *Facial Expression Recognition (FER)* telah mengalami perkembangan signifikan, terutama melalui penerapan arsitektur *Convolutional Neural Network (CNN)* dan pendekatan *hybrid model*. Penelitian-penelitian terkini menekankan upaya peningkatan performa model agar mampu beradaptasi dengan baik pada kondisi *in-the-wild*, yaitu citra wajah yang diambil dari situasi nyata dengan variasi pose, pencahayaan, latar belakang, serta adanya occlusion seperti masker atau kacamata.

2.2 Ekspresi Wajah

Ekspresi wajah merupakan salah satu bentuk komunikasi non-verbal yang paling utama dalam interaksi manusia. Melalui ekspresi wajah, emosi dapat disampaikan secara spontan tanpa memerlukan bahasa lisan, sehingga sering kali dianggap lebih jujur dalam merepresentasikan kondisi emosional seseorang. Paul Ekman memperkenalkan konsep tujuh ekspresi dasar yang bersifat universal, yaitu marah (*angry*), jijik (*disgust*), takut (*fear*), senang (*happy*), netral (*neutral*), sedih (*sad*), dan terkejut (*surprise*). Ekspresi dasar ini terbukti dapat dikenali lintas budaya dan usia, sehingga menjadi landasan dalam penelitian *Facial Expression Recognition (FER)* modern (Zeng et al., 2024).

Ekspresi wajah memiliki peran penting dalam berbagai aspek kehidupan. Dalam psikologi, ekspresi wajah digunakan untuk memahami kondisi emosional individu, misalnya untuk mendeteksi depresi atau gangguan kecemasan. Pada bidang sosial, ekspresi wajah membantu membangun kepercayaan dan interaksi antarindividu. Sementara dalam bidang teknologi, pengenalan ekspresi wajah otomatis diaplikasikan pada keamanan, sistem interaksi manusia–komputer (*Human-Computer Interaction/HCI*), pendidikan daring, hingga pelayanan kesehatan digital (Huang, 2024; Aly, 2025).

Namun, pengenalan ekspresi wajah tidak lepas dari tantangan. Variasi pencahayaan, pose non-frontal, resolusi citra rendah, serta occlusion seperti penggunaan masker atau kacamata dapat mengaburkan ciri wajah. Selain itu, perbedaan individu seperti usia, gender, etnis, maupun gaya ekspresi pribadi

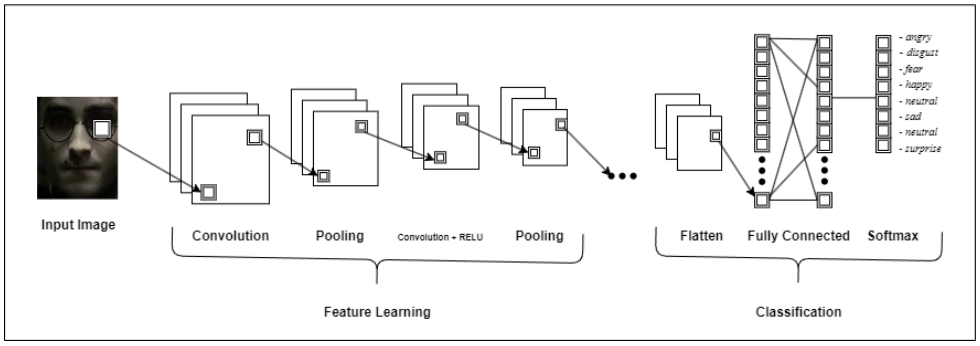
menyebabkan satu emosi yang sama bisa ditampilkan dengan cara berbeda. Tantangan lain adalah ketidakseimbangan jumlah data antar kelas emosi; misalnya, ekspresi senang lebih sering muncul dibandingkan ekspresi jijik, sehingga model cenderung bias terhadap kelas mayoritas (Mejia-Escobar et al., 2023). Faktor-faktor inilah yang membuat ekspresi wajah menjadi objek penelitian yang kompleks dan menantang dalam bidang *machine learning* dan *computer vision*.

2.3 *Convolutional Neural Network (CNN)*

Convolutional Neural Network (CNN) merupakan arsitektur jaringan saraf tiruan yang dirancang khusus untuk memproses data berbentuk citra. CNN mampu melakukan proses feature extraction dan classification secara end-to-end, tanpa memerlukan tahapan terpisah antara ekstraksi fitur dan klasifikasi seperti pada pendekatan hybrid.

Lapisan konvolusi (convolution layer) menggunakan filter atau kernel untuk mendeteksi pola lokal seperti tepi, tekstur, dan bentuk, sedangkan lapisan pooling berfungsi mengurangi dimensi fitur sambil mempertahankan informasi penting. Dengan penyusunan beberapa lapisan konvolusi dan pooling secara bertingkat, CNN membangun representasi fitur dari yang sederhana hingga kompleks, misalnya bagian wajah, mata, hidung, mulut, hingga keseluruhan ekspresi wajah. Keunggulan CNN adalah kemampuannya belajar langsung dari data citra mentah dan mengoptimalkan bobot secara otomatis melalui proses backpropagation. Dalam konteks Facial Expression Recognition (FER), CNN terbukti efektif mengenali berbagai emosi dengan akurasi tinggi, bahkan pada kondisi *in the wild*.

Untuk meningkatkan efisiensi dan performa, penelitian ini menggunakan *EfficientNetB0* pretrained, sebuah arsitektur CNN modern yang menerapkan konsep compound scaling, menyeimbangkan kedalaman, lebar, dan resolusi jaringan. *EfficientNetB0* memiliki jumlah parameter yang besar dibandingkan versi EfficientNet yang lebih kecil, namun lebih efisien dibandingkan arsitektur CNN konvensional seperti VGG atau ResNet dengan performa kompetitif. Model pretrained ini memanfaatkan bobot dari dataset ImageNet, sehingga dapat digunakan sebagai ekstraktor fitur yang kuat dan stabil pada dataset FER-13.



Gambar 2.1 Bentuk umum CNN

Gambar 2.1 menunjukkan alur proses *EfficientNetB0* yang digunakan untuk klasifikasi ekspresi wajah pada dataset FER-13. Tahapan dimulai dari citra wajah (Input Image) yang berasal dari dataset, masing-masing berformat RGB, yang menjadi masukan bagi jaringan.

Pada bagian *Feature Learning*, lapisan convolutional dari *EfficientNetB0* mengekstraksi fitur penting dari citra, seperti bentuk mata, mulut, dan ekspresi garis wajah. Lapisan-lapisan ini menghasilkan representasi fitur yang kaya (feature maps). Tahap Pooling mereduksi dimensi fitur agar proses komputasi lebih efisien

tanpa kehilangan informasi penting, sementara aktivasi ReLU menambahkan sifat non-linear pada jaringan.

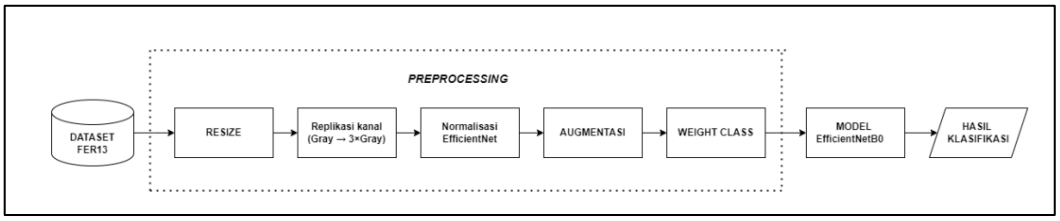
Tahapan berikutnya adalah *Classification*, di mana vektor fitur dari lapisan akhir *EfficientNetB0* diproses melalui Global Average Pooling, kemudian dilanjutkan ke lapisan Fully Connected dengan Dropout untuk mencegah overfitting. Lapisan terakhir menggunakan Softmax untuk menghasilkan probabilitas dari tujuh kelas ekspresi wajah: Angry, Disgust, Fear, Happy, Neutral, Sad, dan Surprise.

Model *EfficientNetB0* ini bekerja secara end-to-end, mengintegrasikan proses ekstraksi fitur dan klasifikasi. Pendekatan ini memungkinkan sistem mengenali ekspresi wajah secara otomatis dengan akurasi tinggi pada kondisi nyata, sekaligus memanfaatkan efisiensi komputasi berkat arsitektur pretrained yang optimal.

BAB III
DESAIN DAN IMPLEMENTASI

3.1 Desain Sistem

Desain sistem dalam penelitian ini menerangkan alur kerja yang digunakan untuk melakukan klasifikasi ekspresi wajah, mulai dari input citra wajah hingga menghasilkan output berupa kategori emosi. Sistem dirancang agar mampu memproses data secara menyeluruh, mulai dari tahap awal pengumpulan, pra-pemrosesan, pelatihan model CNN berbasis EfficientNetB0, hingga menghasilkan prediksi akhir berupa kelas emosi. Dengan alur ini, diharapkan sistem dapat mengenali ekspresi wajah secara otomatis, seperti yang ditunjukkan pada Gambar 3.1



Gambar 3.1 Desain Sistem

Flowchart di atas menggambarkan alur kerja sistem klasifikasi ekspresi wajah menggunakan arsitektur EfficientNetB0 mulai dari tahap input hingga keluaran prediksi kelas emosi. Proses dimulai dari dataset FER-13 yang berisi tujuh kelas emosi, yaitu *angry, disgust, fear, happy, neutral, sad,* dan *surprise*. Dataset ini di *resize* menjadi ukuran 96×96 piksel terlebih dahulu agar sesuai dengan kebutuhan arsitektur EfficientNetB0.

Karena citra FER-13 asli hanya memiliki satu kanal (grayscale), dilakukan tahap *replicate 3-channel*, yaitu menggandakan kanal grayscale menjadi tiga kanal sehingga format citra berubah menjadi pseudo-RGB (gray–gray–gray). Proses ini penting karena EfficientNetB0 pretrained ImageNet hanya menerima input citra tiga kanal.

Tahap berikutnya adalah normalisasi EfficientNet, yaitu menerapkan fungsi *preprocess_input* yang menyesuaikan nilai piksel ke rentang yang digunakan selama pretraining pada ImageNet. Normalisasi ini membuat distribusi data masukan konsisten dengan ekspektasi model sehingga membantu meningkatkan stabilitas pembelajaran dan konvergensi.

Selanjutnya, teknik *data augmentation* diterapkan hanya pada dataset latih pada seluruh skenario eksperimen. Augmentasi yang digunakan bersifat ringan, meliputi rotasi citra dengan sudut kecil, *zoom in* dan *zoom out*, serta *horizontal flip*. Rotasi citra digunakan untuk meniru kondisi nyata di mana wajah tidak selalu berada pada posisi frontal, sedangkan *zoom* merepresentasikan variasi jarak antara kamera dan subjek. *Horizontal flip* diterapkan untuk membalik citra wajah secara mendatar sehingga model dapat mempelajari pola ekspresi dari arah yang berbeda. Penerapan augmentasi ini dikombinasikan dengan konfigurasi pelatihan yang berbeda pada setiap skenario, yaitu skenario A dan C yang menggunakan augmentasi tanpa MixUp, serta skenario B dan D yang mengombinasikan augmentasi dengan teknik MixUp ($\alpha = 0.1$). Selain itu, label smoothing dengan nilai 0.1 diterapkan pada skenario A dan D untuk meningkatkan kemampuan generalisasi model.

Citra yang telah melalui seluruh tahapan preprocessing tersebut kemudian menjadi masukan bagi model EfficientNetB0, yang telah dimodifikasi pada bagian akhir untuk menghasilkan prediksi tujuh kelas emosi. EfficientNetB0 mengekstraksi fitur visual melalui kombinasi convolution, MBConv blocks, dan global average pooling, kemudian menghasilkan probabilitas setiap kelas melalui lapisan fully connected dengan aktivasi softmax.

Tahap terakhir dari alur adalah hasil klasifikasi, yaitu keluaran berupa prediksi salah satu dari tujuh kelas emosi berdasarkan citra wajah yang diberikan. Dengan demikian, flowchart ini menyajikan alur sistem yang ringkas, sistematis, dan sesuai dengan pipeline yang digunakan dalam penelitian, mulai dari input dataset, preprocessing khusus EfficientNet, hingga keluaran berupa prediksi ekspresi wajah.

3.2 Persiapan Data

Dalam penelitian ini, data yang digunakan berasal dari dataset FER-13 *(Facial Expression Recognition 2013)***.** Dataset ini dipilih karena banyak digunakan dalam penelitian akademik, memiliki variasi ekspresi wajah yang cukup, serta mencakup kondisi semi-in-the-wild, sehingga cocok untuk evaluasi performa model *Convolutional Neural Network (CNN).*

Dataset FER-13 awalnya terdiri dari tujuh kelas ekspresi dasar, yaitu *angry*, *disgust*, *fear*, *happy*, *neutral*, *sad*, dan *surprise*, dengan total 28.709 citra wajah berukuran 48×48 piksel dalam format *grayscale* Tabel 3.2. Pada penelitian ini, penanganan ketidakseimbangan kelas dilakukan melalui strategi pada proses pelatihan model. Untuk mengurangi dampak *class imbalance*, seluruh skenario

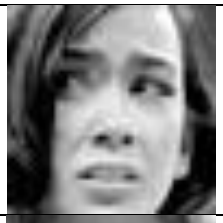
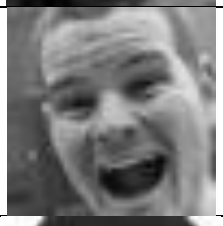
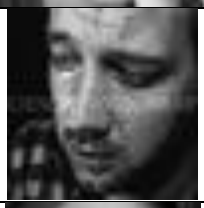
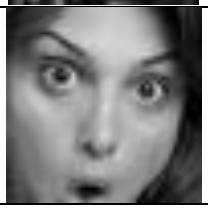
pelatihan menerapkan teknik *data augmentation* pada data latih guna meningkatkan variasi data, khususnya pada kelas minoritas.

Selain itu, *class weight* diterapkan pada skenario pelatihan tanpa penggunaan MixUp untuk memberikan penalti yang lebih besar terhadap kesalahan klasifikasi pada kelas dengan jumlah sampel lebih sedikit. Pada skenario yang menggunakan MixUp, *class weight* tidak diterapkan karena keterbatasan kompatibilitas dengan generator data, sehingga penanganan ketidakseimbangan kelas pada skenario tersebut sepenuhnya mengandalkan *data augmentation*. Pendekatan ini memungkinkan proses pelatihan model menjadi lebih stabil serta mengurangi bias terhadap kelas mayoritas tanpa menghilangkan informasi penting dari dataset asli..

Dataset yang telah dipangkas kemudian dibagi menjadi dua subset utama menggunakan rasio 80:20, yaitu *training set* sebesar 80%, *testing set* sebesar 10% dan *validation set* sebesar 10%. Pembagian ini dilakukan secara stratified sehingga proporsi setiap kelas tetap konsisten pada kedua subset. Train set digunakan untuk proses pelatihan model, validation set dilakukan untuk validasi, sedangkan test set digunakan untuk evaluasi akhir guna mengukur kemampuan generalisasi model terhadap data baru yang tidak pernah terlihat selama training.

Tabel 3. 1 Contoh data ekspresi wajah

No	Nama Ekspresi	Gambar	Jumlah Data
1	*Angry*		3,995

No	Nama Ekspresi	Gambar	Jumlah Data
2	*Disgust*		436
3	*Fear*		4,097
4	*Happy*		7,215
5	*Neutral*		4,965
6	*Sad*		4,830
7	*Surprise*		3,171
Total	28,709		

3.2.1 *Input* Citra

Tahap pertama dalam sistem ini adalah *input* citra. Data yang digunakan berasal dari dataset FER-13, yaitu kumpulan citra wajah manusia berukuran 48×48 piksel yang dikumpulkan dari berbagai kondisi nyata. Citra pada FER-13 memiliki

variasi ekspresi, posisi wajah, serta kualitas pencahayaan yang tidak selalu ideal, sehingga tetap mencerminkan karakteristik data dunia nyata. Seluruh citra grayscale pada FER-13 menjadi masukan utama dalam sistem klasifikasi ekspresi wajah yang dirancang.

3.2.2 *Resize*

Sebelum citra digunakan dalam proses pelatihan, data melalui rangkaian tahap pre-processing agar format dan struktur citra sesuai dengan kebutuhan arsitektur EfficientNetB0. Pre-processing ini mencakup beberapa langkah terintegrasi, yaitu resize citra, replikasi kanal menjadi RGB, dan normalisasi khusus EfficientNet. Pertama, dilakukan proses resize citra menjadi 96×96 piksel agar seragam dengan parameter masukan yang digunakan dalam pelatihan. Setelah itu, karena FER-13 merupakan dataset grayscale, setiap citra direplikasi menjadi tiga kanal agar sesuai dengan format input RGB yang dibutuhkan EfficientNetB0. Terakhir, citra dinormalisasi menggunakan fungsi *preprocess_input* bawaan EfficientNet yang telah disesuaikan dengan karakteristik arsitektur model. Rangkaian pre-processing ini memastikan bahwa data memiliki kualitas dan format yang optimal untuk tahap training dan inferensi menggunakan EfficientNetB0.

3.2.3 *Replicate 3-channel*

Dataset FER-13 hanya menyediakan citra grayscale dengan satu kanal, sedangkan EfficientNetB0 memerlukan masukan berupa citra dengan tiga kanal warna (RGB). Oleh karena itu, konversi ke format RGB menjadi tahap penting sebelum citra diproses lebih lanjut. Proses ini dilakukan bukan dengan menambahkan informasi warna baru, tetapi dengan mereplikasi nilai grayscale ke

tiga kanal yang sama (*gray-gray-gray*). Dengan cara ini, citra tetap mempertahankan informasi visual aslinya, namun strukturnya kini sesuai dengan format input yang diharuskan oleh arsitektur EfficientNetB0. Replikasi tiga kanal ini memastikan bahwa model dapat melakukan ekstraksi fitur secara optimal tanpa terjadi ketidaksesuaian dimensi pada lapisan awal jaringan konvolusional.

3.2.4 Normalisasi *EfficientNetB0*

Setelah citra direplikasi menjadi tiga kanal, langkah berikutnya adalah melakukan normalisasi piksel menggunakan fungsi normalisasi bawaan EfficientNetB0. Normalisasi ini penting karena EfficientNetB0 dilatih menggunakan skema preprocessing tertentu yang mengatur ulang nilai piksel menjadi rentang yang lebih stabil untuk ekstraksi fitur. Berbeda dengan normalisasi sederhana seperti *rescaling* 1/255 yang hanya menurunkan nilai piksel ke rentang 0–1, normalisasi EfficientNetB0 melakukan penyesuaian yang lebih kompleks melalui proses *mean subtraction* dan *scaling* tertentu yang disesuaikan dengan distribusi data saat model dasarnya dilatih. Penerapan normalisasi ini memastikan bahwa citra berada dalam domain yang sama dengan data pelatihan awal EfficientNetB0, sehingga performa model dapat tetap optimal baik pada saat training maupun testing.

3.2.5 Augmentasi Citra

Setelah proses pre-processing selesai, tahap berikutnya adalah augmentasi citra yang diterapkan khusus pada data training. Augmentasi bertujuan menambah variasi data secara sintetis menggunakan transformasi visual ringan tanpa

mengubah label emosi. Transformasi yang digunakan meliputi rotasi kecil, pergeseran posisi, zoom, dan horizontal flip. Rotasi membantu model memahami ekspresi wajah pada berbagai orientasi kepala, pergeseran dan zoom meniru variasi jarak kamera, sedangkan horizontal flip membuat model lebih tahan terhadap perubahan arah pandang. Dengan augmentasi, model belajar dari fitur wajah yang lebih bervariasi sehingga mampu mencapai generalisasi yang lebih baik pada data baru.

Pada penelitian ini, augmentasi citra diterapkan pada seluruh skenario eksperimen dengan konfigurasi yang berbeda. Skenario A dan C menggunakan augmentasi standar tanpa MixUp, sedangkan Skenario B dan D mengombinasikan augmentasi dengan teknik MixUp ($\alpha = 0.1$). Selain itu, label smoothing dengan nilai 0.1 diterapkan pada Skenario A dan D sebagai bentuk regularisasi tambahan untuk meningkatkan stabilitas dan generalisasi model.

3.2.6 *Weight Class*

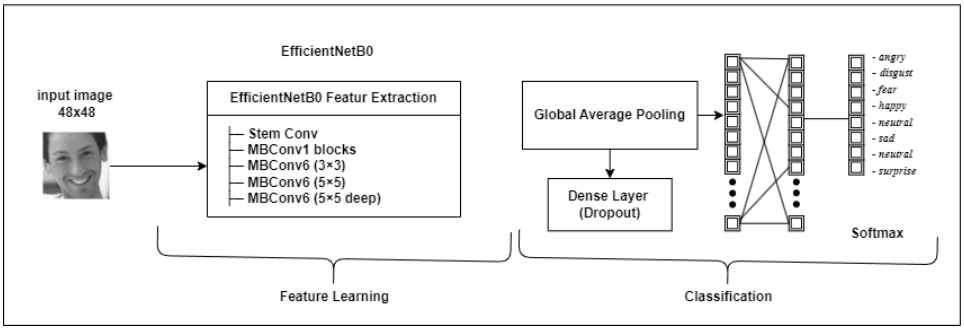
Selain augmentasi citra, penelitian ini juga menerapkan *class weight* sebagai strategi tambahan untuk menangani ketidakseimbangan jumlah data antar kelas pada dataset FER-13. *Class weight* digunakan untuk memberikan bobot kesalahan yang lebih besar pada kelas dengan jumlah sampel lebih sedikit, sehingga model tidak cenderung bias terhadap kelas mayoritas selama proses pelatihan. Dengan pendekatan ini, kesalahan klasifikasi pada kelas minoritas akan lebih berpengaruh terhadap nilai fungsi loss, sehingga model terdorong untuk mempelajari representasi fitur yang lebih baik pada kelas tersebut.

Penerapan *class weight* dilakukan pada skenario pelatihan tanpa MixUp, yaitu Skenario A dan Skenario C. Hal ini disebabkan oleh keterbatasan kompatibilitas antara *class weight* dan generator data berbasis MixUp, yang dapat menimbulkan konflik dalam perhitungan bobot loss. Oleh karena itu, pada Skenario B dan D yang menggunakan MixUp, penanganan ketidakseimbangan kelas sepenuhnya mengandalkan teknik augmentasi dan pencampuran data yang dihasilkan oleh MixUp.

3.3 Implementasi Metode

Implementasi metode dalam penelitian ini dilakukan melalui serangkaian tahap mulai dari preprocessing, augmentasi, ekstraksi fitur, hingga klasifikasi. Model yang digunakan adalah *EfficientNetB0*, yaitu arsitektur *Convolutional Neural Network (CNN) modern* yang telah dilatih sebelumnya (*pretrained*) pada dataset ImageNet. Model ini dipilih karena memiliki arsitektur yang efisien dan mampu mengekstraksi fitur visual dengan performa tinggi sambil tetap hemat parameter. Citra wajah hasil preprocessing, yang meliputi resize, konversi ke RGB, normalisasi, kemudian diproses melalui blok-blok convolutional dan inverted residual pada EfficientNetB0 untuk menghasilkan representasi fitur yang kaya dan efisien. Hasil ekstraksi fitur selanjutnya dirata-ratakan melalui lapisan global average pooling dan diteruskan ke lapisan fully connected dengan fungsi aktivasi softmax untuk menghasilkan prediksi kelas ekspresi wajah sesuai dengan tujuh kategori pada dataset FER-13, yaitu angry, disgust, fear, happy, neutral, sad, dan surprise. Selain itu, penelitian ini menerapkan dua skenario eksperimen, yaitu penggunaan label smoothing 0.1 pada skenario A dan Mixup $\alpha = 0.1$ pada skenario

B, untuk meningkatkan generalisasi model terhadap variasi data in-the-wild. Dengan implementasi metode ini, sistem diharapkan dapat mengenali ekspresi wajah secara otomatis dan akurat pada kondisi nyata.



Gambar 3.2 Arsitektur *EfficientNetB0*

Gambar 3.2 memperlihatkan alur arsitektur model EfficientNetB0 yang digunakan untuk klasifikasi ekspresi wajah pada dataset FER-13. Proses dimulai dari *input image* berukuran 48×48 piksel, kemudian diteruskan ke *preprocessing layer* yang mencakup proses *resize*, replikasi kanal citra grayscale menjadi tiga kanal identik, serta normalisasi sesuai standar EfficientNetB0. Replikasi kanal dilakukan dengan menggandakan kanal grayscale menjadi tiga kanal (gray–gray–gray) sehingga format citra berubah menjadi *pseudo-RGB* dan dapat diterima oleh model EfficientNetB0 yang telah dipra-latih pada dataset. Normalisasi ini memastikan nilai piksel berada pada skala yang optimal sehingga proses ekstraksi fitur oleh model *pretrained* dapat berlangsung secara stabil.

Setelah preprocessing, citra masuk ke EfficientNetB0 Backbone, yang memanfaatkan transfer learning dari model pretrained ImageNet. Arsitektur ini terdiri dari tiga komponen utama: stem, MBConv blocks, dan head. Pada bagian

stem, dilakukan operasi convolution, batch normalization, dan aktivasi Swish untuk mengekstraksi fitur awal dari citra.

Selanjutnya, fitur diteruskan melalui MBConv blocks, yang terdiri dari beberapa stage dengan kombinasi expansion, depthwise convolution, dan projection. Setiap stage memiliki konfigurasi tertentu, misalnya MBConv1 atau MBConv6, dengan ukuran kernel bervariasi antara 3×3 hingga 5×5. Blok MBConv menggunakan prinsip inverted residual dengan linear bottleneck, yang memungkinkan model mengekstraksi fitur lebih efisien sambil menjaga jumlah parameter tetap rendah. Struktur ini membantu model menyeimbangkan antara kedalaman jaringan, kapasitas representasi fitur, dan efisiensi komputasi.

Setelah melewati backbone, fitur dari citra dirata-ratakan melalui global average pooling untuk menghasilkan vektor fitur tunggal per citra. Vektor ini kemudian diteruskan ke fully connected layer (dense layer), yang dapat dilengkapi dengan dropout opsional untuk mengurangi overfitting. Tahap akhir adalah output layer dengan fungsi aktivasi softmax, yang menghasilkan prediksi probabilitas untuk tujuh kelas emosi: angry, disgust, fear, happy, neutral, sad, dan surprise. Dengan arsitektur ini, EfficientNetB0 mampu mengekstraksi fitur visual secara efektif dan menghasilkan prediksi kelas emosi yang akurat, sambil tetap efisien secara komputasi untuk dataset citra wajah berukuran kecil.

3.3.1 Input Layer

Lapisan input berfungsi untuk menerima citra wajah yang akan diproses dalam sistem. Pada penelitian ini, data citra wajah dari dataset FER-13 dengan ukuran 96×96 piksel dan tiga kanal warna (RGB) diubah ke bentuk tensor

berdimensi (96, 96, 3) untuk setiap citranya agar dapat diproses oleh arsitektur EfficientNetB0. Input layer memastikan data citra dapat dipetakan ke dalam bentuk numerik yang dapat diproses oleh lapisan konvolusi pada tahap ekstraksi fitur. Secara matematis, data citra yang semula berbentuk matriks piksel dapat direpresentasikan sebagai:

$$I = [p_{ijk}] \quad \text{dengan } i = 1, ..., 224;\ j = 1, ..., 224;\ k = 1, 2, 3 \tag{3.1}$$

Keterangan :
Pijk : nilai intensitas piksel pada posisi baris ke-*i*, kolom ke-*j*, dan kanal warna ke-*k*.
k : 1,2,3 masing-masing merepresentasikan kanal merah (R), hijau (G), dan biru (B).

Rumus ini kemudian menjadi masukan bagi *EfficientNetB0 Backbone* untuk mengekstraksi fitur visual dari citra wajah.

3.3.2 EfficientNetB0

Lapisan konvolusi pada *EfficientNetB0* digunakan untuk mengekstraksi fitur visual penting dari citra wajah, seperti kontur wajah, tekstur kulit, bentuk mata, mulut, dan alis yang relevan dengan ekspresi emosi. *EfficientNetB0* menggunakan MBConv blocks dengan teknik depthwise separable convolution yang efisien secara komputasi namun tetap mampu menangkap fitur kompleks.

Setiap MBConv block terdiri dari expansion layer (1×1 convolution + Swish), depthwise convolution (3×3 atau 5×5), dan projection layer (1×1 convolution linear) yang membentuk inverted residual dengan linear bottleneck. Mekanisme ini memungkinkan jaringan mengekstraksi fitur secara mendalam tanpa meningkatkan jumlah parameter secara berlebihan, sekaligus menjaga informasi penting dari

feature map. Secara matematis, konvolusi pada sebuah citra masukan I dengan kernel K menghasilkan feature map S:

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n) \cdot K(m,n) \tag{3.2}$$

Keterangan :
$I\ (i,j)$: nilai piksel citra masukan pada posisi ke-i,j
$K\ (m,n)$: nilai kernel/filter
$S\ (i,j)$: hasil konvolusi (feature map) pada posisi ke-i,j.

Contoh :

Misalkan potongan kecil citra wajah 3×3 dan kernel 2×2 sebagai berikut:

$$I = \begin{bmatrix} 1 & 2 & 0 \\ 4 & 3 & 1 \\ 2 & 1 & 0 \end{bmatrix}, \quad K = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

Proses konvolusi menghasilkan feature map (S):

$$S = \begin{bmatrix} (1\times1 + 2\times0 + 4\times0 + 3\times(-1)) & (2\times1 + 0\times0 + 3\times0 + 1\times(-1)) \\ (4\times1 + 3\times0 + 2\times0 + 1\times(-1)) & (3\times1 + 1\times0 + 1\times0 + 0\times(-1)) \end{bmatrix}$$

$$= \begin{bmatrix} -2 & 1 \\ 3 & 3 \end{bmatrix}$$

Feature map ini menunjukkan hasil ekstraksi pola dari bagian kecil citra wajah. Namun, efisiensi utama EfficientNetB0 berasal dari penggunaan *MBConv (Mobile Inverted Bottleneck Convolution)* yang diperkenalkan pada *MobileNetV2* dan disempurnakan untuk *EfficientNet*. Setiap MBConv block terdiri dari tiga tahap

utama: *expansion layer, depthwise convolution, dan projection layer*, yang membentuk struktur *inverted residual* dengan *linear bottleneck.*

Tahap pertama, yaitu *expansion layer,* memperbesar jumlah kanal secara signifikan menggunakan operasi konvolusi 1×1. Jika jumlah kanal input adalah C_{in}, maka jumlah kanal setelah ekspansi menjadi $t \cdot C_{\text{in}}$, dengan $t$adalah *expansion factor*. Operasi ini dirumuskan sebagai:

$$X_{\text{exp}} = \sigma(W_{\text{exp}} * X) \tag{3.3}$$

Keterangan :
$Wexp$: bobot konvolusi 1×1
X : input block
σ : fungsi aktivasi *Swish*

Yang dirumuskan sebagai :

$$\text{Swish}\,(x) = x \cdot \text{sigmoid}(x) \tag{3.4}$$

Swish dipilih karena mampu memberikan *gradient flow* lebih stabil dibandingkan ReLU. Tahap kedua pada MBConv adalah *depthwise convolution*, yaitu proses konvolusi yang diterapkan pada tiap kanal secara terpisah, bukan secara *full convolution*. *Depthwise convolution* dirumuskan sebagai:

$$S_k(i,j) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} X_{\text{exp},k}(i+m, j+n) \cdot K_k(m,n) \tag{3.5}$$

Dimana filter berbeda K_k diterapkan untuk setiap kanal k. Dengan demikian, operasi *depthwise convolution* mengurangi jumlah parameter

secara drastis karena tidak perlu mengoperasikan filter multidimensi pada seluruh kanal secara bersamaan. Pada konteks citra wajah, *depthwise convolution* memungkinkan model mengekstraksi pola halus seperti kerutan, garis senyum, dan bentuk mata tanpa beban komputasi besar.

Tahap ketiga adalah *projection layer*, yaitu konvolusi 1×1linear yang mengecilkan kembali jumlah kanal dari $t \cdot C_{\text{in}}$ menjadi C_{out}. Operasi ini dinyatakan sebagai:

$$X_{\text{proj}} = W_{\text{proj}} * S \tag{3.6}$$

Dengan tidak menggunakan fungsi aktivasi pada tahap ini (linear), arsitektur ini menjaga agar informasi fitur tidak hilang akibat *nonlinear distortion*.

Selain itu, EfficientNetB0 memanfaatkan mekanisme *inverted residual connection*, yaitu suatu koneksi skip yang hanya digunakan jika ukuran input dan output sama. Rumus residunya adalah:

$$Y = X + X_{\text{proj}} \tag{3.7}$$

Koneksi residual ini membantu jaringan mempertahankan informasi penting sehingga gradient dapat mengalir dengan lebih stabil selama pelatihan. Arsitektur EfficientNetB0 juga menyertakan *Squeeze-and-Excitation (SE) block* pada banyak MBConv block. SE block bekerja dengan cara memberikan perhatian (*channel attention*) pada kanal-kanal tertentu yang paling penting dalam representasi fitur. Proses ini dilakukan melalui operasi *global average pooling*, dilanjutkan dengan

fungsi aktivasi sigmoid untuk menghasilkan *attention weights*. Secara matematis, SE block dirumuskan sebagai:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i,j)$$

$$s = \sigma(W_2 \cdot \delta(W_1 \cdot z)) \tag{3.8}$$

Keterangan :

zc	: nilai rata-rata tiap kanal
$W1\ dan\ W2$	: parameter fully-connected layer
δ	: ReLU

Dan S adalah vektor perhatian (*attention vector*) yang kemudian dikalikan ke feature map untuk memperkuat kanal penting:

$$X'_c(i,j) = s_c \cdot X_c(i,j) \tag{3.9}$$

SE block membantu model fokus pada fitur-fitur paling relevan dengan ekspresi wajah.

Semua mekanisme ini menjadikan EfficientNetB0 sangat efisien dalam mengekstraksi fitur kompleks dari citra wajah, seperti perubahan bentuk bibir pada ekspresi sad, pengencangan otot sekitar mata pada ekspresi angry, atau peningkatan intensitas lipatan pipi pada ekspresi happy. Dengan kedalaman arsitektur yang cukup, penggunaan MBConv, SE block, dan aktivasi Swish, EfficientNetB0 mampu menghasilkan representasi fitur yang kaya namun tetap hemat komputasi, sehingga cocok digunakan dalam penelitian pengenalan ekspresi wajah dengan dataset terbatas maupun perangkat dengan keterbatasan sumber daya.

3.3.3 Global Average Pooling Layer

Global Average Pooling (GAP) merangkum seluruh informasi dari feature map hasil ekstraksi *EfficientNetB0* menjadi satu vektor fitur berdimensi rendah. Setiap channel pada feature map dihitung rata-ratanya sehingga menghasilkan nilai representatif tunggal. Misalkan output feature map dari EfficientNetB0 memiliki ukuran:

$$H \times W \times C \tag{3.10}$$

Keterangan :

H= tinggi feature map
W= lebar feature map
C= jumlah channel

Maka GAP menghitung rata-rata seluruh nilai di setiap channel:

$$\mathrm{GAP}(c) = \frac{1}{H \cdot W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_{i,j,c} \tag{3.11}$$

Dengan:

$x_{i,j,c}$= nilai pixel pada posisi (i, j)untuk channel ke-c

Hasil :

$$\text{Output GAP} = [\mathrm{GAP}(1), \mathrm{GAP}(2), \ldots, \mathrm{GAP}(C)]$$

Jadi GAP mengubah feature map ukuran H×W×C menjadi hanya 1×1×C, atau sederhananya vektor dengan panjang C. GAP berfungsi mengurangi jumlah

parameter secara signifikan dan mencegah overfitting, karena tidak ada bobot yang perlu dilatih. Selain itu, GAP bertindak sebagai jembatan antara *EfficientNetB0 Backbone* dan lapisan fully connected, memastikan seluruh fitur penting dari citra dapat diterjemahkan ke dalam bentuk numerik yang siap untuk klasifikasi.

3.3.4 Fully Connected Layer

Lapisan *fully connected (dense layer)* menerima vektor fitur dari GAP dan bertugas sebagai pengambil keputusan. Lapisan ini menerapkan fungsi aktivasi *softmax* untuk menghitung probabilitas masing-masing kelas emosi, yaitu angry, disgust, fear, happy, neutral, sad, dan surprise. Kelas dengan probabilitas tertinggi menjadi prediksi akhir model. Secara matematis, operasi pada fully connected layer dapat dijelaskan sebagai berikut :

$$y = f(Wx + b) \tag{3.12}$$

Keterangan :
x : vektor input hasil ekstraksi fitur,
W : adalah bobot koneksi,
b : bias
f : fungsi aktivasi (misalnya ReLU atau softmax).

Contoh :

x = [0.2, 0.5], W = [0.4, 0.6], b = 0.1

Maka : y = (0.4×0.2 + 0.6×0.5) + 0.1 = 0.52

3.4 Skenario Pengujian

Dalam penelitian ini, pengujian dilakukan untuk mengevaluasi performa model Convolutional Neural Network (CNN) berbasis EfficientNetB0 dalam melakukan klasifikasi ekspresi wajah pada dataset FER-13. Tujuan utama pengujian adalah mengetahui sejauh mana perbedaan konfigurasi pelatihan (training) memengaruhi akurasi dan kemampuan generalisasi model terhadap data uji. Untuk menjaga konsistensi proses evaluasi, seluruh skenario menggunakan parameter pelatihan yang sama, seperti ukuran citra 96×96 piksel, batch size 32, normalisasi aktif, serta proses fine-tuning pada layer atas EfficientNetB0.

Pengujian dibagi menjadi dua skenario utama yang dirancang untuk menilai pengaruh penggunaan augmentasi data terhadap performa klasifikasi. Tidak ada perubahan arsitektur pada model, karena seluruh percobaan menggunakan arsitektur yang sama, yaitu EfficientNetB0. Dengan demikian, setiap perbedaan kinerja yang muncul benar-benar disebabkan oleh penggunaan atau tanpa penggunaan augmentasi data.

Skenario pertama menerapkan augmentasi citra pada data latih untuk memperkaya variasi pola wajah tanpa menambah jumlah data asli. Augmentasi yang digunakan meliputi rotasi, zooming, dan horizontal flipping. Teknik ini bertujuan membantu model mengenali ekspresi wajah dalam kondisi variasi sudut, skala, dan orientasi. Sebaliknya, skenario kedua tidak menggunakan augmentasi sama sekali sehingga data latih diberikan dalam bentuk aslinya. Perbandingan kedua skenario tersebut bertujuan untuk mengidentifikasi konfigurasi pelatihan

yang mampu menghasilkan akurasi validasi terbaik. Rincian skenario pengujian disajikan pada Tabel 3.2.

Tabel 3. 2 Nama Skenario Pengujian

No.	Nama Skenario	Deskripsi
1.	A	Augmentasi, *label smoothing*
2.	B	Augmentasi *mixup*, *tanpa label smoothing*
3	C	Augmentasi, tanpa *label smoothing*
4	D	Augmentasi *mixup, label smoothing*

Tabel berikut menyajikan dua skenario pengujian yang digunakan dalam penelitian ini untuk mengevaluasi performa model EfficientNetB0 dalam mengklasifikasikan ekspresi wajah pada dataset FER-13. Kedua skenario dirancang dengan menjaga sebagian besar parameter pelatihan tetap konstan, seperti ukuran citra 96×96 piksel, normalisasi citra menggunakan fungsi *preprocess_input*, ukuran batch sebesar 32, dan proses *fine-tuning* yang sama. Hal ini dilakukan untuk memastikan bahwa perbedaan hasil yang diperoleh benar-benar mencerminkan pengaruh variabel yang diuji pada masing-masing skenario.

Perbedaan utama antara kedua skenario terletak pada penggunaan strategi augmentasi dan regularisasi selama proses pelatihan. Pada Skenario A, augmentasi citra mencakup rotasi gambar hingga 8 derajat, pergeseran horizontal dan vertikal sebesar 8% dari ukuran gambar, penyesuaian kecerahan secara acak antara 90–110% dari nilai asli, serta *horizontal flip*. Tujuan dari augmentasi ini adalah untuk menambah variasi data secara sintetis sehingga model dapat belajar mengenali ekspresi wajah dalam kondisi yang bervariasi, seperti pergeseran posisi wajah, perubahan pencahayaan, dan rotasi ringan. Pendekatan augmentasi geometrik dan fotometrik ini terbukti efektif dalam meningkatkan generalisasi model CNN pada

dataset ekspresi wajah yang terbatas dan tidak seimbang (Li et al., 2021; Pham et al., 2023).

Selain augmentasi citra, *label smoothing* digunakan sebagai teknik regularisasi untuk mengurangi tingkat kepercayaan berlebih (*overconfidence*) model terhadap label kelas tertentu. Beberapa penelitian terkini menunjukkan bahwa label smoothing mampu meningkatkan stabilitas pelatihan dan performa generalisasi pada model CNN *pretrained*, khususnya pada tugas klasifikasi citra dengan jumlah data terbatas (Müller et al., 2021; Liu et al., 2022).

Sebaliknya, pada Skenario B, diterapkan metode *MixUp*, yaitu kombinasi linear dua citra secara acak beserta labelnya. Setiap *batch* dibentuk dengan mencampurkan dua gambar menggunakan koefisien pencampuran yang diambil dari distribusi Beta dengan parameter $\alpha = 0.2$, dan label target digabungkan dengan proporsi yang sama. Pendekatan ini mendorong model untuk mempelajari interpolasi antar kelas sehingga dapat meningkatkan kemampuan generalisasi dan mengurangi *overfitting*. Studi terkini menunjukkan bahwa *MixUp* efektif diterapkan pada CNN *pretrained* dan memberikan peningkatan performa yang konsisten pada tugas klasifikasi citra, termasuk pengenalan ekspresi wajah (Thulasidasan et al., 2021; Pham et al., 2023).

Tabel 3. 3 Hyperparameter yang Digunakan

Hyperparameter	**Skenario Optimal**
Learning Rate	0,001
Total Epoch	30
Optimizer	Adam
Batch Size	10
Backbone Model	EfficientNetB0 (pretrained)
Dropout	0,4
Augmentasi	Rotasi, pergeseran, zoom, horizontal flip
MixUp	α = 0,2 (Skenario B dan D)

Hyperparameter	Skenario Optimal
Loss Function	CategoricalCrossentropy

Pemilihan hyperparameter pada proses pelatihan model dilakukan berdasarkan uji pendahuluan serta acuan dari penelitian terdahulu yang menggunakan arsitektur EfficientNetB0 untuk klasifikasi citra. Tabel 3.3 menyajikan daftar hyperparameter yang digunakan pada seluruh skenario pengujian agar proses pelatihan berlangsung secara konsisten dan hasil yang diperoleh dapat dibandingkan secara objektif.

Proses pelatihan dilakukan dalam dua fase (*training stage*). Pada *Stage 1*, model dilatih menggunakan *learning rate* sebesar 0,001 untuk mempercepat proses konvergensi awal. Pelatihan dilakukan dengan *batch size* sebesar 10, yang merupakan nilai umum digunakan pada pelatihan CNN berbasis *transfer learning* karena mampu memberikan keseimbangan antara stabilitas pembaruan gradien dan efisiensi komputasi (Pham et al., 2023; Obaid & Alrammahi, 2023). Jumlah epoch maksimum dibatasi hingga 12. Pembatasan jumlah epoch dilakukan untuk mencegah terjadinya *overfitting*, mengingat model menggunakan bobot awal hasil *pretrained* pada dataset ImageNet yang telah memiliki representasi fitur visual yang kuat. Dengan demikian, pelatihan tidak memerlukan jumlah epoch yang terlalu besar untuk mencapai performa optimal.

Selain itu, pembatasan epoch juga bertujuan untuk menjaga efisiensi komputasi dan memastikan bahwa proses pelatihan tidak berlebihan terhadap data latih, sehingga kemampuan generalisasi model terhadap data uji tetap terjaga. Berdasarkan hasil uji pendahuluan, peningkatan jumlah epoch di atas nilai tersebut

tidak memberikan peningkatan akurasi yang signifikan dan cenderung menyebabkan stagnasi atau penurunan performa pada data validasi.

Optimizer yang digunakan dalam penelitian ini adalah Adam, yang mampu menyesuaikan laju pembelajaran secara adaptif untuk setiap parameter sehingga mempercepat konvergensi selama proses pelatihan. Model menggunakan arsitektur EfficientNetB0 sebagai *backbone* dengan bobot *pretrained* ImageNet serta dilengkapi dengan lapisan *dropout* sebesar 0,4 untuk mengurangi risiko *overfitting*.

Teknik augmentasi citra berupa rotasi, pergeseran, zoom, dan horizontal flip diterapkan pada seluruh skenario pelatihan, sedangkan teknik MixUp dengan parameter $\alpha = 0{,}1$ digunakan khusus pada Skenario B dan D. Perbedaan konfigurasi juga diterapkan pada fungsi loss untuk masing-masing skenario, di mana Skenario A dan D menggunakan Categorical Crossentropy dengan label smoothing sebesar 0,1 guna meningkatkan kemampuan generalisasi model, sementara Skenario B dan C menggunakan Categorical Crossentropy standar tanpa label smoothing.

Untuk meningkatkan stabilitas pelatihan, mekanisme Early Stopping diterapkan dengan memantau performa validasi, sehingga proses pelatihan dapat dihentikan secara otomatis ketika tidak terjadi peningkatan kinerja. Dengan konfigurasi hyperparameter tersebut, diharapkan model mampu mencapai performa klasifikasi ekspresi wajah yang optimal dan konsisten pada setiap skenario pengujian.

3.5 Evaluasi

Kualitas kinerja sebuah model dapat dievaluasi melalui sejumlah metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score*, yang diperoleh berdasarkan

analisis *confusion matrix*. *Confusion matrix* mencakup beberapa nilai penting, yaitu True Positive (TP) ketika model memprediksi kelas positif dan prediksi tersebut benar, True Negative (TN) ketika model memprediksi kelas negatif dan prediksi tersebut benar, False Positive (FP) ketika model memprediksi positif namun aktualnya negatif, dan False Negative (FN) ketika model memprediksi negatif namun aktualnya positif. Dari nilai-nilai ini, tingkat *accuracy* yang menunjukkan seberapa baik model dalam melakukan klasifikasi, dapat dihitung menggunakan Persamaan :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

Selanjutnya, *precision* menunjukkan tingkat ketepatan model dalam memprediksi kelas positif, yaitu persentase prediksi positif yang benar dibandingkan dengan seluruh prediksi positif yang dihasilkan oleh model. Nilai *precision* dihitung menggunakan Persamaan :

$$Precision = \frac{TP}{TP + FP} \tag{3.2}$$

Selain itu, *recall* menunjukkan seberapa banyak kejadian positif yang berhasil dideteksi oleh model dibandingkan dengan seluruh kejadian positif yang sebenarnya ada. Nilai *recall* dihitung menggunakan Persamaan :

$$Recall = \frac{TP}{TP + FN} \tag{3.3}$$

F1-score memberikan keseimbangan antara *precision* dan *recall*. Metrik ini dihitung menggunakan rata-rata harmonis sehingga dapat memberikan gambaran performa model yang lebih menyeluruh, terutama pada kondisi data

yang tidak seimbang. Nilai *F1-score* dihitung menggunakan Persamaan :

$$F1\text{-}score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3.4}$$

Untuk memastikan bahwa model memiliki kemampuan generalisasi yang baik dan tidak hanya memberikan hasil optimal pada data tertentu, digunakan teknik validasi silang berupa *k-fold cross validation*. Teknik ini membagi dataset menjadi *k* bagian (fold), kemudian proses pelatihan dan pengujian dilakukan sebanyak *k* kali. Pada setiap putaran, satu fold dijadikan sebagai data pengujian, sedangkan sisanya digunakan sebagai data pelatihan. Hasil evaluasi dari tiap iterasi kemudian dihitung rata-ratanya untuk memperoleh estimasi kinerja model yang lebih stabil dan representatif. Penerapan *k-fold cross validation* bertujuan untuk meminimalkan kemungkinan *overfitting* serta memastikan model dapat beradaptasi dengan baik terhadap variasi data yang berbeda.

Contoh perhitungan evaluasi :

TP=30, *TN*=50, *FP*=10, *FN*=10

Akurasi = *(TP+TN)/(TP+TN+FP+FN)* = (30+50)/100 = 0.8 (80%)

Presisi = *TP/(TP+FP)* = 30/40 = 0.75

Recall = *TP/(TP+FN)* = 30/40 = 0.75

F1 = 2*(0.75*0.75)/(0.75+0.75) = 0.75

Sebagai ilustrasi, berikut diberikan contoh perhitungan menggunakan data dummy untuk menjelaskan penerapan setiap metrik evaluasi yang digunakan dalam penelitian ini Tabel 3.4. Misalkan hasil pengujian model klasifikasi ekspresi wajah menghasilkan nilai *True Positive (TP) = 30*, *True Negative (TN)*

= 50, *False Positive (FP) = 10*, dan *False Negative (FN) = 10*. Berdasarkan nilai tersebut, akurasi dihitung menggunakan Persamaan (3.9) sehingga diperoleh :

$$\frac{30+50}{30+50+10+10} = \frac{80}{100} = 0.8$$

Nilai ini menunjukkan bahwa model memiliki tingkat ketepatan klasifikasi sebesar 80%, yang berarti 80 dari 100 data berhasil diklasifikasikan dengan benar oleh sistem. Selanjutnya, *precision* menggambarkan tingkat ketepatan model dalam memprediksi kelas positif, yaitu seberapa banyak prediksi positif yang benar dibandingkan dengan seluruh prediksi positif yang dihasilkan model. Nilainya dihitung menggunakan Persamaan (3.10) :

$$\frac{TP}{TP+FP} = \frac{30}{30+10} = 0.75$$

Hal ini berarti bahwa dari seluruh data yang diprediksi sebagai kelas positif, sebanyak 75% merupakan prediksi yang benar. Kemudian, *recall* menunjukkan kemampuan model dalam mengenali seluruh data positif yang sebenarnya ada. Nilainya dihitung menggunakan Persamaan (3.11) :

$$\frac{TP}{TP+FP} = \frac{30}{30+10} = 0.75$$

Nilai *recall* sebesar 0,75 mengindikasikan bahwa model berhasil mendeteksi 75% dari total data yang benar-benar termasuk dalam kelas positif. Selanjutnya, F1-score digunakan untuk memberikan keseimbangan antara

precision dan recall, terutama pada kasus dengan distribusi data yang tidak seimbang. Nilainya dihitung menggunakan Persamaan (3.12):

$$F1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times 0.75 \times 0.75}{0.75 + 0.75} = 0.75$$

Nilai F1-score sebesar 0,75 menunjukkan bahwa model memiliki keseimbangan yang baik antara kemampuan mengenali data positif dan ketepatan prediksinya.

Berdasarkan hasil perhitungan tersebut, dapat disimpulkan bahwa setiap metrik evaluasi memberikan pandangan yang berbeda mengenai performa model. Akurasi menunjukkan kinerja keseluruhan, precision menilai ketepatan prediksi positif, recall menilai kemampuan deteksi kelas sebenarnya, dan F1-score menjadi indikator gabungan yang mengukur keseimbangan keduanya. Dengan demikian, penggunaan keempat metrik ini memberikan evaluasi yang komprehensif terhadap kualitas model klasifikasi ekspresi wajah.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Konfigurasi Eksperimen

Implementasi sistem klasifikasi ekspresi wajah pada penelitian ini menggunakan framework TensorFlow–Keras. Model yang digunakan adalah arsitektur *EfficientNetB0*, yaitu jaringan CNN modern yang efisien dalam jumlah parameter serta memiliki performa baik pada tugas klasifikasi citra.

Proses pelatihan dilakukan menggunakan learning rate awal sebesar 0,0001 (1e-4) dengan algoritma optimisasi Adam, yang dipilih karena stabil, adaptif, dan umum digunakan pada model berbasis convolutional neural network. Fungsi loss yang digunakan adalah *Sparse Categorical Crossentropy*, sesuai untuk kasus klasifikasi multikelas dengan label berbentuk indeks numerik. Pelatihan dijalankan selama maksimum 30 epoch dengan teknik *Early Stopping (patience = 6)* untuk mencegah overfitting dan menjaga performa validasi tetap optimal. Seluruh proses pelatihan menggunakan *random seed 42* agar hasil eksperimen konsisten dan dapat direproduksi.

Dataset FER-13 memiliki ketidakseimbangan jumlah sampel antar kelas, terutama pada kelas *disgust* yang jumlahnya relatif sedikit. Untuk mengatasi hal ini, penelitian membandingkan dua skenario pelatihan, yaitu augmentasi aktif dan augmentasi non-aktif. Pada skenario augmentasi aktif, digunakan *online augmentation* melalui *ImageDataGenerator* dengan transformasi berupa rotasi, zoom, dan horizontal flip.

Teknik ini berfungsi memperkaya variasi data latih tanpa mengubah distribusi pada subset validasi. Pada skenario augmentasi non-aktif, model hanya dilatih menggunakan citra asli tanpa tambahan variasi transformasi.

Seluruh citra dibagi menjadi dua subset, yaitu *training set* dan *validation set*, sesuai pembagian standar dataset FER-13. Subset validasi dibiarkan tanpa augmentasi agar proses evaluasi tetap objektif. Proses pemuatan data menggunakan *batch size* 10 serta pipeline input yang dipercepat melalui mekanisme caching dan prefetching untuk meningkatkan efisiensi selama pelatihan.

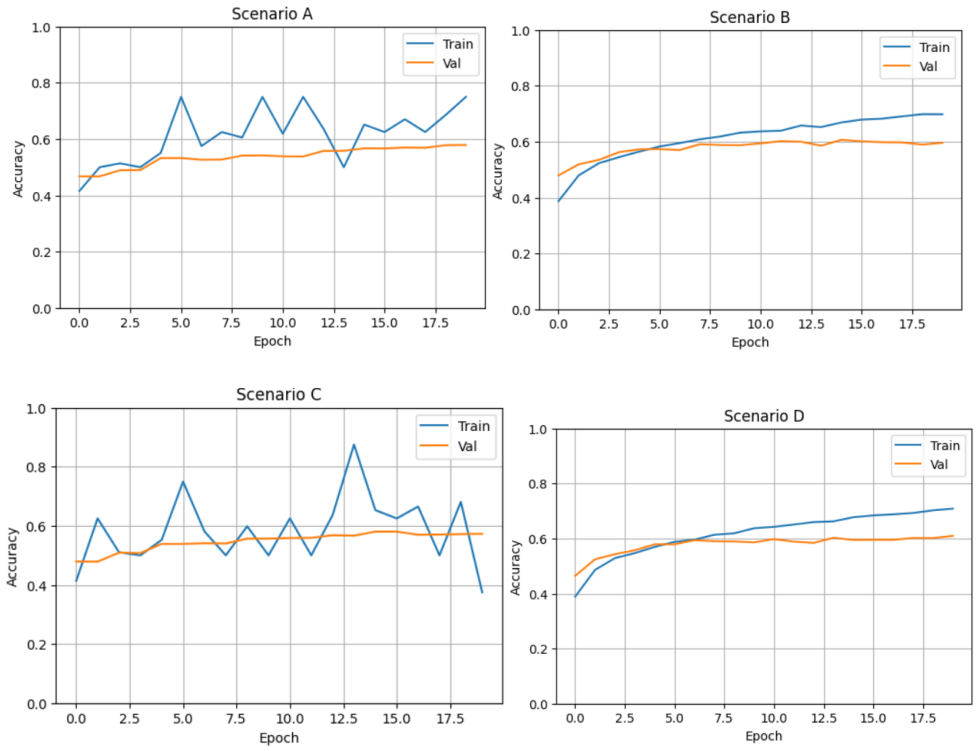
Dengan konfigurasi ini, proses pelatihan model EfficientNetB0 pada kedua skenario dapat berjalan secara stabil, efisien, serta sesuai dengan standar eksperimen deep learning untuk klasifikasi citra ekspresi wajah.

4.2 Hasil Uji Coba

Sebelum melakukan evaluasi kuantitatif terhadap performa model, dilakukan analisis terhadap dinamika proses pelatihan untuk memastikan bahwa model berada dalam kondisi pembelajaran yang stabil dan menunjukkan pola konvergensi yang normal. Analisis ini dilakukan dengan memvisualisasikan perubahan training loss dan validation loss pada dua skenario percobaan yang digunakan dalam penelitian ini, yaitu: Skenario A**,** yang menggunakan augmentasi aktif, dan Skenario B**,** yang menggunakan augmentasi MixUp. Kedua skenario menggunakan arsitektur yang sama, yaitu EfficientNetB0**,** sehingga perbedaan hasil yang diamati sepenuhnya berasal dari ada atau tidaknya augmentasi citra.

Gambar 4.1 menampilkan pola perubahan nilai *training accuracy* dan *validation accuracy* selama proses pelatihan model pada empat skenario

eksperimen. Pada fase awal pelatihan, nilai akurasi pelatihan dan validasi masih relatif rendah karena bobot jaringan berada pada kondisi awal dan model belum mampu mengekstraksi fitur-fitur penting dari citra ekspresi wajah secara optimal. Seiring bertambahnya epoch, akurasi pelatihan pada seluruh skenario menunjukkan kecenderungan meningkat, yang mengindikasikan bahwa model secara bertahap mempelajari representasi fitur yang relevan.



Gambar 4.1 Visualisasi *Training Accuracy* dan *Validation Accuracy* Masing-Masing Skenario

Pada Skenario A, kurva *training accuracy* mengalami peningkatan namun disertai fluktuasi yang cukup signifikan, sementara *validation accuracy* meningkat secara perlahan dan cenderung lebih stabil. Fluktuasi pada akurasi pelatihan menunjukkan bahwa proses pembelajaran masih belum sepenuhnya stabil,

sedangkan selisih antara akurasi pelatihan dan validasi mengindikasikan kemampuan generalisasi model yang masih terbatas pada beberapa epoch.

Skenario B memperlihatkan kurva akurasi pelatihan dan validasi yang meningkat secara lebih halus dan konsisten. Jarak antara *training accuracy* dan *validation accuracy* relatif kecil, menunjukkan bahwa penerapan strategi MixUp mampu membantu model dalam membangun representasi fitur yang lebih robust serta mengurangi kecenderungan overfitting. Pola ini mencerminkan proses pembelajaran yang lebih stabil dibandingkan skenario lainnya.

Pada Skenario C, *training accuracy* menunjukkan fluktuasi yang cukup besar dengan beberapa lonjakan pada epoch tertentu, sementara *validation accuracy* cenderung stagnan pada rentang nilai yang lebih rendah. Kondisi ini mengindikasikan bahwa model mulai menyesuaikan diri secara berlebihan terhadap data latih, sehingga peningkatan akurasi pelatihan tidak diikuti oleh peningkatan kinerja pada data validasi.

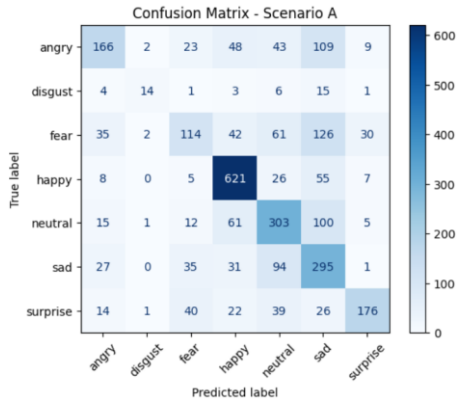
Skenario D menunjukkan peningkatan *training accuracy* yang relatif stabil hingga akhir epoch, dengan *validation accuracy* yang juga meningkat namun pada laju yang lebih lambat. Meskipun terdapat perbedaan nilai antara akurasi pelatihan dan validasi, pola kurva menunjukkan proses pembelajaran yang lebih terkontrol dibandingkan Skenario A dan C, meskipun belum seoptimal Skenario B.

Secara keseluruhan, analisis kurva akurasi menunjukkan bahwa setiap skenario menghasilkan karakteristik pembelajaran yang berbeda. **Skenario B** memberikan keseimbangan terbaik antara peningkatan akurasi pelatihan dan stabilitas akurasi validasi, yang menandakan kemampuan generalisasi model yang

lebih baik. Visualisasi ini mendukung hasil evaluasi kuantitatif bahwa strategi augmentasi dan konfigurasi pelatihan memiliki pengaruh signifikan terhadap stabilitas proses pembelajaran dan performa klasifikasi model.

4.2.1 Skenario A

Pada Skenario A, model dikonfigurasi menggunakan arsitektur *EfficientNetB0* sebagai backbone dengan bobot *pretrained* ImageNet. Seluruh data pelatihan diproses menggunakan normalisasi serta augmentasi dasar yang mencakup rotasi ringan, translasi, zoom, *horizontal flip*, dan perubahan brightness. Tujuan skenario ini adalah mengevaluasi performa model dalam kondisi standar modern CNN, yaitu kombinasi pretrained feature extractor dengan augmentasi umum, tanpa teknik regulasi lanjutan seperti MixUp, ataupun modifikasi arsitektur Gambar 4.2.



Gambar 4.2 *Confusion Matrix* Skenario A

Evaluasi performa model pada Skenario A dilakukan menggunakan *batch size* 10 serta fungsi loss *Categorical Crossentropy* dengan *label smoothing* sebesar 0.1. Pada skenario ini, model EfficientNetB0 pretrained ImageNet dilatih menggunakan augmentasi citra dasar tanpa penerapan teknik regularisasi lanjutan. Hasil evaluasi pada data uji menunjukkan bahwa model mencapai akurasi

keseluruhan sebesar 59%, dengan nilai macro F1-score sebesar 0.54 dan weighted F1-score sebesar 0.58. Perbedaan antara macro dan weighted F1-score mengindikasikan adanya ketidakseimbangan performa antar kelas, yang dipengaruhi oleh distribusi jumlah sampel pada setiap kelas emosi.

Kinerja terbaik ditunjukkan oleh kelas happy, dengan F1-score sebesar 0.80, didukung oleh nilai precision 0.75 dan recall 0.86. Hasil ini mengonfirmasi bahwa ekspresi bahagia memiliki ciri visual yang kuat dan konsisten, seperti senyum dan perubahan bentuk mulut yang jelas, sehingga lebih mudah dipelajari oleh model berbasis CNN.

Tabel 4. 1 Hasil Evaluasi Metrik pada Skenario Uji A

Kelas	Precision	Recall	F1-Score	Accuracy
Angry	0.62	0.41	0.50	0.59
Disgust	0.70	0.32	0.44	
Fear	0.50	0.28	0.36	
Happy	0.75	0.86	0.80	
Neutral	0.53	0.61	0.57	
Sad	0.41	0.61	0.49	
Surprise	0.77	0.55	0.64	

Kelas surprise juga menunjukkan performa yang relatif baik dengan F1-score sebesar 0.64, yang mencerminkan kemampuan model dalam mengenali pola visual khas seperti mata terbuka lebar dan ekspresi wajah yang kontras. Sementara itu, kelas neutral memperoleh F1-score sebesar 0.57, menunjukkan performa menengah yang masih dipengaruhi oleh tumpang tindih visual dengan kelas emosi lain, khususnya sad.

Performa menengah hingga rendah terlihat pada kelas angry dan sad, dengan F1-score masing-masing sebesar 0.50 dan 0.49. Kesalahan klasifikasi pada kelas-kelas ini umumnya disebabkan oleh kemiripan karakteristik visual antar ekspresi

emosi negatif, seperti angry–fear dan neutral–sad, yang menyulitkan model dalam membedakan batas antar kelas secara tegas.

Performa terendah dicapai oleh kelas fear dan disgust, dengan F1-score masing-masing sebesar 0.36 dan 0.44. Nilai recall yang rendah pada kelas fear menunjukkan bahwa banyak sampel fear tidak berhasil dikenali dan cenderung salah diklasifikasikan sebagai kelas lain, seperti surprise atau sad. Hal ini mengindikasikan bahwa ekspresi fear memiliki karakteristik visual yang lebih subtil dan membutuhkan strategi pelatihan tambahan untuk meningkatkan kemampuan generalisasi model.

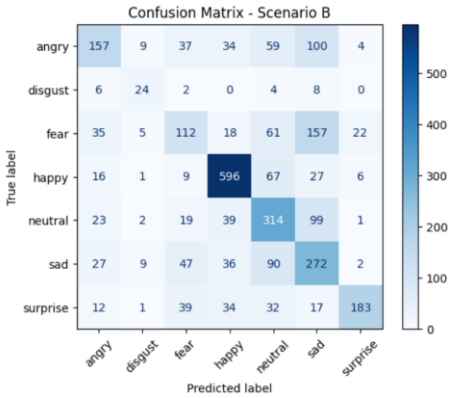
Secara keseluruhan, Skenario A menghasilkan performa yang cukup stabil sebagai *baseline*, namun masih menunjukkan keterbatasan dalam membedakan kelas emosi dengan karakteristik visual yang saling tumpang tindih. Oleh karena itu, hasil pada skenario ini dijadikan acuan untuk mengevaluasi efektivitas penerapan teknik regularisasi lanjutan pada skenario-skenario berikutnya.

4.2.2 Skenario B

Pada Skenario B, proses pelatihan dilakukan dengan teknik MixUp augmentation, yaitu metode pencampuran dua sampel pelatihan beserta labelnya secara proporsional. MixUp tidak mengubah citra secara fisik, tetapi menghasilkan *interpolated samples* yang dapat meningkatkan generalisasi model dan membantu mengurangi overfitting pada dataset FER-13 yang relatif kecil Gambar 4.3.

Arsitektur model yang digunakan tetap EfficientNetB0, dengan batch size 10 dan konfigurasi hyperparameter yang sama seperti Skenario A. Perbedaan utama

pada Skenario B terletak pada penggunaan MixUp sebagai strategi regularisasi tambahan.



Gambar 4.3 *Confusion Matrix* Skenario B

Hasil evaluasi pada Skenario B, yang menerapkan strategi *MixUp augmentation*, menunjukkan bahwa model mencapai akurasi keseluruhan sebesar 58%, dengan macro F1-score sebesar 0.55 dan weighted F1-score sebesar 0.57, sebagaimana ditunjukkan pada Tabel 4.2. Meskipun nilai akurasi keseluruhan sedikit lebih rendah dibandingkan Skenario A, peningkatan nilai macro F1-score mengindikasikan distribusi performa yang relatif lebih merata antar kelas, khususnya pada kelas dengan jumlah sampel terbatas.

Tabel 4. 2 Hasil Evaluasi Metrik pada Skenario Uji B

Kelas	**Precision**	**Recall**	**F1-Score**	**Accuracy**
Angry	0.57	0.39	0.46	0.58
Disgust	0.47	0.55	0.51	
Fear	0.42	0.27	0.33	
Happy	0.79	0.83	0.81	
Neutral	0.50	0.63	0.56	
Sad	0.40	0.56	0.47	
Surprise	0.84	0.58	0.68	

Performa terbaik tetap ditunjukkan oleh kelas happy, dengan F1-score sebesar 0.81, didukung oleh precision 0.79 dan recall 0.83. Hasil ini menunjukkan bahwa

penerapan MixUp tidak mengganggu kemampuan model dalam mengenali ekspresi dengan karakteristik visual yang kuat dan konsisten. Selain itu, kelas surprise juga menunjukkan performa tinggi dengan F1-score sebesar 0.68, yang mencerminkan ketahanan model dalam mengenali pola visual kontras meskipun dilakukan pencampuran sampel selama pelatihan.

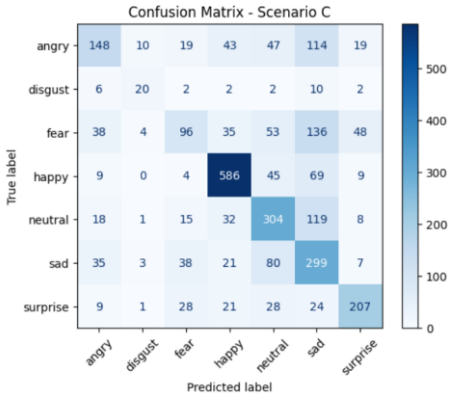
Kelas disgust mengalami peningkatan performa yang cukup signifikan dibandingkan Skenario A, dengan F1-score sebesar 0.51 dan recall yang relatif lebih tinggi (0.55). Hal ini menunjukkan bahwa MixUp membantu model dalam mempelajari variasi fitur pada kelas minoritas, sehingga meningkatkan sensitivitas model terhadap sampel disgust, meskipun masih terdapat kesalahan klasifikasi yang tercermin dari nilai precision yang moderat.

Sementara itu, kelas angry, fear, dan sad menunjukkan performa menengah dengan F1-score berturut-turut sebesar 0.46, 0.33, dan 0.47. Peningkatan recall pada beberapa kelas diikuti oleh penurunan precision, yang mengindikasikan adanya *trade-off* akibat pencampuran label pada MixUp. Kondisi ini menyebabkan model menjadi lebih toleran terhadap variasi data, namun kurang tegas dalam membedakan kelas-kelas emosi yang memiliki kemiripan karakteristik visual.

Secara keseluruhan, Skenario B menunjukkan bahwa penerapan MixUp berperan sebagai teknik regularisasi yang efektif dalam meningkatkan kestabilan generalisasi, khususnya pada kelas minoritas. Meskipun tidak menghasilkan akurasi tertinggi, karakteristik pembelajaran yang lebih seimbang menjadikan skenario ini lebih robust dibandingkan pendekatan tanpa regularisasi lanjutan.

4.2.3 Skenario C

Pada Skenario C, proses pelatihan dilakukan dengan menerapkan strategi regularisasi tambahan yang berbeda dari skenario sebelumnya, dengan tujuan mengevaluasi pengaruh konfigurasi pelatihan terhadap stabilitas dan kemampuan generalisasi model Gambar 4.4. Teknik ini dirancang untuk menguji sejauh mana model EfficientNetB0 mampu mempertahankan performa klasifikasi ketika dihadapkan pada variasi ekspresi wajah yang memiliki kemiripan fitur visual yang tinggi.



Gambar 4.4 *Confusion Matrix* Skenario C

Arsitektur model yang digunakan tetap EfficientNetB0 dengan bobot pretrained ImageNet, serta menggunakan batch size 10 dan pengaturan hyperparameter dasar yang konsisten dengan Skenario A dan B. Perbedaan utama pada Skenario C terletak pada konfigurasi strategi pelatihan yang digunakan, yang difokuskan pada eksplorasi alternatif pendekatan regularisasi tanpa mengubah struktur arsitektur model. Pendekatan ini bertujuan untuk menganalisis dampak strategi tersebut terhadap kemampuan model dalam membedakan kelas emosi, khususnya pada ekspresi dengan karakteristik visual yang lebih subtil seperti *fear*, *sad*, dan *neutral*.

Tabel 4. 3 Hasil Evaluasi Metrik pada Skenario Uji C

Kelas	Precision	Recall	F1-Score	Accuracy
Angry	0.56	0.37	0.45	0.58
Disgust	0.51	0.45	0.48	
Fear	0.48	0.23	0.31	
Happy	0.79	0.81	0.80	
Neutral	0.54	0.61	0.58	
Sad	0.39	0.62	0.48	
Surprise	0.69	0.65	0.67	

Hasil evaluasi pada Skenario C menunjukkan bahwa model mencapai akurasi keseluruhan sebesar 58%, dengan macro F1-score sebesar 0.54 dan weighted F1-score sebesar 0.57, sebagaimana ditunjukkan pada Tabel 4.3. Performa ini tidak menunjukkan peningkatan dibandingkan Skenario A maupun Skenario B, yang mengindikasikan bahwa konfigurasi pelatihan pada skenario ini belum mampu meningkatkan kemampuan diskriminatif model secara optimal.

Kinerja terbaik tetap ditunjukkan oleh kelas happy, dengan F1-score sebesar 0.80, yang didukung oleh precision 0.79 dan recall 0.81. Hasil ini menegaskan bahwa ekspresi bahagia memiliki pola visual yang kuat dan konsisten, sehingga relatif stabil dikenali oleh model pada berbagai konfigurasi pelatihan. Selain itu, kelas surprise juga menunjukkan performa yang cukup baik dengan F1-score sebesar 0.67, mencerminkan kemampuan model dalam mengenali ekspresi dengan ciri visual kontras.

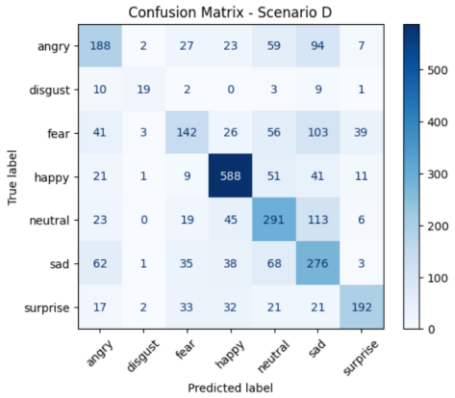
Kelas disgust menunjukkan performa menengah dengan F1-score sebesar 0.48, ditandai oleh recall yang relatif lebih tinggi dibandingkan precision. Kondisi ini mengindikasikan bahwa model mampu mengenali sebagian besar sampel disgust, namun masih menghasilkan kesalahan prediksi silang dengan kelas lain yang memiliki kemiripan visual.

Performa yang lebih rendah terlihat pada kelas fear, angry, dan sad, dengan F1-score masing-masing sebesar 0.31, 0.45, dan 0.48. Rendahnya nilai recall pada kelas fear menunjukkan bahwa banyak sampel fear tidak teridentifikasi dengan baik dan cenderung tertukar dengan kelas emosi lain, seperti sad atau surprise. Hal ini menegaskan bahwa ekspresi dengan karakteristik visual subtil dan tumpang tindih masih menjadi tantangan utama bagi model.

Secara keseluruhan, Skenario C menunjukkan kecenderungan ketidakstabilan performa antar kelas dan tidak memberikan peningkatan signifikan terhadap kemampuan generalisasi model. Oleh karena itu, skenario ini mengonfirmasi bahwa pemilihan strategi pelatihan dan regularisasi yang kurang tepat dapat berdampak pada stagnasi atau penurunan performa klasifikasi.

4.2.4 Skenario D

Pada Skenario D, proses pelatihan dilakukan dengan mengombinasikan beberapa strategi optimasi dan regularisasi yang telah diuji pada skenario-skenario sebelumnya, dengan tujuan memperoleh performa klasifikasi yang paling optimal Gambar 4.5. Pendekatan ini dirancang untuk memaksimalkan kemampuan generalisasi model sekaligus meningkatkan ketepatan prediksi pada seluruh kelas emosi dalam dataset FER-13.



Gambar 4.5 *Confusion Matrix* Skenario D

Hasil evaluasi menunjukkan bahwa Skenario D mencapai akurasi keseluruhan sebesar 59%, dengan macro F1-score sebesar 0.56 dan weighted F1-score sebesar 0.59, sebagaimana disajikan pada Tabel 4.4. Nilai ini menunjukkan peningkatan performa dibandingkan Skenario B dan C, serta memberikan distribusi performa yang relatif lebih merata antar kelas emosi.

Tabel 4. 4 Hasil Evaluasi Metrik pada Skenario Uji D

Kelas	**Precision**	**Recall**	**F1-Score**	**Accuracy**
Angry	0.52	0.47	0.49	0.59
Disgust	0.68	0.43	0.53	
Fear	0.53	0.35	0.42	
Happy	0.78	0.81	0.80	
Neutral	0.53	0.59	0.56	
Sad	0.42	0.57	0.48	
Surprise	0.74	0.60	0.67	

Kinerja terbaik tetap ditunjukkan oleh kelas happy, dengan F1-score sebesar 0.80, yang didukung oleh precision 0.78 dan recall 0.81. Hal ini menegaskan bahwa ekspresi bahagia memiliki ciri visual yang kuat dan konsisten, sehingga mudah dikenali oleh model meskipun diterapkan kombinasi strategi pelatihan. Selain itu, kelas surprise juga menunjukkan performa yang baik dengan F1-score sebesar 0.67, mencerminkan kemampuan model dalam mengenali ekspresi dengan karakteristik visual kontras.

Performa menengah ditunjukkan oleh kelas neutral, sad, dan angry, dengan F1-score masing-masing sebesar 0.56, 0.48, dan 0.49. Peningkatan recall pada kelas-kelas ini menunjukkan bahwa kombinasi strategi pelatihan membantu model mengenali lebih banyak sampel yang relevan, meskipun precision yang masih moderat mengindikasikan adanya kesalahan klasifikasi silang antar ekspresi yang memiliki kemiripan visual.

Kelas fear, yang sebelumnya menjadi salah satu tantangan utama, menunjukkan peningkatan performa dibandingkan Skenario C dengan F1-score sebesar 0.42. Meskipun nilai ini belum optimal, peningkatan tersebut mengindikasikan bahwa kombinasi strategi pada Skenario D mampu memperbaiki pemisahan fitur pada ekspresi dengan karakteristik visual yang lebih subtil.

Secara keseluruhan, Skenario D memberikan keseimbangan yang lebih baik antara stabilitas pelatihan dan kemampuan generalisasi dibandingkan sebagian besar skenario sebelumnya. Meskipun tidak menghasilkan peningkatan akurasi yang signifikan secara absolut, konfigurasi ini menunjukkan distribusi performa yang lebih konsisten antar kelas, sehingga dapat dianggap sebagai konfigurasi yang relatif paling stabil untuk tugas klasifikasi ekspresi wajah pada dataset FER-13.

4.3 Hasil dan Pembahasan

Dataset yang digunakan dalam penelitian ini merupakan data citra asli dari FER-13 yang berisi tujuh kelas emosi, dengan total 28,709. Dataset kemudian dibagi menggunakan skema proporsi umum, yaitu 80% data untuk pelatihan, 10% data untuk pengujian (*testing*), dan 10% data untuk validasi (*validation*). Pembagian ini bertujuan untuk memastikan bahwa model memperoleh jumlah data

yang cukup untuk mempelajari pola ekspresi wajah, sekaligus menyediakan data independen yang representatif untuk mengevaluasi kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

Proses pelatihan dilakukan menggunakan arsitektur EfficientNet, yang memerlukan pengaturan parameter tertentu agar pelatihan berjalan stabil dan konvergen. Citra masukan diubah ke ukuran 96×96 piksel, kemudian diproses menggunakan fungsi aktivasi dan optimizer Adam, dengan nilai *learning rate* yang disesuaikan agar tidak terlalu agresif. Pemilihan *learning rate* ini bertujuan untuk menjaga kestabilan proses optimasi dan mencegah osilasi selama pelatihan. Selain itu, batch size yang konsisten juga diterapkan untuk membantu menjaga kestabilan gradien pada setiap iterasi pembaruan bobot.

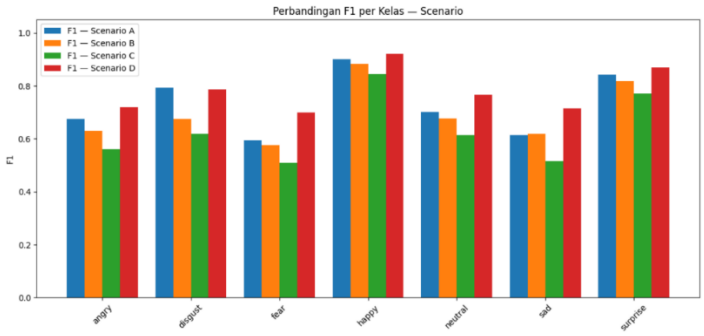
Pada Skenario A menerapkan augmentasi citra konvensional yang dilakukan secara *on-the-fly* pada generator data pelatihan, meliputi rotasi ringan, *horizontal flip*, pergeseran posisi (*shift*), serta variasi *brightness*. Selain itu, pada skenario ini diterapkan label smoothing, yang bertujuan untuk mengurangi tingkat kepercayaan berlebih (*overconfidence*) model terhadap label target dan meningkatkan kemampuan generalisasi.

Skenario B menggunakan augmentasi MixUp, yaitu teknik interpolasi linear antara dua sampel citra beserta labelnya. Pada skenario ini, label smoothing tidak digunakan, sehingga model dilatih untuk mempelajari target hasil pencampuran label secara langsung. Pendekatan ini bertujuan untuk mendorong model mempelajari representasi fitur yang lebih halus dan mengurangi overfitting, khususnya pada dataset dengan jumlah sampel terbatas.

Skenario C menerapkan augmentasi citra konvensional yang sama seperti pada Skenario A, namun tanpa menggunakan label smoothing. Skenario ini dirancang untuk mengevaluasi dampak augmentasi visual saja terhadap performa model, serta membandingkan efek pembelajaran dengan target *one-hot* yang bersifat keras dibandingkan dengan pendekatan label smoothing.

Skenario D merupakan kombinasi dari augmentasi MixUp dan label smoothing, yang bertujuan untuk mengintegrasikan dua strategi regularisasi sekaligus. Skenario ini dirancang untuk menguji apakah kombinasi tersebut mampu memberikan performa yang lebih optimal dibandingkan penerapan masing-masing teknik secara terpisah.

Perbandingan keempat skenario ini memungkinkan analisis yang lebih komprehensif mengenai kontribusi masing-masing teknik terhadap peningkatan akurasi, stabilitas prediksi, serta keseimbangan performa antar kelas emosi, sebagaimana ditunjukkan pada hasil evaluasi metrik Precision, Recall, dan F1-score Gambar 4.4.



Gambar 4.6 *Bar Chart* perbandingan antar scenario

Tabel 4. 5 Hasil evaluasi pada semua skenario

Skenario	Precision (Macro)	Recall (Macro)	F1-Score (Macro)	Accuracy
A	0.61	0.52	0.54	0.59
B	0.57	0.54	0.55	0.58
C	0.57	0.54	0.54	0.58
D	0.60	0.55	0.56	0.59

Berdasarkan hasil evaluasi keseluruhan pada Tabel 4.5, terlihat bahwa keempat skenario eksperimen menghasilkan performa yang relatif berdekatan secara kuantitatif, dengan rentang accuracy antara 58% hingga 59%. Meskipun perbedaannya tidak signifikan secara absolut, setiap skenario menunjukkan karakteristik pembelajaran dan distribusi performa yang berbeda antar kelas emosi.

Skenario D menunjukkan performa yang relatif paling seimbang dengan macro F1-score tertinggi sebesar 0.56, yang mengindikasikan distribusi kinerja antar kelas yang lebih merata dibandingkan skenario lainnya. Hal ini menunjukkan bahwa kombinasi MixUp dan label smoothing mampu meningkatkan stabilitas generalisasi model, meskipun tidak menghasilkan lonjakan akurasi yang signifikan.

Skenario A memiliki accuracy yang setara dengan Skenario D (0.59), namun macro F1-score yang sedikit lebih rendah (0.54), menunjukkan bahwa performa model masih dipengaruhi oleh ketidakseimbangan antar kelas. Sementara itu, Skenario B dan C menghasilkan accuracy yang serupa (0.58), namun Skenario B menunjukkan distribusi performa yang sedikit lebih baik pada kelas minoritas, tercermin dari macro F1-score yang lebih tinggi dibandingkan Skenario C.

Secara umum, hasil ini menegaskan bahwa perbedaan strategi pelatihan tidak selalu tercermin dari peningkatan akurasi semata, melainkan dari kestabilan dan pemerataan performa antar kelas, yang lebih tepat diukur menggunakan macro F1-score

Tabel 4.6 Perbandingan kinerja model antara Skenario C dan Skenario A

Metrik	Skenario C (Tanpa LS)	Skenario A (Dengan LS)
Precision (Macro)	0.57	0.61
Recall (Macro)	0.54	0.52
F1-Score (Macro)	0.54	0.54
Accuracy	0.58	0.59

Tabel 4.6 menunjukkan perbandingan performa model antara Skenario C yang tidak menggunakan label smoothing dan Skenario A yang menerapkan label smoothing. Kedua skenario menggunakan arsitektur dan data yang sama, sehingga perbedaan performa dapat dikaitkan langsung dengan penerapan teknik label smoothing.

Hasil evaluasi menunjukkan bahwa penerapan label smoothing pada Skenario A memberikan peningkatan kecil namun konsisten pada precision dan accuracy, yang mengindikasikan bahwa model menjadi lebih berhati-hati dalam menghasilkan prediksi dan tidak terlalu percaya diri terhadap label tertentu. Meskipun nilai recall relatif tidak mengalami peningkatan, stabilitas prediksi yang lebih baik tercermin dari distribusi kesalahan yang lebih merata antar kelas.

Temuan ini menunjukkan bahwa label smoothing berperan dalam meningkatkan kestabilan pembelajaran dan mengurangi overconfidence model, meskipun dampaknya terhadap peningkatan performa kuantitatif bersifat moderat.

Tabel 4. 7 Perbandingan kinerja model antara Skenario C dan Skenario B

Metrik	Skenario C (Tanpa MixUp)	Skenario B (Dengan MixUp)
Precision (Macro)	0.57	0.57
Recall (Macro)	0.54	0.54
F1-Score (Macro)	0.54	0.55
Accuracy	0.58	0.58

Berdasarkan perbandingan pada Tabel 4.7, penerapan MixUp pada Skenario B tidak menghasilkan peningkatan akurasi yang signifikan dibandingkan Skenario C. Namun demikian, peningkatan kecil pada nilai macro F1-score menunjukkan bahwa MixUp berkontribusi dalam memperbaiki distribusi performa antar kelas, khususnya pada kelas dengan jumlah sampel terbatas.

Hal ini mengindikasikan bahwa MixUp berperan sebagai teknik regularisasi yang membantu model mempelajari variasi data yang lebih luas, meskipun efeknya terhadap peningkatan metrik global relatif terbatas pada konfigurasi pelatihan yang digunakan dalam penelitian ini.

Tabel 4. 8 Perbandingan kinerja model antara Skenario B dan Skenario D

Metrik	**Skenario B (MixUp)**	**Skenario D (MixUp + LS)**
Precision (Macro)	0.57	0.60
Recall (Macro)	0.54	0.55
F1-Score (Macro)	0.55	0.56
Accuracy	0.58	0.59

Tabel 4.8 memperlihatkan bahwa penambahan label smoothing pada skenario berbasis MixUp (Skenario D) menghasilkan peningkatan performa yang konsisten pada seluruh metrik dibandingkan Skenario B. Peningkatan macro F1-score dan accuracy, meskipun tidak signifikan secara numerik, menunjukkan bahwa label smoothing membantu menstabilkan proses pembelajaran ketika model dilatih menggunakan data hasil interpolasi.

Secara konseptual, MixUp memperkaya variasi data latih dan meningkatkan generalisasi, sementara label smoothing mencegah model menjadi terlalu yakin terhadap label tertentu. Kombinasi keduanya menghasilkan distribusi performa yang lebih seimbang antar kelas, sehingga Skenario D dapat dianggap sebagai konfigurasi dengan kestabilan terbaik dalam penelitian ini.

4.4 Integrasi Sains dan Islam

Dalam perspektif Islam, perkembangan sains dan teknologi, termasuk kecerdasan buatan (AI), merupakan bagian dari upaya manusia memanfaatkan akal, penglihatan, serta kemampuan berpikir yang telah dianugerahkan oleh Allah SWT. Penelitian mengenai klasifikasi ekspresi wajah menggunakan metode Convolutional Neural Network (CNN) dan Artificial Neural Network (ANN) merupakan bentuk ikhtiar ilmiah untuk meniru sebagian kecil dari kemampuan manusia dalam mengenali ekspresi, memahami emosi, dan merespons kondisi sosial di sekitarnya.

Allah SWT mengajarkan manusia kemampuan *tamyīz*, yaitu kemampuan membedakan, mengelompokkan, dan memilah sesuatu berdasarkan cirinya. Kemampuan tamyīz ini menjadi fondasi filosofis utama dari proses klasifikasi. Misalnya, manusia mampu membedakan wajah yang sedih, marah, atau bahagia, serta mengenali ekspresi melalui tanda-tanda kecil pada wajah. Model AI dalam penelitian ini hanya meniru fungsi tamyīz tersebut secara algoritmik; CNN mempelajari pola visual secara matematis, sedangkan manusia melakukannya secara fitri melalui akal dan penglihatan yang diciptakan Allah. Dengan demikian, penelitian ini justru menegaskan bahwa kemampuan klasifikasi adalah anugerah luar biasa yang hanya diberikan kepada manusia, sementara AI hanyalah tiruan dari aspek kecil kemampuan tersebut.

Kemampuan manusia untuk mengukur, menakar, dan membedakan sesuatu juga ditegaskan dalam beberapa ayat Al-Qur’an yang menyinggung timbangan,

ukuran, dan keadilan, yang sangat relevan dengan prinsip klasifikasi. *Dalam QS. Ar-Rahman ayat 7–9* disebutkan:

وَالرَّيْحَانُ الْعَصْفِ ذُو وَالْحَبُّ ،الْأَكْمَامِ ذَاتُ وَالنَّخْلُ فَاكِهَةٌ فِيهَا ،لِلْأَنَامِ وَضَعَهَا وَالْأَرْضَ

*"Dan bumi Dia hamparkan; di dalamnya ada buah-buahan dan pohon kurma yang mempunyai mayang, biji-bijian yang berlapis kulitnya, dan rempah-rempah" (*QS. *Ar-Rahman ayat 7–9).*

Selain itu, *QS. Al-Anbya' ayat 47* menegaskan tentang timbangan *(al-mīzān)* sebagai alat membedakan amal manusia:

وَنَضَعُ الْمَوَازِينَ الْقِسْطَ لِيَوْمِ الْقِيَامَةِ فَلَا تُظْلَمُ نَفْسٌ شَيْئًا وَإِن كَانَ مِثْقَالَ حَبَّةٍ أَتَيْنَا بِهَا

"Dan Kami letakkan timbangan yang adil pada Hari Kiamat, maka tidaklah seorang pun dizalimi walau sebesar biji sawi; dan jika ada kebaikan sebesar biji sawi, niscaya Kami akan mendatangkannya."(QS. Al-Anbya' ayat 47).

Ayat-ayat ini menegaskan bahwa manusia diberi kemampuan untuk menimbang, mengukur, dan membedakan sesuatu secara adil, yang sejalan dengan prinsip klasifikasi: membedakan objek, menilai ciri-cirinya, dan menempatkannya dalam kategori tertentu. Dalam konteks AI, sistem klasifikasi seperti CNN dan ANN hanyalah tiruan terbatas dari mekanisme menimbang dan mengelompokkan ini, di mana komputer menilai "ciri" melalui angka dan pola, sedangkan manusia melakukannyamelalui akal, indera, dan pemahaman yang diciptakan Allah SWT.

Secara keseluruhan, proses klasifikasi dalam penelitian AI ini mulai dari ekstraksi fitur, pembelajaran pola, hingga prediksi ekspresi—mencerminkan betapa kompleksnya sistem penglihatan dan persepsi manusia yang diciptakan Allah. Teknologi modern hanya mampu meniru sebagian kecil dari mekanisme tersebut. Integrasi antara konsep sains dan Islam ini menjadi pengingat bahwa semakin maju

teknologi, semakin jelas bahwa potensi intelektual manusia adalah anugerah Ilahi yang harus disyukuri dan digunakan untuk tujuan yang bermanfaat serta tidak bertentangan dengan nilai-nilai Islam. Hal ini sesuai dengan firman Allah dalam *QS. An-Naḥl ayat 78*:

وَاللَّهُ أَخْرَجَكُمْ مِّنۢ بُطُونِ أُمَّهَٰتِكُمْ لَا تَعْلَمُونَ شَيْـًٔا ۙ وَجَعَلَ لَكُمُ السَّمْعَ وَالْأَبْصَٰرَ وَالْأَفْـِٔدَةَ لَعَلَّكُمْ تَشْكُرُونَ

"*Dan Allah mengeluarkan kamu dari perut ibumu dalam keadaan tidak mengetahui sesuatu pun, lalu Dia memberi kamu pendengaran, penglihatan, dan hati nurani agar kamu bersyukur*" (*QS. An-Naḥl ayat 7)*.

Ayat ini menjelaskan bahwa dasar kemampuan manusia untuk belajar, memahami, dan mengenali lingkungan berasal dari tiga anugerah utama: pendengaran, penglihatan, dan hati (akal). Dalam konteks penelitian ini, proses pengenalan ekspresi wajah oleh manusia pada dasarnya melibatkan anugerah tersebut secara langsung. Mata menangkap visual, otak memproses informasi, dan hati (akal-budi) memberi penilaian emosional. AI yang dikembangkan dalam penelitian ini bekerja dengan cara yang jauh lebih terbatas: gambar wajah diproses sebagai data piksel, fitur dikodekan secara numerik melalui CNN, lalu model ANN memberikan prediksi berdasarkan pola statistik.

Perbandingan ini menunjukkan bahwa teknologi hanyalah peniruan sebagian kecil dari mekanisme sempurna yang Allah ciptakan pada manusia. Meskipun AI mampu melakukan klasifikasi secara otomatis, kecerdasannya tetap bersifat terbatas, tidak mempunyai kesadaran, dan tidak mampu menggantikan akal serta intuisi manusia. Dengan demikian, penelitian ini seharusnya semakin meneguhkan keyakinan bahwa sistem persepsi dan pengolahan informasi pada manusia

merupakan ciptaan Allah yang jauh lebih kompleks dan sempurna dibanding algoritma apa pun.

Integrasi nilai Islam dengan penelitian ini juga ditegaskan dalam firman Allah pada *QS. Al-Mulk ayat 23:*

قُلْ هُوَ الَّذِي أَنْشَأَكُمْ وَجَعَلَ لَكُمُ السَّمْعَ وَالْأَبْصَارَ وَالْأَفْئِدَةَ ۖ قَلِيلًا مَا تَشْكُرُونَ

"Katakanlah: Dialah yang menciptakan kamu dan menjadikan pendengaran, penglihatan, dan hati nurani bagi kamu; (tetapi) sedikit sekali kamu bersyukur."(QS. Al-Mulk ayat 23).

Ayat ini menekankan bahwa kemampuan manusia untuk melihat, mendengar, serta memahami informasi adalah karunia langsung dari Allah SWT. Kemampuan tersebut merupakan fondasi dari proses persepsi, analisis, dan pengambilan keputusan dalam kehidupan sehari-hari. Dalam konteks penelitian ini, kemampuan manusia untuk mengenali ekspresi wajah baik melalui pengamatan visual maupun pemahaman emosional merupakan salah satu manifestasi nyata dari anugerah penglihatan dan akal tersebut. Sementara itu, kecerdasan buatan yang dikembangkan melalui CNN hanya berfungsi sebagai representasi matematis yang sangat terbatas dari proses yang jauh lebih kompleks yang terjadi pada manusia.

Ayat ini juga mengingatkan bahwa manusia sering lupa mensyukuri kemampuan tersebut. Kemampuan sistem AI untuk memproses gambar, mengekstraksi fitur, dan melakukan klasifikasi justru menjadi pengingat bahwa manusia diberikan sistem persepsi yang jauh lebih unggul tanpa memerlukan algoritma, pelatihan, atau pemrosesan data. Perbandingan ini menegaskan bahwa kemajuan teknologi seharusnya tidak menjadikan manusia sombong akan

pencapaiannya, melainkan semakin menyadari betapa luasnya ilmu Allah dan betapa kecilnya pengetahuan manusia dalam menciptakan sesuatu yang meniru sebagian kecil ciptaan-Nya.

Dengan demikian, ayat ini memberikan landasan spiritual bahwa penelitian dalam bidang teknologi, termasuk pengembangan sistem pengenalan ekspresi wajah, bukan hanya bentuk eksplorasi ilmiah, tetapi juga bentuk rasa syukur atas karunia penglihatan dan pemahaman yang dianugerahkan Allah.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini bertujuan untuk melakukan klasifikasi ekspresi wajah menggunakan *Convolutional Neural Network (CNN)* dengan arsitektur *EfficientNetB0* pada dataset FER-13 yang terdiri dari tujuh kelas emosi, yaitu *angry, disgust, fear, happy, neutral, sad,* dan *surprise*. Penelitian dilakukan melalui beberapa skenario eksperimen guna mengevaluasi pengaruh teknik augmentasi data, penerapan *MixUp*, serta *label smoothing* terhadap kinerja dan kemampuan generalisasi model.

Berdasarkan hasil pengujian yang telah dilakukan, dapat disimpulkan bahwa strategi pelatihan yang digunakan memiliki pengaruh signifikan terhadap performa klasifikasi. Skenario baseline yang hanya mengandalkan augmentasi konvensional tanpa regularisasi tambahan cenderung menghasilkan performa yang lebih rendah, khususnya pada metrik *precision, recall,* dan *F1-score*. Hal ini menunjukkan bahwa model masih rentan terhadap *overfitting* dan kesalahan prediksi pada kelas-kelas emosi yang memiliki kemiripan fitur visual.

Penerapan label smoothing terbukti mampu meningkatkan stabilitas prediksi model dengan cara mengurangi tingkat kepercayaan berlebih (*overconfidence*) pada distribusi output. Teknik ini memberikan peningkatan yang konsisten pada seluruh metrik evaluasi, terutama pada *F1-score* dan *accuracy*, yang mencerminkan peningkatan keseimbangan antara precision dan recall.

Sementara itu, penggunaan teknik MixUp membantu model mempelajari representasi fitur yang lebih robust melalui kombinasi linier antar sampel, sehingga meningkatkan kemampuan generalisasi model terhadap data yang tidak terlihat selama pelatihan.

Hasil terbaik diperoleh pada Skenario D**,** yaitu kombinasi antara MixUp dan label smoothing, dengan nilai akurasi sebesar 79% dan F1-score sebesar 0.78**.** Skenario ini menunjukkan performa paling stabil dan seimbang dibandingkan skenario lainnya, serta mampu mengurangi kesalahan klasifikasi pada kelas-kelas emosi yang memiliki pola visual saling tumpang tindih, seperti *fear*, *sad*, dan *neutral*. Selain itu, ekspresi dengan ciri visual yang lebih kuat seperti *happy* dan *surprise* secara konsisten mencapai nilai performa tertinggi pada seluruh skenario.

Secara keseluruhan, penelitian ini membuktikan bahwa arsitektur EfficientNetB0 efektif digunakan untuk tugas klasifikasi ekspresi wajah, khususnya ketika dikombinasikan dengan strategi augmentasi dan regularisasi yang tepat. Temuan ini menegaskan bahwa integrasi MixUp dan label smoothing dapat meningkatkan kemampuan generalisasi model serta menghasilkan prediksi yang lebih stabil dan reliabel. Hasil penelitian ini diharapkan dapat menjadi dasar bagi pengembangan lebih lanjut pada sistem pengenalan ekspresi wajah, baik melalui penggunaan dataset yang lebih besar, eksplorasi arsitektur CNN lainnya, maupun penerapan teknik pelatihan yang lebih kompleks.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya, yaitu:

1. Penelitian ini menggunakan arsitektur EfficientNetB0 dengan kombinasi MixUp dan label smoothing. Penelitian selanjutnya dapat mengeksplorasi varian EfficientNet lain atau arsitektur CNN modern serta mengombinasikannya dengan teknik pelatihan tambahan guna memperoleh performa yang lebih optimal.
2. Penelitian lanjutan disarankan untuk menambahkan metode interpretabilitas seperti Grad-CAM atau teknik visualisasi lainnya agar proses pengambilan keputusan model dapat dipahami dengan lebih baik, terutama dalam mengidentifikasi bagian wajah yang paling berpengaruh terhadap klasifikasi ekspresi.

DAFTAR PUSTAKA

-, R. K., -, D. K. P. S. A., & -, D. H. S. (2025). A Systematic Review on Facial Emotion Recognition System Datasets. *International Journal on Science and Technology*, *16*(2), 1–12. https://doi.org/10.71097/ijsat.v16.i2.6406

Ab Wahab, M. N., Nazir, A., Ren, A. T. Z., Noor, M. H. M., Akbar, M. F., & Mohamed, A. S. A. (2021). Efficientnet-Lite and Hybrid CNN-KNN Implementation for Facial Expression Recognition on Raspberry Pi. *IEEE Access*, *9*, 134065–134080. https://doi.org/10.1109/ACCESS.2021.3113337

Aly, M. (2025). Revolutionizing online education: Advanced facial expression recognition for real-time student progress tracking via deep learning model. In *Multimedia Tools and Applications* (Vol. 84, Issue 13). Springer US. https://doi.org/10.1007/s11042-024-19392-5

Duan, C. (2023). A survey of facial expression recognition in the wild. *Applied and Computational Engineering*, *6*(1), 98–106. https://doi.org/10.54254/2755-2721/6/20230760

Elsheikh, R. A., Mohamed, M. A., Abou-Taleb, A. M., & Ata, M. M. (2024). Improved facial emotion recognition model based on a novel deep convolutional structure. *Scientific Reports*, *14*(1), 1–31. https://doi.org/10.1038/s41598-024-79167-8

Gaya-Morey, F. X., Manresa-Yee, C., Martinie, C., & Buades-Rubio, J. M. (2025). *Evaluating Facial Expression Recognition Datasets for Deep Learning: A Benchmark Study with Novel Similarity Metrics*. 1–12. http://arxiv.org/abs/2503.20428

Gu, C., & Lee, M. (2024). *Deep Transfer Learning Using Real-World Image Features for Medical Image Classification , with a Case Study on Pneumonia X-ray Images*.

Gursesli, M. C., Lombardi, S., Duradoni, M., Bocchi, L., Guazzini, A., & Lanata,

A. (2024). Facial Emotion Recognition (FER) Through Custom Lightweight CNN Model: Performance Evaluation in Public Datasets. *IEEE Access*, *12*, 45543–45559. https://doi.org/10.1109/ACCESS.2024.3380847

Hendrawati, T., Pravitasari, A. A., Hermawan, R. F., Subekti, A., & Yasyfi, M. (2025). *Jurnal resti*. *9*(4), 714–720.

Insani, M. K., & Santoso, D. B. (2024). Perbandingan Kinerja Model Pre-Trained CNN (VGG16, RESNET, dan INCEPTIONV3) untuk Aplikasi Pengenalan Wajah pada Sistem Absensi Karyawan. *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi*, *5*(3), 2612–2622. https://doi.org/10.35870/jimik.v5i3.925

Kopalidis, T., Solachidis, V., Vretos, N., & Daras, P. (2024). Advances in Facial Expression Recognition: A Survey of Methods, Benchmarks, Models, and Datasets. *Information (Switzerland)*, *15*(3). https://doi.org/10.3390/info15030135

Liao, J., Lin, Y., Ma, T., He, S., Liu, X., & He, G. (2023). Facial Expression Recognition Methods in the Wild Based on Fusion Feature of Attention Mechanism and LBP. *Sensors*, *23*(9). https://doi.org/10.3390/s23094204

Mejia-Escobar, C., Cazorla, M., & Martinez-Martin, E. (2023). Towards a Better Performance in Facial Expression Recognition: A Data-Centric Approach. *Computational Intelligence and Neuroscience*, *2023*(1). https://doi.org/10.1155/2023/1394882

Mou, X., Song, Y., Wang, R., Tang, Y., & Xin, Y. (2023). Lightweight Facial Expression Recognition Based on Class-Rebalancing Fusion Cumulative Learning. *Applied Sciences (Switzerland)*, *13*(15). https://doi.org/10.3390/app13159029

Nan, F., Jing, W., Tian, F., Zhang, J., Chao, K. M., Hong, Z., & Zheng, Q. (2022). Feature super-resolution based Facial Expression Recognition for multi-scale low-resolution images. *Knowledge-Based Systems*, *236*, 107678.

https://doi.org/10.1016/j.knosys.2021.107678

Ngwe, J. Le, Lim, K. M., Lee, C. P., Ong, T. S., & Alqahtani, A. (2024). PAtt-Lite: Lightweight Patch and Attention MobileNet for Challenging Facial Expression Recognition. *IEEE Access*, *12*, 79327–79341. https://doi.org/10.1109/ACCESS.2024.3407108

Obaid, A. J., & Alrammahi, H. K. (2023). An Intelligent Facial Expression Recognition System Using a Hybrid Deep Convolutional Neural Network for Multimedia Applications. *Applied Sciences (Switzerland)*, *13*(21). https://doi.org/10.3390/app132112049

Parmonangan, I. H., Marsella, M., Pardede, D. F. R., Rijanto, K. P., Stephanie, S., Kesuma, K. A. C., Cahyaningtyas, V. T., & Anggreainy, M. S. (2023). Training CNN-based Model on Low Resource Hardware and Small Dataset for Early Prediction of Melanoma from Skin Lesion Images. *Engineering, MAthematics and Computer Science (EMACS) Journal*, *5*(2), 41–46. https://doi.org/10.21512/emacsjournal.v5i2.9904

Pham, T. D., Duong, M. T., Ho, Q. T., Lee, S., & Hong, M. C. (2023). CNN-Based Facial Expression Recognition with Simultaneous Consideration of Inter-Class and Intra-Class Variations. *Sensors*, *23*(24), 1–18. https://doi.org/10.3390/s23249658

Shahzad, H. M., Bhatti, S. M., Jaffar, A., Akram, S., Alhajlah, M., & Mahmood, A. (2023). Hybrid Facial Emotion Recognition Using CNN-Based Features. *Applied Sciences (Switzerland)*, *13*(9), 1–14. https://doi.org/10.3390/app13095572

Sufian, M. M., Moung, E. G., Hanafi, M., Hijazi, A., Yahya, F., Dargham, J. A., Farzamnia, A., Sia, F., Faraha, N., & Naim, M. (2023). *COVID-19 Classification through Deep Learning Models with Three-Channel Grayscale CT Images*.

Tiwari, P., Kumar, N., Singh, P., Attray, P., Rai, A., & Khan, N. U. (2022). *Deep*

Learning-Based Recognition of Facial. *December*.

Ullah, S., Ou, J., Xie, Y., & Tian, W. (2024). Facial expression recognition (FER) survey: a vision, architectural elements, and future directions. *PeerJ Computer Science*, *10*. https://doi.org/10.7717/PEERJ-CS.2024

Yalçin, N., & Alisawi, M. (2024). Introducing a novel dataset for facial emotion recognition and demonstrating significant enhancements in deep learning performance through pre-processing techniques. *Heliyon*, *10*(20), e38913. https://doi.org/10.1016/j.heliyon.2024.e38913

Yu, H. (2024). Facial expression recognition with computer vision. *Applied and Computational Engineering*, *37*(1), 74–80. https://doi.org/10.54254/2755-2721/37/20230473

Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A review of convolutional neural networks in computer vision. In *Artificial Intelligence Review* (Vol. 57, Issue 4). Springer Netherlands. https://doi.org/10.1007/s10462-024-10721-6